

World Bank

تقييم الأثر في الواقع العملي

بول جيه جيرتلا، سبيستيان
مارتينيز، باتريك بريماند، لورا بي
رولينغز، كريستل ام جيه فيرميرش

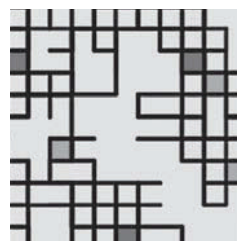
تمت الترجمة والمراجعة من خلال



مركز الأثر
IMPACT CENTER
مستقبل أفضل
BETTER FUTURE

تقييم الأثر في الواقع العملي

الطبعة الثانية



٢٠ مركز الأثر للبحوث والدراسات المحدودة ، ١٤٤٦ هـ

بريماند، باتريك
تقديم الأثر في الواقع العملي. / بريماند، باتريك ؛ فريتلر ،
بول جي ؛ مارتينيز ، سباستيان -. الرياض ، ١٤٤٦هـ

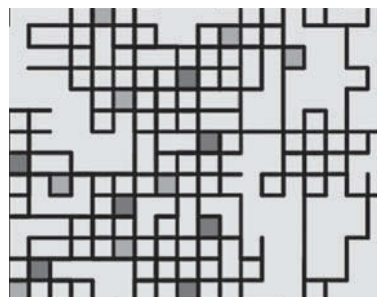
رقم الإيداع: ١٤٤٦/٩٥٦٥
ردمك: ٩٧٨-٦٠٣-٩٢٢٦٥-٢-٩

يرجى زيارة الموقع الإلكتروني لكتاب «تقييم الأثر في الواقع العملي»

على <http://www.worldbank.org/ieinpractice>. يحتوي الموقع الإلكتروني على مواد مرفقة، بما في ذلك حلول لأسئلة دراسة حالة برنامج معونة التأمين الصحي (HISP) المذكورة في الكتاب، بالإضافة إلى مجموعة البيانات ذات الصلة ورمز التحليل في البرنامج الحاسوبي Stata، ودليل فني يوفر معالجة أكثر رسمية لتحليل البيانات، وعروض PowerPoint ذات صلة بالفصول، ونسخة إلكترونية من الكتاب مع روابط إلى مواقع إلكترونية، وروابط إلى مواد إضافية.

لم يكن لهذا الكتاب أن يرى النور إلا بفضل الدعم السخي من صندوق تقييم الأثر الإستراتيجي (SIEF). وقد شهد عام ٢٠١٢ تأسيس الصندوق بدعم من وزارة التنمية الدولية في المملكة المتحدة، وهو برنامج شراكة يعزز صنع السياسات القائمة على الأدلة. ويسلط الصندوق اهتمامه حاليًا على أربعة مجالات بالغة الأهمية للتنمية البشرية السليمة صحيًا؛ ألا وهي التعليم الأساسي، والنظم الصحية وتقديم الخدمات، وتنمية الطفولة المبكرة والتغذية، والمياه والصرف الصحي. وتمتد أعمال صندوق تقييم الأثر الإستراتيجي (SIEF) إلى جميع أنحاء العالم، وبالأخص الدول منخفضة الدخل، حيث ينقل الخبرة في تقييم الأثر والأدلة إلى مجموعة من البرامج وفرق صنع السياسات.

تقييم الأثر في الواقع العملي الطبعة الثانية



بول جيه جيرتير، وسيباستيان مارتينيز، وباتريك
بريمان، ولورا بي رولينغز، وكريستل إم جيه
فيرميرش

حقوق الطبع والنشر © لعام ٢٠١٦ محفوظة لصالح البنك الدولي للإنشاء والتعمير / البنك الدولي

1818 H Street NW, Washington, DC 20433

الهاتف: ٢٠٢-٤٧٣-١٠٠٠؛ الموقع الإلكتروني: www.worldbank.org

بعض الحقوق محفوظة

١٦ ١٧ ١٨ ١٩ ٤ ٣ ٢ ١

لا تعكس النتائج، والتفسيرات، والاستنتاجات الواردة في هذا العمل بالضرورة آراء البنك الدولي، أو مجلس مديريه التنفيذيين، أو الحكومات التي يمثلونها. ولا يضمن البنك الدولي أو بنك التنمية للبلدان الأمريكية دقة البيانات الواردة في هذا العمل. ولا تشير الحدود والألوان والرموز والتصنيفات والمعلومات الأخرى الموضحة على أي خريطة في هذا العمل إلى أي حكم من جانب البنك الدولي أو بنك التنمية للبلدان الأمريكية فيما يتعلق بالوضع القانوني لأي إقليم أو الموافقة على مثل هذه الحدود أو قبولها. ولا يشكل أي نص وارد في هذه الوثيقة أو يعتبر قيداً على امتيازات وحصانات البنك الدولي أو بنك التنمية للبلدان الأمريكية أو تنازلاً عنها، وهي امتيازات وحصانات محفوظة بشكل خاص.

الحقوق والأذونات



هذا العمل مُتاح بموجب ترخيص المشاع الإبداعي، تُسب المُصنّف ٣,٠ منظمة حكومية دولية (CC BY 3.0 IGO) المبين بهذا الرابط <http://creativecommons.org/licenses/by/3.0/igo>. بموجب ترخيص المشاع الإبداعي، يحق لك نسخ هذا العمل، وتوزيعه، ونقله، والاقتباس منه، لأغراض تجارية، وفقاً للشروط التالية:

الإسناد—يرجى الاستشهاد بالعمل على النحو التالي: Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. *Impact Evaluation in Practice, second edition*. Washington, DC: Inter-American Development Bank and World Bank. doi:10.1596/978-1-4648-0779-4. License: Creative Commons Attribution CC BY 3.0 IGO

الترجمات—إذا قمت بترجمة هذا العمل، فيرجى إضافة إخلاء المسؤولية التالي مع الإسناد: لم يتم إنشاء هذه الترجمة من قبل البنك الدولي، ولن تعتبر ترجمة رسمية للبنك الدولي. ولن يتحمل البنك الدولي المسؤولية عن أي محتوى أو خطأ في هذه الترجمة.

الاقتباسات—إذا اقتبست من هذا العمل، فيرجى إضافة إخلاء المسؤولية التالي مع الإسناد: هذا الاقتباس من عمل أصلي للبنك الدولي. ووجهات النظر والآراء المعبر عنها في الاقتباس مسؤولية مؤلف أو مؤلفي الاقتباس وحدهم، ولا تتبناها البنك الدولي.

محتوى الأطراف الثالثة—لا يمتلك البنك الدولي بالضرورة كل عنصر من المحتوى المتضمن في العمل. ومن ثم، لا يضمن البنك الدولي أن استخدام أي عنصر أو جزء فردي مملوك لطرف ثالث وارد في العمل لن ينتهك حقوق تلك الأطراف الثالثة. وتقع مخاطر المطالبات الناجمة عن مثل هذا الانتهاك على عاتقك وحدك. وإذا كنت ترغب في إعادة استخدام أحد عناصر العمل، فأنت المسؤول عن تحديد ما إذا كان يلزم الحصول على إذن لإعادة الاستخدام والحصول على إذن من مالك حقوق الطبع والنشر. ويمكن أن تشمل أمثلة العناصر، على سبيل المثال لا الحصر، الجداول، أو الأشكال، أو الصور.

يجب توجيه جميع الاستفسارات المتعلقة بالحقوق والتراخيص إلى Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA. البريد الإلكتروني: pubrights@worldbank.org.

الرقم الدولي الموحد للكتاب (النسخة الورقية): ٩٧٨-١-٤٦٤٨-٠٧٧٩-٤

الرقم الدولي الموحد للكتاب (النسخة الإلكترونية): ٩٧٨-١-٤٦٤٨-٠٧٨٠-٠

مُعَرِّف الكائن الرقمي: ١٠,١٥٩٦/٩٧٨-١-٤٦٤٨-٠٧٧٩-٤

الرسوم التوضيحية: أندريس جوميز بينا وميكايلا فيزر

تصميم الغلاف: صحيفة Critical Stages

مكتبة الكونغرس، بيانات الفهرسة في النشر

الأسماء: Gertler, Paul, 1955- author. | World Bank

العنوان: Impact evaluation in practice / Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch

الوصف: Second Edition. | Washington, D.C.: World Bank, 2016. | Revised

edition of Impact evaluation in practice, 2011

المُعَرِّفات: رقم الضبط في مكتبة الكونغرس ٢٠١٦٠٢٩٠٦١ (النسخة المطبوعة) | رقم الضبط في مكتبة الكونغرس ٢٠١٦٠٢٩٤٦٤ (الكتاب الإلكتروني) | الرقم الدولي

الموحد للكتاب ٩٧٨١٤٦٤٨٠٧٧٩٤ (pdf) | الرقم الدولي الموحد للكتاب ٩٧٨١٤٦٤٨٠٧٨٠٠ | الرقم الدولي الموحد للكتاب ٩٧٨١٤٦٤٨٠٧٨٠٠ ()

الموضوعات: عناوين الموضوعات في مكتبة الكونغرس: Evaluation | Economic development projects—Evaluation

research (Social action programs)

التصنيف: تصنيف مكتبة الكونغرس HD75.9.G478 2016 (النسخة المطبوعة) | تصنيف مكتبة الكونغرس HD75.9 (الكتاب الإلكتروني) | DDC 338.91—dc23

إقرار

إن دليل تقييم الأثر في الواقع العملي الذي نشره البنك الدولي في الأصل باللغة الإنجليزية تحت عنوان: Impact Evaluation in Practice, Second Edition,

© 2016 البنك الدولي.

جميع الحقوق محفوظة.

متاح على العنوان التالي :

<https://openknowledge.worldbank.org/handle/10986/25030>

هذا الدليل متاح ومرخص بموجب ترخيص المشاع الإبداعي نسب المصنف - غير موطنة 3.0 دولي، وفقاً للشروط التالية: نسب المصنف – يجب عليك نسب العمل لصاحبه بطريقة مناسبة، وتوفير رابط للتريخ، وبيان إذا ما قد أُجريت أي تعديلات على العمل. يمكنك القيام بهذا بأي طريقة مناسبة، ولكن على ألا يتم ذلك بطريقة توهي بأن المؤلف أو المرخص مؤيد لك أو لعملك. منع القيود الإضافية – يجب عليك ألا تطبق أي شروط قانونية أو تدابير تكنولوجية تقيد الآخرين من ممارسة الصلاحيات التي تسمح بها الرخصة. وللإطلاع على نسخة من هذا الترخيص

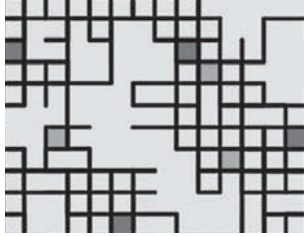
، يمكنك زيارة: <https://creativecommons.org/licenses/by/3.0/deed.ar>

جودة الترجمة باللغة العربية وتماسكها مع النص الأصلي هي المسؤولية الوحيدة لمركز الأثر. الأصل الإنجليزي لهذا العمل هو النسخة الرسمية الوحيدة.

“Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J.. 2016. Impact Evaluation in Practice, Second Edition. Washington, DC: Inter-American Development Bank and World Bank. © World Bank.

<https://openknowledge.worldbank.org/handle/10986/25030> License: CC BY 3.0 IGO.”

“لم يتم ترجمة هذا الدليل من قبل أي مؤسسة عضو في مجموعة البنك الدولي ولا ينبغي اعتبارها ترجمة رسمية لمجموعة البنك الدولي. لن تتحمل أي مؤسسة عضو في مجموعة البنك الدولي المسؤولية عن أي محتوى أو خطأ في هذه الترجمة.”



المحتويات

xv	مقدمة
xxi	شكر وتقدير
xxiii	نبذة عن المؤلفين
xxvii	الاختصارات
I	الجزء الأول. مقدمة إلى تقييم الأثر
٣	الفصل الأول. ما فائدة التقييم؟
٣	صنع السياسات القائمة على الأدلة
٧	ما المقصود بتقييم الأثر؟
٩	تقييم الأثر السابق مقابل اللاحق
I	دراسات النجاعة ودراسات الفعالية
١٣	المقاربات التكميلية
٢٠	الاعتبارات الأخلاقية المتعلقة بتقييم الأثر
٢١	تقييم الأثر لاتخاذ القرارات المتعلقة بالسياسات
٢٦	اتخاذ قرار بشأن إجراء تقييم الأثر
٣١	الفصل الثاني. الإعداد لإجراء التقييم
٣١	الخطوات الأولية
٣٢	بناء نظرية التغيير
٣٤	وضع سلسلة النتائج
٣٦	تحديد أسئلة التقييم
٤١	اختيار مؤشرات الحاصل والأداء
٤٢	قائمة المراجعة: جمع البيانات لمؤشراتك
٤٥	الجزء الثاني. كيفية التقييم
٤٧	الفصل الثالث. الاستدلال السببي والواقع المضاد
٤٧	الاستدلال السببي
٤٨	الواقع المضاد
٥٤	تقييمان زائفان للواقع المضاد
VIII	

الفصل الرابع. التعيين العشوائي

تقييم البرامج استنادًا إلى قواعد التعيين

التعيين العشوائي للمعالجة

قائمة المراجعة: التعيين العشوائي

الفصل الخامس. المتغيرات المساعدة

تقييم البرامج عندما لا يمثل الجميع لتعييناتهم

أنواع تقديرات الأثر

الامتثال غير المثالي

الترويج العشوائي كمتغير مساعد

قائمة المراجعة: الترويج العشوائي كمتغير مساعد

الفصل السادس. تصميم انقطاع الانحدار

تقييم البرامج التي تستخدم مؤشر الأهلية

تصميم انقطاع الانحدار المشوش

التحقق من صلاحية تصميم انقطاع الانحدار

القيود على طريقة تصميم انقطاع الانحدار وتفسيراتها

قائمة المراجعة: تصميم انقطاع الانحدار

الفصل السابع. الاختلاف في الاختلافات

تقييم البرنامج عندما تكون قاعدة التعيين أقل وضوحًا

طريقة الاختلاف في الاختلافات

كيف تكون طريقة الاختلاف في الاختلافات مفيدة؟

افتراض "الاتجاهات المتعادلة" في الاختلاف في الاختلافات

قيود طريقة الاختلاف في الاختلافات

قائمة المراجعة: الاختلاف في الاختلافات

الفصل الثامن. المطابقة

تشكيل مجموعة مقارنة صورية

مطابقة درجة الميل

دمج المطابقة مع طرق أخرى

قيود طريقة المطابقة

قائمة المراجعة: المطابقة

الفصل التاسع. معالجة التحديات المنهجية

تأثيرات المعالجة غير المتجانسة

التأثيرات السلوكية غير المتعمدة

الامتثال غير المثالي

١٦٣	التأثيرات غير المباشرة
١٦٩	التناقص
١٧١	توقيت التأثيرات واستمرارها
١٧٥	الفصل العاشر. تقييم البرامج متعددة الأوجه
١٧٥	تقييم البرامج التي تجمع بين العديد من خيارات المعالجة
١٧٦	تقييم البرامج ذات مستويات المعالجة المتباينة
١٧٩	تقييم التدخلات المتعددة
١٨٥	الجزء الثالث. كيفية تنفيذ تقييم الأثر
١٨٧	الفصل الحادي عشر. اختيار أسلوب تقييم الأثر
١٨٧	تحديد الأسلوب الذي يجب استخدامه لبرنامج معين
١٨٨	كيف يمكن لقواعد تشغيل البرنامج المساعدة على اختيار أسلوب تقييم الأثر
١٩٣	مقارنة بين أساليب تقييم الأثر
١٩٧	العثور على أصغر وحدة تدخل قابلة للتنفيذ
٢٠١	الفصل الثاني عشر. إدارة تقييم الأثر
٢٠١	إدارة فريق التقييم، والوقت، والميزانية
٢٠٢	أدوار ومسؤوليات فرق البحث والسياسات
٢٠٨	إقامة التعاون
٢١٣	كيفية تحديد وقت إجراء التقييم
٢١٦	كيفية تحديد ميزانية للتقييم
٢٣١	الفصل الثالث عشر. أخلاقيات وعلم تقييم الأثر
٢٣١	إدارة تقييمات أخلاقية وذات مصداقية
٢٣٢	أخلاقيات إجراء تقييمات الأثر
٢٣٧	ضمان إجراء تقييمات موثوقة وذات مصداقية من خلال العلم المفتوح
٢٤٣	قائمة المراجعة: تقييم أثر أخلاقي وذو مصداقية
٢٤٧	الفصل الرابع عشر. نشر النتائج وتحقيق أثر السياسة
٢٤٧	قاعدة أدلة قوية للسياسات
٢٥٠	تصميم إستراتيجية إعلام لأنواع الجمهور المختلفة
٢٥٤	نشر النتائج
٢٥٩	الجزء الرابع. كيفية الحصول على البيانات لإجراء تقييم الأثر

الفصل الخامس عشر. اختيار العينة

أخذ العينات وحسابات القوة الإحصائية

سحب العينة

التحديد القائم على حجم العينة لتقييم الأثر: حسابات القوة الإحصائية

الفصل السادس عشر. العثور على مصادر بيانات مناسبة

أنواع البيانات المطلوبة

استخدام البيانات الكمية الحالية

جمع بيانات مسح جديدة

الفصل السابع عشر. الخلاصة

تقييمات الأثر: تدريبات جديرة بالاهتمام ولكن معقدة

قائمة المراجعة: العناصر الأساسية لتقييم أثر جيد التصميم

قائمة المراجعة: نصائح لتخفيف المخاطر الشائعة المصاحبة لإجراء تقييم الأثر

قائمة المصطلحات

المربعات

- ١-١ كيف يمكن لتقييم ناجح أن يعزز الاستدامة السياسية لبرنامج تنمية: برنامج التحويلات النقدية المشروطة في المكسيك
- ٢-١ أثر سياسة نموذج مبتكر لمرحلة رياض الأطفال: مرحلة رياض الأطفال وتنمية الطفولة المبكرة في موزمبيق
- ٣-١ التحقق من مدى قابلية تعميم النتائج: تقييم متعدد المواقع لمقاربة "التخرج" في سبيل تخفيف حدة الفقر المدقع
- ٤-١ محاكاة التأثيرات المحتملة للمشروع من خلال النمذجة الهيكلية: بناء نموذج لاختبار التصاميم البديلة باستخدام بيانات برنامج Progresa في المكسيك
- ٥-١ تقييم الأسلوب المختلط في التطبيق العملي: الجمع بين تجربة عشوائية مضبوطة ودراسة إثنوغرافية في الهند
- ٦-١ توجيه التوسع في نطاق البرنامج الوطني من خلال تقييم الصيرورة في تنزانيا
- ٧-١ تقييم فعالية التكلفة: مقارنة تقييمات البرامج التي تؤثر في التعليم في المدارس الابتدائية
- ٨-١ تقييم البرامج المبتكرة: فريق الرؤى السلوكية في المملكة المتحدة
- ٩-١ تقييم بدائل تصميم البرنامج: سوء التغذية والنمو المعرفي في كولومبيا
- ١٠-١ المقاربة العنقودية لتقييم الأثر: جمع الأدلة بشكل إستراتيجي لسد الفجوات المعرفية
- ١-٢ صياغة نظرية التغيير: من الأرضيات الإسمنتية إلى السعادة في المكسيك
- ٢-٢ تجارب الآلية

٣٨	٣-٢	إصلاح منهج الرياضيات للمرحلة الثانوية: وضع سلاسل النتائج وأسئلة التقييم
٥٠	١-٣	إشكالية الواقع المضاد: «الطفلة المميزة» وبرنامج التحويل النقدي
٦٥	١-٤	التعيين العشوائي كأداة تشغيلية
٧٠	٢-٤	قيّمة التعيين العشوائي كقاعدة لتخصيص الأموال للبرنامج: التحويلات النقدية المشروطة والتعليم في المكسيك
٧٠	٣-٤	التعيين العشوائي للمنح لتحسين فرص عمل الشباب في شمال أوغندا
٧١	٤-٤	التعيين العشوائي لتدخلات المياه والصرف الصحي في ريف بوليفيا
٧٢	٥-٤	التعيين العشوائي لحماية مياه الينابيع في سبيل تحسين الصحة في كينيا
٧٢	٦-٤	التعيين العشوائي للمعلومات عن مخاطر فيروس نقص المناعة البشرية (HIV) للحد من الحمل في سن المراهقة في كينيا
٩١	١-٥	استخدام المتغيرات المساعدة لتقدير أثر عالم سمس على الاستعداد للمدرسة
٩٩	٢-٥	استخدام المتغيرات المساعدة للتعامل مع عدم الامتثال في برنامج قسيمة المدرسة في كولومبيا
١٠٧	٣-٥	الترويج العشوائي للاستثمارات في البنية التحتية للتعليم في بوليفيا
١١٤	١-٦	استخدام تصميم انقطاع الانحدار لتقييم أثر خفض المصروفات الدراسية على معدلات الالتحاق بالمدارس في كولومبيا
١١٨	٢-٦	شبكات الأمان الاجتماعي المستندة إلى مؤشر الفقر في جامايكا
١٢٠	٣-٦	تأثير تصنيف الطلاب حسب درجات الاختبار على الأداء المدرسي في كينيا
١٣١	١-٧	استخدام الاختلاف في الاختلافات لفهم تأثير الحوافز الانتخابية على معدلات التسرب من المدارس في البرازيل
١٣٥	٢-٧	استخدام الاختلاف في الاختلافات لدراسة تأثيرات نشر أفراد الشرطة لمكافحة الجريمة في الأرجنتين
١٣٨	٣-٧	اختبار افتراض الاتجاهات المتعادلة: خصصة شبكات المياه ووفيات الرضع في الأرجنتين
١٣٩	٤-٧	اختبار افتراض الاتجاهات المتعادلة: تشييد المدارس في إندونيسيا
١٤٩	١-٨	الاختلاف في الاختلافات المتطابق: تطوير الطرق الريفية والسوق المحلية في فيتنام
١٤٩	٢-٨	الاختلاف في الاختلافات المتطابق: الأراضي الإسمنتية، وصحة الأطفال، وسعادة الأمهات في المكسيك
١٥١	٣-٨	طريقة الوحدات الضابطة الاصطناعية: الآثار الاقتصادية للصراع الإرهابي في إسبانيا
١٦٠	١-٩	القصص الشعبية عن تقييم الأثر: تأثير هوثورن وتأثير جون هنري
١٦٤	٢-٩	التأثيرات غير المباشرة السلبية نتيجة آثار التوازن العام: المساعدة في التوظيف وحصائل سوق العمل في فرنسا
١٦٦	٣-٩	العمل مع التأثيرات غير المباشرة: القضاء على الديان، والعوامل الخارجية، والتعليم في كينيا
١٦٨	٤-٩	تقييم التأثيرات غير المباشرة: التحويلات النقدية المشروطة والتأثيرات غير المباشرة في المكسيك
١٧٠	٥-٩	التناقص في الدراسات مع المتابعة طويلة المدى: تنمية الطفولة المبكرة والهجرة في جامايكا
١٧٢	٦-٩	تقييم التأثيرات طويلة المدى: المعونات واعتماد الناموسيات المعالجة بمبيدات حشرية في كينيا
١٧٨	١-١٠	اختبار كثافة البرنامج لزيادة الالتزام بالمعالجة المضادة للفيروسات العكوسة
١٧٩	٢-١٠	اختبار بدائل البرنامج لمتابعة الفساد في إندونيسيا

٢٠٠	١-١١	برامج التحويلات النقدية والمستوى الأدنى للتدخل
٢٠٥	١-١٢	المبادئ التوجيهية للتعاون بين فرق السياسات والتقييم
٢٠٧	٢-١٢	مخطط عام لخطة تقييم الأثر
٢١١	٣-١٢	أمثلة على نماذج فريق البحث-السياسات
٢٤٠	١-١٣	سجلات تجارب العلوم الاجتماعية
٢٤٩	١-١٤	أثر سياسة نموذج مبتكر لمرحلة رياض الأطفال في موزمبيق
٢٥٤	٢-١٤	أدوات التوعية والنشر
٢٥٥	٣-١٤	نشر تقييمات الأثر بفعالية
٢٥٦	٤-١٤	نشر تقييمات الأثر عبر الإنترنت
٢٥٧	٥-١٤	مدونات تقييمات الأثر
٢٦٥	١-١٥	أخذ العينات العشوائية ليس كافيًا لإجراء تقييم الأثر
٢٩٧	١-١٦	إنشاء مجموعة بيانات في تقييم Plan Nacer في الأرجنتين
٢٩٨	٢-١٦	استخدام بيانات التعداد لإعادة تقييم برنامج PRAF في هندوراس
٣٠٥	٣-١٦	تصميم وصياغة الاستبيانات
٣٠٧	٤-١٦	بعض إيجابيات وسلبيات جمع البيانات إلكترونياً
٣١٢	٥-١٦	جمع البيانات لتقييم برنامج Atención a Crisis التجريبي في نيكاراغوا
٣١٤	٦-١٦	المبادئ التوجيهية لتوثيق البيانات وحفظها

الأشكال

٣٥	١-٢ عناصر سلسلة النتائج
٣٧	ب-٢-١ تحديد تجربة الآلية من سلسلة نتائج أطول
٣٩	ب-٢-١٠ سلسلة نتائج لإصلاح منهج الرياضيات للمرحلة الثانوية
٤٠	٢-٢ سلسلة نتائج برنامج معونة التأمين الصحي (HISP)
٥١	١-٣ القرين المثالي
٥٣	٢-٣ مجموعة المقارنة الصالحة
٥٥	٣-٣ التقديرات قبل وبعد لبرنامج التمويل الصغير
٦٨	١-٤ خصائص المجموعات في ظل التعيين العشوائي للمعالجة
٧٣	٢-٤ أخذ العينات العشوائية والتعيين العشوائي للمعالجة
٧٦	٣-٤ خطوات التعيين العشوائي للمعالجة
٧٨	٤-٤ استخدام جدول بيانات لجعل التعيين للمعالجة عشوائيًا
٨١	٥-٤ تقدير الأثر في ظل التعيين العشوائي
٩٥	١-٥ التعيين العشوائي مع الامتثال غير المثالي
٩٧	٢-٥ تقدير متوسط تأثير المعالجة المكاني في ظل التعيين العشوائي مع الامتثال غير المثالي
١٠٥	٣-٥ الترويج العشوائي
١٠٦	٤-٥ تقدير متوسط تأثير المعالجة المكاني في ظل الترويج العشوائي
١١٦	١-٦ محصول الأرز، المزارع الصغيرة مقابل المزارع الكبيرة (خط الأساس)
١١٧	٢-٦ محصول الأرز، المزارع الصغيرة مقابل المزارع الكبيرة (المتابعة)
١١٩	٣-٦ الامتثال للتعيين
١٢٠	٤-٦ التلاعب بمؤشر الأهلية
١٢٢	٥-٦ برنامج HISP: كثافة الأسر المعيشية، حسب مؤشر خط الأساس للفقر
١٢٢	٦-٦ المشاركة في برنامج HISP، حسب مؤشر خط الأساس للفقر
١٢٣	٧-٦ مؤشر الفقر والنفقات الصحية، برنامج HISP، بعد عامين
١٣٢	١-٧ طريقة الاختلاف في الاختلافات
١٣٦	٢-٧ الاختلاف في الاختلافات عند اختلاف اتجاهات الحصائل
١٤٤	١-٨ مطابقة تامة في الخصائص الأربع
١٤٦	٢-٨ مطابقة درجة الميل والدعم المشترك
١٥٣	٣-٨ المطابقة لبرنامج HISP: الدعم المشترك

١٦٧	١-٩	مثال تقليدي على التأثيرات غير المباشرة: العوامل الخارجية الإيجابية للقضاء على الديدان لدى أطفال المدارس
١٧٧	١-١٠	خطوات التعيين العشوائي لمستويين من المعالجة
١٨١	٢-١٠	خطوات التعيين العشوائي لتدخلين
١٨١	٣-١٠	تصميم متقاطع لبرنامج يتضمن تدخلين
٢٦٢	١-١٥	استخدام عينة للاستدلال على متوسط خصائص المجتمع الإحصائي محل الاهتمام
٢٦٣	٢-١٥	إطار تحديد عينة صالح يغطي المجتمع الإحصائي محل الاهتمام بالكامل
٢٦٥	١١-١٥	أخذ العينات العشوائية فيما بين المجموعات غير القابلة للمقارنة للمشاركين وغير المشاركين
٢٦٦	٢-١٥-١	التعيين العشوائي لفوائد البرنامج بين مجموعة معالجة ومجموعة مقارنة
٢٦٩	٣-١٥	العينة الأكبر على الأرجح أن تمثل المجتمع الإحصائي محل الاهتمام

الجدول

٥٧	١-٣	تقييم برنامج HISP: المقارنة قبل وبعد
٥٨	٢-٣	تقييم برنامج HISP: قبل وبعد مع تحليل الانحدار
٦٠	٣-٣	تقييم برنامج HISP: مقارنة المتوسطات للمشاركين وغير المشاركين
٦١	٤-٣	تقييم برنامج HISP: تحليل انحدار المشاركين وغير المشاركين
٨٣	١-٤	تقييم برنامج HISP: تحقيق التوازن بين قرى المعالجة والمقارنة على مستوى بيانات خط الأساس
٨٤	٢-٤	تقييم برنامج HISP: التعيين العشوائي من خلال مقارنة المتوسطات
١٠٨	٣-٤	تقييم برنامج HISP: التعيين العشوائي من خلال تحليل الانحدار
١٠٩	١-٥	تقييم برنامج HISP: الترويج العشوائي من خلال مقارنة المتوسطات
١٢٣	٢-٥	تقييم برنامج HISP: الترويج العشوائي من خلال تحليل الانحدار
١٣٣	١-٦	تقييم برنامج HISP: تصميم انقطاع الانحدار من خلال تحليل الانحدار
	١-٧	طريقة حساب الاختلاف في الاختلافات (DD)
١٤٠	٢-٧	تقييم برنامج HISP: مقارنة المتوسطات للاختلاف في الاختلافات
١٤٠	٣-٧	تقييم برنامج HISP: الاختلاف في الاختلافات من خلال تحليل الانحدار
١٥٢	١-٨	تقدير درجة الميل بناءً على الخصائص المرصودة لخط الأساس
١٥٤	٢-٨	تقييم برنامج HISP: مطابقة خصائص خط الأساس ومقارنة المتوسطات
١٥٤	٣-٨	تقييم برنامج HISP: مطابقة خصائص خط الأساس وتحليل الانحدار
١٥٤	٤-٨	تقييم برنامج HISP: دمج الاختلاف في الاختلافات مع مطابقة خصائص خط الأساس
١٧٨	ب-١٠-١	ملخص تصميم البرنامج
١٩١	١-١١	العلاقة بين القواعد التشغيلية للبرنامج وأساليب تقييم الأثر

١٩٤	٢-١١	المقارنة بين أساليب إجراء تقييم الأثر
٢١٧	١-١٢	تكلفة تقييمات الأثر لمجموعة مختارة من المشاريع المدعومة من البنك الدولي
٢١٨	٢-١٢	التكاليف المفصلة لمجموعة مختارة من تقييمات الأثر المدعومة من البنك الدولي
٢٢٤	٣-١٢	نموذج ميزانية تقييم الأثر
٢٣٨	١-١٣	التأكد من توفير معلومات موثوقة وذات مصداقية للسياسات من خلال العلم المفتوح
٢٥١	١-١٤	إشراك الفئات المستهدفة الرئيسية لمعرفة أثر السياسة: لماذا، ومتى، وكيف
٢٧٣	١-١٥	أمثلة على المجموعات العنقودية
	٢-١٥	تقييم برنامج HISP+: حجم العينة المطلوب للكشف عن الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف، مع افتراض القوة الإحصائية = ٠,٩
٢٧٨	٣-١٥	تقييم برنامج HISP+: حجم العينة المطلوب للكشف عن الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف، مع افتراض القوة الإحصائية = ٠,٨
٢٧٨	٤-١٥	تقييم برنامج HISP+: حجم العينة المطلوب لاكتشاف الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف (الزيادة في معدل الاستشفاء)
٢٧٩	٥-١٥	تقييم برنامج HISP+: حجم العينة المطلوب لاكتشاف الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف (خفض النفقات الصحية التي تتحملها الأسر)
٢٨٢	٦-١٥	تقييم برنامج HISP+: حجم العينة المطلوب لاكتشاف الحد الأدنى البالغ دولارين
٢٨٣		لأعداد مختلفة من المجموعات العنقودية

مقدمة

يوفر هذا الكتاب مقدمة سهلة الفهم إلى موضوع تقييم الأثر وتطبيقه في مجال التنمية. ويقدم إرشادات عملية عن تصميم تقييمات الأثر وتنفيذها، فضلاً عن نظرة عامة غير تقنية عن أساليب تقييم الأثر.

هذه الطبعة الثانية من دليل تقييم الأثر في الواقع العملي. وقد نُشر الدليل لأول مرة في عام ٢٠١١، واستخدمته المجتمعات الإنمائية والأكاديمية على نطاق واسع في جميع أنحاء العالم. وتتوفر الطبعة الأولى باللغات الإنجليزية، والفرنسية، والبرتغالية، والإسبانية.

تغطي الطبعة المحدثة آخر الأساليب المستخدمة في تقييم البرامج وتشمل نصائح حول أحدث سبل التنفيذ، بالإضافة إلى مجموعة كبيرة من الأمثلة ودراسات الحالة التي تستند إلى التدخلات الإنمائية الحديثة. وتتضمن أيضاً مواداً جديدة معنية بأخلاقيات البحث والشراكات لإجراء تقييم الأثر. وفي جميع أجزاء الكتاب، توضح دراسات الحالة تطبيقات تقييمات الأثر. ويرتبط الكتاب بمواد تعليمية تكميلية متاحة على الإنترنت.

يتسم مقارنة تقييم الأثر في هذا الكتاب بالسهولة إلى حد كبير. وقد حاولنا تقليل استخدام المصطلحات التقنية المتخصصة. وأخذت الأساليب مباشرةً من البحث التطبيقي في العلوم الاجتماعية، وتتشارك في العديد من السمات المشتركة مع طرق البحث المستخدمة في العلوم الطبيعية. ومن هذا المنطلق، يجمع تقييم الأثر أدوات البحث التجريبية المستخدمة على نطاق واسع في الاقتصاد والعلوم الاجتماعية الأخرى مع وقائع الاقتصاد التشغيلي والسياسي المعنية بتنفيذ السياسات والممارسة الإنمائية.

إن مقاربتنا لتقييم الأثر عملي أيضاً؛ إذ نعتقد أنه يجب تحديد أنسب أساليب التقييم التي تتلاءم مع السياق التشغيلي، وليس العكس. ويتحقق ذلك على النحو الأفضل في بداية البرنامج، من خلال تصميم تقييمات الأثر المرتقبة المدمجة في عملية تنفيذ المشروع. ونزعم أن التوصل إلى إجماع في الآراء بين الأطراف المعنية وتحديد تصميم التقييم الذي يتناسب مع السياق السياسي والتشغيلي لا يقل أهمية عن أسلوب التقييم ذاته. ونعتقد أيضاً أن تقييمات الأثر يجب

أن تعبر بصراحة عن قيودها ومحاذيرها. وأخيرًا، نشجع بشدة صانعي السياسات ومديري البرامج على تناول تقييمات الأثر باعتبارها جزءًا من نظرية التغيير الموضوعية بشكل جيد، والتي تحدد بوضوح المسارات السببية التي يعمل بموجبها البرنامج على وضع مخرجات والتأثير في الحاصلات النهائية، ونحثهم على دمج تقييمات الأثر مع مقاربات المتابعة والتقييم التكميلية للحصول على صورة كاملة عن النتائج.

إن خبراتنا ودروسنا المستفادة حول كيفية إجراء تقييم أثر في الواقع العملي مستمدة من التدريس والعمل مع مئات الشركاء المؤهلين من الحكومة، والمجال الأكاديمي، ومجال التنمية. وإجمالاً، جاء مضمون الكتاب خلاصة عشرات السنين من الخبرة في العمل مع تقييمات الأثر في كل ركن في العالم تقريبًا، ويخاطب الأجيال القادمة من الممارسين وصانعي السياسات.

نأمل أن يصبح الكتاب مصدرًا قيّمًا للمعلومات لمجتمع التنمية الدولي، والجامعات، وصانعي السياسات الذين يتطلعون إلى إيجاد أدلة أفضل حول ما يصلح في مجال التنمية. وسيساعد التوسع في تقييمات الأثر كمًا ونوعًا على تعزيز قاعدة الأدلة المعنية بسياسات وبرامج التنمية حول العالم. ويحدونا الأمل في أنه إذا تمكنت الحكومات والعاملون في مجال التنمية من اتخاذ قرارات السياسات بناءً على الأدلة، بما في ذلك الأدلة المستخلصة من تقييم الأثر، فستُنقذ موارد التنمية بشكل أكثر فعالية للحد من الفقر وتحسين حياة الناس.

خارطة الطريق لمحتويات الكتاب

يناقش الجزء الأول -مقدمة إلى تقييم الأثر (الفصلان الأول والثاني) سبب إجراء تقييم الأثر ومتى يكون ذلك مفيدًا. ونراجع مختلف الأهداف التي يستطيع تقييم الأثر تحقيقها، ونسلط الضوء على الأسئلة الأساسية للسياسة التي يمكن للتقييم التعامل معها وحلها. ونصر على ضرورة البحث الدقيق لنظرية التغيير التي تشرح القنوات التي تستطيع من خلالها البرامج التأثير على الحاصلات النهائية. ونحث على إجراء دراسة متأنية لمؤشرات الحاصلات وأحجام التأثير المتوقعة. يراجع الجزء الثاني -كيفية التقييم (الفصل الثالث إلى العاشر) الأساليب المختلفة التي تسفر عن مجموعات مقارنة يمكن استخدامها لتقدير آثار البرنامج. ونبدأ بتقديم مفهوم الواقع المضاد بصفته جوهر أي تقييم للأثر، وشرح الخصائص التي يجب أن يتصف بها تقدير الواقع المضاد، وتقديم أمثلة على التقديرات غير الصحيحة للواقع المضاد. ومن ثم، نستعرض قائمة من خيارات تقييم الأثر التي يمكن أن تقدم تقديرات صحيحة للواقع المضاد. وعلى وجه الخصوص،

نناقش المنطق البديهي الأساسي وراء أساليب تقييم الأثر الخمس: *التعيين العشوائي Randomized Assignment*، *والمتغيرات المساعدة Instrumental Variables*، *وتصميم انقطاع الانحدار Regression Discontinuity Design*، *والاختلاف في الاختلافات Difference-In-Differences*، *والمطابقة Matching*. ونناقش الأسباب التي تتيح لكل أسلوب أن يقدم تقديرًا صالحًا للواقع المضاد وكيف يُمكن القيام بذلك، حيث يمكن تنفيذ سياق السياسة، والقيود الرئيسية التي يواجهها كل أسلوب مستخدم في التقييم.

يتضمن في هذا الجزء من الكتاب دراسة حالة، برنامج معونة التأمين الصحي (HISP)، لبيان كيفية تطبيق هذه الأساليب. بالإضافة إلى ذلك، نقدم أمثلة محددة على تقييمات الأثر التي استخدمت كل أسلوب. ويُختتم الجزء الثاني بمناقشة كيفية دمج الأساليب ومعالجة المشكلات التي يمكن أن تنشأ خلال التنفيذ، مع إدراك أن تصميمات تقييم الأثر غالبًا ما لا تُنفذ على النحو المخطط له في الأصل. وفي هذا السياق، نراجع التحديات المشتركة التي تمت مواجهتها في أثناء التنفيذ، بما في ذلك الامتثال غير المثالي أو التأثيرات غير المباشرة، ونناقش كيفية معالجة هذه المشكلات. ويُختتم الفصل العاشر بإرشادات حول تقييمات البرامج متعددة الأوجه، وخصوصًا تلك التي تضم مستويات معالجة مختلفة وتصميمات متقاطعة.

يركز الجزء الثالث- كيفية تنفيذ تقييم الأثر (الفصل الحادي عشر إلى الرابع عشر) على كيفية تنفيذ تقييم الأثر، بدءًا من الفصل الحادي عشر مع كيفية استخدام قواعد تشغيل البرنامج، وتحديثًا الموارد المتاحة للبرنامج، ومعايير اختيار المستفيدين، وتوقيت التنفيذ، باعتبارها الأساس لاختيار أسلوب تقييم الأثر. وقد وُضع إطار عمل بسيط لتحديد أي من أساليب تقييم الأثر التي استُعرضت في الجزء الثاني الأنسب لبرنامج معين، اعتمادًا على قواعده التشغيلية. ويناقش الفصل الثاني عشر العلاقة بين فريق البحث وفريق السياسات وأدوارهما في تشكيل فريق التقييم على نحو مشترك. ونراجع الفرق بين الاستقلالية وعدم الانحياز، ونسلط الضوء على الجوانب التي قد تكون حساسة عند تنفيذ تقييم الأثر. ونقدم إرشادات حول كيفية إدارة التوقعات، ونبرز بعض المخاطر المشتركة التي ينطوي عليها إجراء تقييمات الأثر، ونقدم اقتراحات حول كيفية التعامل مع هذه المخاطر. ويُختتم الفصل بنظرة عامة على كيفية إدارة أنشطة تقييم الأثر، بما في ذلك تشكيل فريق التقييم، وتوقيت إجراء التقييم، وتحديد الميزانية، وجمع الأموال، وجمع البيانات. ويقدم الفصل الثالث عشر نظرة عامة على أخلاقيات وعلم تقييم الأثر، بما في ذلك أهمية عدم حرمان المستفيدين المؤهلين من الفوائد من أجل التقييم، ويوضح الفصل دور مجالس المراجعة المؤسسية التي تعتمد وترصد الأبحاث التي تشمل بشرًا، ويناقش أهمية تسجيل التقييمات بعد ممارسة العلم المفتوح؛ حيث تتم إتاحة البيانات للجمهور في سبيل إجراء المزيد من الأبحاث وتكرار النتائج. ويستعرض الفصل الرابع عشر رؤى حول كيفية استخدام تقييمات الأثر في توجيه السياسات، بما في ذلك نصائح حول كيفية جعل النتائج ذات صلة، ومناقشة أنواع المنتجات التي يمكن لتقييمات الأثر، بل وينبغي لها، أن تقدمها، وإرشادات حول كيفية استخلاص الاستنتاجات ونشرها لتعظيم أثر السياسة.

يناقش الجزء الرابع- كيفية الحصول على البيانات لإجراء تقييم الأثر (الفصل الخامس عشر إلى السابع عشر) كيفية جمع البيانات لإجراء تقييم الأثر، بما في ذلك اختيار العينة وتحديد الحجم المناسب لعينة التقييم (الفصل الخامس عشر)، بالإضافة إلى العثور على مصادر كافية للبيانات (الفصل السادس عشر). ويختتم الجزء بالفصل السابع عشر الذي يقدم بعض القوائم المرجعية.

المواد التكميلية عبر الإنترنت

توجد مواد مرفقة على موقع «تقييم الأثر في الواقع العملي» الإلكتروني (<http://www.worldbank.org/ieinpractice>)، بما في ذلك حلول لأسئلة دراسة حالة برنامج معونة التأمين الصحي (HISP) المذكورة في الكتاب، بالإضافة إلى مجموعة البيانات ذات الصلة ورمز التحليل في برنامج Stata الحاسوبي، علاوة على مُرافق فني يوفر معالجة أكثر رسمية لتحليل البيانات. وتشمل المواد أيضًا عروض PowerPoint ذات صلة بالفصول، ونسخة إلكترونية من الكتاب مع روابط إلى المواقع الإلكترونية، وروابط إلى مواد إضافية.

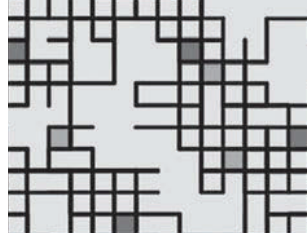
ويتضمن موقع "تقييم الأثر في الواقع العملي" الإلكتروني أيضًا روابط لمواد ذات صلة من المواقع الإلكترونية لصندوق تقييم الأثر الإستراتيجي (SIEF)، وتقييم الأثر الإنمائي (DIME)، ومجموعة أدوات تقييم الأثر التابعة للبنك الدولي، بالإضافة إلى بوابة تقييم الأثر التابعة لبنك التنمية للبلدان الأمريكية ودورة تدريبية عن أساليب تقييم الأثر المطبقة في جامعة كاليفورنيا، بيركلي.

تحديث كتاب تقييم الأثر في الواقع العملي

استندت الطبعة الأولى من كتاب تقييم الأثر في الواقع العملي إلى مجموعة أساسية من المواد التعليمية التي وضعت من أجل ورشة عمل «تحويل الوعود إلى أدلة» (Turning Promises to Evidence) التي نظمها مكتب كبير الخبراء الاقتصاديين للتنمية البشرية، بالشراكة مع وحدات إقليمية ومجموعة أبحاث اقتصاديات التنمية في البنك الدولي. في وقت كتابة الطبعة الأولى، كانت ورشة العمل قد قُدمت أكثر من ٢٠ مرة في جميع بقاع العالم. ولم تكن لورش العمل وكل من الطبعتين الأولى والثانية من هذا الدليل أن ترى النور إلا بفضل المنح السخية من الحكومة الإسبانية، ووزارة التنمية الدولية في المملكة المتحدة (DFID)، ومؤسسة صندوق الاستثمار من أجل الأطفال (CIFF UK)، ومن خلال الإسهامات المُقدمة إلى صندوق تقييم الأثر الاستراتيجي (SIEF).

وقد استفادت الطبعة الثانية أيضًا من الدعم المقدم من مكتب التخطيط الإستراتيجي وفعالية التنمية التابع لبنك التنمية للبلدان الأمريكية (IDB).

وتم تحديث هذه الطبعة الثانية لتغطية أحدث التقنيات المستخدمة في تقييم البرامج ونصائح حول أحدث سبل التنفيذ عقب التطورات التي شهدتها المجال في السنوات الأخيرة. وقمنا أيضًا بتوسيع نطاق مجموعة الأمثلة ودراسات الحالة بحيث تعكس التطبيقات واسعة المدى لتقييم الأثر في عمليات التنمية وللتأكيد على علاقتها بالسياسة. وأخيرًا، قمنا بتضمين تطبيقات تقنيات تقييم الأثر مع برنامج Stata، باستخدام مجموعة بيانات دراسة حالة برنامج HISP، كجزء من المواد التكميلية عبر الإنترنت.



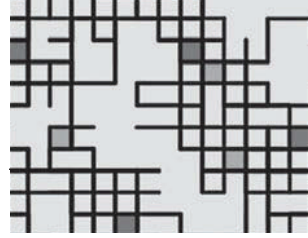
مرّت المواد التعليمية التي يستند إليها الكتاب بالعديد من التجسيّدات وتمّ تدريسها على يد باقة من أعضاء هيئة التدريس الموهوبين، الذين تركوا جميعًا بصماتهم على أساليب ومقاربات تقييم الأثر التي يتبناها الكتاب. ونودّ أن نتقدم بالشكر والتقدير إلى عدد من أعضاء هيئة التدريس الذين شاركوا في تدريس ورش العمل التي قامت عليها الطبعة الأولى من هذا الكتاب على إسهاماتهم ومشاركاتهم الجوهرية، بما في ذلك بالوما أسيفيدو ألاميدا، وفيليب باريرا، وسيرجيو باوتيسستا-أريدوندو، وستيفانو بيرتوزي، وباربارا برونز، وبيدرو كارنيرو، وجيشنو داس، وداميان دي والكي، وديفيد إيفانز، وكلاوديو فيراز، وديون فيلمر، وجيد فريدمان، وإيمانويلا جالاسو، وسيباستيان غالياني، وأريانا ليجوفيني، وفيليب لايت، وجونزالو هيرنانديز ليكونا، وماتياس لوندبيرج، وكارين ماكورس، وخوان مونوز، وبلادامين نيكولوف، وبيرك أوزلر، ونانسي تشيان، وجلوريا إم روبيو، ونوربيرت شادي، وجوليتا ترياس، وسيغريد فيغو جوزمان. ونحن ممتنون للتعليقات التي قدّمها زملاؤنا المراجعون على الطبعة الأولى من الكتاب (باربرا برونز، وأريانا ليجوفيني، ودان ليفي، وإيمانويل سكوفياس) والطبعة الثانية (ديفيد إيفانز، وفرانيسكو جاليجو، ودان ليفي، وداميان دي والكي). بالإضافة إلى جيليت هول. ونعرب أيضًا عن امتناننا وتقديرنا لجهود فريق تنظيم ورشة العمل الموهوب، بما في ذلك هولي بالجريف، وتيريزا أدوبيا بامبو، وفيبي مكي، وسيلفيا باروزولو، وتاتيانا رينجلاند، وآدم روس، وجينيفر ستوردي.

نتقدم بالشكر إلى جميع من شاركوا في كتابة نصوص ورشة العمل في يوليو ٢٠٠٩ في بكين، الصين، والتي تستند إليها أجزاء من هذا الكتاب، ونخص بالشكر بالوما أسيفيدو ألاميدا، وكارلوس أسينجو رويز، وسيباستيان باهوف، وبرادلي تشين، وتشانجشينج سونج، وجين زانج، وشوفانغ تشانغ. ونشكر غاريت كريستنسن ومبادرة بيركلي للشفافية في العلوم الاجتماعية، فضلًا عن جينيفر ستوردي، وإليسا روتنولر، على إسهاماتهم في الفصل الثالث عشر. ونحن ممتنون أيضًا لمارينا تولشينسكي، وكريستين كرونين على المساعدة البحثية الممتازة، وكامبيرون بريسلين، وريستيتوتو كارديناس على تحديد مواعيد الدعم، وماركو جوزمان، ومارتن روجينبيرج على تصميم الأشكال التوضيحية، ونانسي موريسون، وسيندي آيه فيشر، وفيونا ماكينتوش، وستيوارت كيه تاكر على دعم التحرير في أثناء إنتاج الطبعتين الأولى والثانية من الكتاب.

نعرب بكل امتنان عن تقديرنا للدعم المستمر والحماس الشديد لهذا المشروع الذي وجدناه من مديرينا في البنك الدولي وبنك التنمية للبلدان الأمريكية، وخاصة من فريق صندوق تقييم الأثر الإستراتيجي (SIEF)، بما في ذلك دافنا بيرمان، وهولي بلاجر، وريستيتوتو كارديناس، وجوست دي لات، وأرييل فيزين، وألاكا هول، وأليزا ماركوس، وديانا لوليانا بيرجول، وراشيل روزنفيلد، وجوليتا ترياس. ونحن ممتنون جدًا للدعم الذي تلقيناه من إدارة SIEF، بما في ذلك لويس بينغينيس، وجوست دي لات، وجوليتا ترياس. ونعرب عن امتناننا أيضًا لأندريه غوميز بينيا، وميكايلا فيزر من بنك التنمية للبلدان الأمريكية، وماري فيسك، وباتريشيا كاتاياما، ومايا ريفزينا من البنك الدولي لمساعدتهم في عمليات التواصل والنشر.

أخيرًا، نوّد أن نتقدم بالشكر للمشاركين في ورش العمل العديدة، ولا سيما تلك التي أُقيمت في أبيدجان، وأكرا، وأديس أبابا، وعمان، وأنقرة، وبكين، وبيركلي، وبوينس آيرس، والقاهرة، وكيب تاون، وكويرنافاكا، وداكار، ودكا، وفورتاليزا، وكاتماندو، وكيفالي، وليما، ومدريد، وماناغوا، ومانيلا، ومكسيكو سيتي، ونيودلهي، وبابايا، وبنما سيتي، وبريتوريا، وريو دي جانيرو، وسان سلفادور، وسانتياغو، وسراييفو، وسيول، وصوفيا، وتونس، والعاصمة واشنطن. لقد تمكّنّا بفضل اهتمامهم، وأسئلتهم الذكية، وحماسهم من تعلّم ما يبحث عنه صانعو السياسات في تقييمات الأثر خطوة بخطوة. نأمل أن يعكس هذا الكتاب أفكارهم.

نبذة عن المؤلفين



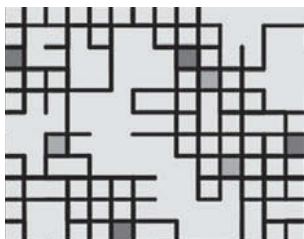
بول جيه جيرتزر هو أستاذ الاقتصاد في جامعة كاليفورنيا بيركلي، حيث يشغل منصبين في كلية هاس لإدارة الأعمال وكلية الصحة العامة، وقد تخرّج في كلية لي كا شينج. فضلاً عن أنه المدير العلمي لمركز العمل العالمي الفعال التابع لجامعة كاليفورنيا. وعمل كبير الخبراء الاقتصاديين لشبكة التنمية البشرية التابعة للبنك الدولي في الفترة بين ٢٠٠٤ و٢٠٠٧ والرئيس المؤسس لمجلس إدارة المبادرة الدولية لتقييم الأثر (3ie) من ٢٠٠٩ إلى ٢٠١٢. في البنك الدولي، قاد الجهود المبذولة لإضفاء الطابع المؤسسي على تقييم الأثر وتوسيع نطاقه لمعرفة ما يناسب مجال التنمية البشرية. وكان الباحث الرئيسي في عدد كبير من تقييمات الأثر متعددة المواقع وواسعة النطاق، بما في ذلك برنامج التحويلات النقدية المشروطة (CCT) في المكسيك، PROGRESA/OPORTUNIDADES، ونظام الرعاية الصحية مقابل الأداء في رواندا. وقد حصل على درجة الدكتوراه في الاقتصاد من جامعة ويسكونسن، وشغل مناصب أكاديمية في جامعة هارفارد، ومؤسسة راند، وجامعة ولاية نيويورك في ستوني بروك.

سيباستيان مارتينيز هو الخبير الاقتصادي الرئيسي في مكتب التخطيط الإستراتيجي وفعالية التنمية التابع لبنك التنمية للبلدان الأمريكية (IDB). ويرتكز عمله على تعزيز قاعدة الأدلة وفعالية التنمية في القطاعات الاجتماعية وقطاعات البنية التحتية، بما في ذلك الصحة، والحماية الاجتماعية، وأسواق العمل، والمياه والصرف الصحي، والإسكان والتنمية الحضرية. ويتأرس فريقاً من الاقتصاديين الذي يجري أبحاثاً حول تأثيرات برامج وسياسات التنمية، ويدعم تنفيذ تقييمات الأثر للعمليات التشغيلية، ويضطلع بمهمة تطوير القدرات للعملاء والموظفين. وقبل الانضمام إلى بنك التنمية للبلدان الأمريكية، قضى ست سنوات في البنك الدولي، حيث قاد تقييمات البرامج الاجتماعية في أمريكا اللاتينية وإفريقيا جنوب الصحراء الكبرى. وحصل على درجة الدكتوراه في الاقتصاد من جامعة كاليفورنيا في بيركلي، وتخصص في التنمية والاقتصاد الجزئي التطبيقي.

باتريك بريماند هو كبير الاقتصاديين في الممارسات العالمية للحماية الاجتماعية والعمل في البنك الدولي. ويجري أعمالاً تحليلية وتشغيلية حول الحماية الاجتماعية وشبكات الأمان، وأسواق العمل، وتوظيف الشباب وريادة الأعمال، بالإضافة إلى تنمية الطفولة المبكرة. وتركز أبحاثه على جمع الأدلة عن فعالية سياسات التنمية من خلال تقييمات الأثر للبرامج الاجتماعية وبرامج التنمية البشرية واسعة النطاق. وقد شغل سابقاً العديد من المناصب الأخرى في البنك الدولي، بما في ذلك في وحدة اقتصاديات التنمية البشرية واسعة النطاق. ومكتب كبير الخبراء الاقتصاديين للتنمية البشرية، ووحدة مكافحة الفقر في أمريكا اللاتينية ومنطقة البحر الكاريبي. وحصل على درجة الدكتوراه في الاقتصاد من جامعة أكسفورد.


لورا بي رولينجز هي كبيرة أخصائيي الحماية الاجتماعية في البنك الدولي، وتمتلك خبرة تزيد عن ٢٠ عامًا في تصميم برامج التنمية البشرية، وتنفيذها، وتقييمها. وتتولى إدارة كل من العمليات التشغيلية والأبحاث، مع التركيز على تطوير مقاربات مبتكرة لأنظمة الحماية الاجتماعية الفعالة والقابلة للتوسع في البيئات ذات الموارد المنخفضة. وكانت قائدة الفريق المسؤول عن تطوير إستراتيجية البنك الدولي للحماية الاجتماعية والعمل في الفترة بين ٢٠١٢ و٢٠٢٢، وكانت سابقاً مدير صندوق تقييم الأثر الإستراتيجي (SIEF). وشغلت أيضاً منصب قائدة القطاع للتنمية البشرية في أمريكا الوسطى، حيث كانت مسؤولة عن إدارة محافظ البنك الدولي للصحة، والتعليم، والحماية الاجتماعية. وبدأت مسيرتها المهنية في البنك الدولي في مجموعة أبحاث التنمية، حيث عملت على إجراء تقييم الأثر للبرامج الاجتماعية. وقد عملت في أمريكا اللاتينية ومنطقة البحر الكاريبي، بالإضافة إلى إفريقيا جنوب الصحراء الكبرى، وهناك قادت العديد من المشاريع ومبادرات البحوث في مجالات التحويلات النقدية المشروطة، والأشغال العامة، والصناديق الاجتماعية، وتنمية الطفولة المبكرة، وأنظمة الحماية الاجتماعية. وقبل الانضمام إلى البنك الدولي، عملت لصالح مجلس تنمية ما وراء البحار، حيث تولت إدارة برنامج تعليمي حول قضايا التنمية للموظفين العاملين في الكونغرس الأمريكي. ونشرت العديد من الكتب والمقالات في مجالات التقييم والتنمية البشرية، وهي أستاذة مساعدة في برنامج التنمية البشرية العالمي في جامعة جورج تاون، في واشنطن العاصمة.

كريستل إم جيه فيرميرش هي كبيرة الاقتصاديين في الممارسات العالمية للصحة، والتغذية، والسكان في البنك الدولي. وتتخصص في القضايا المتعلقة بتمويل قطاع الصحة، والتمويل القائم على النتائج، والمتابعة والتقييم، وتقييم الأثر. وقد عملت سابقًا في مجالات التعليم، وتنمية الطفولة المبكرة، والمهارات. وشاركت في تأليف دراسات تقييم الأثر لبرامج التمويل القائم على النتائج في الأرجنتين ورواندا، ومتابعة طويلة المدى لدراسة تحفيز الطفولة المبكرة في جامايكا، بالإضافة إلى مجموعة أدوات البنك الدولي لتقييم الأثر في قطاع الصحة. وقبل الانضمام إلى البنك الدولي، كانت الزميلة المختارة لأبحاث ما بعد الدكتوراه في جامعة أكسفورد. وحصلت على درجة الدكتوراه في الاقتصاد من جامعة هارفارد.

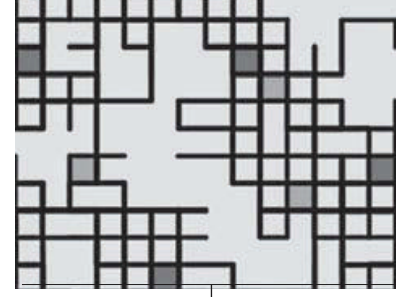


الاختصارات

المبادرة الدولية لتقييم الأثر	3IE
متوسط تأثير المعالجة	ATE
التحويل النقدي المشروط	CCT
مبادرة التدريب المؤسسي التعاوني	CITI
الاختلاف في الاختلافات أو الاختلافات المزدوجة	DD
تقييم الأثر الإنمائي (البنك الدولي)	DIME
برنامج معونة التأمين الصحي	HISP
رقم التعريف	ID
بنك التنمية للبلدان الأمريكية	IDB
الشبكة الدولية لمسوحات الأسر المعيشية	IHSN
مجلس المراجعة المؤسسية	IRB
النية للعلاج	ITT
المتغيرات المساعدة	IV
معمل عبد اللطيف جميل لمكافحة الفقر	J-PAL
متوسط تأثير المعالجة المكاني	LATE
الحد الأدنى للتأثيرات القابلة للاكتشاف	MDE
منظمة غير حكومية	NGO
المعاهد الوطنية للصحة (الولايات المتحدة)	NIH
معهد التنمية لما وراء البحار	ODI
إطار عمل العلوم المفتوحة	OSF
تجربة عشوائية مضبوطة	RCT
تصميم انقطاع الانحدار	RDD
سجل تقييمات الآثار الإنمائية على الصعيد الدولي	RIDIE



صندوق تقييم الأثر الإستراتيجي (البنك الدولي)	SIEF
محددة، وقابلة للقياس، ومنسوبة، وواقعية، ومستهدفة	SMART
افتراض قيمة معالجة الوحدة المستقرة	SUTVA
تأثير المعالجة على المعالجين	TOT
الأمم المتحدة	UN
الوكالة الأمريكية للتنمية الدولية	USAID
منظمة الصحة العالمية	WHO



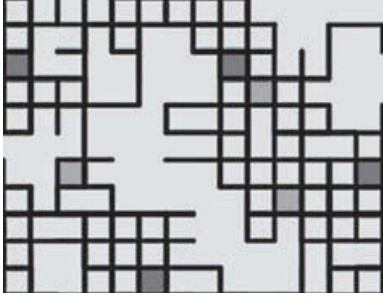
الجزء الأول

مقدمة إلى تقييم الأثر

يقدم الجزء الأول من الكتاب نظرة عامة على تقييم الأثر. ويناقش الفصل الأول سبب أهمية تقييم الأثر وكيف يتلاءم مع سياق صنع السياسات الأخلاقية القائمة على الأدلة. ونعقد مقارنة بين تقييم الأثر والمتابعة، ونقدم السمات المميزة لتقييم الأثر، ونناقش المقاربات التكميلية، بما في ذلك تحليل التكلفة والمنفعة وتحليل فعالية التكلفة. ونطرح أحد المحاور الرئيسية للكتاب؛ ألا وهو كيف تسهم الموارد المتاحة للبرنامج، ومعايير الأهلية لاختيار المستفيدين، وتوقيت التنفيذ في وضع الخيارات التي تُنتقى منها أساليب تقييم الأثر. وأخيرًا، نقدم طرقًا مختلفة لتقييم الأثر، مثل التقييم السابق والتقييم اللاحق، وتجارب النجاعة مقابل الفعالية، ونختتم الفصل بنقاش حول الوقت المناسب لاستخدام تقييمات الأثر.



يُناقش الفصل الثاني كيفية صياغة أسئلة التقييم والفرضيات المفيدة للسياسات. وتحدد هذه الأسئلة والفرضيات محور تركيز التقييم. ونقدم أيضًا المفهوم الأساسي لنظرية التغيير والاستخدام ذي الصلة لسلاسل النتائج ومؤشرات الأداء. ويقدم الفصل الثاني المقدمة الأولى إلى دراسة الحالة الخيالية، برنامج معونة التأمين الصحي (HISP)، المستخدمة في جميع أجزاء الكتاب وفي المواد المرفقة الموجودة على موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice).



ما فائدة التقييم؟

صنع السياسات القائمة على الأدلة

يهدف تصميم برامج وسياسات التنمية في العادة إلى تغيير الحوائل، مثل زيادة الدخل، أو تحسين مستوى التعليم، أو الحد من الأمراض. وتظل مسألة ما إذا تحققت هذه التغييرات بالفعل أم لا تساؤلًا مهمًا مطروحًا في السياسة العامة، غير أنه يتعرض للتجاهل في كثير من الأحيان. في حين أن الأمر الأكثر شيوعًا هو تركيز مديري البرامج وصانعي السياسات على قياس مدخلات البرنامج ومخرجاته الآنية ورفع تقارير عنها—مقدار الأموال التي يتم إنفاقها، وعدد الكتب المدرسية التي يتم توزيعها، وعدد الأشخاص المشاركين في برنامج التوظيف—عوضًا عن تقييم ما إذا نجحت البرامج في تحقيق الغايات المرجوة منها لتحسين الحوائل أم لا.

تمثل تقييمات الأثر جزءًا من مخطط أوسع نطاقًا لصنع سياسات قائمة على الأدلة. ويتسم هذا الاتجاه العالمي المتنامي بتحوّل التركيز من المدخلات إلى الحوائل والنتائج، ويعتمد إلى إعادة تشكيل السياسة العامة. ولا يقتصر التركيز على استخدام النتائج في وضع الأهداف الوطنية والدولية وتتبعها، وإنما يزداد استخدام مديري المشاريع لها، ومطالبتهم بها، لتعزيز المساءلة، وتحديد مخصصات الميزانية، وتوجيه تصميم البرنامج والقرارات المتعلقة بالسياسة. تعد المتابعة والتقييم جزءًا لا يتجزأ من عملية صنع السياسات القائمة على الأدلة. ويوفران مجموعة أساسية من الأدوات التي تستطيع الأطراف المعنية استخدامها للتحقق من جودة السياسات والبرامج، وكفاءتها، وفعاليتها، والعمل على تحسينها، في مراحل مختلفة من عملية التنفيذ، أو بعبارة أخرى، التركيز على النتائج. وعلى مستوى إدارة البرنامج، هناك حاجة إلى فهم خيارات تصميم البرنامج الأكثر فعالية من حيث التكلفة، أو تقديم البراهين لصانعي القرارات التي تفيد بأن البرامج تحقق نتائجها المرجوة من أجل الحصول على مخصصات الميزانية في سبيل مواصلة تنفيذ البرامج أو التوسع فيها. وعلى المستوى القطري، تتنافس الوزارات مع بعضها البعض للحصول على تمويل من وزارة المالية. وأخيرًا، تتحمل الحكومات المسؤولية أمام المواطنين لإطلاعهم على أداء البرامج العامة. ويمكن أن تشكل الأدلة أساسًا قويًا للشفافية والمساءلة.



يتزايد استخدام الأدلة القوية الناجمة عن تقييمات الأثر كأساس لمزيد من المساءلة، والابتكار، والتعلّم. وفي ظل ظروف يطالب فيها صانعو السياسات والمجتمع المدني بنتائج ومساءلة من البرامج العامة، يمكن أن يوفر تقييم الأثر أدلة قوية وموثوقة على الأداء، والأهم، حول ما إذا كان برنامج معين قد حقق الحاصلات المرجوة منه أم ما زال يحققها. وعلاوة على ذلك، تُستخدم تقييمات الأثر بكثرة للتحقق من الابتكارات في تصميم البرنامج أو تقديم الخدمة. وعلى مستوى العالم، فإن تقييمات الأثر محورية في بناء المعرفة حول مدى فعالية برامج التنمية من خلال تسليط الضوء على ما ينجح وما لا ينجح في الحد من الفقر وتحسين الرفاهية.

ببساطة، يعتمد تقييم الأثر إلى تقدير التغييرات التي تطرأ على رفاهية الأفراد التي يمكن أن تُسند إلى مشروع، أو برنامج، أو سياسة معينة. وبمثل هذا التركيز على الإسهام السمة المميزة لتقييمات الأثر. وفي المقابل، يكمن التحدي الرئيسي أمام إجراء تقييمات الأثر الفعالة في تحديد العلاقة السببية بين البرنامج أو السياسة والحاصلات محل الاهتمام. وبشكل عام، تعمل تقييمات الأثر على تقدير متوسط آثار البرنامج، أو أشكال البرنامج، أو ابتكار التصميم. على سبيل المثال، هل نجح برنامج المياه والصرف الصحي في زيادة إمكانية الوصول إلى مياه صالحة للشرب وتحسين الحاصلات الصحية؟ وهل أدى المنهج الدراسي الجديد إلى زيادة درجات الاختبار بين الطلاب؟ هل نجح ابتكار تضمين المهارات غير المعرفية كجزء من برنامج تدريب الشباب في تعزيز ريادة الأعمال وزيادة الدخل؟ في كل واحدة من هذه الحالات، يقدم تقييم الأثر معلومات عما إذا كان البرنامج قد أسفر عن التغييرات المرجوة في الحاصلات، على نقيض دراسات حالة أو حكايات محددة، والتي يمكن أن تقدم معلومات جزئية فقط ولا تمثل إجمالي آثار البرنامج. ومن هذا المنطلق، فإن تقييمات الأثر جيدة التصميم والتنفيذ قادرة على تقديم أدلة مقنعة وشاملة يمكن استخدامها في توجيه قرارات السياسة، وتشكيل الرأي العام، وتحسين العمليات التشغيلية للبرنامج.

تبحث تقييمات الأثر التقليدية مدى فعالية البرنامج مقابل غيابه. ويتناول المربع ١-١ تقييم الأثر المعروف لبرنامج التحويلات النقدية المشروطة (CCT) في المكسيك، إذ يبين كيف أسهم التقييم في مناقشات السياسات المتعلقة بتوسيع نطاق البرنامج.



المربع ١-١: كيف يمكن لتقييم ناجح أن يعزز الاستدامة السياسية لبرنامج إنمائي: برنامج التحويلات النقدية المشروطة في المكسيك

بمتوسط ٠,٧ سنوات دراسة إضافية. وتوصل جيرتler (Gertler, ٢٠٠٤) إلى أن معدلات الإصابة بالمرض بين الأطفال انخفضت بمعدل ٢٣ في المائة، في حين انخفض عدد أيام الإجازات المرضية أو العجز عن العمل بمعدل ١٩ في المائة بين البالغين. ومن حيث حصائل التغذية، وجد بيرمان وهودينوت (Behrman and Hoddinott, ٢٠٠١) أن البرنامج تمكّن من تقليل احتمال التقزم بنحو سنتيمتر واحد في العام لدى الأطفال في الفئة العمرية الحرجة من ١٢ إلى ٣٦ شهرًا.

أسفرت نتائج التقييم هذه عن دعم إقامة حوار سياسي قائم على الأدلة، وأسهمت في اتخاذ الإدارة الجديدة قرار مواصلة البرنامج. وعمدت الحكومة إلى توسيع نطاق البرنامج؛ حيث قدمت منحة دراسية للمدارس العليا والإعدادية، وقامت بتحسين البرامج الصحية للمراهقين. في الوقت ذاته، استُخدمت النتائج لتعديل برامج المساعدة الاجتماعية الأخرى، مثل برنامج معونة التورتيللا الضخم والأقل جودة في تحديد الأهداف، والذي تم تقليصه.

أسهم التقييم الناجح لبرنامج Progresa أيضًا في اعتماد برنامج التحويلات النقدية المشروطة بسرعة في جميع أنحاء العالم، فضلًا عن تبني المكسيك لتشريع يطالب بخضوع جميع المشاريع الاجتماعية للتقييم.

أطلقت الحكومة المكسيكية في تسعينيات القرن العشرين برنامج التحويلات النقدية المشروطة (CCT) المبتكر، والذي عُرف في البداية باسم برنامج Progresa (تم تغيير الاسم، مع بعض عناصر البرنامج، إلى Oportunidades، ثم Prospera). وكانت أهداف البرنامج تزويد الأسر الفقيرة بدعم الدخل قصير المدى وخلق حوافز للاستثمارات في رأس المال البشري للأطفال، في المقام الأول من خلال تقديم التحويلات النقدية إلى الأمهات في الأسر الفقيرة بشرط ذهاب أطفالهن بانتظام إلى المدرسة وزيارة المراكز الصحية.

من البداية، اعتبرت الحكومة أنه من الضروري متابعة البرنامج وتقييمه. ومن ثم، تعاقب مسؤولو البرنامج مع مجموعة من الباحثين لتصميم تقييم الأثر ودمجه في عملية توسيع البرنامج في الوقت نفسه الذي تم تقديمه فيه بصورة متتابعة في المجتمعات المشاركة.

أسفرت الانتخابات الرئاسية عام ٢٠٠٠ عن تغيير الحزب الحاكم. وفي عام ٢٠٠١، قدّم المقيمون الخارجيون لبرنامج Progresa استنتاجاتهم إلى الإدارة المُنتخبة حديثًا. وجاءت نتائج البرنامج مبهرة؛ حيث أظهرت أن البرنامج استهدف الفقراء بشكل جيد وأحدث تغييرات واعدة في رأس المال البشري للأسر. وجد شولتز (Schultz, ٢٠٠٤) أن البرنامج أدى بشكل ملحوظ إلى تحسين معدلات التسجيل في المدارس،

المصادر: Behrman and Hoddinott 2001; Fiszbein and Schady 2009; Gertler 2004; Levy and Rodríguez 2005; Schultz 2004; Skoufias and McClafferty 2001.

يوضح المربع ٢-١ كيف يؤثر تقييم الأثر على سياسة التعليم في موزمبيق عن طريق إثبات أن مؤسسات رياض الأطفال المجتمعية يمكن أن تكون وسيلة ميسورة التكلفة وفعالة لمعالجة مشكلات التعليم المبكر وتشجيع الأطفال على التسجيل في المدرسة الابتدائية في السن المناسب.

علاوة على الإجابة عن السؤال الأساسي حول ما إذا كان البرنامج فعالاً أم لا، يمكن أيضاً استخدام تقييمات الأثر لاختبار أشكال البرنامج البديلة أو ابتكارات التصميم بوضوح. ومع تزايد تركيز صانعي السياسات على اكتساب فهم أفضل لكيفية تحسين عملية التنفيذ والحصول على قيمة مقابل المال، تكتسب مقاربات اختبار بدائل التصميم مكانة لها على أرض الواقع سريعاً. على سبيل المثال، يمكن أن يقارن التقييم بين أداء برنامج تدريبي وحملة ترويجية لمعرفة

المربع ٢-١: أثر سياسة نموذج مبتكر لمرحلة رياض الأطفال: مرحلة رياض الأطفال وتنمية الطفولة المبكرة في موزمبيق

كما زادت احتمالية بدئهم الدراسة في السن المناسب بنسبة ١٠ في المائة عن الأطفال الآخرين في مجموعة المقارنة. في المدرسة الابتدائية، قضى الأطفال الذين التحقوا برياض الأطفال ما يقرب من ٥٠ في المائة وقتاً إضافياً في أداء الفروض المنزلية وغيرها من الأنشطة المدرسية عن أولئك الذين لم يفعلوا ذلك. وأظهر التقييم أيضاً مكاسب إيجابية في الاستعداد للمدرسة؛ حيث كان أداء الأطفال الذين التحقوا برياض الأطفال أفضل في اختبارات النمو المعرفي، والاجتماعي العاطفي، والمهارات الحركية الدقيقة بالمقارنة مع مجموعة المقارنة.

علاوة على ذلك، استفاد أفراد الأسرة الآخرون من تسجيل الأطفال في رياض الأطفال؛ إذ توفر لديهم المزيد من الوقت للمشاركة في أنشطة مثمرة. ومن الأرجح أن يزداد معدل التحاق الأشقاء الأكبر سناً بالمدرسة بنسبة ٦ في المائة، وزاد احتمال تمكن مقدمي الرعاية من العمل بمعدل ٢٦ في المائة في الثلاثين يوماً الماضية عند التحاق طفل صغير في الأسرة برياض الأطفال.

وكشف هذا التقييم عن أنه حتى في بيئة الدخل المنخفض، يمكن أن تكون مؤسسات رياض الأطفال وسيلة فعالة لتعزيز النمو المعرفي، وتأهيل الأطفال للمدرسة الابتدائية، وزيادة احتمالية أن يبدأ الأطفال المدرسة الابتدائية في السن المناسب.

بينما يُنظر إلى رياض الأطفال على أنها استثمار جيد ومقاربة فعال لتأهيل الأطفال للمدرسة وحياتهم القادمة، عانت الدول النامية مع التساؤل المتعلق بكيفية تقديم نموذج قابل للتوسع وفعال من حيث التكلفة لمرحلة رياض الأطفال. في موزمبيق، يرتاد حوالي ٤ في المائة فقط من الأطفال مؤسسات رياض الأطفال. وعند بلوغ سن المدرسة الابتدائية، تظهر على بعض الأطفال من المجتمعات الريفية علامات تأخر النمو، وغالباً ما يكونون غير مستعدين لمتطلبات نظام التعليم. علاوة على ذلك، على الرغم من أن معدل التسجيل في المدارس الابتدائية يقارب ٩٥٪ في المائة، لم يتم تسجيل ثلث الأطفال في السن المناسب.

في عام ٢٠٠٦، أطلقت منظمة "إنقاذ الطفل" (Save the Children) برنامج رياض أطفال مجتمعية تجريبي في المجتمعات الريفية في موزمبيق يهدف إلى تحسين النمو المعرفي، والاجتماعي، والعاطفي، والبدني لدى الأطفال. وفيما يُعتقد أنه التقييم العشوائي الأول لبرنامج رياض الأطفال في المناطق الريفية في إفريقيا، أجرى فريق بحثي تقييم أثر للبرنامج في عام ٢٠٠٨. بناءً على النتائج الإيجابية للتقييم، اعتمدت حكومة موزمبيق وقررت توسيع نطاق تطبيق نموذج رياض الأطفال المجتمعية من منظمة "إنقاذ الطفل" في ٦٠٠ مجتمع محلي.

وجد التقييم أن الأطفال الذين التحقوا برياض الأطفال زادت احتمالية تسجيلهم في المدرسة الابتدائية بنسبة ٢٤ في المائة،

المصدر: Martinez, Nadeau, and Pereira 2012.

أيهما أكثر فعالية في زيادة محو الأمية المالية. ويمكن أن يختبر تقييم الأثر مزيّجًا من مقاربات التغذية وتحفيز الطفل الذي له أكبر أثر على نمو الطفل. أو قد يختبر التقييم ابتكار التصميم لتحسين برنامج قائم، مثل استخدام الرسائل النصية للحض على تناول الادوية الموصوفة.

ما المقصود بتقييم الأثر؟

تقييم الأثر هو أحد المقاربات العديدة التي تدعم السياسات القائمة على الأدلة، بما في ذلك المتابعة وأنواع التقييم الأخرى.

المتابعة هي عملية مستمرة تستهدف تتبع ما يحدث داخل البرنامج وتستخدم المعلومات التي تم جمعها في توجيه تنفيذ البرنامج والإدارة والقرارات اليومية. وباستخدام بيانات إدارية في الغالب، تتبع عملية المتابعة الإنفاق المالي وأداء البرنامج مقابل النتائج المتوقعة، وتحلل الاتجاهات مع مرور الوقت.^١ ولا غنى عن المتابعة في جميع البرامج، والتي تعد مصدرًا مهمًا للمعلومات عن أداء البرنامج، بما في ذلك التنفيذ والتكاليف. في العادة، تتبع المتابعة المدخلات، والأنشطة، والمخرجات، على الرغم من أنها يمكن أن تشمل أحيانًا حصائل، مثل التقدم نحو تحقيق غايات التنمية الوطنية.

المفهوم الأساسي

التقييمات هي تقويمات دورية وموضوعية لمشروع، أو برنامج، أو سياسة مخطط لها، أو قائمة، أو مكملة. وتستخدم التقييمات بشكل انتقائي للإجابة عن أسئلة محددة ذات صلة بالتصميم، والتنفيذ، والنتائج. وعلى خلاف المتابعة المستمرة، تُنفذ التقييمات في أوقات زمنية منفصلة، وغالبًا ما تلتزم الحصول على منظور خارجي من الخبراء الفنيين. ويختلف تصميمها، وأسلوبها، وتكلفتها اختلافًا كبيرًا بناءً على نوع السؤال الذي يحاول التقييم الإجابة عنه. ومجمل القول، يمكن أن تتناول التقييمات ثلاثة أنواع من الأسئلة (Imas and Rist 2009):^٣

• *الأسئلة الوصفية* التي تسأل عما يحدث. وتتعلق بالعمليات، والظروف، والعلاقات التنظيمية، وآراء الأطراف المعنية.

• *الأسئلة المعيارية* التي تقارن بين ما يحدث وما ينبغي أن يحدث. وتعتمد إلى تقييم الأنشطة وما إذا تحققت الأهداف أم لا. ويمكن أن تنطبق الأسئلة المعيارية على المدخلات، والأنشطة، والمخرجات.

• *أسئلة السبب والنتيجة* التي تركز على الإسناد. وتسأل عن الاختلاف الذي طرأ على الحصائل بسبب التدخل.

توجد العديد من أنواع التقييمات وأساليب التقييم التي تعتمد على كل من البيانات الكمية والنوعية. ولا يتم التعبير عن البيانات النوعية بالأرقام، وإنما بوسائل اللغة أو الصور في بعض الأحيان. في حين أن البيانات الكمية هي قياسات عددية، وترتبط عادةً بالمقاييس أو القياسات المترية. ويمكن استخدام كل من البيانات الكمية والنوعية للإجابة عن أنواع الأسئلة المطروحة أعلاه. في الواقع العملي، تعتمد العديد من التقييمات على كلا نوعي البيانات. وهناك العديد من مصادر البيانات التي يمكن استخدامها في التقييمات، والتي تستند إلى البيانات الأولية التي تم جمعها لغرض التقييم أو البيانات الثانوية المتاحة (انظر الفصل السادس عشر عن مصادر البيانات). ويركز هذا الكتاب على تقييمات الأثر التي تستخدم البيانات الكمية، ولكنه يشدد على قيمة المتابعة، وأساليب التقييم التكميلية، واستخدام كل من البيانات الكمية والنوعية.

المفهوم الأساسي

تسعى تقييمات الأثر إلى الإجابة عن نوع واحد محدد من الأسئلة: ما الأثر أو المفعول السببي للبرنامج على الحصيلة محل الاهتمام؟

تقييمات الأثر هي نوع معين من التقييم الذي يسعى إلى الإجابة عن سؤال محدد عن السبب والنتيجة: ما الأثر أو المفعول السببي للبرنامج على الحصيلة محل الاهتمام؟ ينطوي هذا السؤال الأساسي على بُعد سببي مهم. وينصب التركيز على الأثر فقط: أي التغييرات التي يمكن إسنادها مباشرةً إلى أحد البرامج، أو طريقة البرنامج، أو ابتكار التصميم.

يمكن تطبيق سؤال التقييم الأساسي—ما الأثر أو المفعول السببي للبرنامج على الحصيلة محل الاهتمام؟—على العديد من السياقات. على سبيل المثال، ما المفعول السببي للمنح الدراسية على الحضور المدرسي والإنجاز الأكاديمي؟ وما أثر التعاقد مع مقدمي خدمات

الرعاية الأولية من القطاع الخاص على تسهيل الوصول إلى الرعاية الصحية؟ وإذا تم استبدال الأراضي الترابية بأرضيات إسمنتية، فما أثر ذلك على صحة الأطفال؟ وهل تعمل الطرق المحسنة على زيادة الوصول إلى أسواق العمل وتزويد دخل الأسر، وإذا كان الأمر كذلك، فما مقدار ذلك؟ هل يؤثر حجم الفصل على إنجاز الطالب، وإذا كان يؤثر، فبأي درجة؟ كما يتضح من هذه الأمثلة، يمكن توسيع نطاق سؤال التقييم الأساسي لفحص أثر طريقة البرنامج أو ابتكار التصميم، وليس البرنامج فحسب.

يمثل التركيز على السببية والإسناد السمة المميزة لتقييمات الأثر. وتتناول جميع أساليب تقييم **الأثر** أحد أشكال سؤال السبب والنتيجة. بينما تحدد مقارنة تناول السببية الأساليب التي يمكن استخدامها. وحتى تتمكن من تقدير المفعول السببي للبرنامج أو أثره على الحاصل، يجب أن يقدّر أي أسلوب مختار لتقييم الأثر ما يدعى الواقع المضاد: أي، ما الحصيلة التي كانت ستتحقق للمشاركين في البرنامج إذا لم يشاركوا فيه. في الواقع العملي، يتطلب تقييم الأثر أن يبحث فريق التقييم عن مجموعة مقارنة لتقدير ما كان سيحدث للمشاركين في البرنامج من دونه، ثم عقد مقارنات مع مجموعة المعالجة التي استفادت من البرنامج. ويصف الجزء الثاني من الكتاب الأساليب الرئيسية التي يمكن استخدامها للعثور على مجموعات مقارنة مناسبة.

المفهوم الأساسي

يعتمد اختيار أسلوب
تقييم الأثر على الخصائص
التشغيلية للبرنامج قيد
التقييم، ولا سيما الموارد
المتاحة، ومعايير الأهلية
لاختيار المستفيدين،
وتوقيت تنفيذ البرنامج.

تتمثل إحدى الرسائل الرئيسية لهذا الكتاب في اختيار أسلوب تقييم أثر قائم على الخصائص التشغيلية للبرنامج قيد التقييم. وعندما تكون قواعد تشغيل البرنامج عادلة وشفافة وتكفل المساواة، يمكن دائمًا تقريبًا إيجاد تصميم جيد لتقييم الأثر، شريطة أن يُخطط لتقييم الأثر في وقت مبكر من عملية تصميم البرنامج أو تنفيذه. ولا يشكل وجود قواعد واضحة ومحددة بشكل جيد لعمليات تشغيل البرنامج قيمة جوهرية للسياسة العامة الرشيدة وإدارة البرنامج فحسب، بل إنها ضرورية أيضًا لتشكيل مجموعات مقارنة جيدة، التي تمثل أساس تقييمات الأثر شديدة الدقة. وعلى وجه التحديد، يتم اختيار أسلوب تقييم الأثر بناءً على الخصائص التشغيلية

للبرنامج، ولا سيما الموارد المتاحة ومعايير الأهلية في اختيار المستفيدين و توقيت تنفيذ البرنامج. وكما سنناقش في الجزأين الثاني والثالث من الكتاب، يمكنك طرح ثلاثة أسئلة عن السياق التشغيلي لبرنامج معين: هل يتضمن برنامجك موارد لخدمة جميع المستفيدين المؤهلين؟ هل برنامجك موجّه أم عام؟ وهل سيُطرح برنامجك لجميع المستفيدين دفعة واحدة أم بالتسلسل؟ ستحدد الإجابة عن هذه الأسئلة الثلاثة أي من الأساليب المقدمة في الجزء الثاني الخيار الأنسب لسياقك التشغيلي: التعيين العشوائي، أو المتغيرات المساعدة، أو انقطاع الانحدار، أو الاختلاف في الاختلافات، أو المطابقة.

تقييم الأثر السابق مقابل اللاحق

المفهوم الأساسي

تُصمّم التقييمات السابقة
وتُطبّق قبل تنفيذ البرنامج.

يمكن تقسيم تقييمات الأثر إلى فئتين: سابقة ولاحقة. توضع التقييمات السابقة في الوقت نفسه الذي يتم فيه تصميم البرنامج وتُدمج في عملية تنفيذ البرنامج. ويتم جمع بيانات خط الأساس قبل تنفيذ البرنامج عن كل من المجموعة التي تتلقى التدخل (تُعرف باسم مجموعة المعالجة) والمجموعة المُستخدمة في المقارنة التي لا تتلقى التدخل (تُعرف باسم مجموعة المقارنة). تعمل التقييمات اللاحقة على تقييم أثر البرنامج بعد تنفيذه، وتدرس مجموعات المعالجة والمقارنة بعد عملية التدخل.

من المرجح أن تؤدي تقييمات الأثر السابقة إلى نتائج تقييم قوية وذات مصداقية لثلاثة أسباب. أولاً، يمكن جمع بيانات خط الأساس لوضع مقاييس للحصائل محل الاهتمام قبل بدء البرنامج. وتتسم بيانات خط الأساس بأهميتها في قياس الحصائل قبل التدخل. وينبغي تحليل بيانات خط الأساس عن مجموعات المعالجة والمقارنة للتأكد من وجود تشابه بين المجموعات. فضلاً عن أنه يمكن استخدام بيانات خط الأساس لتقييم فعالية الاستهداف؛ بمعنى معرفة ما إذا كان البرنامج قادراً على الوصول إلى المستفيدين المستهدفين أم لا.

ثانيًا، يؤدي تحديد مقاييس نجاح البرنامج في مرحلة التخطيط للبرنامج إلى تركيز اهتمام كل من البرنامج والتقييم على النتائج المرجوة. ومثلما سنرى، تترسخ تقييمات الأثر في نظرية التغيير للبرنامج أو سلسلة النتائج. ويساعد تصميم تقييم الأثر على إبراز أهداف البرنامج، ولا سيما لأنها تتطلب وضع مقاييس محددة بدقة لنجاح البرنامج. ويجب على صانعي السياسات وضع غايات واضحة يحققها البرنامج، وأسئلة واضحة يجب عنها التقييم، وذلك لضمان أن تكون النتائج وثيقة الصلة بالسياسة. وبالفعل، يمثل الدعم الكامل من صانعي السياسات شرطًا أساسيًا لتنفيذ تقييم ناجح، وينبغي عدم إجراء تقييمات الأثر إلا إذا كان صانعو السياسات مقتنعين بصحة التقييم وقيمته في توجيه القرارات المهمة للسياسة.

ثالثًا والأهم، في التقييم السابق، يتم تحديد مجموعات المعالجة والمقارنة قبل تنفيذ التدخل الذي يتم تقييمه. ومثلما سنوضح بمزيد من العمق في الفصول التالية، هناك العديد من الخيارات الأخرى لإجراء تقييمات صالحة عند التخطيط للتقييمات من البداية قبل التنفيذ. وندقق في الجزأين الثاني والثالث أنه من الممكن دائمًا تقريبًا العثور على تقدير صالح للواقع المضاد لأي برنامج من خلال قواعد التعيين الواضحة والشفافة، شريطة أن يتم تصميم التقييم في وقت لاحق في المستقبل. باختصار، تحظى التقييمات السابقة بأفضل فرصة لتوفير وقائع مضادة صالحة. في مرحلة التصميم، يمكن النظر في طرق بديلة لتقدير الواقع المضاد الصالح. ويمكن أيضًا مواءمة تصميم تقييم الأثر بالكامل مع قواعد تشغيل البرنامج، بالإضافة إلى مسار طرح البرنامج أو توسيع نطاقه.

على النقيض، في التقييمات اللاحقة، غالبًا ما يكون لدى الفريق الذي يجري التقييم معلومات محدودة لدرجة أنه من الصعب تحليل ما إذا تم تنفيذ البرنامج بنجاح أم لا وما إذا كان المشاركون قد استفادوا منه حقًا. ولا تجمع العديد من البرامج بيانات خط الأساس ما لم يتم تضمين التقييم في البرنامج من البداية، وبمجرد أن يدخل البرنامج حيز التطبيق، يكون الوقت قد فات كثيرًا على القيام بذلك.

تكتسب التقييمات اللاحقة باستخدام البيانات الموجودة أهمية بالغة لدورها في تقييم البرامج التي تم إنشاؤها في الماضي. وخيارات الحصول على تقدير صالح للواقع المضاد أكثر محدودية بكثير في مثل هذه المواقف. ويستند التقييم إلى قواعد واضحة لتشغيل البرنامج فيما يخص تعيين الفوائد. فضلًا عن أنه يعتمد على توفر البيانات مع التغطية الكافية لمجموعات المعالجة والمقارنة قبل تنفيذ البرنامج وبعده. ونتيجة لذلك، تستند إمكانية تنفيذ التقييم اللاحقة إلى السياق، ولا يمكن ضمانها إطلاقًا. وحتى عندما تكون التقييمات اللاحقة ممكنة التنفيذ، فإنها غالبًا ما تستخدم أساليب شبه تجريبية وتعتمد على افتراضات أقوى، ومن ثم، يمكنها تقديم أدلة أكثر قابلية للنقاش.^٤

يتمحور الدور الرئيسي لتقييم الأثر حول تقديم أدلة عن أداء البرنامج لاستخدام المسؤولين الحكوميين، ومديري المشاريع، والمجتمع المدني، والأطراف المعنية الأخرى. ويتضح نفع نتائج تقييم الأثر بشكل خاص عندما يمكن تطبيق الخلاصات على قطاع أكبر من المجتمع الإحصائي محل الاهتمام. وتعد مسألة التعميم أساسية لصانعي السياسات؛ حيث إنها تحدد ما إذا كان من الممكن تكرار النتائج المحددة في التقييم على مجموعات أخرى بخلاف تلك التي خضعت للدراسة في التقييم في حال توسيع نطاق تطبيق البرنامج.

في الأيام الأولى لتقييمات الأثر المتعلقة ببرامج التنمية، استند جزء كبير من الأدلة إلى *دراسات النجاعة*: دراسات يتم تنفيذها في بيئة محددة في ظل ظروف مضبوطة بدقة لضمان دقة الأداء بين تصميم التقييم وتنفيذ البرنامج. ونظرًا لأنه غالبًا ما يتم إجراء دراسات النجاعة كمشاريع تجريبية بمشاركة تقنية مكثفة من الباحثين في الوقت الذي يتم تنفيذ البرنامج فيه، فقد لا تكون آثار هذه المشاريع التجريبية عن النجاعة، والتي تكون على نطاق صغير في كثير من الأحيان، مفيدة بالضرورة فيما يخص أثر مشروع مماثل تم تنفيذه على نطاق أوسع في ظروف طبيعية. تستكشف دراسات النجاعة إثبات صحة المفهوم، ويكون هذا غالبًا للتحقق من جدوى برنامج جديد أو نظرية تغيير محددة. وإذا لم ينتج عن البرنامج الآثار المتوقعة في ظل هذه الظروف المُدارة بعناية، فمن غير المرجح أن ينجح إذا تم طرحه في ظل

الظروف الطبيعية. على سبيل المثال، قد ينجح التدخل التجريبي الذي يقدم بروتوكولات علاج طبي جديدة في مستشفى يضم فريق ممتاز من المديرين وأفراد الطاقم الطبي، ولكن قد لا ينجح التدخل نفسه في مستشفى متوسط المستوى يعمل به فريق أقل يقظة من المديرين وموظفين محدودي الإمكانيات. بالإضافة إلى ذلك، ستختلف حسابات التكلفة والمنفعة؛ حيث إنه يمكن ألا تُسجل التكاليف الثابتة ووفورات الحجم في الدراسات الصغيرة عن النجاعة. ونتيجة لذلك، بينما يمكن أن تكون الأدلة المستقاة من دراسات النجاعة مفيدة في اختبار مقاربة مبتكرة، غالبًا ما تتسم النتائج بإمكانية تعميم محدودة ولا تمثل بشكل كافٍ دائمًا البيئات الأعم، التي عادةً ما تكون الشاغل الرئيسي لصانعي السياسات.

وفي المقابل، تقدم *دراسات الفعالية* أدلة من التدخلات التي تتم في ظروف طبيعية، باستخدام قنوات التنفيذ المنتظم، وتهدف إلى التوصل إلى نتائج يمكن تعميمها على مجتمعات إحصائية كبيرة. وعند تنفيذ تقييمات الفعالية وتنفيذها كما ينبغي، قد تكون النتائج قابلة للتعميم على مستفيدين مستهدفين بخلاف عينة التقييم، طالما أن التوسع يستخدم هياكل التنفيذ نفسها ويصل إلى مجتمعات إحصائية مماثلة للعينة المشاركة في التقييم. وتتسم هذه الصلاحية الخارجية بأهمية بالغة لدى صانعي السياسات، وذلك لأنها تسمح لهم باستخدام نتائج التقييم لتوجيه القرارات على مستوى البرنامج، والتي تنطبق على المستفيدين المستهدفين بخلاف عينة التقييم (انظر المربع ١-٣).

المفهوم الأساسي

تقيّم دراسات النجاعة ما إذا كان يمكن للبرنامج النجاح في ظل الظروف المثالية، في حين تقيّم دراسات الفعالية ما إذا كان البرنامج يعمل في ظل الظروف الطبيعية.

المربع ٣-١: التحقق من مدى قابلية تعميم النتائج: تقييم متعدد المواقع لمقاربة "التخرج" في سبيل تخفيف حدة الفقر المدقع

البرنامج أفقر الأسر في القرى الواقعة في أشد المناطق فقرًا في كل دولة. لمدة ٢٤ شهرًا، حصلت الأسر المستفيدة على أصول منتجة، وتدريب، ودعم، وإرشاد عن المهارات الحياتية، وأموال نقدية، ومعلومات صحية، ومساعدة في الشمول المالي. وعمد تقييم الأثر إلى تقييم مدى فعالية توفير هذه الحزمة من الفوائد.

قيمت الدراسة آثار البرنامج على ١٠ مجموعات من الحاصلين. وبعد مرور عام على انتهاء البرنامج في الدول الست، كانت هناك تحسينات كبيرة في ٨ من أصل ١٠ مجموعات من الحاصلين: نصيب الفرد من الاستهلاك، والأمن الغذائي، وقيمة الأصول، والشمول المالي، والوقت المنقضي في العمل، والدخل والإيراد، والصحة النفسية، والمشاركة السياسية. وتباين حجم الآثار فيما بين الدول، مع حدوث آثار جوهرية في قيمة الأصول في جميع الدول، ما عدا دولة واحدة. ولم تكن هناك آثار ذات دلالة إحصائية في مؤشر الصحة البدنية.

تفاوتت النتائج من دولة لأخرى. ولم تكن التحسينات في نصيب الفرد من الاستهلاك كبيرة في هندوراس وبيرو، علاوة على أن التحسينات في قيمة الأصول لم تكن كبيرة في هندوراس. ومع ذلك، أشار التقييم في المجمل إلى أن هذا النوع من التدخل متعدد الأوجه واعد في تحسين حياة الناس شديدي الفقر على مستوى مجموعة من البيئات.

من خلال تقييم برنامج في سياقات متعددة، يمكن للباحثين فحص ما إذا كانت النتائج المستمدة من تقييم الأثر قابلة للتعميم. وتسهم هذه التقييمات التي يُطلق عليها *التقييمات متعددة المواقع* في تزايد مجموعة الأدلة حول ما يصلح وما لا يصلح في عملية التنمية، ويمكن أن تقدم رؤى مهمة لصانعي السياسات عبر الدول.

على سبيل المثال، في عام ٢٠٠٧، بدأ بانيرجي (Banerjee) وآخرون تقييمًا متعدد المواقع لمقاربة "التخرج" في سبيل تخفيف حدة الفقر المدقع. وحظي النموذج باهتمام كبير في جميع أنحاء العالم بعد تحقيق نتائج مبهره في بنغلاديش. وطورت لجنة تطوير الريف في بنغلاديش (Bangladesh Rural Advancement Committee, BRAC)، وهي منظمة إنمائية عالمية كبرى، النموذج الذي يهدف إلى مساعدة شديدي الفقر على "الخروج" من الفقر المدقع من خلال منحهم التحويلات النقدية، والأصول الإنتاجية، والتدريب المكثف. سعى بانيرجي (Banerjee) وزملاؤه إلى استكشاف ما إذا كانت مقاربة التخرج سينجح عبر الدول من خلال ستة تقييمات أثر عشوائية متزامنة في إثيوبيا، وغانا، وهندوراس، والهند، وباكستان، وبيرو. وفي كل دولة، عمل الباحثون مع المنظمات غير الحكومية (NGO) المحلية على تنفيذ برنامج تخرج مماثل. وبينما تم تعديل البرنامج ليتناسب مع السياقات المختلفة في كل دولة، ظلت المبادئ الرئيسية كما هي. وقد استهدف

المصادر: Banerjee and others 2015; BRAC 2013.

كما ذكر، تجيب تقييمات الأثر عن أسئلة محددة حول السبب والنتيجة. ويمكن أن تكون المقاربات الأخرى—بما في ذلك *المتابعة الدقيقة* للبرنامج، فضلاً عن الاستخدام التكميلي لمقاربات التقييم الأخرى، مثل عمليات المحاكاة السابقة وتحليل الأسلوب المختلط القائم على كل من البيانات الكمية والنوعية، و*تقييمات الصيرورة*—بمثابة مكملات قيّمة لتقييمات الأثر. وتتميز هذه المقاربات الأخرى بالعديد من التطبيقات المفيدة، مثل تقدير تأثير الإصلاحات قبل تنفيذها، والمساعدة على تركيز أسئلة تقييم الأثر الأساسية، ومتابعة تنفيذ البرنامج، وتفسير نتائج تقييمات الأثر. تعاني تقييمات الأثر التي تُجرى بمعزل عن مصادر المعلومات الأخرى من ضعف فيما يتعلق بجودتها التقنية ومدى ملاءمتها للسياسات. وبينما يمكن أن توفر نتائج تقييم الأثر أدلة قوية على ما إذا كان هناك تأثير، فإنها غالباً ما تكون محدودة في تقديم رؤى حول القنوات التي أثرت من خلالها السياسة أو البرنامج في النتائج المرصودة. و يمكن أن يؤدي انعدام المعلومات من تقييمات الصيرورة حول طبيعة ومحتوى البرنامج (بهدف وضع نتائج التقييم في سياقها) إلى شعور صانعي السياسات بالحيرة بشأن تحقق نتائج معينة من عدمها. بالإضافة الى ذلك، يمكن أن يغفل التقييم عن تلقي المستفيدين المستهدفون للفوائد من عدمها ومتى تم ذلك ، أو ما إذا كانت الفوائد قد وصلت إلى مجموعة المقارنة من غير قصد.

المتابعة

تعد *متابعة* تنفيذ البرنامج، غالباً من خلال استخدام البيانات الإدارية، أمراً بالغ الأهمية في تقييم الأثر. وترجع تلك الأهمية إلى أنه يتيح لفريق التقييم التثبت مما إذا كانت الأنشطة يتم تنفيذها على النحو المخطط لها: أي المشاركين تلقوا البرنامج، ومدى سرعة توسع البرنامج، وكيف يتم إنفاق الموارد. وهذه المعلومات ضرورية لتنفيذ التقييم، على سبيل المثال، لضمان جمع بيانات خط الأساس قبل تقديم البرنامج في عينة التقييم وللتأكد من نزاهة مجموعات المعالجة والمقارنة. والمتابعة أمر جوهري للتحقق من أن المستفيد يشارك بالفعل في البرنامج، وأن غير المستفيد لا يشارك. بالإضافة إلى ذلك، يمكن أن توفر البيانات الإدارية معلومات عن تكلفة تنفيذ البرنامج، وهي لازمة أيضاً لتحليلات التكلفة والمنفعة وفعالية التكلفة.

عمليات المحاكاة السابقة هي تقييمات تستخدم البيانات المتاحة لمحاكاة التأثيرات المتوقعة لبرنامج أو إصلاح السياسات على الحاصل محل الاهتمام. ويمكن أن تكون مفيدة للغاية في تقدير الفعالية المتوقعة النسبية على النتائج وذلك لمجموعة من الخيارات في تصميم البرنامج البديلة. وهذه الأساليب شائعة الاستخدام وتستند إلى توفر بيانات وافية عالية الجودة يمكن استخدامها لتطبيق نماذج المحاكاة المناسبة على الأسئلة المطروحة (انظر المربع ٤-١). على عكس تقييمات الأثر، تُستخدم هذه الأساليب لمحاكاة التأثيرات المستقبلية المحتملة، بدلاً من قياس الآثار الفعلية للبرامج المنفذة. ويمكن أن تكون هذه الأنواع من الأساليب غاية في النفع في قياس التأثيرات المرتقبة للبرنامج ووضع أهداف واقعية، فضلاً عن تقدير التكاليف، ومعدلات المنفعة، والمعايير الاقتصادية الأخرى. وغالباً ما تُستخدم كأساس للتحليل الاقتصادي للمشاريع، ولا سيما قبل إجراء الإصلاح أو تنفيذ البرنامج.

المربع ٤-١: محاكاة التأثيرات المحتملة للمشروع من خلال النمذجة الهيكلية: بناء نموذج لاختبار التصاميم البديلة باستخدام بيانات برنامج Progresa في المكسيك

الدراسة واستخدم الأموال لزيادة الحوافز النقدية للطلاب في الصفوف العليا، فمن المرجح أن تكون التأثيرات على متوسط سنوات الدراسة المستكملة أكبر. في هذه الحالة، أُجريت الإسقاطات باستخدام مسح خط الأساس لتقييم الأثر الذي تم الانتهاء منه. ويمكن اختبار نتائج التنبؤات لمعرفة ما إذا كانت قد أسفرت عن الآثار نفسها كتجربة البرنامج الفعلية. ومع ذلك، فهذا الأمر ليس بهدف ممكنًا بشكل عام. غالباً ما تُستخدم هذه الأنواع من أساليب المحاكاة قبل التنفيذ الفعلي للبرنامج من أجل دراسة التأثيرات المحتملة لتصميمات البرنامج البديلة المختلفة. ومن ثم، يمكنها توفير أساس لتضييق نطاق مجموعة الخيارات التي ستخضع للاختبار في الواقع العملي.

يمكن استخدام نوع معين من عمليات المحاكاة السابقة—النمذجة الهيكلية—لتقدير تأثيرات البرنامج في ظل مجموعة من التصميمات البديلة. في تقييم برنامج Progresa/ Oportunidades/Prospera المربع ١-١، كانت البيانات التي تم جمعها غنية بما فيه الكفاية ليتمكن الباحثون من بناء نموذج قادر على محاكاة التأثيرات المتوقعة لتصميمات البرنامج البديلة. استخدم تود وولبين (Todd and Wolpin، ٢٠٠٦) بيانات خط الأساس من تقييم الأثر لبناء نموذج لقرارات الآباء المتعلقة بأطفالهم، بما في ذلك التعليم المدرسي للأطفال. وقاما بمحاكاة التأثيرات في ظل تصاميم البرنامج المختلفة. ووجدا أنه إذا ألغى البرنامج الحوافز النقدية الممنوحة للصفوف الدنيا بهدف الانتظام في

المصدر: Todd and Wolpin 2006.

ملاحظة: للحصول على مثال آخر على النمذجة الهيكلية، انظر Bourguignon, Ferreira, and Leite (2003).

تمثل مقاربات الأساليب المختلطة التي تجمع بين البيانات الكمية والنوعية مكملًا أساسيًا لتقييمات الأثر القائمة على استخدام البيانات الكمية وحدها، ولا سيما للمساعدة على إنشاء فرضيات وتركيز أسئلة البحث قبل جمع البيانات الكمية وتقديم وجهات نظر ورؤى حول أداء البرنامج خلال مرحلة تنفيذ البرنامج وبعدها. وهناك العديد من الأساليب النوعية، والتي تشكل مجال بحثها الخاص.^٥ وبشكل عام، تستخدم أساليب توليد البيانات النوعية مقاربات مفتوحة لا تعتمد على إجابات محددة مسبقًا من هؤلاء الذين تُجرى معهم المقابلة. وتُستخلص البيانات من خلال مجموعة من المقاربات، بما في ذلك مجموعات التركيز، وسيرة الحياة، والمقابلات مع المستفيدين المختارين والمجيبين الرئيسيين الآخرين (Rao and Woolcock 2003). ويمكن أن تتضمن أيضًا العديد من تقييمات الملاحظة وتقديرات الإثنوغرافية. ومع أن الملاحظات، ووجهات النظر، والآراء التي جُمعت خلال العمل النوعي لا تُقدم في العادة تمثيلًا إحصائيًا عن المستفيدين من البرنامج، ومن ثمّ ليست قابلة للتعميم، إلا أنها مفيدة في فهم سبب تحقيق نتائج معينة أو عدم تحقيقها (انظر المربع ٥-١).

وتتميز التقييمات التي تدمج التحليل الكمي والنوعي بأنها تستخدم أساليب مختلطة (Bamberger, Rao, and Woolcock 2010).

المربع ٥-١: تقييم الأسلوب المختلط في التطبيق العملي: الجمع بين تجربة عشوائية مضبوطة ودراسة إثنوغرافية في الهند

في العينة لإجراء الدراسة النوعية. وتم تعيين باحث ميداني مخضرم للعيش في كل قرية ودراسة آثار البرنامج على الهياكل الاجتماعية والسياسية للقرية. استمرت الدراسة الإثنوغرافية لمدة عامين بعد انتهاء التجربة العشوائية المضبوطة، وهو ما أتاح الفرصة لملاحظة التأثيرات طويلة المدى. وبينما توصلت التجربة العشوائية المضبوطة إلى أن التدخل ليس له أثر ذي دلالة إحصائية، قدمت الدراسة النوعية رؤى حول سبب فشل التدخل. وقد حدّد البحث النوعي العديد من العوامل التي أعاققت فعالية التدخل: التباينات في جودة تيسير البرنامج، وغياب الدعم من القمة إلى القاعدة، وهياكل السلطة المحلية المستحكمة.

(تابع)

يمكن أن تكون مقاربات الأساليب المختلطة مفيدة بشكل خاص عند تقييم البرامج ذات الحوائل التي يصعب قياسها باستخدام المسوحات الكمية. وأحد الأمثلة على ذلك البرامج المتعلقة بالديمقراطية والحوكمة. عند تصميم إستراتيجية تقييم لبرنامج حملة الناس (People's Campaign)، الذي يهدف إلى زيادة مشاركة المواطنين في الحكومات القروية، عمد أنانثابور ومالك وراو (Ananthpur, Malik, and Rao, 2014) إلى دمج تجربة عشوائية مضبوطة (RCT)، انظر قائمة المصطلحات) مع دراسة إثنوغرافية أُجريت على مجموعة فرعية تضم ١٠ في المائة من عينة التقييم المستخدمة في التجربة العشوائية المضبوطة. وتم استخدام أساليب مطابقة لضمان وجود خصائص مماثلة بين قرى المعالجة والمقارنة

المربع ١-٥: تقييم الأسلوب المختلط في التطبيق العملي: الجمع بين تجربة عشوائية مضبوطة ودراسة إثنوغرافية في الهند (تابع)

دون الفهم الدقيق للسياق والديناميكيات المحلية التي يقدمها المكون النوعي، لم يكن ليتمكن الباحثون من فهم سبب عدم وجود أي آثار للبيانات الكمية. وتمكنت الدراسة الإثنوغرافية من تقديم تقييم أكثر ثراءً، مع توفير رؤى حول العناصر المفيدة في تحسين البرنامج.

كشفت الأدلة النوعية أيضاً عن بعض آثار البرنامج الملموسة القليلة و غير المتوقعة. في قرى المعالجة (التي تعرضت لتدخل البرنامج)، نجح البرنامج في تحسين تسوية النزاعات المتعلقة بتقديم الخدمة وزاد من معدل مشاركة المرأة في أنشطة تنمية القرية. علاوة على ذلك، لاحظ الباحثون الميدانيون أن الحكومات القروية تعمل بشكل أفضل في قرى المعالجة.

المصدر: Ananthpur, Malik, and Rao 2014.

عند وضع مقارنة الأسلوب المختلط، يحدّد كريسويل (Creswell, ٢٠١٤) ثلاث مقاربات أساسية:

١. *المتوازي المتقارب*. تُجمّع كلُّ من البيانات الكمية والنوعية في الوقت نفسه، وتستخدم هذه المقاربة لتثليث النتائج أو لتقديم نتائج مبكرة عن كيفية تنفيذ البرنامج وفهمه من قبل المستفيدين.
٢. *التفسيري المتتابع*. تقدم البيانات النوعية سياقاً وتفسيرات للنتائج الكمية، في سبيل استكشاف حالات النجاح والفشل غير العادية، ولوضع تفسيرات منهجية لأداء البرنامج كما تبين في النتائج الكمية. بهذه الطريقة، يمكن أن يساعد العمل النوعي في تفسير سبب ملاحظة نتائج معينة في التحليل الكمي، ويمكن استخدامه للدخول إلى "الصندوق الأسود" ومعرفة ما حدث في البرنامج (Bamberger, Rao, and Woolcock 2010).
٣. *الاستكشافي المتتابع*. يمكن لفريق التقييم استخدام مجموعات التركيز، والقوائم، والمقابلات مع المجيبين الرئيسيين، والمقاربات النوعية الأخرى لوضع فرضيات حول كيفية عمل البرنامج والسبب في ذلك، وتوضيح أسئلة البحث التي يلزم الإجابة عنها في عمل تقييم الأثر الكمي، بما في ذلك بدائل تصميم البرنامج الأكثر صلة التي ستخضع للاختبار من خلال تقييم الأثر.

تقييمات الصيرورة

تركز *تقييمات الصيرورة* على كيفية تنفيذ البرنامج وتشغيله، وتقييم ما إذا كان يتوافق مع تصميمه الأصلي، وتوثيق تطويره وتشغيله. وفي العادة، يمكن إجراء تقييمات الصيرورة بسرعة نسبية وبتكلفة معقولة. في المشاريع التجريبية وفي المراحل الأولى من البرنامج، يمكن أن تكون التقييمات مصدراً قيماً للمعلومات حول كيفية تحسين تنفيذ البرنامج، وغالباً ما تُستخدم كخطوات أولى في تطوير البرنامج بحيث يمكن إجراء التعديلات التشغيلية قبل الانتهاء من تصميم البرنامج. ويمكنها اختبار ما إذا كان البرنامج يعمل طبقاً للتصميم ومتوافقاً مع نظرية التغيير للبرنامج (المربع ٦-١).

المربع ٦-١: توجيه التوسع في نطاق البرنامج الوطني من خلال تقييم الصيرورة في تنزانيا

في العلاقات بين أفراد الأسرة أو ديناميكيات المجتمع. وكان هدف تقييم الصيرورة فهم كيفية عمل البرنامج في الواقع العملي وتقديم توصيات للتحسينات.

وجد تقييم الأثر أن البرنامج كان له آثار إيجابية ودلالية إحصائية على الحاصلات التعليمية والصحية الرئيسية. وكان الأطفال في الأسر المشاركة أكثر عرضة بحوالي ١٥ في المائة لإكمال المدرسة الابتدائية وأقل عرضة بنسبة ١١ في المائة للإصابة بالمرض. وكشفت مجموعات التركيز التي تضم معلمين أيضًا عن أن الطلاب في مجموعات المعالجة كانوا أكثر استعدادًا وبقية.

ومع ذلك، أشارت مجموعات التركيز التي تضم أعضاء من المجتمع إلى أنه كان هناك مستوى من عدم الرضا على عملية اختيار المستفيدين. وقد اشتكى المشاركون من غياب الشفافية في اختيار المستفيدين وتأخير في المدفوعات. وأتاح تقييم الصيرورة لمديري البرنامج معالجة هذه المشكلات، وتحسين عمليات تشغيل البرنامج.

وكان عمل التقييم موجهًا لقرار الحكومة التنزانية لتوسيع نطاق البرنامج. ومن المتوقع أن تصل التحويلات النقدية المشروطة (CCT) المجتمعية إلى قرابة مليون أسرة بحلول عام ٢٠١٧، بالاستناد إلى الدروس المستفادة من هذا التقييم.

توجد العديد من الجوانب المتعلقة بأداء البرنامج. ويمكن أن تكمل الأدلة المستمدة من تقييمات الصيرورة نتائج تقييم الأثر وتقدم صورة أكثر اكتمالاً عن أداء البرنامج. وقد يكون ذلك مهمًا بشكل خاص للبرامج التجريبية حتى تسلط الضوء على كيفية عمل المؤسسات والعمليات الجديدة.

في عام ٢٠١٠، قررت الحكومة التنزانية تجربة برنامج تحويلات نقدية مشروطة (CCT) مجتمعي في ثلاث مناطق. ووقّر البرنامج تحويلات نقدية للأسر الفقيرة بناءً على الامتثال لمتطلبات تعليمية وصحية معينة. وساعدت المجموعات المجتمعية في تعيين التحويلات النقدية للأسر الأكثر ضعفًا في مجتمعاتها. ولتقييم ما إذا كان هذا النظام الذي يقوده المجتمع المحلي قد نجح في السياق التنزاني، قررت مجموعة من الباحثين في البنك الدولي دمج تقييم الصيرورة في تقييم الأثر التقليدي.

استخدم تقييم الصيرورة كل من البيانات الكمية والنوعية. وبعد مرور عام على إجراء مسح خط الأساس في المناطق الخاضعة للتجربة، نظم الباحثون تمرين سجل الأداء المجتمعي لتقييم جوانب البرنامج، بالاستناد إلى مجموعات التركيز المكونة من أفراد المجتمع. وتمت الاستعانة بمجموعات التركيز أيضًا في إجراء مناقشات متعمقة عن آثار البرنامج التي قد يكون من الأصعب تحديدها كمياً، مثل التغييرات

المصادر: Berman 2014; Evans and others 2014.

يجب أن يتضمن تقييم الصيرورة العناصر التالية، والتي غالبًا ما تُستخلص من سلسلة النتائج أو النموذج المنطقي (انظر الفصل الثاني)، وتكتمل بوثائق البرنامج والمقابلات مع المجيبين الرئيسيين ومجموعات التركيز التي تضم المستفيدين:^٦

- أهداف البرنامج والسياق الذي يعمل فيه البرنامج
 - وصف العملية المستخدمة في تصميم البرنامج وتنفيذه
 - وصف العمليات التشغيلية للبرنامج، بما في ذلك أي تغييرات فيها
 - البيانات الأساسية عن العمليات التشغيلية للبرنامج، بما في ذلك المؤشرات المالية ومؤشرات التغطية
 - تحديد ووصف الأحداث المتداخلة التي ربما أثرت في التنفيذ والحوائل
 - الدعم بالوثائق، مثل مذكرات المفاهيم، وكتيبات العمليات التشغيلية، ومحاضر الاجتماعات، والتقارير، والمذكرات.
- يشكل تطبيق تقييم الأثر على برنامج لم يتم التثبت من صحة عملياته التشغيلية خطراً يتمثل في احتمال تبديد موارد تقييم الأثر بينما كان إجراء تقييم صيرورة أكثر بساطة كافياً أو أنه تم إدخال التعديلات اللازمة على تصميم البرنامج بمجرد إجراء تقييم الأثر وبالتالي يغيّر طبيعة البرنامج قيد التقييم وفائدة تقييم الأثر.

تحليل التكلفة والمنفعة وتحليل فعالية التكلفة

من الأهمية بمكان أن يُستكمل تقييم الأثر بمعلومات عن تكلفة المشروع، أو البرنامج، أو السياسة قيد التقييم.

بمجرد أن تتوفر نتائج تقييم الأثر، يمكن دمجها مع معلومات عن تكاليف البرنامج للإجابة عن سؤالين إضافيين. أولاً، بالنسبة إلى الشكل الأساسي من تقييم الأثر، سيتيح لك إضافة معلومات عن التكاليف إجراء تحليل التكلفة والمنفعة، الذي سيجيب عن السؤال التالي: ما الفائدة التي يقدمها البرنامج مقابل تكلفة محددة؟ *يقدّر تحليل التكلفة والمنفعة إجمالي المنافع المتوقعة للبرنامج، مقارنة بإجمالي تكلفته المتوقعة. ويسعى إلى تحديد جميع تكاليف ومنافع البرنامج من الناحية المالية، ويقيّم ما إذا كانت المنافع تفوق التكاليف أم لا.*^٧

في عالم مثالي، لن يكون تحليل التكلفة القائم على أدلة تقييم الأثر متافاً لبرنامج معين فحسب، وإنما أيضاً لمجموعة من البرامج أو بدائل البرنامج، بحيث يمكن لصانعي القرارات تقييم أي برنامج أو بديل أكثر فعالية من حيث التكلفة في الوصول إلى غاية معينة. عندما يختبر تقييم الأثر بدائل البرنامج، تتيح لك إضافة معلومات عن التكاليف الإجابة عن السؤال الثاني: كيف يمكن مقارنة العديد من بدائل تنفيذ البرنامج من حيث فعالية التكلفة؟ يقارن *تحليل فعالية التكلفة* هذا التكلفة النسبية لاثنتين أو أكثر من البرامج أو بدائل البرنامج للوصول إلى حصة مشتركة، مثل المحاصيل الزراعية أو درجات اختبار الطالب.

المفاهيم الأساسية

يقدّر تحليل التكلفة والمنفعة إجمالي المنافع المتوقعة من البرنامج، مقارنةً بإجمالي تكاليفه المتوقعة. يعقد تحليل فعالية التكلفة مقارنة بين التكلفة النسبية لاثنتين أو أكثر من البرامج أو بدائل البرنامج للوصول إلى حصة مشتركة.

في تحليل التكلفة والمنفعة أو فعالية التكلفة، يقدّر تقييم الأثر جانب المنفعة أو الفعالية، ويقدم تحليل التكلفة معلومات عن التكاليف. ينصب تركيز هذا الكتاب على تقييم الأثر ولا يناقش بالتفصيل كيفية جمع بيانات التكاليف أو إجراء تحليل التكلفة والمنفعة أو فعالية التكلفة.^٧ ومع ذلك، من الأهمية بمكان أن يُستكمل تقييم الأثر بمعلومات عن تكلفة المشروع، أو البرنامج، أو السياسة قيد التقييم. ومتى توفرت معلومات عن الأثر والتكاليف لمجموعة متنوعة من البرامج، يمكن لتحليل فعالية التكلفة تحديد أي الاستثمارات التي ستدر أعلى معدل منفعة ويتيح لصانعي السياسات اتخاذ قرارات مستنيرة بشأن التدخل الذي ينبغي الاستثمار فيه. يوضح المربع ٧-١ كيف يمكن استخدام تقييمات الأثر لتحديد البرامج الأكثر فعالية من حيث التكلفة وتحسين توزيع الموارد.

المربع ٧-١: تقييم فعالية التكلفة: مقارنة تقييمات البرامج التي تؤثر في التعليم في المدارس الابتدائية

من حيث الانحرافات المعيارية، التي يمكن الحصول عليها مقابل كل ١٠٠ دولار أمريكي يُنفق على البرنامج. ومع أنه من المحتمل أن تنخفض التكاليف إذا تم تنفيذ البرامج على نطاق واسع، فقد استخدم الباحثون التكاليف كما وردت في التقييمات من أجل الحفاظ على الاتساق. ووجدوا أن الإصلاحات والتدخلات التربوية التي تحسّن المساءلة وتزيد الحوافز المقدمة للمعلمين تميل إلى أن تكون الأكثر فعالية من حيث التكلفة. ومن جهة أخرى، خلص الباحثون إلى أن توفير المزيد من نفس المدخلات دون تغيير أصول التدريس أو المساءلة كانت آثاره محدودة على درجات الاختبار. على سبيل المثال، لم يكن لأحد البرامج في كينيا الذي أدى إلى زيادة عدد المعلمين في المدارس تأثيرًا كبيرًا على درجات اختبار الطلاب.

(تابع)

بموجب تقييم عدد من البرامج ذات الأهداف المتشابهة، من المحتمل مقارنة فعالية التكلفة النسبية للمقاربات المختلفة من أجل تحسين الحاصل، مثل التعليم في المدارس الابتدائية. وحتى يكون ذلك ممكنًا، يجب ألا يوفر المقيّمون نتائج تقييم الأثر فحسب، ولكن أيضًا معلومات تفصيلية عن التكاليف المتعلقة بالتدخلات. وفي التحليل البعدي لحاصل التعليم في الدول النامية، استخدم كريمر، وبرانين، وجلينبرستر (Kremer, Brannen, and Glenner, ٢٠١٣) المعلومات عن التكاليف من ٣٠ تقييم أثر لتحليل فعالية التكلفة لأنواع مختلفة من التدخلات التعليمية. قارن المؤلفون بين أنواع عديدة من التدخلات التعليمية، بما في ذلك الحصول على التعليم، ومدخلات مقارنة العمل المعتاد، والابتكارات التربوية، ومساءلة المعلم، والإدارة المدرسية. وبشكل خاص، تحققوا من تحسّن درجات الاختبار،

المربع ١-٧: تقييم فعالية التكلفة: مقارنة تقييمات البرامج التي تؤثر في التعليم في المدارس الابتدائية (تابع)

وجلينرستر (Kremer, Brannen, and Glennerster, ٢٠١٣)، فإن مقارنة تقييمات التدخلات ذات الأهداف المتشابهة يمكن أن يسלט الضوء على فعالية التدخلات المختلفة على مستوى مختلف السياقات. ومع ذلك، يلزم على الباحثين إدراك أن السياقات تختلف بصورة كبيرة عبر البرامج والبيئات. ويظل من النادر نسبيًا أيضًا أن تتوفر بيانات غنية عبر البرامج مع مقاييس حاصلة قابلة للمقارنة، وتقييمات أثر، ومعلومات عن التكاليف.

يبدو أن البرامج التي مكّنت المجتمعات المحلية من خلال تدخلات الإدارة المدرسية الأكثر نجاحًا وفعالية من حيث التكلفة، وخصوصًا عند إضفاء الطابع الرسمي على هذه الإصلاحات. على سبيل المثال، بينما لم يكن لإنشاء لجان مدرسية محلية في إندونيسيا وتدريبها تأثيرًا كبيرًا على درجات الاختبار، فقد كان تشكيل لجان أكثر تمثيلًا من خلال الانتخابات بالغ الفعالية من حيث التكلفة.

مثلما توضح الدراسة التي أجراها كريم، وبرانين،

المصدر: Kremer, Brannen, and Glennerster 2013.

الاعتبارات الأخلاقية المتعلقة بتقييم الأثر

عند اتخاذ قرار تصميم تقييم الأثر، يجب وضع بعض المسائل الأخلاقية المهمة في الاعتبار. وحتى أنه أُثيرت أسئلة حول ما إذا كان تقييم الأثر أخلاقيًا في حد ذاته أم لا. وتتمثل إحدى نقاط استهلال هذا النقاش في دراسة أخلاقيات استثمار قدر كبير من الموارد العامة في برامج فاعليتها غير معلومة. في هذا السياق، يمكن النظر إلى غياب التقييم نفسه على أنه عمل غير أخلاقي. ويمكن أن تؤدي المعلومات عن فعالية البرنامج التي تنتج عن تقييمات الأثر إلى استثمار أكثر فعالية وأخلاقية للموارد العامة.

تتعلق الاعتبارات الأخلاقية الأخرى بالقواعد المستخدمة في تعيين فوائد البرنامج، والأساليب المستخدمة في دراسة البشر، والشفافية في توثيق خطط البحث، والبيانات، والنتائج. وتمت مناقشة هذه المسائل بالتفصيل في الفصل الثالث عشر.

إن أبسط مبدأ أخلاقي أساسي في التقييم هو أنه ينبغي ألا يتم رفض تقديم التدخلات ذات الفوائد المعروفة أو تأخيرها لغرض إجراء التقييم فحسب. وفي هذا الكتاب، نرى أنه ينبغي ألا تفرض التقييمات كيفية تعيين الفوائد، وإنما يجب أن تتلاءم التقييمات مع قواعد تعيين البرنامج، والتي تتسم بأنها عادلة وشفافة. وفي هذا السياق، لا تنبع أي مخاوف أخلاقية بشأن قواعد تعيين البرنامج من تقييم الأثر نفسه، ولكن من القواعد التشغيلية للبرنامج مباشرة. ويمكن أن تكون تقييمات التخطيط مفيدة في توضيح القواعد التشغيلية للبرنامج والمساعدة على إجراء مراجعة لمعرفة ما إذا كانت عادلة وشفافة أم لا، وذلك استنادًا إلى معايير أهلية واضحة.

غالبًا ما يتسبب التعيين العشوائي لفوائد البرنامج في إثارة مخاوف أخلاقية بشأن حرمان المستفيدين المؤهلين من الحصول على فوائد البرنامج. ومع ذلك، تعمل معظم البرامج في سياقات تشغيلية ذات موارد مالية وإدارية محدودة؛ ما يجعل من المستحيل الوصول إلى جميع المستفيدين المؤهلين في وقت واحد. من وجهة نظر أخلاقية، يجب أن يحظى جميع الأشخاص المؤهلين للمشاركة في أي نوع من أنواع البرامج الاجتماعية بفرصة متكافئة في تلقي البرنامج. ومن الجدير بالذكر أن التعيين العشوائي يلبي هذا المطلب الأخلاقي. وفي الأوضاع التي يتم فيها تنفيذ البرنامج على مراحل، يمكن أن يستند طرح البرنامج على الترتيب العشوائي الذي سيتلقى بموجبه المستفيدون المستحقون البرنامج. في هذه الحالات، يمكن استخدام المستفيدين المشاركين لاحقًا كمجموعة مقارنة للمستفيدين السابقين؛ مما يؤدي إلى وضع تصميم متين للتقييم و طريقة شفافة وعادلة لتوزيع الموارد النادرة.

وتتخطى أخلاقيات تقييم الأثر أخلاقيات قواعد تعيين البرنامج. وتشمل أيضًا أخلاقيات إجراء الأبحاث على البشر، فضلًا عن أخلاقيات إجراء أبحاث شفافة، وموضوعية، وقابلة للتكرار، كما هو موضح في الفصل الثالث عشر. في العديد من الدول والمؤسسات الدولية، تم إنشاء مجالس مراجعة أو لجان أخلاقيات لتنظيم الأبحاث التي تتضمن بشرًا وإرساء قواعد لها. وتتولى هذه المجالس مسؤولية تقييم الدراسات البحثية، والموافقة عليها، ومتابعتها، مع الغايات الأساسية المتمثلة في حماية الحقوق وتعزيز رفاهية جميع الأفراد. وعلى الرغم من أن تقييمات الأثر مهام تشغيلية في الأساس، فإنها تشكل أيضًا دراسات بحثية، ومن ثم، يجب أن تلتزم بالمبادئ التوجيهية للأبحاث المتعلقة بالبشر.

إن جعل تقييم الأثر موضوعيًا، وشفافًا، وقابلًا للتكرار عنصر أخلاقي لا يقل أهمية عن إجراء البحث. وحتى يتسم البحث بالشفافية، يمكن تضمين خطط تقييم الأثر في خطة التحليل المسبق وتقديمها إلى سجل الدراسة. وعند الانتهاء من البحث، يمكن إتاحة البيانات والرموز المستخدمة في التحليل للعامة حتى يمكن للآخرين تكرار العمل، مع حماية سرية الهوية.

تقييم الأثر لاتخاذ القرارات المتعلقة بالسياسات

ثمة حاجة إلى تقييمات الأثر لتوجيه صانعي السياسات بشأن مجموعة من القرارات، بدءًا من إنهاء البرامج منعدمة النجاعة، ووصولًا إلى توسيع نطاق التدخلات الفعالة، وتعديل فوائد البرنامج، والاختيار من بين العديد من بدائل البرنامج. وتصبح تقييمات الأثر أكثر فعالية عند تطبيقها بانتقائية للإجابة عن أسئلة مهمة عن السياسات، وغالبًا ما تُطبق على البرامج التجريبية المبتكرة التي تختبر مقارنة غير مثبتة، ولكنه واعد.

أصبح تقييم التحويلات النقدية المشروطة المكسيكي الموضح في المربع ١-١ مؤثرًا ليس فقط بسبب الطبيعة المبتكرة للبرنامج، وإنما أيضًا بسبب تقييم الأثر الذي قدّم أدلة موثوقة وقوية لا يمكن غض الطرف عنها عند اتخاذ قرارات السياسات اللاحقة. وقد تأثر اعتماد البرنامج و توسيع نطاقه على الصعيدين الوطني والدولي بدرجة كبيرة على نتائج التقييم.

يمكن استخدام تقييمات الأثر لاستكشاف مختلف أنواع الأسئلة المتعلقة بالسياسات. وسيختبر الشكل الأساسي لتقييم الأثر مدى فعالية برنامج بعينه. بعبارة أخرى، سيجيب عن السؤال التالي: ما هي فعالية تنفيذ برنامج أو تدخل معين مقارنة بغياب البرنامج؟ كما ناقش الجزء الثاني، فإن تقييم الأثر من هذا النوع يعتمد على مقارنة مجموعة المعالجة التي حازت على الابتكار، أو البرنامج، أو السياسة بمجموعة المقارنة التي لم تحصل عليه لتقدير مدى فاعليته. ويتمثل التحدي الأساسي الذي يواجه تقييم الأثر في تكوين مجموعة مقارنة مشابهة قدر الإمكان لمجموعة المعالجة. وتعد درجة قابلية المقارنة بين مجموعات المعالجة والمقارنة محورية بالنسبة إلى *الصلاحية الداخلية* للتقييم، وبناءً عليه، فهي أساسية لتقييم الأثر السببي للبرنامج.

يتزايد استخدام تقييمات الأثر لاختبار ابتكارات التصميم داخل البرنامج دون اختيار مجموعة مقارنة خالصة من خارجه. وغالبًا ما تُجرى هذه الأنواع من التقييمات لمعرفة ما إذا كان بوسع ابتكار تصميم معين تعزيز فعالية البرنامج أو خفض التكاليف (انظر المربع ٨-١).

يمكن استخدام التقييمات أيضًا لاختبار مدى فعالية بدائل تنفيذ البرنامج. على سبيل المثال، يمكنها الإجابة عن السؤال التالي: عندما يمكن تنفيذ برنامج بعدة طرق، أيّ منها هي الطريقة الأكثر فعالية أو الفعالة من حيث التكلفة لتنفيذ البرنامج؟ في هذا النوع من التقييم، يمكن مقارنة اثنين أو أكثر من المقاربات أو ميزات التصميم داخل البرنامج مع بعضها البعض للتوصل إلى دليل حول البديل الأكثر فعالية من حيث التكلفة أو تحقيق غاية معينة. وغالبًا ما يُشار إلى بدائل البرنامج هذه باسم أذرع المعالجة. على سبيل المثال، قد يرغب برنامج في اختبار حملات توعية بديلة واختيار مجموعة واحدة لتلقي حملة رسائل بريد إلكتروني، بينما تستقبل مجموعة أخرى زيارات منزلية، وتتلقي أخرى رسائل نصية عبر خدمة الرسائل القصيرة (SMS)، لتقييم أيّ البدائل أكثر فعالية من حيث التكلفة. وفي العادة، تتضمن تقييمات الأثر التي تختبر معالجات البرامج البديلة مجموعة معالجة واحدة لكل واحد من أذرع المعالجة، بالإضافة إلى مجموعة مقارنة خالصة لا تتلقى أي تدخل من البرنامج. وتسمح هذه الأنواع من التقييمات لصانعي القرارات بالاختيار من بين بدائل التنفيذ، ويمكن أن تكون مفيدة جدًا في تحسين أداء البرنامج وتوفير التكاليف (المربع ٩-١).

بالإضافة إلى ذلك، يمكن عقد المقارنات بين المجموعات الفرعية من المتلقين داخل تقييم معين، للإجابة عن السؤال التالي: هل البرنامج أكثر فعالية لمجموعة فرعية مقارنةً بمجموعة فرعية أخرى؟ على سبيل المثال، هل نجح إدخال منهج دراسي جديد في زيادة درجات الاختبار بين الطالبات الإناث أكثر من الطلاب الذكور؟ يسعى هذا النوع من أسئلة تقييم الأثر إلى توثيق ما إذا كان هناك قدر من عدم التجانس في آثار البرنامج عبر المجموعات الفرعية. وتحتاج مثل هذه الأسئلة إلى النظر فيها مقدمًا؛ حيث يلزم دمجها في مرحلة تصميم تقييم الأثر وتتطلب عينات كبيرة بما فيه الكفاية لإجراء تحليل المجموعات الفرعية المختلفة محل الاهتمام.

المربع ٨-١: تقييم البرامج المبتكرة: فريق الرؤى السلوكية في المملكة المتحدة

من المال؛ حيث كانت تعتمد الحكومة سابقًا على استقطاب أصوات الناخبين من خلال الزيارات المنزلية المكلفة لزيادة معدل تسجيل الناخبين.

في تقييم مبتكر آخر، دخلت مؤسسة فريق الرؤى السلوكية في شراكة مع هيئة الخدمات الصحية الوطنية ووزارة الصحة لدراسة كيفية تشجيع الناس بأسلوب فعال من حيث التكلفة على التسجيل كمتبرعين بالأعضاء. وكانت تلك واحدة من أكبر التجارب العشوائية المضبوطة التي شهدتها القطاع العام في المملكة المتحدة على الإطلاق. وقد توصل الباحثون إلى نتائج مشجعة من تدخل اختبار استخدام رسائل مختلفة على صفحة انترنت حكومية ذات معدلات زيارة عالية. وكانت تستند العبارة القصيرة الأفضل أداءً على فكرة الأخذ والعطاء وسألت، إذا احتجت إلى زراعة عضو، فهل ستفعل؟ إذا كان الأمر كذلك، فساعد الآخرين.

يتشارك في ملكية مؤسسة فريق الرؤى السلوكية وتمويلها كلٌّ من الحكومة البريطانية والمؤسسة الخيرية Nesta المعنية بالابتكارات، والموظفين أنفسهم. وقد امتد النموذج إلى خارج المملكة المتحدة، بفضل فتح المؤسسة مكاتب لها في أستراليا والولايات المتحدة الأمريكية. علاوة على ذلك، اتبعت الولايات المتحدة نموذج مؤسسة فريق الرؤى السلوكية لإنشاء مبادرة العلوم الاجتماعية والسلوكية في البيت الأبيض في عام ٢٠١٥.

أنشأت الحكومة البريطانية مؤسسة فريق الرؤى السلوكية (Behavioural Insights Team, BIT) في عام ٢٠١٠، وكانت أول مؤسسة حكومية مخصصة لتحسين الخدمات العامة من خلال تطبيق العلم السلوكي. وتضمنت أهداف المؤسسة تحسين فعالية تكلفة الخدمات العامة، وتقديم نماذج واقعية على سلوك البشر لتحليل السياسة، وتمكين الأفراد من القيام بخيارات أفضل. وتحقيقًا لهذا الهدف، تستخدم مؤسسة فريق الرؤى السلوكية تجارب متضمنة تقييمات الأثر لاختبار الأفكار المبتكرة في السياسة العامة. وقد نفذت المؤسسة منذ إنشائها أكثر من ١٥٠ تجربة عشوائية مضبوطة في تشكيلة كبيرة من مجالات السياسات الداخلية، غالبًا باستخدام بيانات إدارية.

أجرت مؤسسة فريق الرؤى السلوكية تقييمات الابتكارات على الخدمات العامة التي تستند إلى مؤلفات العلم السلوكي. وقد تعاونت المؤسسة مع مقاطعة لندن لتقديم حافز من نوع الياصيب بهدف زيادة معدل تسجيل الناخبين قبل الانتخابات. تم توزيع السكان عشوائيًا إلى ثلاث مجموعات—وكانت الأولى بدون يانصيب، والثانية مع يانصيب بجائزة قدرها ١٠٠٠ جنيه إسترليني إذا سجلوا قبل تاريخ معين، والثالثة مع يانصيب بجائزة قدرها ٥٠٠٠ جنيه إسترليني إذا سجلوا قبل نفس التاريخ. وجدت المؤسسة أن الحافز من نوع الياصيب عمل على زيادة معدل تسجيل الناخبين بشكل كبير. علاوة على ذلك، فقد وُفّر على الحكومة المحلية الكثير

المصدر: فريق الرؤى السلوكية، <http://www.behaviouralinsights.co.uk>

المربع ٩-١: تقييم بدائل تصميم البرنامج: سوء التغذية والنمو المعرفي في كولومبيا

التسلسل الذي بدأوا به العلاج، وبناءً عليه، في مقدار الوقت الذي قضوه في البرنامج. فقد كانت المجموعة ٤ أول من بدأت، وتعرضت للعلاج لأطول مدة، تبعثها المجموعتين ٣ و ٢، ثم ١. وتألف العلاج ذاته من ست ساعات من الرعاية الصحية والأنشطة التعليمية في اليوم، فضلاً عن الأغذية الإضافية والمكملات الغذائية. وفي فترات منتظمة على مدار البرنامج، استخدم المقيّمون اختبارات معرفية لتتبع التقدم الذي أحرزه الأطفال في المجموعات الأربع كافة. وجد المقيّمون أن الأطفال الذين شاركوا في البرنامج لأطول وقت أظهروا أعظم المكاسب في التحسن المعرفي. في مقياس ستانفورد بينيه للذكاء، الذي يقدّر العمر العقلي مطروحاً منه العمر الزمني، كان متوسط عمر أطفال المجموعة (٤) ٥ أشهر ومتوسط عمر أطفال المجموعة (١) ١٥ شهراً.

يوضح هذا المثال كيف تمكن منفذو البرنامج وصانعو السياسات من استخدام تقييمات أذرع العلاج المتعددة لتحديد بديل البرنامج الأكثر فعالية.

في أوائل سبعينيات القرن العشرين، نفذت محطة أبحاث علم البيئة البشرية، بالتعاون مع وزارة التعليم الكولومبية، برنامجاً تجريبياً لمعالجة مشكلة سوء التغذية لدى الأطفال في كولومبيا، وذلك من خلال توفير الرعاية الصحية والأنشطة التعليمية، فضلاً عن الأغذية والمكملات الغذائية. وكجزء من البرنامج التجريبي، تم تكليف فريق من المقيّمين بتحديد المدة اللازمة للاستمرار في تقديم مثل هذا البرنامج حتى ينجح في الحد من سوء التغذية بين الأطفال في سن ما قبل المدرسة من الأسر ذات الدخل المنخفض، وما إذا كانت التدخلات تؤدي إلى تحسينات في النمو المعرفي.

في النهاية، أصبح البرنامج متاحاً لجميع الأسر المستهدفة، ولكن خلال المرحلة التجريبية، كان بمقدور المقيّمين مقارنة مجموعات مماثلة من الأطفال الذين تلقوا فترات علاج مختلفة. واستخدم المقيّمون لأول مرة عملية فحص لتحديد مجموعة مستهدفة تتضمن ٣٣٣ طفلاً يعاني من سوء التغذية. ومن ثم، تم تصنيف هؤلاء الأطفال إلى ٢٠ قطاعاً، ثم تعيين كل قطاع عشوائياً إلى واحدة من أربع مجموعات معالجة. وكان وجه الاختلاف الوحيد بين المجموعات في

المصدر: McKay and others 1978.

بخلاف ميزات التصميم المختلفة التي تمت مناقشتها بالفعل، من المفيد تأمل القنوات التي تؤثر من خلالها تقييمات الأثر على السياسة. ويمكن أن يحدث ذلك داخل البرنامج مع القرارات المتعلقة بمواصلة البرنامج، أو إصلاحه، أو إنهائه. ويمكن أن تفيد نتائج تقييم الأثر في توجيه عملية توسيع نطاق البرامج التجريبية، كما يتضح من حالة موزمبيق في المربع ٢-١.

يمكن أن تقدم التقييمات أيضاً أدلة من دولة إلى أخرى، أو يمكن استخدامها لاستكشاف الأسئلة الأساسية، مثل تلك المتعلقة بالسلوك. وتثير المغامرة بالخروج عن حدود تقييم البرنامج الفردي مسألة قابلية التعميم. ومثلما يناقش الفصل الرابع في سياق تقييم معين، تهدف عينة التقييم إلى أن تكون ممثلة إحصائياً لمجتمع الوحدات المؤهلة الإحصائي الذي أخذت منه عينة التقييم، ومن ثم، تكون صالحة خارجياً. فضلاً عن الصلاحية الخارجية، تتعلق قابلية التعميم بما إذا كانت نتائج التقييم المنفذ محلياً تنطبق على بيانات أخرى وبين مجموعات مجتمع إحصائي آخر. ويعتمد هذا المفهوم الأكثر شمولية وطموحاً على تراكم الأدلة التجريبية الموثوقة عبر مجموعة من البيئات.

يسعى مجال تقييم الأثر على نحو متزايد إلى زيادة المخزون المتنامي من التقييمات الموثوقة لتحقيق نتائج قابلة للتعميم على نطاق واسع. ويركز هذا الجهد على اختبار ما إذا كانت نظرية تغيير معينة تنطبق على سياقات مختلفة، وعلى استكشاف ما إذا أسفر برنامج مماثل خضع للاختبار في بيئات مختلفة عن نتائج مماثلة (انظر المربع ١٠-١). ويشهد استخدام العديد من التقييمات للإجابة عن أسئلة جوهرية أو تجميع الأدلة من خلال التحليلات البعدية، والمراجعات المنهجية، وسجلات التقييم نموًا سريعًا ويفتح آفاقًا جديدة في عمل التقييم. وإذا جاءت النتائج متسقة عبر البيئات المتعددة، فإن هذا يمنح صانعي السياسات ثقة أكبر في جدوى البرنامج عبر مجموعة من السياقات ومجموعات المجتمع الإحصائي. وهذا اعتبار مهم؛ حيث تعتبر المناقشات حول القدرة على تكرار النتائج في جوهر المسائل المتعلقة بالفعالية وقابلية التوسع لبرنامج معين.

المربع ١٠-١: المقاربة العنقودية لتقييم الأثر: جمع الأدلة بشكل إستراتيجي لسد الفجوات المعرفية

كيفية تطبيق تدخلات البرنامج والسياسة على نطاق أوسع، باستخدام المقاربة العنقودية للأبحاث.

غالبًا ما تتمحور طلبات استدراج العروض حول مجموعة من أسئلة البحث التي تهدف إلى توجيه تصميم البرنامج والسياسة، من أجل الخروج بتقييمات أثر تسهم في وضع قاعدة من الأدلة المنسقة. ويتمثل الهدف في توجيه البحث وتقديم أدلة حول أنواع التدخلات أو أنواع الحاصلات.

(تابع)

على الرغم من أن قابلية تعميم تقييم أثر واحد قد تكون منخفضة، بالتزامن مع تقييمات مماثلة عبر سياقات مختلفة، فإن العاملين في مجال التنمية قادرون على وضع خلاصات قابلة للتطبيق على نطاق أوسع حول ما يصلح وما لا يصلح. وعلى نحو متزايد، تهدف مبادرات تقييم الأثر، مثل صندوق تقييم الأثر الإستراتيجي (SIEF) وتقييم الأثر الإنمائي (DIME) من البنك الدولي، بالإضافة إلى المبادرة الدولية لتقييم الأثر (3IE)، إلى تزويد صانعي السياسات برؤى حول

المربع ١٠-١: المقاربة العنقودية لتقييم الأثر: جمع الأدلة بشكل إستراتيجي لسد الفجوات المعرفية (تابع)

يساعد تجميع التقييمات حول مجموعة مشتركة من أسئلة البحث واستخدام مجموعة أساسية من المقاييس لقياس الحاصلات صانعي السياسات والعاملين في مجال التنمية على معرفة أي أنواع البرامج التي تصلح في بيئات متعددة. بعد ذلك، يمكنهم مراجعة تصاميم سياساتهم وبرامجهم بإدراك أفضل للسباقات التي نجحت فيها برامج معينة أو لم تنجح، أو فيما يتعلق بكيفية تحقيق حائل معينة عبر العديد من الحالات.

وفي إطار هذه المقاربات العنقودية، توضع التقييمات لسد الفجوات الموجودة في مجموعة الأدلة القائمة. على سبيل المثال، هناك دليل دامغ يثبت أن الأطفال الذين يتلقون مزيجًا من التغذية والتحفيز المعرفي، والدعم الصحي في أول ١٠٠٠ يوم من حياتهم هم أكثر عرضة لتجنب تأخر النمو. ومع ذلك، هناك نقص في الأبحاث التي تتناول أفضل السبل لتقديم هذا الدعم المشترك بطرق قابلة للتوسع وفعالة من حيث التكلفة. ويدعم صندوق تقييم الأثر الإستراتيجي الأبحاث لاستكشاف هذا السؤال في بنغلاديش، وكولومبيا، والهند، وإندونيسيا، ومدغشقر، وموزمبيق، ونيبال، والنيجر.

المصادر: تقييم الأثر الإنمائي (<http://www.worldbank.org/dime>)؛ صندوق تقييم الأثر الإستراتيجي (<http://www.3ieimpact.org>)؛ المبادرة الدولية لتقييم الأثر (<http://www.worldbank.org/en/programs/sief-trust-fund>)

اتخاذ قرار بشأن إجراء تقييم الأثر

لا تكفل جميع البرامج إجراء تقييم للأثر. ويجب استخدام تقييمات الأثر بشكل انتقائي عندما يستدعي السؤال المطروح إجراء فحص قوي للسببية. ويمكن أن تكون تقييمات الأثر مكلفة إذا جمعت بياناتك الخاصة، وينبغي استخدام ميزانية التقييم بشكل إستراتيجي. إذا كنت تبدأ برنامجًا جديدًا، أو تفكر في توسيع نطاقه، وتتساءل عما إذا كان عليك المضي قدمًا في إجراء تقييم الأثر أم لا، فسيساعدك طرح بعض الأسئلة الأساسية على اتخاذ القرار.

السؤال الأول الذي يجب طرحه هو "ما الأمور المعرضة للخطر؟" هل ستوجه الأدلة حول نجاح البرنامج، أو طريقة البرنامج، أو ابتكار التصميم عملية اتخاذ قرارات مهمة؟ وغالبًا ما تتضمن هذه القرارات مخصصات الميزانية ونطاق البرنامج. وإذا كانت هناك آثار محدودة على الميزانية أو إذا كان تأثير النتائج ينحصر في عدد قليل من الأشخاص، فقد لا يستحق الأمر إجراء تقييم أثر. على سبيل المثال، قد لا يستحق الأمر إجراء تقييم أثر لبرنامج في عيادة صغيرة تقدم المشورة لمرضى المستشفيات بالاستعانة بمتطوعين. وعلى النقيض من ذلك، فإن إصلاح نظام أجور المعلمين سيؤثر في النهاية في جميع معلمي المرحلة الابتدائية في البلاد سيكون برنامجًا ينطوي على مخاطر أكبر بكثير.

إذا قررت أن المخاطر عالية، فإن السؤال التالي هو "هل هناك أي دليل يثبت أن البرنامج يعمل؟" على وجه الخصوص، هل تعلم كم سيكون حجم تأثير البرنامج؟ هل يوجد دليل متاح من برامج مماثلة في ظل ظروف مشابهة؟ وإذا لم يكن هناك دليل متاح عن إمكانيات نوع البرنامج قيد البحث حاليًا، فقد ترغب في البدء ببرنامج تجريبي يتضمن تقييم الأثر. وفي المقابل، إذا توفرت أدلة من ظروف مشابهة، فمن المحتمل أن تكون تكلفة تقييم الأثر مبررة فقط إذا كان بإمكانها الإجابة عن سؤال مهم وجديد حول السياسة. وسيكون هذا هو الحال إذا كان برنامجك يتضمن بعض الابتكارات المهمة التي لم يتم اختبارها بعد.

لتبرير حشد الموارد التقنية والمالية اللازمة لإجراء تقييم أثر عالي الجودة، يجب أن يكون التدخل المراد تقييمه:

- *مبتكراً*. ستختبر مقارنة جديدة وواعدة.
- *قابلاً للتكرار*. يمكن توسيع نطاقه أو تطبيقه في بيئة مختلفة.
- *ملائماً من الناحية الإستراتيجية*. سيؤجّه الدليل الذي يقدمه تقييم الأثر عملية اتخاذ قرار مهم بشأن التدخل. ويمكن أن يرتبط ذلك بتوسيع نطاق البرنامج، أو الإصلاح، أو مخصصات الميزانية.
- *لم يُختبر*. لا يُعرف إلا القليل عن فعالية بدائل البرنامج أو التصميم، عالمياً أو في سياق محدد.
- *مؤثراً*. ستستخدم النتائج لتوجيه قرارات السياسات.

السؤال الأخير الذي يجب طرحه هو "هل تتوفر لدينا الموارد اللازمة لإجراء تقييم أثر جيد؟" تتعلق هذه الموارد بالعناصر التقنية، مثل البيانات والوقت المناسبين، والموارد المالية لتنفيذ التقييم، فضلاً عن الموارد المؤسسية فيما يتعلق بالفرق المشاركة واهتماماتها والتزامها بجمع الأدلة السببية واستخدامها. وكما نوقش على نحو أكثر تعمقاً في الفصل الثاني عشر، يمثل فريق التقييم في الأساس شراكة بين مجموعتين: فريق من صانعي السياسات وفريق من الباحثين. وتحتاج الفرق إلى العمل على تحقيق الغاية المشتركة المتمثلة في ضمان تنفيذ تقييم جيد التصميم وفعال من الناحية التقنية، وتقديم نتائج ذات صلة بالأسئلة المتعلقة بتصميم البرنامج والسياسة الرئيسية. وسيساعد الفهم الواضح لفرضيات وعود تقييم الأثر من قبل فريق التقييم على ضمان نجاحه. إذا قررت أن تقييم الأثر منطقي بالنظر إلى الأسئلة المطروحة والحاجة ذات الصلة لفحص السببية، والمخاطر المرتبطة بالنتائج، والحاجة إلى أدلة عن أداء برنامجك، فاستمر في القراءة—فهذا الكتاب من أجلك أنت وفريق التقييم.

مصادر إضافية

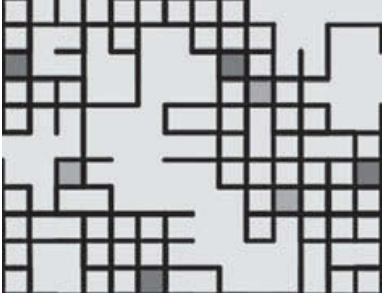
- للحصول على المواد المرفقة مع هذا الفصل وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice).
- للحصول على معلومات إضافية عن تقييمات الأثر، انظر Khandker, Shahidur R., Gayatri B. Koolwal, and Hussain Samad. 2009. *Handbook on Quantitative Methods of Program Evaluation*. Washington, DC: World Bank.
- للحصول على نظرة عامة جيدة عن التجارب العشوائية المضبوطة، انظر Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.
- تشمل المصادر الأخرى عن التجارب العشوائية المضبوطة ما يلي:
 - Duflo, E., R. Glennerster, and M. Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, volume 4, edited by T. Paul Schultz and John Strauss, 3895–962. Amsterdam: Elsevier.
 - Duflo, Esther, and Michael Kremer. 2008. "Use of Randomization in the *Evaluation of Development Effectiveness*." In Vol. 7 of *Evaluating Development Effectiveness*. Washington, DC: World Bank.
- تشمل مصادر تقييم الأثر المفيدة الأخرى ما يلي:
 - Leeuw, Frans, and Jos Vaessen. 2009. *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Washington, DC: NONIE.
 - Ravallion, Martin. 2001. "The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation." *World Bank Economic Review* 15 (1): 115–40.
 - ——. 2007. "Evaluating Anti-Poverty Programs." In Vol. 4 of *Handbook of Development Economics*, edited by T. Paul Schultz and John Strauss. Amsterdam: North Holland.
 - ——. 2009. "Evaluation in the Practice of Development." *World Bank Research Observer* 24 (1): 29–53.

ملاحظات

1. للحصول على نظرة عامة على برامج التحويلات النقدية المشروطة والدور المؤثر الذي يؤديه برنامج المكسيك وتقييم أثره، انظر Fiszbein and Schady (2009).
2. البيانات الإدارية هي تلك البيانات التي يتم جمعها بشكل روتيني كجزء من إدارة البرنامج وتشتمل على معلومات عن التكاليف، والتسجيل، والمعاملات، عادةً كجزء من تقديم الخدمة.
3. للاطلاع على العديد من أنواع التقييمات وأسئلة التقييم، انظر Berk and Rossi (1998) and Rossi, Lipsey, and Freeman (2003).
4. *الأساليب شبه التجريبية* هي أساليب تقييم الأثر التي تستخدم الواقع المضاد، ولكنها تختلف عن الأساليب التجريبية في أن الأساليب شبه التجريبية لا تستند إلى التعيين العشوائي للتدخل. انظر القسم الثاني للاطلاع على نقاش يتناول كلا نوعي الأساليب.
5. للحصول على نظرة عامة على أساليب البحث النوعي، انظر Patton (1990).
6. مقتبس من مكتب مساعدة العدل (١٩٩٧، ٩٧-٩٨ و ٢٠١٠-٣).
7. للاطلاع على مناقشة تفصيلية عن تحليل التكلفة والمنفعة، انظر Zerbe and Dively (1994); Brent (1996); Belli and others (2001); and Boardman and others (2001).

- Ananthpur, Kripa, Kabir Malik, and Vijayendra Rao. 2014. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." Policy Research Working Paper 6958, World Bank, Washington, DC.
- Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper 5245, World Bank, Washington, DC.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, and others. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236). doi:10.1126/science.1260799.
- Behrman, Jere R., and John Hoddinott. 2001. "An Evaluation of the Impact of PROGRESA on Pre-school Child Height." FCND Briefs 104, International Food Policy Research Institute, Washington, DC.
- Belli, Pedro, Jock Anderson, Howard Barnum, John Dixon, and Jee-Peng Tan. 2001. *Handbook of Economic Analysis of Investment Operations*. Washington, DC: World Bank.
- Berk, Richard A., and Peter Rossi. 1998. *Thinking about Program Evaluation*, second edition. Thousand Oaks, CA: Sage Publications.
- Berman, Daphna. 2014. "Tanzania: Can Local Communities Successfully Run Cash Transfer Programs?" Human Development Network, World Bank, Washington, DC.
- Boardman, Anthony, Aidan Vining, David Greenberg, and David Weimer. 2001. *Cost-Benefit Analysis: Concepts and Practice*. New Jersey: Prentice Hall.
- Bourguignon, François, Francisco H. G. Ferreira, and Phillippe G. Leite. 2003. "Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's Bolsa Escola Program." *The World Bank Economic Review* 17 (2): 229–54.
- BRAC (Bangladesh Rural Advancement Committee). 2013. "An End in Sight for Ultra-poverty." BRAC Briefing Note, November. <http://www.brac.net/sites/default/files/BRAC%20Briefing%20-%20TUP.pdf>.
- Brent, Robert. 1996. *Applied Cost-Benefit Analysis*. Cheltenham, U.K.: Edward Elgar.
- Bureau of Justice Assistance. 1997. Urban Street Gang Enforcement. Report prepared by the Institute for Law and Justice, Inc. Washington, DC: Office of Justice Programs, Bureau of Justice Assistance, U.S. Department of Justice.
- Creswell, John W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications.
- Evans, David K., Stephanie Hausladen, Katrina Kosec, and Natasha Reese. 2014. "Community-based Conditional Cash Transfers in Tanzania: Results from a Randomized Trial." World Bank, Washington, DC.

- Fiszbein, Ariel, and Norbert Schady. 2009. *Conditional Cash Transfers, Reducing Present and Future Poverty*. Policy Research Report 47603. Washington, DC: World Bank.
- Gertler, Paul J. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *American Economic Review* 94 (2): 336–41.
- Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.
- Imas, Linda G. M., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.
- Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.
- Khandker, Shahidur, Gayatri B. Koolwal, and Hussain A. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, DC: World Bank.
- Levy, Santiago, and Evelyne Rodríguez. 2005. *Sin Herencia de Pobreza: El Programa Progres-Oportunidades de México*. Washington, DC: Inter-American Development Bank.
- Martinez, Sebastian, Sophie Nadeau, and Vitor Pereira, 2012. "The Promise of Preschool in Africa: A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique." Washington, DC: World Bank and Save the Children.
- McKay, Harrison, Arlene McKay, Leonardo Siniestra, Hernando Gomez, and Pascuala Lloreda. 1978. "Improving Cognitive Ability in Chronically Deprived Children." *Science* 200 (21): 270–78.
- Patton, M. Q. 1990. *Qualitative Evaluation and Research Methods*, second edition. Newbury Park, CA: Sage.
- Rao, Vijayendra, and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, edited by F. J. Bourguignon and L. Pereira da Silva, 165–90. New York: Oxford University Press.
- Rossi, Peter, Mark W. Lipsey, and Howard Freeman. 2003. *Evaluation: A Systematic Approach*, seventh edition. Thousand Oaks, CA: Sage Publications.
- Schultz, Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progres-Oportunidades Program." *Journal of Development Economics* 74 (1): 199–250.
- Skoufias, Emmanuel, and Bonnie McClafferty. 2001. "Is Progres-Oportunidades Working? Summary of the Results of an Evaluation by IFPRI." International Food Policy Research Institute, Washington, DC.
- Todd, Petra, and Kenneth Wolpin. 2006. "Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico." *American Economic Review* 96 (5): 1384–417.
- Zerbe, Richard, and Dwight Dively. 1994. *Benefit Cost Analysis in Theory and Practice*. New York: Harper Collins Publishing.



الفصل الثاني

الإعداد لإجراء التقييم

الخطوات الأولى

يستعرض هذا الفصل الخطوات الأولى في عملية إعداد التقييم. وتشمل الخطوات بناء نظرية التغيير التي تحدد كيف يُفترض بالمشروع تحقيق النتائج المرغوبة، ووضع سلسلة نتائج لتكون أداة مفيدة في تحديد نظرية التغيير، وتحديد سؤال (أسئلة) التقييم، واختيار مؤشرات لتقييم الأداء. هذه الخطوات ضرورية للتحضير لإجراء التقييم، ومن الأفضل اتخاذها في بداية البرنامج أو الإصلاح الذي يخضع للتقييم، وذلك عند استهلال مرحلة التصميم. وتتضمن الخطوات إشراك مجموعة من الأطراف المعنية، من صانعي السياسات إلى منفذي البرنامج، لصياغة رؤية مشتركة لغايات البرنامج وكيفية تحقيقها. وتعمل هذه المشاركة على بناء توافق في الآراء بشأن المسائل محط تركيز التقييم والأسئلة الرئيسية التي يجب الإجابة عنها، وستعزز الروابط بين التقييم، وتنفيذ البرنامج، والسياسة. ويضفي تطبيق الخطوات وضوحًا ودقة مفيدتين في وضع تقييم أثر جيد وتصميم برنامج فعال وتنفيذه. وقد تم تحديد كل خطوة وصياغتها بوضوح داخل النموذج المنطقي المُجسد في سلسلة النتائج، من تحديد دقيق للغايات والأسئلة، إلى صياغة الأفكار المنصوص عليها في نظرية التغيير، وتحديد الحاصل التي يهدف البرنامج إلى توفيرها. وثمة حاجة إلى تحديد واضح للمؤشرات المعينة التي ستستخدم لقياس مدى نجاح البرنامج، ليس فقط لضمان أن يكون التقييم مركز الهدف، ولكن للتأكد أيضًا من أن البرنامج يتضمن أهدافًا محددة جيدًا. علاوة على أنها توفر أساسًا لتحديد أحجام التأثير المتوقعة من البرنامج. ولا غنى عن هذه المعلومات في ترسيخ العناصر التقنية للتقييم، بما في ذلك حجم العينة المطلوبة للتقييم وحسابات القوة الإحصائية، كما تمت مراجعتها في الفصل الخامس عشر.

في معظم تقييمات الأثر، سيكون من المهم تضمين تقدير للتكلفة والمنفعة وفعالية التكلفة، على النحو الموضح في الفصل الأول. ولا يقتصر اهتمام صانعي السياسات دومًا على معرفة البرامج أو الإصلاحات الفعالة فقط، ولكن يهتمون أيضًا بمعرفة التكلفة. وهذا اعتبار غاية في الأهمية في توجيه القرارات المعنية بالبرنامج الذي يمكن توسيع نطاقه وتكراره، وشاغل محوري لقرارات السياسة.

بناء نظرية التغيير

نظرية التغيير هي وصف للكيفية التي من المفترض أن يحقق بها التدخل النتائج المرجوة. وتصف المنطق السببي لكيفية وأسباب قدرة برنامج معين، أو طريقة البرنامج، أو ابتكار التصميم على الوصول إلى الحاصلات المستهدفة. وتعد نظرية التغيير ركيزة أساسية لأي تقييم أثر، وذلك بالنظر إلى تركيز البحث على السبب والنتيجة. وباعتبارها واحدة من الخطوات الأولى في تصميم التقييم، يمكن أن يساعد بناء نظرية التغيير على تحديد أسئلة البحث. تصور نظريات التغيير سلسلة الأحداث التي تقود إلى الحاصلات، وتكشف الظروف والافتراضات اللازمة لحدوث التغيير، وتوضح المنطق السببي وراء تنفيذ البرنامج، وتخطط لتدخلات البرنامج على طول المسارات السببية المنطقية. ويمكن أن يساعد العمل مع الأطراف المعنية في البرنامج لوضع نظرية التغيير على توضيح تصميم البرنامج وتحسينه. ولهذا الأمر أهمية خاصة في البرامج التي تسعى إلى التأثير على السلوك: يمكن أن تساعد نظريات التغيير على فصل مدخلات وأنشطة التدخل، والمخرجات المقدمة، والحاصلات النابعة من التغييرات السلوكية المتوقعة بين المستفيدين. يتمثل أفضل وقت لوضع نظرية التغيير لأحد البرامج في بداية عملية التصميم، عندما يكون من الممكن تجميع الأطراف المعنية لوضع رؤية مشتركة للبرنامج، وغاياته، ومسار تحقيق هذه الغايات. ومن ثم، يمكن للأطراف المعنية بدء تنفيذ البرنامج من منطلق فهم مشترك للبرنامج، وأهدافه، وكيفية عمله.

على مصممي البرنامج أيضًا مراجعة المؤلفات لوضع التجارب مع البرامج المشابهة في الاعتبار، والتحقق من السياقات والافتراضات وراء المسارات السببية في نظرية التغيير التي يحدونها. على سبيل المثال، في حالة المشروع في المكسيك (الموضح في المربع ١-٢) الذي استبدل الأرضيات الترابية بأرضيات إسمنتية، قدمت المؤلفات معلومات قيّمة حول كيفية انتقال الطفيليات وكيف تتسبب العدوى بالطفيليات في إصابة الأطفال بالإسهال.

المربع ١-٢: صياغة نظرية التغيير: من الأرضيات الإسمنتية إلى السعادة في المكسيك

الحفاظ عليها نظيفة. تعيش الطفيليات وتتكاثر في البراز ويمكن أن يبتلعها البشر عندما تدخل إلى المنزل بواسطة الحيوانات أو البشر. وتشير الأدلة إلى أن الأطفال الصغار الذين يعيشون في منازل ذات أرضيات ترابية أكثر عرضة للإصابة بطفيليات معوية، والتي يمكن أن تسبب الإسهال وسوء التغذية، وغالبًا ما تؤدي إلى ضعف النمو المعرفي أو حتى الوفاة. توقف الأرضيات الإسمنتية انتقال العدوى الطفيلية. فضلًا عن أنها تتحكم بشكل أفضل في درجة الحرارة، وهي أكثر تحقيقًا للبهجة من الناحية الجمالية. توجّه هذه الحقائق المتوقعة أسئلة البحث التي تناولها كاتانيو وآخرون (٢٠٠٩) في التقييم. وقد افترضوا أن استبدال الأرضيات الترابية بأخرى إسمنتية سيقول من حدوث الإسهال، وسوء التغذية، ونقص المغذيات الدقيقة. وفي المقابل، يجب أن يؤدي تحسين الصحة والتغذية إلى تحسين النمو المعرفي لدى الأطفال الصغار. بالإضافة إلى ذلك، توقع الباحثون وكشفوا عن حدوث تحسينات في رفاهية البالغين، مقاسًا بزيادة رضا الأشخاص عن وضعهم السكني وانخفاض معدلات الاكتئاب والضغط النفسي الملحوظ.

في تقييم كاتانيو وآخرون (Cattaneo and others, ٢٠٠٩) لمشروع Piso Firme أو "الأرض الصلبة"، بحثوا في تأثير تحسين الإسكان على الصحة والرفاهية. وكانت نظرية تغيير واضحة المحرك لكل من المشروع والتقييم. ويهدف مشروع Piso Firme إلى تحسين مستويات المعيشة، ولا سيما صحة المجموعات المستضعفة التي تعيش في مناطق مكتظة بالسكان ومنخفضة الدخل في المكسيك. وقد انطلق البرنامج في ولاية كواويلا الشمالية، واستند إلى تقدير للحالة أجرته حكومة الولاية. وتتسم سلسلة نتائج البرنامج بالوضوح. وتخضع الأحياء المؤهلة لمسح من الباب للباب، وتُمنح الأسر ما يصل إلى ٥٠ مترًا مربعًا من الإسمنت؛ إذ تشتري الحكومة الإسمنت وتوصله، ويوفر المتطوعون من الأسر والمجتمع المحلي الأيدي العاملة اللازمة لتثبيت الأرضية. والمحصلة هي صب أرضية من الإسمنت، والتي يمكن الانتهاء منها في غضون يوم واحد تقريبًا. وتشمل الحقائق المتوقعة من تحسين بيئة المنزل النظافة، والصحة، والسعادة. يتمثل الأساس المنطقي لسلسلة النتائج هذه في أن الأرضيات الترابية حامل للطفيليات؛ لأنه من الصعب

المصدر: Cattaneo and others 2009.

تمثل سلسلة النتائج إحدى طرق تصوير نظرية التغيير. وتشمل المقاربات الأخرى النماذج النظرية، والنماذج المنطقية، والأطر المنطقية، ونماذج الحصائل. ويتضمن كل واحد من هذه النماذج العناصر الأساسية لنظرية التغيير: سلسلة سببية، وتحديد الظروف والمؤثرات الخارجية، والافتراضات الرئيسية. في هذا الكتاب، سنستخدم نموذج سلسلة النتائج؛ لأننا وجدنا أنه النموذج الأبسط والأوضح لعرض نظرية التغيير في السياق التشغيلي لبرامج التنمية.

ترسي سلسلة النتائج المنطق السببي من بداية البرنامج، بدءًا من الموارد المتاحة، وانتهاءً بأخذ الغايات طويلة الأجل في الاعتبار. وتحدد مخططًا منطقيًا ومعقولًا لكيفية تفاعل تسلسل من المدخلات، والأنشطة، والمخرجات المسؤول عنها البرنامج مسؤولية مباشرة مع السلوك لوضع مسارات يمكن من خلالها تحقيق الآثار (الشكل ٢،١). وتحدد سلسلة النتائج الأساسية العناصر التالية:

المفهوم الأساسي

تحدد سلسلة النتائج
تتابعًا من المدخلات،
والأنشطة، والمخرجات
التي من المتوقع أن
تعمل على تحسين
الحصائل والحصائل
النهائية.

• **المدخلات.** الموارد المتاحة للمشروع، بما في ذلك الموظفين والميزانية.

• **الأنشطة.** الإجراءات المتخذة أو العمل المنجز لتحويل المدخلات إلى مخرجات.

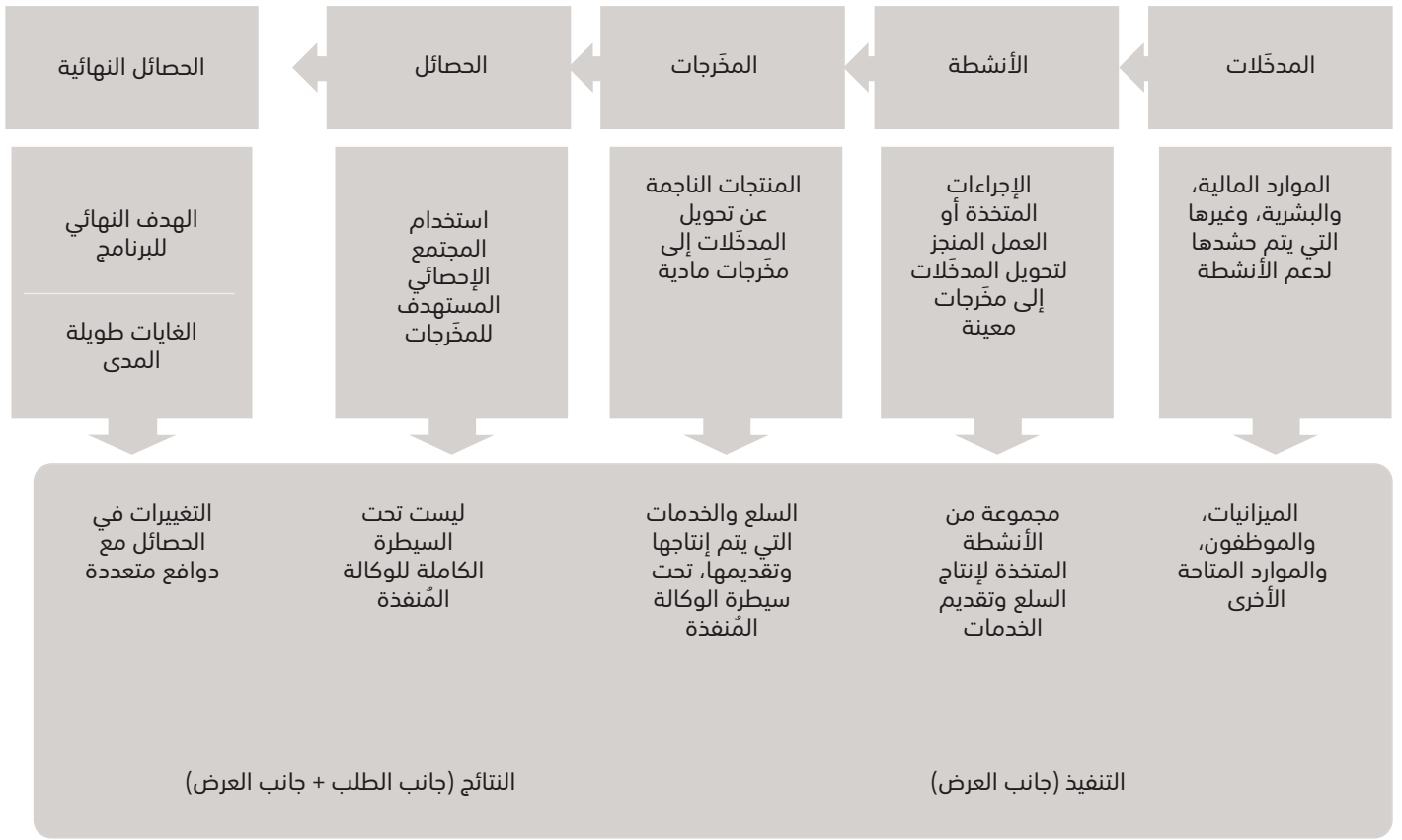
• **المخرجات.** السلع والخدمات المادية التي تنتجها أنشطة المشروع، والتي تخضع للسيطرة المباشرة للوكالة المنفذة.

• **الحصائل.** النتائج المحتملة تحققها بمجرد استخدام الفئة المستفيدة لمخرجات المشروع، وعادة ما تتحقق هذه النتائج على المدى القصير إلى المتوسط ولا تخضع في العادة لسيطرة الوكالة المنفذة.

• **النتائج النهائية** المنجزة والتي تدل إذا تحققت غايات المشروع أم لا. وفي العادة، يمكن أن تتأثر الحصائل النهائية بعدة عوامل، ويمكن تحقيقها على مدى فترة زمنية أطول.

تغطي سلسلة النتائج كلاً من التنفيذ والنتائج. ويتعلق *التنفيذ* بالعمل الذي يقدمه المشروع، بما في ذلك المدخلات، والأنشطة، والمخرجات. وتقع هذه الجوانب تحت المسؤولية المباشرة للمشروع، والتي تخضع عادةً للمتابعة للتحقق مما إذا كان المشروع يقدم السلع والخدمات على النحو المنشود. وتتكون النتائج من الحصائل والحصائل النهائية، والتي لا تخضع للسلطة المباشرة للمشروع وتتوقف على التغييرات السلوكية للمستفيدين من البرنامج. بعبارة أخرى، فإنها تعتمد على التفاعلات بين جانب العرض (التنفيذ) وجانب الطلب (المستفيدون). وهذه هي المجالات التي تخضع في العادة لتقييم الأثر من أجل قياس الفعالية.

الشكل ١-٢ عناصر سلسلة النتائج



ستساعد سلسلة النتائج الجيدة على ظهور الافتراضات والمخاطر الكامنة في نظرية التغيير. ويعد صانعو السياسات الأقدر على توضيح المنطق السببي والافتراضات التي تستند إليها، فضلاً عن المخاطر التي قد تؤثر في تحقيق النتائج المرجوة. ويجب أن يستخلص الفريق الذي يجري التقييم هذه الافتراضات والمخاطر الكامنة بالتشاور مع صانعي السياسات. وستشير سلسلة النتائج الجيدة أيضاً إلى أدلة من المؤلفات بشأن أداء برامج مماثلة.

إن سلاسل النتائج مفيدة لجميع البرامج، بغض النظر عما إذا كانت ستتضمن تقييم أثر أم لا؛ ويرجع ذلك إلى أنها تتيح لصانعي السياسات ومديري البرامج الإعلان صراحةً عن غايات البرنامج، ومن ثمّ، المساعدة على توضيح المنطق السببي وتسلسل الأحداث وراء البرنامج. علاوة على أنه بإمكانها تحديد الثغرات والروابط الضعيفة في تصميم البرنامج، وهكذا يمكنها المساعدة على تحسين تصميم البرنامج. وتسهّل سلاسل النتائج أيضاً عمليتي المتابعة والتقييم من خلال إبراز المعلومات التي تلزم متابعتها على امتداد كل رابط في السلسلة في سبيل تتبع عملية تنفيذ البرنامج، فضلاً عن إبراز مؤشرات النتائج التي يلزم تضمينها عند تقييم المشروع.

سؤال التقييم الواضح هو نقطة البداية لأي تقييم فعال. وتركز صياغة سؤال التقييم على البحث للتأكد من أن السؤال يتناسب مع اهتمامات السياسة المطروحة. في حالة تقييم الأثر، يجب وضعه كفرضية قابلة للاختبار. ومن ثم، يقدم تقييم الأثر أدلة موثوقة للإجابة عن ذلك السؤال. وكما تذكر، فإن سؤال تقييم الأثر الأساسي هو "ما الأثر أو المفعول السببي للبرنامج على الحصيلة محل اهتمام؟" ينصب التركيز على الأثر: بمعنى التغييرات التي يمكن إسنادها مباشرة إلى أحد البرامج، أو طريقة البرنامج، أو ابتكار التصميم.

يجب أن يسترشد سؤال التقييم بالاهتمامات الجوهرية للسياسة المطروحة. وعلى نحو ما نوقش في الفصل الأول، يمكن أن تستطلع تقييمات الأثر مجموعة من الأسئلة. ويجب أن يوضح فريق التقييم الأسئلة التي ستخضع للدراسة كخطوة أولى، بالاستناد إلى نظرية التغيير قبل النظر في كيفية إجراء التقييم.

عادةً ما تركز تقييمات الأثر على تأثير برنامج مُنفذ بالكامل على الحاصلات النهائية، مقارنةً بالنتائج التي لوحظت في مجموعة المقارنة التي لم تستفد من البرنامج. ويتسع نطاق استخدام تقييمات الأثر. ويستطيع فريق التقييم طرح الأسئلة التالية: هل سؤال التقييم الرئيسي هو السؤال "التقليدي" عن فعالية البرنامج في تغيير الحاصلات النهائية؟ أم أنه يتمحور حول اختبار ما إذا كان أحد أشكال البرنامج أكثر فعالية من حيث التكلفة من غيرها؟ أم أنه يتعلق بتقديم تصميم مبتكر للبرنامج من المتوقع أن يغيّر السلوكيات، مثل الالتحاق بالمدارس؟ ويجري تقديم مقاربات جديدة لتقييم الأثر بطرق إبداعية لتناول الأسئلة محل اهتمام السياسة عبر مجموعة من التخصصات (انظر المربع ٢-٢).

في تقييم الأثر، تلزم صياغة سؤال التقييم على أنه فرضية محددة جيداً وقابلة للاختبار. ويجب أن تكون قادرًا على صياغة السؤال بطريقة تتيح لك تحديد الفرق بين النتائج المحرزة عندما تقارن بين مجموعتي المعالجة والمقارنة. ويمكن استخدام سلسلة النتائج كأساس لصياغة الفرضية التي تودّ اختبارها باستخدام تقييم الأثر. مثلما يتضح في المربع ٢-٣، غالبًا ما يكون هناك عدد قليل من الفرضيات المرتبطة بالبرنامج، غير أنه لا يمكن أو يجب تناولها كلها في تقييم الأثر. في مثال منهج الرياضيات المذكور في المربع ٢-٢، تم استخلاص سؤال التقييم من العناصر الأساسية لنظرية التغيير وصياغته كفرضية واضحة، وقابلة للاختبار، وقابلة للقياس الكمي: ما تأثير منهج الرياضيات الجديد على درجات الاختبار؟ في المثال الذي سنطبقه على جميع أجزاء الكتاب على برنامج معونة التأمين الصحي (HISP)، سؤال التقييم هو "ما تأثير برنامج معونة التأمين الصحي على النفقات الصحية التي تدفعها الأسر الفقيرة من مالها الخاص؟"

المربع ٢-٢: تجارب الآلية

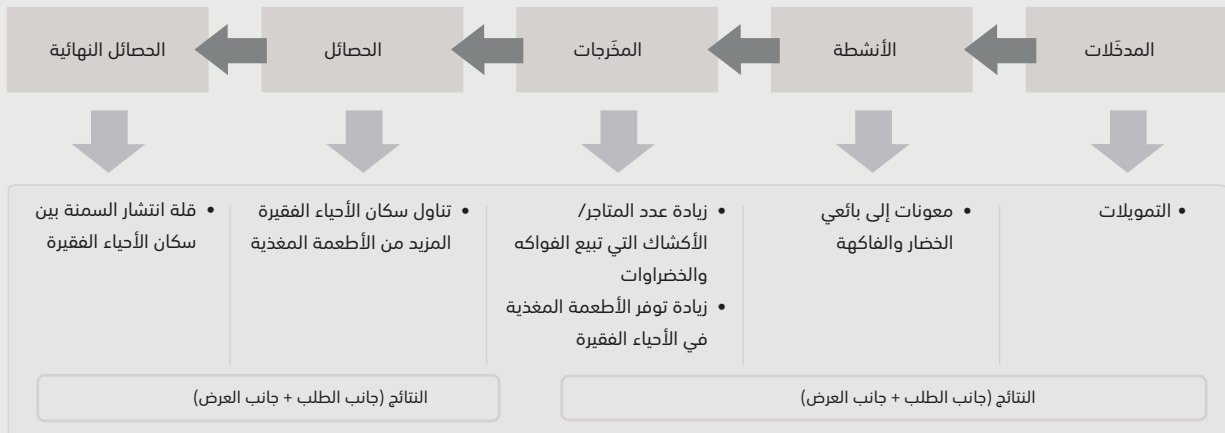
الإصابة بالسمنة، وأنك قد تكون قادرًا على علاج هذا الوضع بتقديم معونات لبائعي الخضار والفاكهة لفتح المزيد من نقاط البيع. ويرد توضيح لسلسلة نتائج بسيطة أدناه (انظر الشكل ب٢-٢-١).

سيركز تقييم البرنامج على اختبار أثر المعونات المقدمة لبائعي الخضار والفاكهة على مجموعة من الأحياء الفقيرة. وعلى النقيض، قد تختبر تجربة الآلية افتراضاتك الضمنية بصورة مباشرة أكثر. على سبيل المثال، يمكن أن تختبر الافتراض التالي: إذا كان سكان الأحياء الفقيرة أكثر قدرة على الوصول إلى الأطعمة المغذية، فسيأكلون منها أكثر. وتتمثل إحدى طرق اختبار ذلك في توزيع سلة أسبوعية مجانية من الفواكه والخضراوات على مجموعة من السكان ومقارنة استهلاكهم للفواكه والخضراوات بهؤلاء الذين لم يحصلوا على السلة المجانية. وإذا لم تجد اختلافات في استهلاكهم من الفواكه والخضراوات في تجربة الآلية هذه، فمن غير المرجح أن يكون لتقديم معونات إلى بائعي الخضار والفاكهة تأثير كبير أيضًا؛ وذلك لأن إحدى الآليات السببية الكامنة معطلة.

تمثل تجربة الآلية تقييم أثر يختبر آلية سببية معينة ضمن نظرية التغيير. لنفترض أنك حددت مشكلة وبرنامجًا محتملاً لعلاج تلك المشكلة. وتفكر في تصميم تقييم لاختبار مدى فعالية البرنامج. هل ينبغي أن يختبر تقييمك أثر البرنامج مباشرة؟ يزعم تيار فكري حديث بأن مثل هذا النوع من تقييم البرنامج قد لا يكون أفضل طريقة دائمًا للبدء، وقد يكون من المفضل في بعض الحالات عدم إجراء تقييم للبرنامج، وإنما اختبار بعض الافتراضات أو الآليات الضمنية. لا تختبر تجارب الآلية البرنامج، ولكنها تختبر الآلية السببية التي تكمن وراء اختيار البرنامج.

على سبيل المثال، قد تكون قلقًا من أن الأشخاص الذين يعيشون في أحياء فقيرة في المدينة يعانون من معدلات سمنة أعلى من هؤلاء الذين يعيشون في مناطق أكثر ثراءً في نفس المدينة. بعد قليل من البحث، لاحظت أن الأحياء الفقيرة تحتوي على عدد أقل من المتاجر والأكشاك التي تباع الخضراوات والفواكه الطازجة والأطعمة المغذية الأخرى. وتفكر أن هذا النقص في الإمداد (العرض) ربما يسهم في

الشكل ب٢-٢-١ تحديد تجربة الآلية من سلسلة نتائج أطول



(تابع)

أرخص بكثير، وسيكون تسجيل بضع مئات من الأسر كافيًا لإجراء التجربة. وإذا أظهرت تجربة الآلية أن تلك الآلية مجدية، فستظل في حاجة إلى تنفيذ تجربة سياسات لتقييم ما إذا كانت المعونات طريقة فعالة لتوفير الفواكه والخضراوات لسكان الأحياء الفقيرة.

ينبغي في العادة أن يكون تنفيذ تجربة الآلية أرخص من إجراء تقييم كامل للبرنامج؛ حيث يمكنك تنفيذها على نطاق أصغر. في مثال السمنة، سيكون تقديم معونات لبائعي الخضر والفاكهة في العديد من الأحياء وإجراء مسح على عدد كبير من سكان هذه الأحياء مكلفًا للغاية. وعلى النقيض من ذلك، ستكون سلال الخضراوات والفواكه المجانية

المصدر: Ludwig, Kling, and Mullainathan 2011.

المربع ٢-٣: إصلاح منهج الرياضيات للمرحلة الثانوية: وضع سلاسل النتائج وأسئلة التقييم

- المعلمين، والتكليف بإصدار الكتب الدراسية الجديدة، وطباعتها، وتوزيعها.
- المخرجات هي عدد المعلمين الذين تم تدريبهم، وعدد الكتب الدراسية التي تم تسليمها إلى الفصول الدراسية، وتعديل الاختبارات الموحدة لتناسب مع المنهج الجديد.
- تتكون الحصائل قصيرة المدى من استخدام المعلمين للأساليب والكتب الدراسية الجديدة في فصولهم الدراسية، واستخدامهم للاختبارات الجديدة.
- الحصائل متوسطة المدى هي تحسينات في مستوى أداء الطلاب في اختبارات الرياضيات الموحدة.
- تتمثل الحصائل النهائية في زيادة معدلات إتمام الدراسة الثانوية ومعدلات توظيف ودخل مكتسب أعلى للخريجين
- تكمّن العديد من الفرضيات وراء نظرية التغيير:
- استخدام المعلمين المدربين للمنهج الجديد بفعالية.

(تابع)

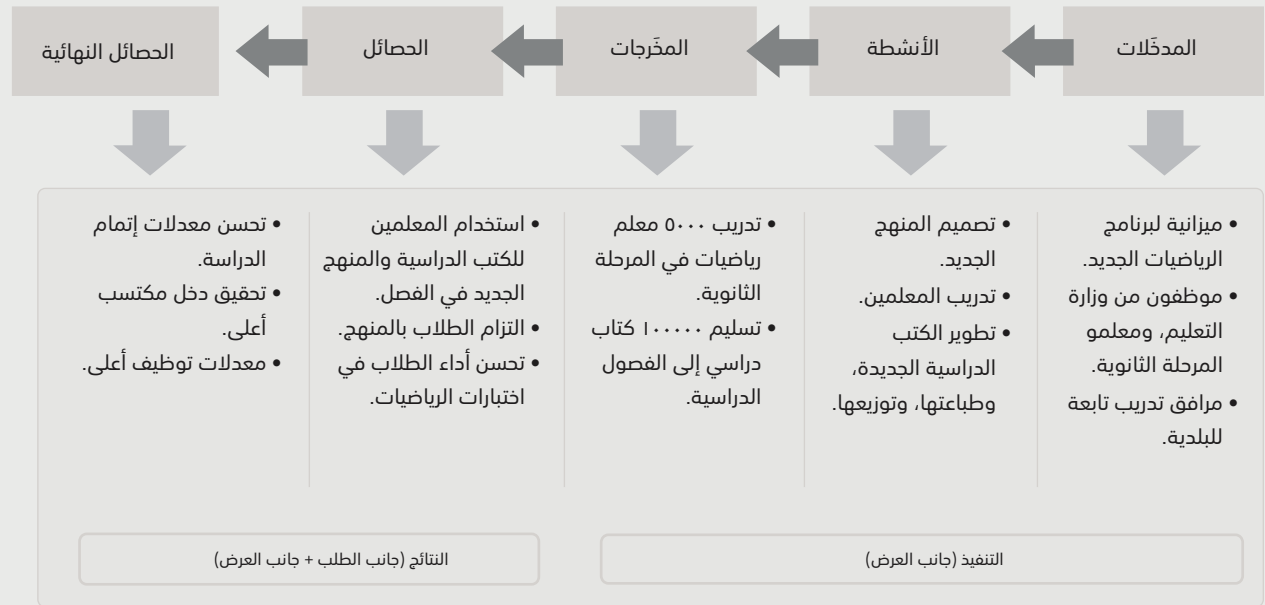
تخيّل أن وزارة التعليم في دولة A تفكر في تقديم منهج رياضيات جديد للمرحلة الثانوية. ويهدف هذا المنهج الدراسي إلى أن يكون أكثر سهولة للمعلمين والطلاب، وتحسين مستوى أداء الطلاب في اختبارات الرياضيات الموحدة، وفي النهاية، تحسين قدرة الطلاب على إتمام الدراسة الثانوية والحصول على وظائف أفضل. وتوضح سلسلة النتائج التالية نظرية التغيير الخاصة بالبرنامج (انظر الشكل ب ٢-٣-١).

- تشمل المدخلات الموظفين من وزارة التعليم لتصدر عملية الإصلاح، ومعلمي الرياضيات في المدارس الثانوية، وميزانية لتطوير منهج رياضيات جديد، والمرافق البلدية حيث سيتلقى معلمو الرياضيات التدريب على المنهج الجديد.
- تتألف أنشطة البرنامج من تصميم منهج الرياضيات الجديد، ووضع برنامج تدريب للمعلمين، وتدريب

المربع ٢-٣: إصلاح منهج الرياضيات للمرحلة الثانوية: وضع سلاسل النتائج وأسئلة التقييم (تابع)

- إذا تم تدريب المعلمين وتوزيع الكتب الدراسية، في معدلات إتمام المرحلة الثانوية، وفرص الحصول على فستستخدم هذه الكتب وسيلتزم الطلاب بالمنهج وظائف، وتحقيق دخل مكتسب.
- الدراسي.
- يتفوق المنهج الجديد على القديم في نقل المعرفة صانعي السياسات في وزارة التعليم والباحثين المشاركين بالرياضيات.
- إذا تم التنفيذ كما هو مخطط له، فستحسن نتائج اختبار الجديد على درجات الاختبار؟" يمس هذا السؤال صميم الرياضيات بمقدار ٥ نقاط في المتوسط.
- يؤثر مستوى الأداء في الرياضيات في المدرسة الثانوية

الشكل ب ٢-٣ سلسلة نتائج لإصلاح منهج الرياضيات للمرحلة الثانوية



برنامج معونة التأمين الصحي (HISP): مقدمة



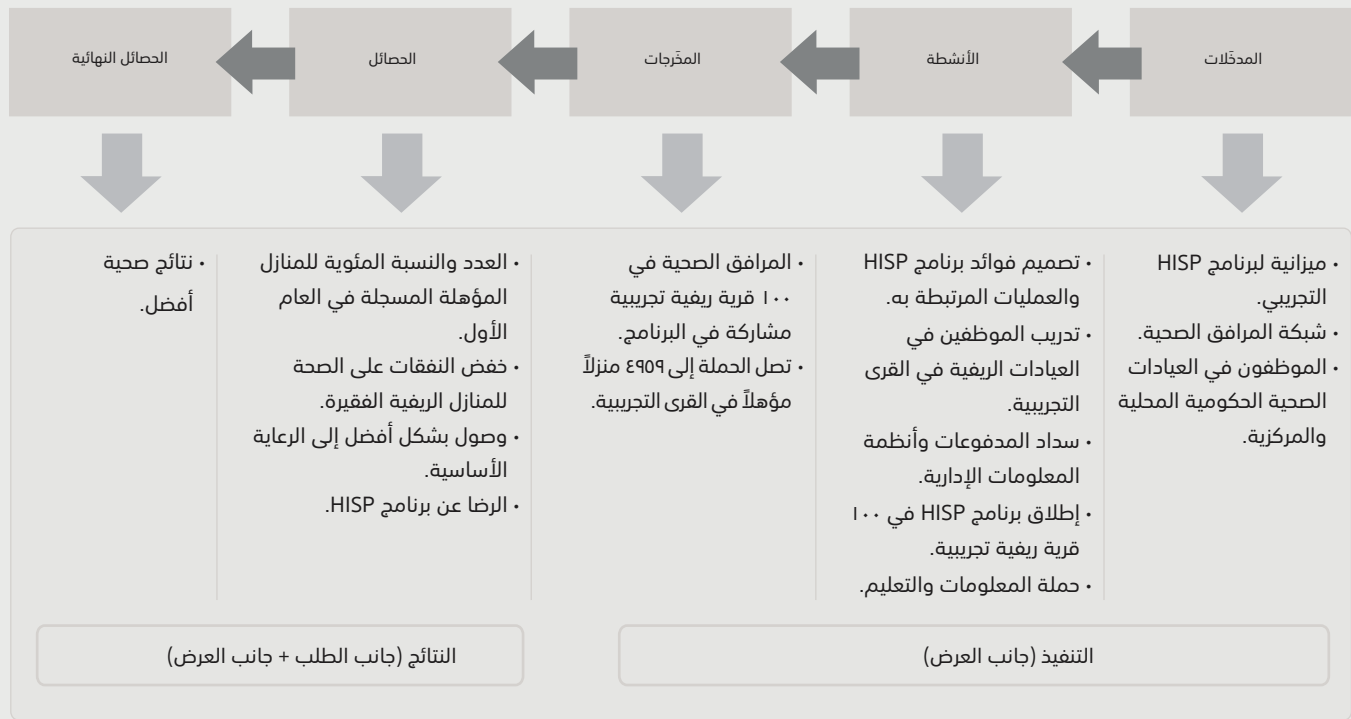
برنامج معونة التأمين الصحي (HISP) هو حالة خيالية لحكومة تضطلع بإصلاحات على نطاق واسع في القطاع الصحي. وستستخدم الأسئلة ذات الصلة بهذه الحالة في جميع أجزاء الكتاب. يحتوي موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice) على حلول لأسئلة دراسة حالة برنامج معونة التأمين الصحي، ومجموعة بيانات، ورمز التحليل في برنامج Stata، بالإضافة إلى رفيق تقني عبر الإنترنت يقدم معالجة أكثر رسمية لتحليل البيانات.

يتمثل الهدف النهائي لبرنامج معونة التأمين الصحي (HISP) في تحسين مستوى صحة سكان البلد. ويجري تنفيذ هذا البرنامج المبتكر، والمكلف على الأرجح، بصورة تجريبية. ويساور الحكومة القلق إزاء عدم قدرة الأسر الريفية الفقيرة على تحمل تكاليف الرعاية الصحية الأساسية، مع ما يترتب على ذلك من عواقب وخيمة على صحتهم. ولحل هذه المشكلة، يقدم برنامج معونة التأمين الصحي الدعم للتأمين الصحي للأسر الريفية الفقيرة؛ حيث يغطي التكاليف المتعلقة بالرعاية الصحية الأساسية والأدوية. ويتمحور الهدف الرئيسي للبرنامج حول تقليل تكلفة الرعاية الصحية للأسر الفقيرة، وفي النهاية، تحسين الحاصلات الصحية. ويدرس صانعو السياسات توسيع نطاق برنامج معونة التأمين الصحي ليغطي البلد بأكمله، والذي سيكلف مئات الملايين من الدولارات.

ترد سلسلة نتائج برنامج معونة التأمين الصحي في الشكل ٢-٢. تفترض الفرضيات المتعلقة بإصلاح برنامج معونة التأمين الصحي التالي: إقدام الأسر على التسجيل في البرنامج بمجرد توفيره، وسيقلل التسجيل في البرنامج من حجم نفقات الأسر على الصحة من مالها الخاص، وتحول التكاليف دون حصول سكان الريف على الرعاية الصحية والأدوية المتاحة، وتمثل النفقات من المال الخاص على التكاليف ذات الصلة بالصحة عاملاً أساسياً له دور في الفقر والحاصلات الصحية السيئة.

سؤال التقييم الرئيسي هو: ما تأثير برنامج معونة التأمين الصحي على النفقات الصحية التي تدفعها الأسر الفقيرة من مالها الخاص؟ على مدار الكتاب وفي المواد عبر الإنترنت، سنجيب عن سؤال التقييم ذاته فيما يتعلق ببرنامج معونة التأمين الصحي مرات عديدة. باستخدام أساليب منهجية مختلفة. وستلاحظ ظهور إجابات مختلفة، وأحياناً متضاربة، وذلك استناداً إلى منهجية التقييم المستخدمة.

الشكل ٢-٢ سلسلة نتائج برنامج معونة التأمين الصحي (HISP)



اختيار مؤشرات الحاصل والأداء

يجب أن يُلَاحَظ سؤال التقييم الواضح تحديد مقاييس الحاصل التي سيتم استخدامها لتقييم النتائج، بما في ذلك في حالة الحاصل المتعددة. وستُستخدم مقاييس الحاصل المختارة لتحديد ما إذا كان برنامج أو إصلاح معين ناجحًا أم لا. علاوة على أنها المؤشرات التي يمكن الرجوع إليها في تطبيق حسابات القوة الإحصائية المستخدمة في تحديد أحجام العينات اللازمة للتقييم، على نحو ما نوقش في الفصل الخامس عشر.

بمجرد اختيار المؤشرات الرئيسية موضع الاهتمام، يجب وضع أهداف واضحة فيما يتعلق بنجاح البرنامج. وتُعد هذه الخطوة تحديد التأثير المتوقع للبرنامج على مؤشرات الحاصل الأساسية التي تم اختيارها. *أحجام التأثير* هي التغييرات المتوقعة نتيجة البرنامج أو الإصلاح، مثل التغيير في درجات الاختبار أو معدل الإقبال على نوع جديد من بوليصة التأمين. وتمثل أحجام التأثير المتوقعة الأساس اللازم لإجراء حسابات القوة الإحصائية.

من الأهمية بمكان أن تتفق الأطراف المعنية الرئيسية في فريق التقييم (كل من فريق البحث وفريق السياسات) على مؤشرات الحاصل الأولية محل الاهتمام في تقييم الأثر وأحجام التأثير المتوقعة الناجمة عن البرنامج أو الابتكار (لمزيد من المعلومات عن فريق التقييم، انظر الفصل الثاني عشر). وهذه هي المؤشرات التي ستُستخدم للحكم على نجاح البرنامج، وتشكل الأساس لحسابات القوة الإحصائية. ويمكن أن تفشل تقييمات الأثر؛ لأنها لا تتضمن أحجام عينات كبيرة بما يكفي لاكتشاف التغييرات التي نتجت عن البرنامج، ويرجع ذلك إلى كونها "منخفضة القوة الإحصائية". ويجب تحديد الحد الأدنى لأحجام التأثير المتوقعة من أجل وضع معايير أساسية لنجاح البرنامج أو الابتكار. وعندما تتوفر البيانات، يمكن إجراء عمليات محاكاة سابقة للنظر في سيناريوهات الحاصل المختلفة من أجل وضع أساس مقارنة لنوع أحجام التأثير التي يمكن توقعها عبر مجموعة من المؤشرات. ويمكن أيضًا استخدام عمليات المحاكاة السابقة للنظر في المقاييس الأولية للتكلفة والمنفعة أو فعالية التكلفة، ولمقارنة التدخلات البديلة في سبيل إحداث تغييرات في الحاصل محل الاهتمام.

تقدم سلسلة النتائج المحددة بوضوح خريطة مفيدة لاختيار المؤشرات التي ستخضع للقياس على طول السلسلة. وتشمل المؤشرات المستخدمة في كل من متابعة تنفيذ البرنامج وتقييم النتائج. مرة أخرى، من المفيد إشراك الأطراف المعنية في البرنامج من فرق السياسات والبحث في اختيار هذه المؤشرات؛ وذلك لضمان أن تكون تلك المؤشرات المختارة مقاييس جيدة لأداء البرنامج. وهناك طريقة مجربة واسعة الاستخدام للتأكد من أن المؤشرات المستخدمة مقاييس جيدة، ويمكن تلخيصها في الاختصار SMART. يجب أن تكون المؤشرات على النحو التالي:

المفهوم الأساسي
تتميز المؤشرات الجيدة بأنها SMART (محددة، وقابلة للقياس، ومنسوبة، وواقعية، ومستهدفة).

- **محددة (S):** لقياس المعلومات المطلوبة قدر الإمكان
- **قابلة للقياس (M):** لضمان إمكانية الحصول على المعلومات بسهولة
- **منسوبة (A):** لضمان أن كل مقياس مرتبط بجهود المشروع
- **واقعية (R):** لضمان إمكانية الحصول على البيانات في الوقت المناسب، وبمعدل تكرار معقول، وبتكلفة معقولة
- **مستهدفة (T):** إلى المجتمع الإحصائي المستهدف.

عند اختيار المؤشرات، تذكر أنه من المهم تحديد المؤشرات على طول سلسلة النتائج، وليس على مستوى الحصائل فحسب؛ بحيث تصبح قادرًا على تتبع المنطق السببي لأي حصائل للبرنامج تتم ملاحظتها. في الواقع، مع تقييمات التنفيذ التي تركز على اختبار بديلين أو أكثر من بدائل التصميم، قد تحدث النتائج محل الاهتمام في وقت مبكر في سلسلة النتائج، في شكل مخرج مسبق أو حصيلة في مرحلة مبكرة. وحتى إذا كنت مهتمًا فقط بمقاييس الحصائل من أجل التقييم، فلا يزال من المهم تتبع مؤشرات التنفيذ؛ بحيث يمكنك تحديد ما إذا تم تنفيذ التدخلات كما هو مخطط لها، وما إذا وصلت إلى المستفيدين المستهدفين، وما إذا وصلت في الوقت المحدد. ودون هذه المؤشرات على طول سلسلة النتائج، يخاطر تقييم الأثر بإنشاء "صندوق أسود" يحدد ما إذا كانت النتائج المتوقعة قد تحققت أم لا، ولكنه لن يكون قادرًا على شرح سبب ذلك.

قائمة المراجعة: جمع البيانات لمؤشراتك

من المفيد بعد اختيار المؤشرات النظر في ترتيبات إنتاج البيانات لقياس المؤشرات، وذلك من خلال استخدام قائمة المراجعة النهائية. وترد مناقشة كاملة حول مكان الحصول على بيانات من أجل تقييمك في القسم الرابع. وتغطي قائمة المراجعة هذه الترتيبات العملية اللازمة لضمان قدرتك على إنتاج كل واحد من المؤشرات بشكل موثوق وفي الوقت المحدد (مقتبس من برنامج الأمم المتحدة الإنمائي ٢٠٠٩):

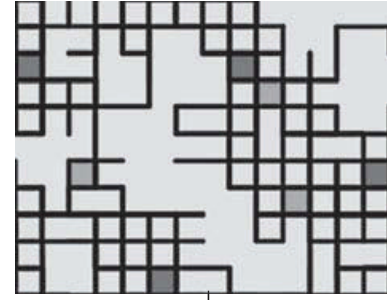
- ✓ هل المؤشرات (المخرجات والحصائل) محددة بوضوح؟ هذه المؤشرات مستمدة من أسئلة التقييم الأساسية، ويجب أن تكون متسقة مع وثائق تصميم البرنامج وسلسلة النتائج.
- ✓ هل يمكن أن نطلق على المؤشرات أنها SMART (محددة، وقابلة للقياس، ومنسوبة، وواقعية، ومستهدفة)؟
- ✓ ما مصدر البيانات لكل واحد من المؤشرات؟ يجب أن يكون هناك وضوح بشأن المصدر الذي ستحصل منه على البيانات، مثل مسح، أو مراجعة، أو بيانات إدارية.
- ✓ ما معدل تكرار جمع البيانات؟ قم بتضمين جدول زمني.
- ✓ من المسؤول عن جمع البيانات؟ حدّد المسؤول عن تنظيم عملية جمع البيانات، والتحقق من جودة البيانات ومصدرها، وضمان الامتثال للمعايير الأخلاقية.
- ✓ من المسؤول عن التحليل ورفع التقارير؟ حدّد معدل تكرار إجراء التحليل، وطريقة التحليل، ومسؤولية رفع التقارير.
- ✓ ما الموارد اللازمة لإنتاج البيانات؟ تأكد من أن الموارد اللازمة واضحة وملزمة بإنتاج البيانات، وغالبًا ما تكون الجزء الأكثر تكلفة في التقييم عند جمع البيانات الأولية.
- ✓ هل هناك توثيق مناسب؟ يجب تطبيق الخطط التي تحدد كيفية توثيق البيانات، بما في ذلك استخدام سجل وضمان عدم الكشف عن الهوية.
- ✓ ما المخاطر المتضمنة؟ ضع في حسابك المخاطر والافتراضات المتضمنة في تنفيذ أنشطة المتابعة والتقييم المخطط لها، وكيف يمكن أن تؤثر على توقيت وجودة البيانات والمؤشرات.

مصادر إضافية

- للحصول على المواد المرفقة مع هذا الفصل وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice).
- يرد في الوحدة الأولى من مجموعة أدوات تقييم الأثر التابعة للبنك الدولي شكلٌ يبين نظرية التغيير، ونموذجٌ لسلسلة النتائج، وأمثلة على مؤشرات التمويل القائم على النتائج (www.worldbank.org/health/impactevaluationtoolkit).
- توجد مراجعة جيدة لنظريات التغيير في Imas, Linda G. M., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.
- للاطلاع على مناقشات حول كيفية اختيار مؤشرات الأداء، ارجع إلى ما يلي: Imas, Linda G. M., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.
- Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: World Bank.

المراجع

- Cattaneo, Matias, Sebastian Galiani, Paul Gertler, Sebastian Martinez, and Rocio Titiunik. 2009. "Housing, Health and Happiness." *American Economic Journal: Economic Policy* 1 (1): 75–105.
- Imas, Linda G. M., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.
- Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: World Bank.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. "Mechanism Experiments and Policy Evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.
- UNDP (United Nations Development Programme). 2009. *Handbook on Planning, Monitoring and Evaluating for Development Results*. New York: UNDP.



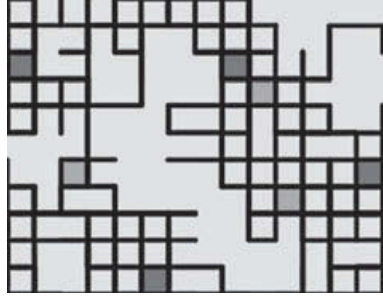
الجزء الثاني كيفية التقييم

يشرح الجزء الثاني من هذا الكتاب ما تفعله تقييمات الأثر، والأسئلة التي تجيب عنها، والطرق المتاحة لإجرائها، ومزايا كل منها وعيوبه. يفضّل مقارنة تقييم الأثر الذي يدعو إليه هذا الكتاب اختيار الطريقة الأكثر صرامة والمتوافقة مع الخصائص التشغيلية للبرنامج. تتضمن قائمة خيارات تقييم الأثر التي نوقشت التعيين العشوائي، والمتغيرات المساعدة، وتصميم انقطاع الانحدار، والاختلاف في الاختلافات، والمطابقة. تشترك جميع هذه المقاربات في هدف إنشاء مجموعات مقارنة صحيحة، بحيث يمكن تقدير الآثار الحقيقية للبرنامج.

نستهل الفصل الثالث بتقديم مفهوم الواقع المضاد بصفته جوهر أي تقييم للأثر، وشرح الخصائص التي يجب أن يمتلكها تقدير الواقع المضاد، وتقديم أمثلة عن تقديرات غير صحيحة أو زائفة عن الواقع المضاد. تناقش الفصول من الرابع إلى الثامن بعد ذلك كل مقارنة، حيث يتم تناول التعيين العشوائي في الفصل الرابع، والمتغيرات المساعدة في الفصل الخامس، وتصميم انقطاع الانحدار في الفصل السادس، والاختلاف في الاختلافات في الفصل السابع، والمطابقة في الفصل الثامن. نناقش لماذا تنتج كل طريقة تقديرًا صالحًا للواقع المضاد وكيف يمكنها ذلك، حيث يمكن تنفيذ سياق السياسة، والقيود الرئيسية لكل طريقة. نوضح استخدام كل طريقة باستخدام أمثلة واقعية محددة لتقييمات الأثر التي استخدمت هذه الطرق، إضافة إلى دراسة حالة برنامج معونة التأمين الصحي (HISP) التي قُدمت في الفصل الثاني. نناقش في الفصل التاسع كيفية معالجة الإشكاليات التي تنشأ خلال التنفيذ، مع ملاحظة أن تقييمات الأثر لا تُنفذ غالبًا كما هي مصممة بالضبط. وفي هذا السياق، نراجع التحديات الشائعة، بما في ذلك التوافق غير المثالي، والتأثيرات غير المباشرة، والتناقص، ونقدم التوجيه بشأن كيفية معالجة هذه المشكلات. يُختتم الفصل العاشر بإرشادات حول تقييمات البرامج متعددة الأوجه، وخصوصًا تلك التي تضم مستويات معالجة مختلفة وأذرع معالجة مختلفة.

خلال الجزء الثاني، ستتاح لك فرصة تطبيق الطرق واختبار فهمك باستخدام دراسة حالة برنامج معونة التأمين الصحي (HISP). تذكر أن سؤال التقييم الرئيسي لصانعي سياسات برنامج معونة التأمين الصحي (HISP) هو: ما أثر برنامج معونة التأمين الصحي (HISP) على النفقات الصحية المدفوعة من المال الخاص لدى الأسر الفقيرة؟ سنستخدم مجموعة بيانات برنامج معونة التأمين الصحي (HISP) لتوضيح كل عملية تقييم ومحاولة الإجابة عن هذا السؤال. يتعين عليك افتراض أن البيانات قد جُمعت بالفعل على النحو الصحيح لنفي أي مشكلات متعلقة بالبيانات. سيقدم إليك الكتاب نتائج التحليل التي سيطلب منك تفسيرها. وبالتحديد، مهمتك هي تحديد السبب في أن تقدير أثر برنامج معونة التأمين الصحي (HISP) يتغير تبعًا لكل طريقة، وتحديد أي نتائج تعد موثوقة على نحو كافٍ كي تبرّر ما إذا كان ينبغي توسيع برنامج معونة التأمين الصحي (HISP) أو لا. تُقدّم حلول الأسئلة في موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice). إذا كنت مهتمًا بتكرار التحليل، فستجد أيضًا مجموعة البيانات، ورمز التحليل في برنامج Stata، ودليلًا فنيًا يوفر معالجة أكثر رسمية لتحليل البيانات على هذا الموقع.

يبدأ الجزء الثالث بكيفية استخدام قواعد تشغيل البرنامج - أي الموارد المتاحة للبرنامج، ومعايير اختيار المستفيدين، وتوقيت التنفيذ - بمثابة أساس لاختيار طريقة تقييم الأثر. وقد وُضع إطار عمل بسيط لتحديد أي من أساليب تقييم الأثر التي استُعرضت في الجزء ٢ هي الأنسب لبرنامج معين اعتمادًا على قواعده التشغيلية.



الفصل الثالث

الاستدلال السببي والواقع المضاد

الاستدلال السببي

نبدأ بدراسة مفهومين يشكلان جزءًا أساسيًا من عملية إجراء تقييمات أثر دقيقة وموثوقة، وهما الاستدلال السببي والواقع المضاد.

تنطوي العديد من مسائل السياسات على علاقات بين السبب والتأثير. هل يحسّن تدريب المعلمين نتائج الطلاب في الاختبارات؟ هل تحسّن برامج التحويل النقدي المشروط الحاصل الصحية لدى الأطفال؟ هل تزيد برامج التدريب المهني دخل المتدربين؟

المفهوم

الأساسي

تحدد تقييمات الأثر مدى تسبّب برنامج ما - ذلك البرنامج فقط - في تغيير في حصة ما.

تسعى تقييمات الأثر إلى الإجابة بدقة عن أسئلة السبب والتأثير. يعد تقييم أثر برنامج ما على مجموعة من الحاصلات مكافئًا لتقييم الأثر السببي للبرنامج على تلك الحاصلات. وعلى الرغم من أن أسئلة السبب والتأثير شائعة إلا أن الإجابة عنها قد تشكل تحديًا. ففي سياق برنامج التدريب المهني على سبيل المثال، لا يكفي مجرد ملاحظة زيادة دخل المتدرب بعد إكماله لبرنامج من هذا القبيل لإثبات السببية. فقد يكون دخل المتدرب قد ازداد حتى ولو لم يتلقَ تدريبًا - بسبب جهوده الخاصة أو نتيجةً للظروف المتغيرة لسوق العمل، أو بفعل عوامل

متعددة يمكن أن تؤثر على الدخل. تساعدنا تقييمات الأثر على التغلب على تحدي تأسيس علاقة سببية عبر التحديد التجريبي لمدى مساهمة برنامج معين - وهذا البرنامج بمفرده - في التغيير في حصة ما. لتحديد العلاقة السببية بين برنامج وحصة ما، نستخدم طرق تقييم الأثر لاستبعاد احتمال أن تفسر أي عوامل أخرى غير البرنامج محل الاهتمام الأثر الملحوظ.

الإجابة عن سؤال تقييم الأثر الأساسي—ما الأثر أو المفعول السببي للبرنامج (P) على الحصيلة محل الاهتمام (Y)؟— يمكننا إيجاد النتيجة من خلال المعادلة الأساسية لتقييم الأثر:

$$\Delta = (Y | P = 1) - (Y | P = 0).$$

تنص المعادلة على أن الأثر السببي (Δ) لبرنامج ما (P) على حصيلة (Y) هو الفرق بين الحصيلة (Y) بوجود البرنامج (بمعنى آخر، عندما تكون $P = 1$) والحصيلة نفسها (Y) بدون البرنامج (أي عندما تكون $P = 0$). على سبيل المثال، إذا كان P يرمز لبرنامج تدريب مهني، و Y يرمز لحصيلة، فإن الأثر السببي لبرنامج التدريب المهني (Δ) هو الفرق بين دخل شخص ما (Y) بعد المشاركة في برنامج التدريب المهني (بمعنى آخر، عندما تكون $P = 1$) وبين دخل الشخص نفسه (Y) في نفس النقطة الزمنية لو أنه لم يشارك في البرنامج (بمعنى آخر، عندما تكون $P = 0$). بمعنى آخر، نوّد قياس الدخل عند النقطة الزمنية نفسها لوحدة الملاحظة نفسها (في هذه الحالة، شخص)، ولكن في حالة مختلفة. لو أمكن ذلك، فإننا سنكون نلاحظ مقدار الدخل الذي كان الشخص سيحققه في نقطة زمنية معينة بوجود البرنامج وبدونه، بحيث يكون التفسير الممكن الوحيد لأي فرق في دخل الفرق هو البرنامج. وبمقارنة نفس الفرد مع نفسه في اللحظة نفسها، سنتمكن من تنحية أي عوامل خارجية قد تكون فسرت أيضًا الفرق في الحصائل. عندها، لنا أن نثق أن العلاقة بين برنامج التدريب المهني والتغير في الدخل هي علاقة سببية. تصلح تلك المعادلة الأساسية لتقييم الأثر لأي وحدة خاضعة للتحليل؛ فرد أو أسرة أو مجتمع أو شركة أو مدرسة أو مستشفى أو أي وحدة مراقبة أخرى قد تتلقى البرنامج أو تتأثر به. والمعادلة صالحة أيضًا لأي حصيلة (Y) متعلقة بالبرنامج المعني. بمجرد أن نقيس المكونين الرئيسيين لهذه المعادلة—الحصيلة (Y) بوجود البرنامج وبدونه—يمكننا الإجابة عن أي سؤال عن أثر البرنامج.

الواقع المضاد

كما ناقشنا، يمكننا التفكير في أثر (Δ) برنامج ما على أنه الفرق في الحصائل (Y) للوحدة نفسها (شخص، أسرة، مجتمع، وما إلى ذلك) مع المشاركة في البرنامج وبدونها. ولكننا نعلم أن قياس الوحدة نفسها في حالتين مختلفتين في الوقت نفسه مستحيل. ففي أي وقت من الأوقات، إما أن تكون الوحدة قد شاركت في البرنامج أو لم تشارك. لا يمكن ملاحظة الوحدة في الوقت نفسه في حالتين مختلفتين (بعبارة أخرى، مع البرنامج وبدونه). يُسمى ذلك إشكالية الواقع المضاد: كيف نقيس ما كان سيحدث لو أن ظروفًا أخرى كانت سائدة؟ على الرغم من أن بإمكاننا ملاحظة الحصيلة (Y) لمشارك في برنامج ($Y | P = 1$) وقياسها، لا توجد بيانات لتحديد حصيلتها في غياب البرنامج ($Y | P = 0$). في المعادلة الأساسية لتقييم الأثر، يمثل الحد ($Y | P = 0$) الواقع المضاد. يمكنك التفكير في ذلك على أنه ما كان سيحدث للحصيلة لو أن شخصًا أو أن وحدة الملاحظة لم تشارك في البرنامج.

على سبيل المثال، تخيل أن "السيد عاثر الحظ" تناول قرص دواء ثم توفي بعد خمسة أيام. إن موت السيد عاثر الحظ بعد تناول قرص الدواء لا يجعلك تستنتج أن قرص الدواء قد تسبب في وفاته. فربما كان مريضاً مرضاً شديداً عندما تناول قرص الدواء، وكان ذلك المرض هو سبب وفاته، وليس قرص الدواء. يتطلب الاستدلال على السببية أن تستبعد العوامل المحتملة الأخرى التي يمكن أن تكون قد أثرت على النتيجة قيد الدراسة. في مثال تحديد ما إذا كان تناول قرص الدواء قد تسبب في وفاة السيد عاثر الحظ، سيحتاج المقيّم إلى تحديد ما كان سيحدث للسيد عاثر الحظ لو أنه لم يتناول قرص الدواء. بما أن السيد عاثر الحظ تناول قرص الدواء بالفعل، فمن غير الممكن أن نلاحظ مباشرة ما كان سيحدث لو أنه لم يفعل ذلك. إن ما كان سيحدث له لو أنه لم يتناول قرص الدواء واقعٌ مضاد. لتحديد أثر قرص الدواء، يتمثل التحدي الرئيسي أمام المقيّم في تحديد ماهية حالة الواقع المضاد للسيد عاثر الحظ (انظر المربع ١-٣ للاطلاع على مثال آخر).

عند إجراء تقييم أثر، من البسيط نسبياً الحصول على الحد الأول للمعادلة الأساسية ($P = 1 \mid Y$)—الحصيلة بوجود البرنامج (تُعرف أيضًا بـ الحصيلة مع إجراء المعالجة). نقيس ببساطة الحصيلة موضع الاهتمام للمشاركة في البرنامج. ولكن، لا يمكننا ملاحظة الحد الثاني في المعادلة ($P = 0 \mid Y$) للمشاركة مباشرة. نحتاج إلى ملء هذه المعلومة المفقودة عبر تقدير الواقع المضاد.

للمساعدة على التفكير في مفهوم تقدير الواقع المضاد الرئيسي هذا، ننتقل إلى مثال افتراضي آخر. سيكون حل مسألة الواقع المضاد ممكناً لو تمكّن المقيّم من التوصل إلى "قرين مثالي" للمشاركة في البرنامج (الشكل ١-٣). على سبيل المثال، لنقل إن السيد فلان بدأ يتلقى ١٢ دولاراً أمريكياً ضمن مصروف الجيب، ونريد قياس أثر هذه المعالجة على استهلاكه من الحلوى. لو استطعت تحديد قرين مثالي للسيد فلان، فسيكون التقييم سهلاً: يمكنك مقارنة عدد قطع الحلوى التي يتناولها السيد فلان (لنقل ٦) عندما يتلقى مصروف الجيب مع عدد قطع الحلوى التي يتناولها قرينه (لنقل ٤) الذي لا يتلقى مصروف الجيب. وفي هذه الحالة، سيكون أثر مصروف الجيب قطعتي حلوى: الفرق بين عدد قطع الحلوى المستهلكة مع إجراء المعالجة (٦) وعدد قطع الحلوى المستهلكة بدون المعالجة (٤). في الواقع، نعلم أن من المستحيل تحديد أقران مثاليين: حتى بين التوائم المتطابقة وراثياً، فهناك فروقات هامة.

المفهوم

الأساسي

الواقع المضاد هو ما كان سيحدث—أي الحصيلة (Y) للمشاركة في البرنامج—بغياب البرنامج (P).

المفهوم

الأساسي

بما أننا لا نستطيع ملاحظة الواقع المضاد مباشرة، علينا تقديره.

المربع (١-٣): إشكالية الواقع المضاد: "الطفلة المميزة" وبرنامج التحويل النقدي

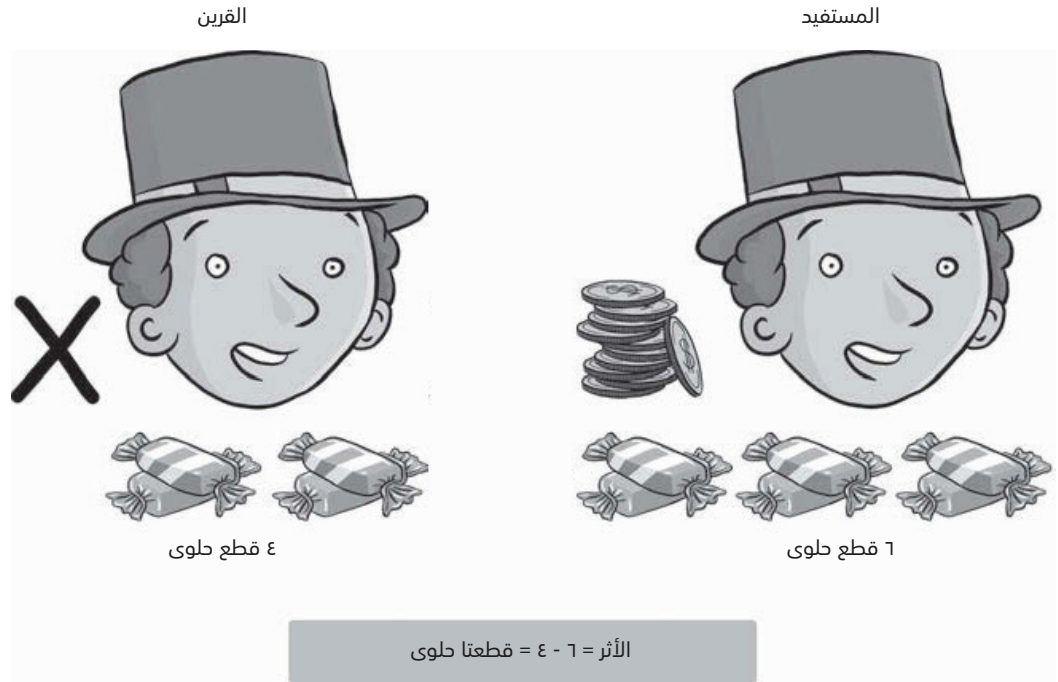
تحقق الأسرة الشروط (الفحوصات واللقاحات ومراقبة النمو) وتتلقى التحويل النقدي أو لا تتلقاه. بمعنى آخر، لا يمكننا ملاحظة ماهية الواقع المضاد. وبما أن والدة الطفلة المميزة كانت تحقق الشروط بالفعل وتتلقى التحويل النقدي، فلا يمكننا معرفة الطول الذي ستكون عليه الطفلة مميزة لو أن والدتها لم تكن تتلقى التحويل النقدي.

سيكون من الصعب العثور على مقارنة مناسبة للطفلة المميزة، لأنها مميزة بالفعل؛ فخلقيتها الاجتماعية والاقتصادية، وسماتها الوراثية، وخصائصها الشخصية والأسرية لا يمكن أن توجد لدى أي شخص آخر. ولو أردنا ببساطة مقارنة الطفلة المميزة مع طفل غير مشارك في برنامج التحويل النقدي، ولنقل "الطفل الفريد"، فقد لا تكون المقارنة كافية. لا يمكن أن تكون الطفلة المميزة مطابقة بالضبط للطفل الفريد. قد لا تبدو الطفلة المميزة والطفل الفريد متماثلين، فربما لا يعيشان في المكان نفسه، وربما ليس لهما الوالدين نفسهما، وقد لا يكون لهما الطول نفسه عندما ولدا. لذلك، إذا لاحظنا أن الطفل الفريد أقصر من الطفلة المميزة في سن الثالثة، فلا يمكننا معرفة ما إذا كان الفرق ناتجاً عن برنامج التحويل النقدي أو عن واحد من الفروقات الكثيرة الأخرى بين هذين الطفلين.

"الطفلة المميزة" هي طفلة مولودة حديثاً. وتتلقى والدتها حوالات نقدية شهرية طالما أنها تضمن أن "الطفلة المميزة" تتلقى فحوصات صحية منتظمة في المركز الصحي المحلي، وأنه يتم تحصينها، ومتابعة نموها. تفترض الحكومة أن التحويل النقدي سيحفز والدة الطفلة المميزة على طلب الخدمات الصحية التي يتطلبها البرنامج وسيساعد الطفلة المميزة على اكتساب القوة والطول. تختار الحكومة الطول بمثابة مؤشر للحصائل للصحة على المدى الطويل بسبب أثره على التحويل النقدي.

افترض أنك تستطيع قياس طول الطفلة المميزة في سن الثالثة. من الناحية المثالية، لتقييم أثر البرنامج، قد يتعين عليك قياس طول الطفلة المميزة في سن الثالثة مع تلقي والدتها التحويل النقدي، وكذلك طول الطفلة المميزة في سن الثالثة إذا كانت والدتها لا تتلقى التحويل النقدي. بعد ذلك، عليك تقييم الطولين لتحديد الأثر. لو استطعنا مقارنة طول الطفلة المميزة في سن الثالثة بوجود البرنامج مع طولها في سن الثالثة بدون البرنامج، فسنعرف أن أي فرق في الطول كان ناتجاً فقط عن برنامج التحويل النقدي. ونظراً لكون جميع الجوانب الأخرى للطفلة المميزة ستكون نفسها، فلن تكون هناك خصائص أخرى يمكن أن تفسر الفرق في الطول.

ولكن، ولسوء الحظ، من المستحيل ملاحظة الطفلة المميزة بوجود برنامج التحويل النقدي وبدونه: فإما أن



تقدير الواقع المضاد

إن المفتاح لتقدير الواقع المضاد للمشاركين في البرنامج هو الانتقال من مستوى الفرد أو مستوى الوحدة إلى مستوى المجموعة. على الرغم من عدم وجود قرين مثالي للوحدة المفردة، يمكننا الاعتماد على الخصائص الإحصائية لتوليد مجموعتين من الوحدات لا يمكن تمييزهما إحصائيًا عن بعضهما البعض على مستوى المجموعة إذا كانت أعدادهما كبيرة بما فيه الكفاية. تُعرّف المجموعة التي تشارك في البرنامج باسم مجموعة المعالجة، وحصيلتها تساوي $(Y | P = 1)$ بعد أن تشارك في البرنامج. مجموعة المقارنة المطابقة إحصائيًا (التي تُسمى أحيانًا المجموعة الضابطة) هي المجموعة التي تبقى غير متأثرة بالبرنامج، وتسمح لنا بتقدير حيلة الواقع المضاد $(Y | P = 0)$: أي الحيلة التي كانت ستسود لو أن مجموعة المعالجة لم تتلق البرنامج.

لذلك، ومن الناحية العملية، فإن تحدي تقييم الأثر هو تحديد مجموعة معالجة ومجموعة مقارنة متطابقتين إحصائيًا، في المتوسط، في غياب البرنامج. إذا كانت المجموعتان متطابقتين، باستثناء أن إحدى المجموعتين تشارك في البرنامج، والأخرى لا تشارك، فيمكننا التحقق من أن أي فرق في الحواصل لا بد أن يكون ناتجًا عن البرنامج. إن العثور على مجموعات المقارنة هذه هو جوهر أي تقييم للأثر، بغض النظر عن نوع البرنامج الخاضع للتقييم. ببساطة، في غياب مجموعة مقارنة تعطي تقديرًا دقيقًا للواقع المضاد، لا يمكن تحديد الأثر الحقيقي لأي برنامج.

المفهوم

الأساسي

بدون مجموعة مقارنة
تعطي تقديرًا دقيقًا للواقع
المضاد، فلا يمكن تحديد
الأثر الحقيقي لأي برنامج.

إذن، التحدي الرئيسي لتحديد الآثار هو العثور على مجموعة مقارنة صالحة لها خصائص مجموعة المعالجة نفسها بغياب البرنامج. وتحديدًا، يجب أن تكون مجموعتا المعالجة والمقارنة متماثلتين بثلاث طرق على الأقل.

أولاً، يجب أن تكون الخصائص المتوسطة لمجموعة العلاج ومجموعة المقارنة متطابقة في غياب البرنامج. وعلى الرغم من أنه لا يلزم أن يكون للوحدات الفردية في مجموعة المعالجة "أقران مثاليون" في مجموعة المقارنة، فإن خصائص مجموعتي المعالجة والمقارنة في المتوسط ينبغي أن تكون نفسها. على سبيل المثال، ينبغي أن يكون العمر المتوسط للوحدات في مجموعة المعالجة هو نفسه في مجموعة المقارنة.

ثانيًا، ينبغي ألا تؤثر المعالجة على مجموعة المقارنة إما بصورة مباشرة أو غير مباشرة. في مثال مصروف الجيب، ينبغي ألا تنقل مجموعة المعالجة الموارد إلى مجموعة المقارنة (مفعول مباشر) أو تؤثر على سعر الحلوى في الأسواق المحلية (مفعول غير مباشر). على سبيل المثال، إذا أردنا عزل أثر مصروف الجيب على استهلاك الحلوى، فينبغي ألا تُمنَح مجموعة المعالجة أيضًا جولات ذهاب إلى متجر الحلويات أكثر من مجموعة المقارنة، وإلا فلن نتأكد من تمييز ما إذا كان الاستهلاك الإضافي للحلوى بسبب مصروف الجيب أو بسبب جولات الذهاب الإضافية إلى المتجر.

ثالثًا، يجب أن تتغير حصائل الوحدات في مجموعة المقارنة بالطريقة نفسها لتغير النتائج

في مجموعة المعالجة، إذا خضعت كلتا المجموعتين للبرنامج (أو لا). وفي هذه الحالة، ينبغي أن تتفاعل مجموعتا المعالجة والمقارنة مع البرنامج بالطريقة نفسها. على سبيل المثال، إذا ازداد دخل الأشخاص في مجموعة المعالجة بمقدار ١٠٠ دولار أمريكي بفضل برنامج تدريبي، فإن دخل الأشخاص في مجموعة المقارنة كان سيزيد أيضًا بمقدار ١٠٠ دولار أمريكي، لو تلقوا التدريب.

عند تحقيق هذه الشروط الثلاثة، فإن وجود البرنامج محل الاهتمام سيفسر أي فروقات في الحصيلة (Y) بين المجموعتين. يُعزى ذلك إلى أن الفرق الوحيد بين مجموعتي المعالجة والمقارنة هو أن أعضاء مجموعة المعالجة يتلقون البرنامج، بينما لا يتلقاه أعضاء مجموعة المقارنة. عندما يمكن إسناد الفرق في الحصيلة بأكمله إلى البرنامج، يكون الأثر السببي للبرنامج قد تم تحديده.

المفهوم الأساسي

لمجموعة المقارنة الصالحة
(١) خصائص مماثلة في
المتوسط لمجموعة
المعالجة بغياب البرنامج،
(٢) حيث إنها تبقى غير متأثرة
بالبرنامج،
(٣) وستتفاعل مع البرنامج
بالطريقة نفسها لتفاعل
مجموعة المعالجة، في حال
تلقاها له.

المفهوم

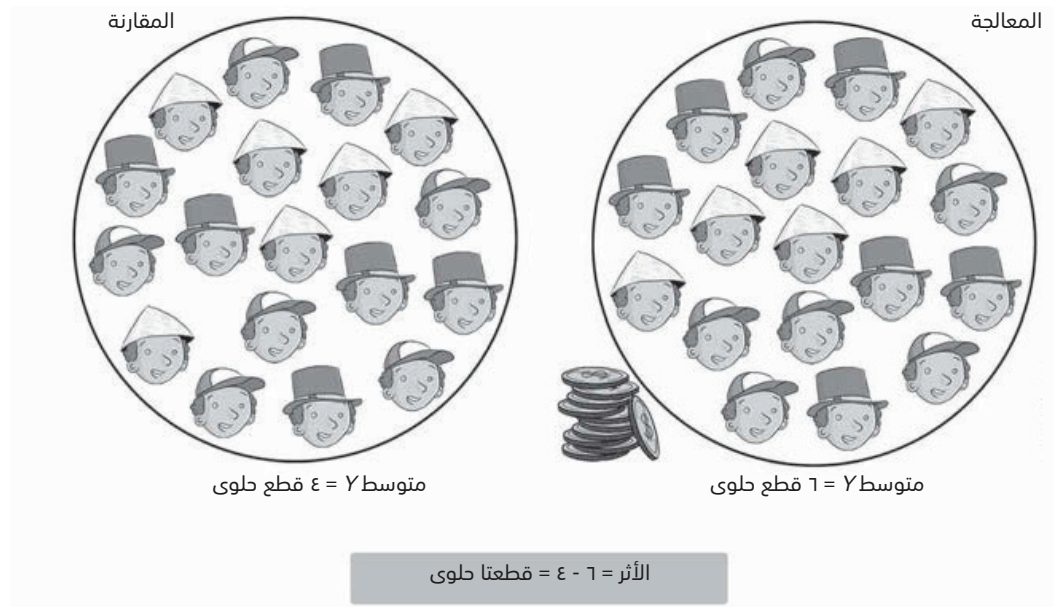
الأساسي

عندما لا تقيّم مجموعة المقارنة بدقة الواقع المضاد الحقيقي، فإن الأثر المقدّر للبرنامج لن يكون صالحًا. بالمعنى الإحصائي، سيكون متحيزًا.

بالعودة إلى حالة السيد فلان، رأينا أن تقدير أثر مصروف الجيب على استهلاكه للحلوى سيستلزم مهمة العثور على قرين السيد فلان المثالي غير المعقولة. بدلاً من النظر إلى الأثر فقط لفرد واحد، من الواقعي أكثر النظر إلى الأثر المتوسط لمجموعة من الأفراد (الشكل ٢-٣). إذا استطعت تحديد مجموعة أخرى من الأفراد تشترك في متوسط العمر، والنوع الاجتماعي، والتعليم، وتفضيل الحلوى، وما إلى ذلك، باستثناء أنها لا تتلقى مصروف جيب إضافيًا، فيمكنك تقدير أثر مصروف الجيب. وهو يساوي ببساطة الفرق بين متوسط استهلاك الحلوى في المجموعتين. وهكذا، إذا كانت مجموعة المعالجة تستهلك في المتوسط ٦ قطع حلوى لكل شخص، بينما تستهلك مجموعة المقارنة في المتوسط ٤ قطع، فإن أثر مصروف الجيب الإضافي على استهلاك الحلوى سيساوي قطعتي حلوى.

بعد تحديد مجموعة مقارنة صالحة، من المهم التفكير فيما سيحدث إذا قررنا المضي في تقييم دون إيجاد مجموعة من هذا القبيل. بديهيًا، مجموعة المقارنة غير الصالحة هي مجموعة تختلف عن مجموعة المعالجة بطريقة ما بخلاف غياب المعالجة. يمكن أن تتسبب هذه الفروقات الإضافية في أن يصبح تقدير الأثر غير صالح، أي متحيزًا بالتعبير الإحصائي: لن يقدّر تقييم الأثر الحقيقي للبرنامج، بل سيقوّم أثر البرنامج ممزوجة مع فروقات أخرى.

الشكل ٢-٣ مجموعة مقارنة صالحة



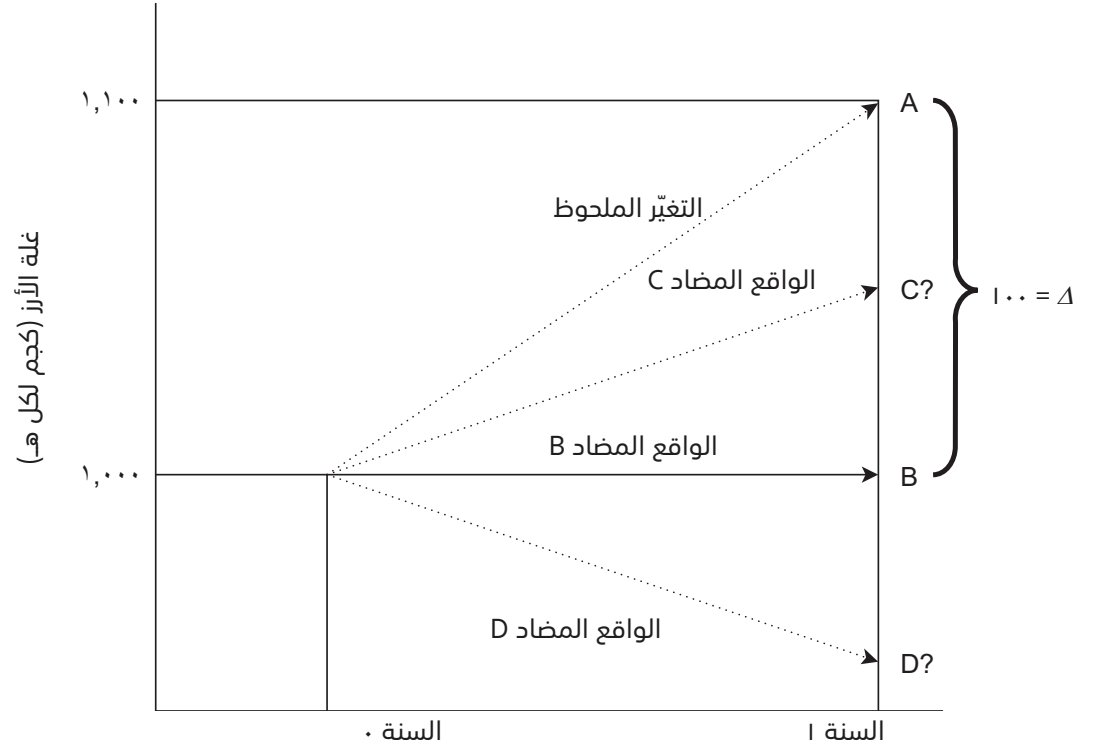
سنناقش فيما تبقى من الجزء الثاني من هذا الكتاب طرقًا متعددة يمكن استخدامها لإنشاء مجموعات مقارنة صالحة ستسمح لنا بتقدير الواقع المضاد. ولكن، قبل فعل ذلك، من المفيد مناقشة طريقتين شائعتين، ولكنهما شديدتا الخطورة، لإنشاء مجموعات مقارنة قد تؤدي في الكثير من الأحيان إلى تقديرات غير صحيحة ("زائفة") للواقع المضاد.

- *المقارنات القبلية والبعديّة (المعروفة أيضًا بالمقارنات السابقة واللاحقة أو الانعكاسية)* والتي تقارن حصائل المجموعة نفسها قبل المشاركة في البرنامج وبعدها.
- *مقارنات المشتركين وغير المشتركين (أو الاختيار الذاتي)* والتي تقارن حصائل مجموعة تختار الاشتراك في برنامج مع مجموعة تختار عدم المشاركة.

التقدير الزائف ١ للواقع المضاد: مقارنة الحصائل قبل البرنامج وبعده

بالعودة إلى المعادلة الأساسية لتقييم الأثر، تساوي حصة مجموعة المعالجة ($Y | P = 1$) ببساطة الحصة بعد المشاركة في البرنامج. ولكن المقارنات السابقة واللاحقة تأخذ الواقع المضاد ($Y | P = 0$) على أنه حصة مجموعة المعالجة قبل بدء التدخل. تفترض هذه المقارنة جوهريًا أنه لو أن البرنامج كان غير موجود أبدًا، فإن الحصة (Y) للمشاركين في البرنامج كانت ستساوي بالضبط ما كانت عليه قبل البرنامج. ولسوء الحظ، بالنسبة إلى غالبية البرامج التي نُفذت على مدى سلسلة من الأشهر أو السنوات، فإن هذا الافتراض ببساطة غير صحيح.

فكّر في تقييم برنامج التمويل الصغير للمزارعين في المناطق الريفية الفقيرة. يقدم البرنامج قروضًا صغيرة للمزارعين لتمكينهم من شراء سماد لزيادة إنتاجهم من الأرز. لقد لاحظت أنه في العام السابق لبدء البرنامج، حصد المزارعون ما معدله ١٠٠٠ كيلوجرام من الأرز لكل هكتار (النقطة B في الشكل ٣-٣). أُطلق برنامج التمويل الجزئي، وبعد عام ازدادت غلة الأرز إلى ١١٠٠ كيلوجرام للهكتار (النقطة A في الشكل ٣-٣). إذا كنت تحاول تقييم الأثر باستخدام المقارنة السابقة واللاحقة، فستستخدم حصة خط الأساس بمثابة تقدير للواقع المضاد. بتطبيق المعادلة الأساسية لتقييم الأثر، ستستنتج أن البرنامج قد زاد من غلة الأرز بمقدار ١٠٠ كيلوجرام للهكتار ($A-B$).



ملاحظة Δ: = التغير في غلة الأرز (كجم)؛ هـ = هكتار؛ كجم = كيلوجرام.

ولكن، تخيل لو أن هطول الأمطار كان طبيعيًا في العام السابق لإطلاق البرنامج، ولكن حدثت موجة جفاف في عام تشغيل البرنامج. بسبب الجفاف، يمكن أن تكون الغلة المتوسطة للمزارعين بدون برنامج التمويل الصغير أقل من B: لنقل عند المستوى D. في هذه الحالة، يساوي الأثر الحقيقي للبرنامج A-D، وهو أكبر من كمية الـ ١٠٠ كيلوجرام المقدرة باستخدام المقارنة قبل وبعد. على النقيض من ذلك، لو تحسّن هطول الأمطار فعليًا بين العامين، ربما كانت ستبلغ الغلة المستوى C. وفي هذه الحالة، سيكون الأثر الحقيقي للبرنامج أقل من ١٠٠ كيلوجرام. بمعنى آخر، إذا لم يكن بإمكان تحليلنا للأثر حساب هطول الأمطار وكل عامل آخر يمكن أن يؤثر على غلة الأرز بمرور الزمن، فلا يمكننا ببساطة حساب الأثر الحقيقي للبرنامج بإجراء مقارنة قبل وبعد.

في مثال التمويل الصغير السابق، كان هطول الأمطار أحد العوامل الخارجية التي لا تعد ولا تحصى والتي قد تؤثر على حصة البرنامج موضع الاهتمام (غلة الأرز) بمرور الزمن. وبالمثل، يتأثر العديد من الحاصلات التي تهدف برامج التنمية إلى تحسينها، كالدخل أو الإنتاجية أو الصحة أو التعليم، بمجموعة من العوامل بمرور الزمن. ولهذا السبب، فإن حواصل خط الأساس تكاد لا تكون تقديرًا جيدًا للواقع المضاد. ولهذا فإننا نعدّ تقدير الواقع المضاد زائفًا.

تقييم أثر برنامج معونة التأمين الصحي (HISP): المقارنة القبلية والبعدية للحصائل



تذكر أن برنامج معونة التأمين الصحي (HISP) هو برنامج جديد في بلدك يدعم شراء التأمين الصحي للأسر الريفية الفقيرة، وأن هذا التأمين يغطي النفقات المتعلقة بالرعاية الصحية والأدوية للمشاركين. هدف برنامج معونة التأمين الصحي (HISP) هو الحد مما تنفقه الأسر الفقيرة على الرعاية الأولية والدواء وتحسين الحصائل الصحية في نهاية المطاف. وعلى الرغم من أن العديد من مؤشرات الحصائل يمكن أخذها في الحسبان لتقييم البرنامج، فإن حكومتك مهتمة تحديدًا بتحليل آثار برنامج معونة التأمين الصحي (HISP) على نصيب الفرد من النفقات الصحية السنوية المدفوعة من المال الخاص (يشار إليها لاحقًا ببساطة باسم النفقات الصحية).

سيمثل برنامج معونة التأمين الصحي (HISP) نسبة كبيرة من الميزانية الوطنية إذا وُسع نطاقه على المستوى الوطني - حتى ١,٥ في المائة من الناتج المحلي الإجمالي وفقًا لبعض التقديرات. علاوة على ذلك، تشارك تعقيدات إدارية ولوجستية كبيرة في إدارة برنامج بهذه الخصائص. ولهذه الأسباب، اتُخذ قرار على أعلى المستويات الحكومية لإدخال برنامج معونة التأمين الصحي (HISP) كبرنامج تجريبي أولاً، ثم توسيع نطاقه تدريجيًا بمرور الزمن اعتمادًا على نتائج المرحلة الأولى. واستنادًا إلى نتائج التحليلات المالية وتحليلات التكلفة والمنفعة، أعلن الرئيس ومجلس وزرائه أنه من أجل أن يكون برنامج معونة التأمين الصحي (HISP) قابلاً للتطبيق، ولكي يتم توسيعه على المستوى الوطني، يجب أن يخفض النفقات الصحية السنوية للفرد في المنازل الريفية الفقيرة بما لا يقل عن ١٠ دولارات أمريكية في المتوسط، بالمقارنة مع ما كانت ستنفقه في غياب البرنامج، ويجب أن يتم ذلك في غضون عامين.

سيتم تنفيذ برنامج معونة التأمين الصحي (HISP) في ١٠٠ قرية ريفية خلال المرحلة التجريبية الأولى. وقبيل بدء البرنامج، تستأجر حكومتك شركة تنظيم للمسوحات لإجراء مسح خط الأساس لجميع الأسر البالغة ٤٩٥٩ أسرة في هذه القرى. يجمع المسح معلومات مفصلة عن كل أسرة، بما في ذلك تكوينها الديموغرافي، وأصولها، ووصولها إلى الخدمات الصحية، ونفقاتها الصحية في العام الماضي. وبعد إجراء مسح خط الأساس بمدة قصيرة، يتم إدخال برنامج معونة التأمين الصحي (HISP) إلى ١٠٠ قرية تجريبية بجلبة كبيرة، بما في ذلك الفعاليات الاجتماعية والحملات الترويجية الأخرى لتشجيع الأسر على التسجيل.

من بين الأسر الـ ٤٩٥٩ في العينة الأساسية، سجّلت ٢٩٠٧ أسر في برنامج معونة التأمين الصحي (HISP)، وعمل البرنامج بنجاح خلال العامين المقبلين. تقبل جميع العيادات الصحية والصيدليات التي تخدم ١٠٠ قرية المرضى الخاضعين لخطة التأمين، وتُظهر المسوحات أن معظم الأسر المشتركة راضية عن البرنامج. في نهاية الفترة التجريبية الممتدة على عامين، تبدأ المرحلة الثانية من جمع بيانات التقييم من عينة الأسر الـ ٤٩٥٩ نفسها.^٣

كلّفك الرئيس ووزير الصحة بمسؤولية الإشراف على تقييم أثر برنامج معونة التأمين الصحي (HISP) والتوصية بما إذا كان ينبغي توسيع البرنامج أو عدم توسيعه على المستوى الوطني. سؤال التقييم الرئيسي موضع الاهتمام بالنسبة إليك هو: ما أثر برنامج معونة التأمين الصحي (HISP) على النفقات الصحية المدفوعة من المال الخاص لدى الأسر الفقيرة؟ تذكر أن المخاطر كبيرة. إذا وُجد أن برنامج معونة التأمين الصحي (HISP) يقلل من النفقات الصحية بمقدار ١٠ دولارات أمريكية أو أكثر، فسيُوسَّع على المستوى الوطني. وإذا لم يبلغ البرنامج الهدف البالغ ١٠ دولارات أمريكية، فستوصي بعدم توسيعه.

يشير الاستشاري "الخبير" الأول الذي توظفه إلى أنه لتقدير أثر برنامج معونة التأمين الصحي (HISP)، عليك حساب التغير في النفقات الصحية بمرور الزمن للأسر المشتركة. حسب رأي الخبير وبما أن برنامج معونة التأمين الصحي (HISP) يغطي كافة النفقات الصحية، فإن أي انخفاض في النفقات بمرور الزمن يجب أن يُعزى إلى أثر برنامج معونة التأمين الصحي (HISP). باستخدام المجموعة الفرعية للأسر المشتركة، تحسب متوسط نفقاتها الصحية قبل تنفيذ البرنامج ثم بعد ذلك بعامين مرة أخرى. بمعنى ذلك، تجري مقارنة قبلية وبعدية (سابقة ولاحقة). النتائج موضحة في الجدول ٣-١. نلاحظ أن مجموعة المعالجة خفضت من نفقاتها الصحية من المال الخاص بمقدار ٦,٦٥ دولارات أمريكية، من ١٤,٤٩ دولارًا أمريكيًا قبل إدخال برنامج معونة التأمين الصحي (HISP) وإلى ٧,٨٤ دولارات أمريكية بعد عامين. وكما تشير قيمة الإحصائي t (t-stat)، فإن الفرق بين النفقات الصحية قبل البرنامج وبعده ذات دلالة إحصائية.^٤ وهذا يعني عثورك على دليل قوي ضد الادعاء بأن الفرق الحقيقي بين النفقات قبل التدخل وبعده يساوي صفر.

على الرغم من أن المقارنة قبل وبعد تخص المجموعة نفسها من الأسر، إلا أنك تشعر بالقلق من أن ظروفًا أخرى قد تغيرت أيضًا بالنسبة لهذه الأسر خلال العامين الماضيين، ما يؤثر على نفقاتها الصحية. على سبيل المثال، أصبح عدد من العقاقير الجديدة متوفرًا مؤخرًا. وينتابك القلق أيضًا من أن الانخفاض في النفقات الصحية قد يكون ناتجًا جزئيًا عن الأزمة المالية التي مر بها بلدك مؤخرًا. لمعالجة بعض هذه المخاوف، يجري الاستشاري تحليل انحدار أكثر تعقيدًا يحاول التحكم في بعض العوامل الإضافية.

الجدول ٣-١ تقييم برنامج معونة التأمين الصحي (HISP): المقارنة "قبل وبعد"

بعد	قبل	الفرق	t-stat
٧,٨٤	١٤,٤٩	-٦,٦٥**	-٣٩,٧٦

النفقات الصحية
للأسرة (دولار أمريكي)

ملاحظة: مستوى الدلالة: ** = ١ بالمائة.

الجدول ٢-٣ تقييم برنامج معونة التأمين الصحي (HISP): قبل وبعد مع تحليل الانحدار

الانحدار الخطي	الانحدار الخطي متعدد المتغيرات
الأثر المقدّر على النفقات الصحية	
-٦,٦٠ (٠,٢٣) **	-٦,٧١ (٠,٢٣) **

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

يستخدم تحليل الانحدار الإحصاء لتحليل العلاقات بين متغير تابع (المتغير المطلوب تفسيره) والمتغيرات التفسيرية. تظهر النتائج في الجدول ٢-٣. الانحدار الخطي هو الشكل الأبسط: المتغير التابع هو النفقات الصحية، وهناك متغير تفسيري واحد فقط: مؤشر ثنائي (٠-١) يأخذ القيمة ٠ إذا أُخذت الملاحظة عند خط الأساس و ١ إذا أُخذت الملاحظة عند المتابعة.

يضيف الانحدار الخطي متعدد المتغيرات متغيرات تفسيرية للتحكم في الخصائص الأخرى التي يتم ملاحظتها للأسر في عينتك أو الاحتفاظ بها ثابتة، بما في ذلك مؤشرات الثروة (الأصول) وتكوين الأسرة وما إلى ذلك.^٥ تلاحظ أن النتيجة من الانحدار الخطي تعادل الفرق البسيط "قبل وبعد" في متوسط النفقات الصحية من الجدول ١-٣ (انخفاض بمقدار ٦,٦٠ دولارات أمريكية في النفقات الصحية). بمجرد أن تستخدم الانحدار الخطي متعدد المتغيرات للتحكم في العوامل الأخرى المتاحة في بياناتك، فإنك ستحصل على نتيجة مماثلة - انخفاض قدره ٦,٧١ دولارات أمريكية في النفقات الصحية.

السؤال ١ الخاص ببرنامج HISP



- هل تتحكم المقارنة قبل وبعد في جميع العوامل التي تؤثر على النفقات الصحية بمرور الزمن؟
- استنادًا إلى هذه النتائج التي يقدمها التحليل قبل وبعد، هل يتعين توسيع برنامج معونة التأمين الصحي (HISP) على المستوى الوطني؟

التقدير الزائف ٢ للواقع المضاد: مقارنة الأسر المشتركة وغير المشتركة (المختارة ذاتيًا)

إن مقارنة مجموعة من الأفراد الذين يسجلون طوعًا في برنامج مجموعة من الأفراد الذين يختارون عدم المشاركة مقارنةً بخطة أخرى لتقييم الأثر. تعطي مجموعة مقارنة تختار ذاتيًا من البرنامج تقدير واقع مضاد زائفًا آخر. يحدث الاختيار عندما تستند المشاركة في البرنامج إلى التفضيلات أو القرارات أو الخصائص غير الملحوظة للمشاركين المحتملين.

تأمل على سبيل المثال برنامج تدريب مهني للشباب العاطلين عن العمل. افترض أنه بعد عامين من إطلاق البرنامج، تمت محاولة تقييم أثره على الدخل من خلال مقارنة متوسط دخل مجموعة من الشباب الذين اختاروا التسجيل في البرنامج بالنسبة إلى مجموعة من الشباب الذين، على الرغم من كونهم مؤهلين، اختاروا عدم التسجيل. افترض أن النتائج توضح أن الشاب الذي اختار التسجيل في البرنامج يحقق ضعف ما يحققه الشاب الذي اختار عدم المشاركة. كيف ينبغي تفسير هذه النتائج؟ في هذه الحالة، يُقدّر الواقع المضاد استنادًا إلى حصائل الأفراد الذين قرروا عدم التسجيل في البرنامج. رغم ذلك، يُرجّح أن تكون المجموعتان مختلفتين جوهريًا. وقد يكون الأفراد الذين اختاروا المشاركة متحمسين للغاية لتحسين سبل معيشتهم وقد يتوقعون عائدًا كبيرًا على التدريب. وعلى النقيض من ذلك، قد يشعر الذين اختاروا عدم التسجيل بالإحباط من الشباب الذين لا يتوقعون الاستفادة من هذا النوع من البرامج. ومن المرجح أن يؤدي هذان النوعان على نحو مختلف تمامًا في سوق العمل وسيكون لهما دخل مختلف حتى بدون برنامج التدريب المهني.

المفهوم

الأساسي

يحدث تحيّز الاختيار عند ربط أسباب مشاركة الفرد في البرنامج بالحصائل. إن ضمان خلو الأثر المقدر من تحيّز الاختيار هو أحد الأهداف والتحديات الكبرى لأي تقييم أثر.

وتبرز المشكلة نفسها عندما يعتمد القبول في برنامج على تفضيلات غير خاضعة للملاحظة لمسؤولي البرنامج. لنقل، على سبيل المثال، إن مسؤولي البرنامج يبنون القبول والتسجيل على المقابلات. قد يكون الأفراد المقبولين في البرنامج هم أولئك الذين يعتقد المسؤولون أن لديهم فرصة جيدة للاستفادة من البرنامج. وقد يبدي غير المقبولين حماسًا أقل في المقابلة، أو تكون مؤهلاتهم أقل، أو يفتقرون إلى مهارات المقابلات الجيدة. ومرة أخرى، من المحتمل أن يكون لهاتين المجموعتين من الشباب دخل مختلف في سوق العمل حتى في حال عدم وجود برنامج تدريب مهني.

ومن ثم فإن المجموعة التي لم تسجّل لم تعطِ تقديرًا جيدًا للواقع المضاد. إذا لاحظت فرقًا في الدخل بين المجموعتين، فلن تتمكن من تحديد ما إذا كان الدخل من برنامج التدريب أو من الفروق الكامنة في التحفيز، والمهارات، وعوامل أخرى توجد بين المجموعتين. وبالتالي، فإن عدم تسجيل الأفراد الأقل تحفيزًا أو تأهيلًا يؤدي إلى تحيّز في أثر البرنامج.^٦ يُسمى هذا التحيّز تحيّز الاختيار. وعلى نحو أعم، سيحدث تحيز الاختيار عندما ترتبط أسباب مشاركة الفرد في برنامج بالنتائج، حتى في غياب البرنامج. إن ضمان خلو الأثر المقدر من تحيّز الاختيار هو أحد الأهداف والتحديات الكبرى لأي تقييم أثر. في هذا المثال، إذا كان الشبان اللذان التحقوا بالتدريب المهني سيحصلان على دخل أعلى حتى في حال عدم وجود البرنامج، فسيكون تحيز الاختيار إيجابيًا؛ بعبارة أخرى، قد تبالغ في تقدير أثر برنامج التدريب المهني عبر نسب الدخل الأعلى التي كان سيحصل عليها المشاركون على أي حال إلى البرنامج.

تقييم أثر برنامج معونة التأمين الصحي (HISP): مقارنة الأسر المشتركة وغير المشتركة

بعد التفكير ملياً في المقارنة قبل وبعد مع فريق التقييم لديك، تدرك أنه ما زال هناك العديد من العوامل الأخرى التي قد تفسر جزءاً من التغيير في النفقات الصحية بمرور الزمن (لإسيما شعور وزير المالية بالقلق من أن الأزمة المالية الأخيرة قد أثرت على دخل الأسرة، وقد تفسر التغيير الملحوظ في النفقات الصحية). يقترح استشاري آخر أنه سيكون أكثر ملائمة تقدير الواقع المضاد في فترة ما بعد التدخل: أي بعد عامين من بدء البرنامج. ينوّه الاستشاري على نحو صحيح إلى أنه من بين الأسر الـ ٤٩٥٩ في عينة خط الأساس، سجّلت ٢٩٠٧ أسر بالفعل في البرنامج فقط، ومن ثم تبقى نسبة ٤١ بالمائة من الأسر في العينة بدون تغطية برنامج معونة التأمين الصحي (HISP). يدّعي الاستشاري أن الأسر ضمن القرى التجريبية الـ ١٠٠ جميعاً كانت مؤهلة للتسجيل. تشترك هذه الأسر جميعاً بالعيادات الصحية نفسها، وهي خاضعة لنفس الأسعار المحلية للأدوية. إضافة إلى ذلك، تشارك معظم الأسر في أنشطة اقتصادية متشابهة. يقول الاستشاري إنه في هذه الظروف، يمكن أن تساعد حصائل المجموعة غير المشتركة بعد التدخل على تقدير حيلة الواقع المضاد للمجموعة المشتركة في برنامج معونة التأمين الصحي (HISP). لذا تقرّر حساب متوسط النفقات الصحية في فترة ما بعد التدخل لكلا صنفَي الأسر المشتركة في البرنامج وغير المشتركة فيه. النتائج موضحة في الجدول ٣-٣. باستخدام متوسط النفقات الصحية للأسر غير المشتركة بصفته تقدير الواقع المضاد، تجد أن البرنامج قد خفض متوسط النفقات الصحية بمقدار ١٤,٤٦ دولارًا أمريكيًا تقريبًا.

عندما تناقش هذه النتيجة بالتفصيل مع الاستشاري، فإنك تسأل عما إذا كانت الأسر التي اختارت عدم التسجيل في البرنامج قد تكون مختلفة بشكل منهجي عن الأسر التي لم تشارك. على سبيل المثال، قد تكون الأسر التي سجلت في برنامج معونة التأمين الصحي (HISP) هي الأسر التي توقعت أن تكون لديها نفقات صحية أعلى، أو الأشخاص الذين كانوا أكثر دراية بالبرنامج، أو الأشخاص الذين يهتمون أكثر بصحة عائلاتهم. بدلاً من ذلك، ربما كانت الأسر التي اشتركت أفقر في المتوسط من تلك التي لم تشترك، بالنظر إلى أن برنامج معونة التأمين الصحي (HISP) كان موجّهاً إلى الأسر الفقيرة. يدعي استشاريك أن تحليل الانحدار يمكن أن يتحكم في الفروق المحتملة بين المجموعتين. لذلك يجري انحدارًا إضافيًا يتحكم في جميع خصائص الأسر التي يمكنه العثور عليها في مجموعة البيانات، ويقدر أثر البرنامج كما هو موضح في الجدول ٣-٤.

الجدول ٣-٣ تقييم برنامج معونة التأمين الصحي (HISP): مقارنة المتوسطات للمشاركين وغير المشاركين

المشاركون	غير المشاركين	الفرق	t-stat
٧,٨٤	٢٢,٣٠	-١٤,٤٦**	-٤٩,٠٨

النفقات الصحية
للأسرة (دولار أمريكي)

ملاحظة: مستوى الدلالة: ** = ١ بالمائة.

الجدول ٤-٣ تقييم برنامج معونة التأمين الصحي (HISP): تحليل انحدار المشتركين وغير المشتركين

الانحدار الخطي	الانحدار الخطي متعدد المتغيرات
الأثر المقدّر على النفقات الصحية الأسرية (دولار أمريكي)	
-١٤,٤٦ (٠,٣٣) **	-٩,٩٨ (٠,٢٩) **

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

من خلال تطبيق الانحدار الخطي البسيط للنفقات الصحية على متغير مؤشر ما إذا كانت الأسرة مشتركة في البرنامج أم لا، يمكنك العثور على أثر مقدّر يساوي سالب ١٤,٤٦ دولار أمريكي؛ بعبارة أخرى، تقدّر أن البرنامج قد خفض متوسط النفقات الصحية بمقدار ١٤,٤٦ دولار أمريكي. ولكن، عند التحكم في خصائص أخرى للبيانات، تقدّر أن البرنامج قد حدّد من النفقات الصحية بمقدار ٩,٩٨ دولارات أمريكية في العام.

السؤال ٢ الخاص ببرنامج HISP



- أ. هل يُحتمل أن يتحكم هذا التحليل في جميع العوامل التي تحدد الفروق في النفقات الصحية بين المجموعتين؟
- ب. استنادًا إلى هذه النتائج التي تنتجها طريقة "المشتركون وغير المشتركين"، هل ينبغي توسيع برنامج معونة التأمين الصحي (HISP) على المستوى الوطني؟

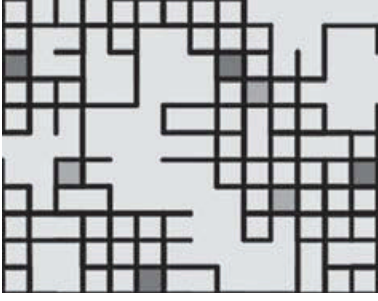
مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice).

١. نستخدم نموذج روبن السببي بمثابة إطار للاستدلال السببي (Imbens and Rubin 2008 ؛ Rubin 1974).
٢. سيُخفف هذا الشرط في بعض طرق تقييم الأثر، ما سيتطلب بدلاً من ذلك أن يكون متوسط التغير في الحصائل (الاتجاهات) هو نفسه بغياب البرنامج.
٣. نفترض أن أي أسرة لم تغادر العينة على مدار عامين (لا يوجد تناقص في العينة). وهذا ليس افتراضاً واقعياً لمعظم المسوحات الأسرية. عملياً، لا يمكن تتبع العائلات التي تنتقل إلى موقعها الجديد، كما تنفصل بعض الأسر ولا تعود موجودة معاً.
٤. لاحظ أن الإحصائي t (t-stat) المساوي ١,٩٦ أو أكثر (بالقيمة المطلقة) ذو دلالة إحصائية عند مستوى ٥ بالمائة.
٥. للمزيد من المعلومات حول التحليل متعدد المتغيرات، راجع الدليل الفني عبر الإنترنت على الموقع الإلكتروني "تقييم الأثر في الواقع العملي" (www.worldbank.org/ieinpractice).
٦. في مثال آخر، إذا كان الشباب الذين يتوقعون استفادة كبيرة من خطة التدريب هم أيضاً أكثر عرضة للاشتراك (على سبيل المثال، لأنهم يتوقعون أجوراً أعلى مع التدريب)، فإن مقارنتهم بمجموعة لم تشترك ذات عوائد متوقعة أقل ستؤدي إلى تقدير متحيز للأثر.

المراجع

- Imbens, Guido W., and Donald B. Rubin. 2008. "Rubin Causal Model." In *The New Palgrave Dictionary of Economics*, second edition, edited by Steven N. Durlauf and Lawrence E. Blume. Palgrave.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.



الفصل ٤

التعيين العشوائي

تقييم البرامج استنادًا إلى قواعد التعيين

بعد أن ناقشنا تقديرين "زائفين" للواقع المضاد من الشائع استخدامهما ينطويان على خطورة تحييز عالية، ألا وهما المقارنات القبلية والبعدية ومقارنة المشتركين وغير المشتركين، نلغيت الآن إلى مجموعة من الطرق التي يمكن تطبيقها لتقدير آثار البرامج بدقة أكبر. وهذا التقدير ليس بالبساطة التي يبدو عليها للوهلة الأولى على الدوام. تُصمّم معظم البرامج وتنفّذ في بيئة معقدة ومتغيرة، حيث يمكن للعديد من العوامل أن تؤثر على النتائج لكل من المشاركين وغير المشاركين في البرنامج. إن حالات الجفاف والزلازل والركود والتغيّرات في الحكومات والتغيّرات في السياسات الدولية والمحلية كلها جزء من العالم الواقعي. نوّد في أي تقييم التأكد من بقاء الأثر المقدر لبرنامجنا ساريًا على الرغم من هذه العوامل التي لا تعد ولا تحصى.

إن القواعد التي يستخدمها البرنامج لاختيار المشاركين هي المعلمة الرئيسية لتحديد طريقة تقييم الأثر. نعتقد أنه في معظم الأحيان، يتعين أن تسعى طرق التقييم إلى ذلك في سياق القواعد التشغيلية للبرنامج (مع بعض التعديلات هنا وهناك) - وليس العكس. ولكننا نبدأ أيضًا من فرضية مفادها أنه ينبغي أن تكون في جميع البرامج قواعد منصفة وشفافة لتعيين تلك البرامج. يبدو أن من أكثر القواعد عدلاً وشفافية لتخصيص الموارد النادرة بين المجموعات السكانية المستحقة على حد سواء إعطاء كل من هو مؤهل فرصة متساوية للمشاركة في البرنامج. إحدى الطرق لعمل ذلك هي إجراء قرعة ببساطة.



سندرس في هذا الفصل طريقة تشبه عمل قرعة تقرر من يدخل في برنامج في وقت معين ومن لا يدخل فيه: طريقة التعيين العشوائي، المعروفة أيضًا باسم *التجارب العشوائية المضبوطة* (RCTs). إن هذه الطريقة لا تزوّد فقط مسؤولي البرامج بقاعدة منصفة وشفافة لتخصيص الموارد النادرة بين *المجتمعات الإحصائية المستحقة على حد سواء*، بل تمثل أيضًا الطريقة الأقوى لتقييم أثر أي برنامج. وهكذا ازداد تطبيق هذه الطريقة لتقييم آثار البرامج الاجتماعية ازديادًا كبيرًا في السنوات الأخيرة.

التعيين العشوائي للمعالجة

عند تعيين برنامج عشوائيًا، أي باستخدام القرعة، على مجموعة كبيرة من المجتمع الإحصائي المستحق، نحصل على تقدير قوي للواقع المضاد. التعيين العشوائي للمعالجة هو المعيار الذهبي لتقييم الأثر. وهو يستخدم عملية عشوائية، أو المصادفة، لتحديد من يُمنَح إمكانية الوصول إلى البرنامج ومن لا يُمنَح إياها. بموجب التعيين العشوائي، فإن لكل وحدة مؤهلة (على سبيل المثال، الفرد أو الأسرة أو الشركة أو المدرسة أو المستشفى أو المجتمع) الاحتمال نفسه التي اختارته للمعالجة بواسطة البرنامج.^١

قبل أن نناقش كيفية تنفيذ التعيين العشوائي عمليًا والسبب في إنتاجه لتقدير قوي للواقع المضاد، لنستغرق بعض اللحظات في التفكير في سبب كون التعيين العشوائي طريقة منصفة وشفافة أيضًا لتعيين موارد البرامج النادرة. بمجرد تحديد مجتمع إحصائي مستهدف (على سبيل المثال، أسر تعيش تحت خط الفقر، أو أطفال دون سن الخامسة، أو طرق في المناطق الريفية في شمال البلاد)، يكون التعيين العشوائي قاعدة منصفة، لأنها تتيح لمديري البرامج التأكد من أن كل وحدة مؤهلة تحظى بفرصة تلقي البرنامج نفسها وضمان عدم تعيينه باستخدام معايير تعسفية أو ذاتية، أو حتى من خلال المحسوبية أو غيرها من الممارسات غير المنصفة. عند وجود طلب زائد على برنامج ما، يكون التعيين العشوائي قاعدة يمكن توضيحها بسهولة من قبل مدير البرامج، ويمكن أن تفهمها الفئات المستهدفة، وتُعد منصفة في الكثير من الظروف. إضافةً إلى ذلك، عند إجراء عملية التعيين بانفتاح وشفافية، لا يمكن التلاعب بها بسهولة، ولذا فإنها تقي مديري البرامج من الاتهامات المحتملة بالمحاباة أو الفساد. ولذا فإن التعيين العشوائي مزاياه، كآلية ترشيد تتجاوز فائدها كأداة لتقييم الأثر.

وفي الواقع، يستخدم عدد من البرامج القُرعات بشكل روتيني بمثابة طريقة لاختيار المشاركين من مجموعة الأفراد المؤهلين، وذلك في المقام الأول بسبب مزاياها في الإدارة والحوكمة.^٣ يقدم المربع ٤-١ حالتين من هذا النوع من إفريقيا.

يمكن اشتقاق التعيين العشوائي غالبًا من القواعد التشغيلية للبرامج. بالنسبة إلى الكثير من البرامج، المجتمع الإحصائي للمشاركين—أي مجموعة جميع الوحدات التي يود البرنامج خدمتها—أكبر من عدد المشاركين الذين يمكن للبرنامج استيعابهم بالفعل في وقت معيّن.

المربع (١-٤): التعيين العشوائي بمثابة أداة تشغيلية قيّمة

الخانات المتوفرة للبرنامج كان N ، فقد اختير المتقدمون الذين سحبوا الأعداد الصغرى ليشتركوا في البرنامج. نُظمت عملية القرعة على نحوٍ منفصل للرجال والنساء. وقد لاقت القرعة العامة قبولاً جيداً لدى المشاركين، وساعدت على توفير صورة عن عدالة البرنامج وشفافيته في بيئات ما بعد النزاع التي تميّزها التوترات الاجتماعية. وبعد عدة سنوات من العمليات، استخدم الباحثون قاعدة التخصيص هذه، المدمجة بالفعل في تشغيل البرنامج، لإجراء تقييم الأثر.

في النيجر، شرعت الحكومة في تنفيذ شبكة مشروع أمان عام ٢٠١١ بدعم من البنك الدولي. وتعتبر النيجر واحدة من أفقر بلاد العالم، وقد تجاوز عدد سكان الأسر الفقيرة التي تستحق البرنامج بشكل كبير المخصصات المتاحة خلال السنوات الأولى من التشغيل. كان منفذو البرنامج يعتمدون على الاستهداف الجغرافي لتحديد الإدارات والبلديات التي سيُنفذ برنامج التحويلات النقدية فيها أولاً. وكان من الممكن فعل ذلك عند وجود بيانات لتحديد حالة الفقر أو الضعف النسبي لمختلف الإدارات أو البلديات. ولكن، كان يتوفر قدر ضئيل جداً من البيانات لتقييم القرى التي ستكون أكثر استحقاقاً من غيرها استناداً إلى معايير موضوعية. (تابع)

يمكن أن يكون التعيين العشوائي قاعدة مفيدة لتخصيص مخصصات البرامج، حتى خارج سياق تقييم الأثر. توضح الحالتان التاليتان من إفريقيا ذلك.

بعد فترة من الأزمة في ساحل العاج، قدمت الحكومة برنامج توظيف مؤقتاً والذي كان يستهدف في البداية المحاربين السابقين، ثم توسّع لاحقاً ليشمل الشباب عمومًا. وقرّ البرنامج للشباب فرص توظيف قصيرة الأجل، معظمها في تنظيف الطرق أو إعادة تأهيلها من خلال الوكالة الوطنية للطرق. ودُعي الشباب في البلديات المشاركة للتسجيل. في ضوء جاذبية المنافع، تقدّم عدد من الشباب يفوق بكثير الأماكن المتاحة. ولتقديم طريقة شفافة ومنصفة لتخصيص المخصصات بين المشاركين، طرح منفذو البرنامج قرعة عامة. وبمجرد إغلاق التسجيل ومعرفة عدد المتقدمين (وليكن N)، نُظمت قرعة عامة. دُعي جميع المشاركين إلى موقع عام، ووضعت أوراق مرقّمة بالأعداد من ١ إلى N في صندوق. بعد ذلك، دُعي المتقدمون واحدًا تلو الآخر كي يسحب كل منهم عددًا من الصندوق أمام بقية المتقدمين. وبمجرد سحب العدد، كان يُقرأ بصوت مرتفع. وبعد استدعاء جميع المتقدمين، تحقّق شخص من الأعداد المتبقية في الصندوق واحدًا تلو الآخر لضمان التحقق من أنها تمثل متقدمين لم يحضروا القرعة. نظرًا إلى أن عدد

الأسر لتحديد الأسر الأفقر، التي سُجّلت لاحقًا بمثابة مستفيدين). كانت الشفافية والإنصاف في القرعة العامة محط تقدير كبير لدى السلطات والقرى المحلية، ولدى منفذي البرنامج أيضًا- لدرجة أنه تواصل استخدام عملية القرعة العامة في الدوريتين الثانية والثالثة من المشروع لاختيار أكثر من ١٠٠٠ قرية في جميع أنحاء البلد. وعلى الرغم من أن القرعة العامة لم تكن ضرورية لتقييم الأثر في تلك المرحلة، إلا أن قيمتها بصفتها أداة تشغيلية شفافة ومنصفة ومقبولة على نطاق واسع لتخصيص المخصصات بين السكان الذين يستحقونها على قدم المساواة تبرر مواصلة استخدامها في أعين منفذي البرنامج والسلطات المحلية.

بالنسبة للمرحلة الأولى من المشروع، قرر منفذو البرنامج استخدام قُرعات عامة لاختيار القرى المستفيدة داخل البلديات المستهدفة. اتُخذ هذا القرار جزئيًا لأن البيانات المتوفرة لتحديد القرى ذات الأولوية كانت محدودة، وجزئيًا لأن تقييم الأثر كان مضمنًا في المشروع. بالنسبة إلى القرعة العامة، دُعي جميع رؤساء القرى إلى مركز البلدية، وكُتب اسم كل قرية من قراهم على ورقة، ووضعت في صندوق. بعد ذلك، سحب طفل عشوائيًا القرى المستفيدة من الصندوق إلى أن مُلئت الحصص. نُفذ الإجراء بشكل منفصل للقرى المستقرة والبدو لضمان تمثيل كل مجموعة. (بعد أن اختيرت القرى، نُفذت آلية منفصلة للاستهداف على مستوى

المصادر: Bertrand and others 2016; Premand, Barry, and Smitz 2016.

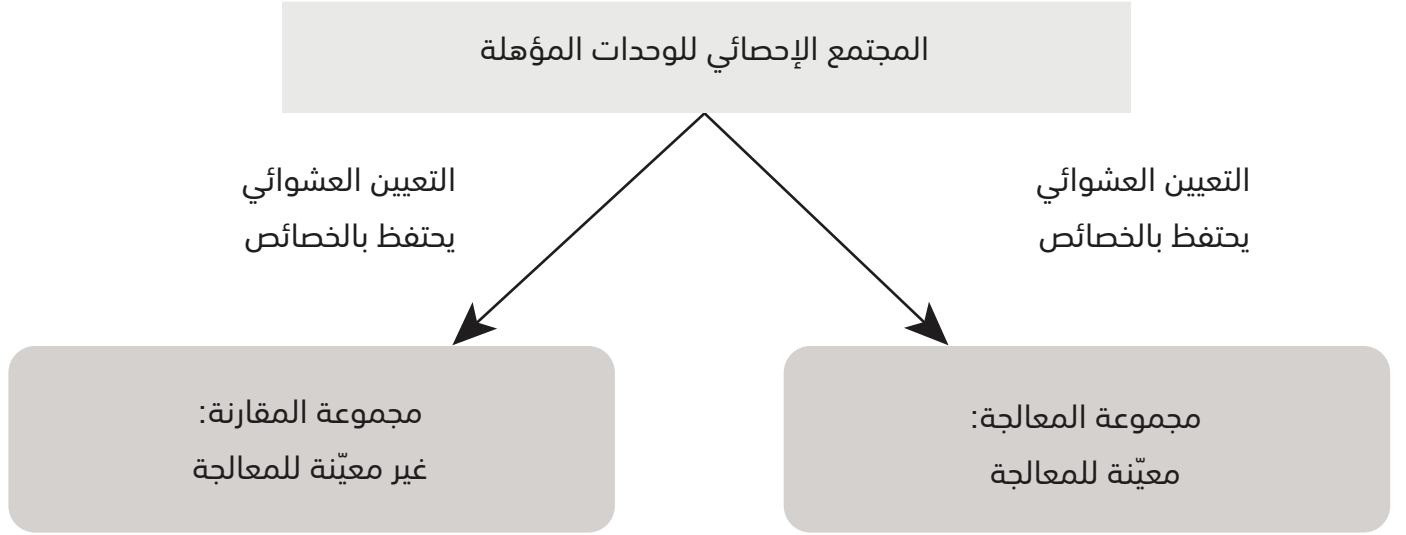
على سبيل المثال، قد يستطيع برنامج تعليمي وخلال عام واحد توفير المواد الدراسية لـ ٥٠٠ مدرسة من بين آلاف المدارس المؤهلة في الدولة. أو قد يستهدف برنامج لتحسين الطرق الريفية تعبيد ٢٥٠ طريقًا ريفيًا، على الرغم من وجود مئات الطرق الأخرى التي يود البرنامج تحسينها. أو قد يستهدف برنامج لتشغيل الشباب الوصول إلى ٢٠٠٠ شاب عاطل عن العمل خلال العام الأول من تشغيله، على الرغم من وجود عشرات الآلاف من الشباب العاطلين عن العمل الذين يرغب البرنامج في خدمتهم في النهاية. قد يتعذر على البرامج الوصول إلى المجتمع الإحصائي محل الاهتمام بأكمله، وذلك لأسباب عديدة. قد تعيق قيود الميزانية ببساطة تقديم المشرفين البرنامج لجميع الوحدات المؤهلة من البداية. وحتى إذا توفرت الميزانيات لتشمل عددًا غير محدود من المشاركين، فإن قيود القوة ستحول أحيانًا من تنفيذ البرنامج للجميع في الوقت نفسه. على سبيل المثال، في حالة برنامج التدريب على توظيف الشباب، قد يكون عدد الشباب الذين يريدون الخضوع للتدريب المهني أكبر من عدد المنافذ المتاحة في الكليات الفنية خلال العام الأول من البرنامج، وقد يحد ذلك من عدد من يمكنهم التسجيل.

عندما يكون عدد المشاركين المؤهلين أكبر من عدد الأماكن المتاحة في البرنامج، يجب على شخص ما اتخاذ قرار بشأن من سيدخل في البرنامج ومن لن يشارك. بمعنى آخر، يجب أن يحدد المشرفون على البرنامج آلية ترشيح لتخصيص خدمات البرنامج. يمكن تعيين البرنامج استنادًا إلى قاعدة الأولوية في التسجيل للحصول على الخدمة، أو استنادًا إلى خصائص ملحوظة (على سبيل المثال، خدمة المناطق الأفقر أولاً)، أو يمكن أن يستند الاختيار إلى خصائص غير ملحوظة (على سبيل المثال، السماح للأفراد بالاشتراك استنادًا إلى حافزهم ومعرفتهم الشخصية) أو بالقرعة. وحتى في السياقات التي يمكن فيها تصنيف المشاركين المحتملين على أساس مقياس الحاجة، قد يكون من المرغوب فيه تخصيص بعض المخصصات عن طريق القرعة. تأمل على سبيل المثال برنامجًا يستهدف أفقر ٢٠ في المائة من الأسر على أساس مقياس الدخل. إذا كان من الممكن قياس الدخل بصورة غير مثالية فقط، فيمكن للبرنامج استخدام هذا المقياس ليشمل جميع المشاركين المحتملين الذين يُحدّد أنهم في حالة فقر مدقع (ال ١٥ بالمائة الأدنى على سبيل المثال). ولكن، بما أن الدخل يُقاس بصورة غير مثالية، فقد تكون الأسر عند مستوى أدنى بقليل من عتبة الأهلية الموجودة عند الشريحة المئوية العشرين أو قد تكون غير مؤهلة في الواقع (إذا تمكنا من قياس الدخل الحقيقي)، في حين أن الأسر التي تزيد قليلًا عن النسبة المئوية العشرين قد تكون أو لا تكون مؤهلة. وفي هذا السياق، يمكن أن يكون إجراء قرعة لتخصيص المخصصات للأسر حول النسبة المئوية الـ ٢٠ (على سبيل المثال، بين النسبة المئوية ١٥ والنسبة المئوية ٢٥ في توزيع الدخل) طريقة منصفة لتخصيص المخصصات في هذه المجموعة من الأسر.

لماذا ينتج التعيين العشوائي تقديرًا ممتازًا للواقع المضاد؟

كما ناقشنا من قبل، ستكون مجموعة المقارنة المثالية مماثلة قدر الإمكان لمجموعة المعالجة من جميع النواحي، باستثناء ما يتعلق بمشاركتها في البرنامج الخاضع للتقييم. عند التعيين العشوائي للوحدات في مجموعات المعالجة والمقارنة، ستؤدي عملية التعيين العشوائي بحد ذاتها إلى تكوين مجموعتين تنطويان على احتمالية عالية بأن تكونا متطابقتين، طالما أن عدد الوحدات المحتملة التي نطبق عليها عملية التعيين العشوائي كبيرة بما فيه الكفاية. وبالتحديد، عند وجود عدد كبير بما يكفي من الوحدات، تنتج عملية التعيين العشوائي مجموعات لها متوسطات متكافئة إحصائيًا لجميع خصائصها.^٤

يوضح الشكل ١-٤ لماذا ينتج التعيين العشوائي مجموعة مقارنة تكافئ إحصائيًا لمجموعة المعالجة. افترض أن المجتمع الإحصائي للوحدات المؤهلة (مجموعة المشاركين، أو المجتمع الإحصائي محل الاهتمام للتقييم) يتكون من ١٠٠٠ شخص. يُعيّن نصف المشاركين عشوائيًا في مجموعة المعالجة، ويُعيّن النصف الآخر عشوائيًا في مجموعة المقارنة. على سبيل المثال، يمكنك أن تتصور أنك كتبت أسماء ١٠٠٠ شخص على أوراق فردية، ثم خلطتها في وعاء، ثم طلبت من أحدهم سحب ٥٠٠ اسم عشوائيًا. إذا كانت الأسماء الـ ٥٠٠ الأولى تمثل مجموعة المعالجة، فيمكن أن تكون لديك مجموعة معالجة معيّنة عشوائيًا (الأسماء الـ ٥٠٠ الأولى المسحوبة)، وعينة مقارنة عشوائية (الأسماء الـ ٥٠٠ المتبقية في الوعاء).



المفهوم

الأساسي

في التعيين العشوائي، يكون لكل وحدة مؤهلة نفس احتمالية اختيارها للمعالجة، وهذا يضمن التكافؤ بين مجموعتي المعالجة والمقارنة في كل من الخصائص المرصودة وغير المرصودة.

افترض الآن أنه من أصل الأشخاص الـ ١٠٠٠ الأصليين، كانت نسبة ٤٠٪ من النساء. ولأن الأسماء اختيرت عشوائيًا، فإن من بين الأسماء الـ ٥٠٠ المسحوبة من الوعاء، ستكون نسبة ٤٠ بالمائة من النساء أيضًا. من بين الأشخاص الـ ١٠٠٠، إذا كان لنسبة ٢٠ بالمائة عيون زرقاء، فإنه ينبغي أن تكون لنسبة ٢٠ بالمائة تقريبًا من كلتا مجموعتي المعالجة والمقارنة عيون زرقاء. عمومًا، إذا كان المجتمع الإحصائي للوحدات المؤهلة كبيرًا بما يكفي، فإن آلية التعيين العشوائي ستضمن أن أي خاصية من خصائص المجتمع الإحصائي ستنتقل إلى كل من مجموعة العلاج ومجموعة المقارنة. مثلما تنتقل خصائص مثل النوع الاجتماعي للشخص أو لون عينيه إلى مجموعة المعالجة ومجموعة المقارنة، فإن الخصائص التي من الأصعب ملاحظتها منطقيًا (المتغيرات غير الملحوظة)، مثل التحفيز أو التفضيلات أو السمات الشخصية الأخرى

التي يصعب قياسها، ستنتقل بالتساوي أيضًا على مجموعتي المعالجة والمقارنة. وبالتالي، فإن مجموعتي المعالجة والمقارنة اللتين تنشآن عن التعيين العشوائي ستكونان متشابهتين ليس فقط في خصائصها الملحوظة ولكن أيضًا في خصائصها غير الملحوظة. يضمن وجود مجموعتين متشابهتين في كل شيء كون الواقع المضاد المقدر يقارب القيمة الحقيقية للحصيلة في غياب المعالجة، وأنه بمجرد تنفيذ البرنامج، لن تخضع الآثار المقدرة لتحيز الاختيار.

عندما يستخدم تقييم ما التعيين العشوائي لمجموعات المعالجة والمقارنة، فيجب من الناحية النظرية أن تنتج العملية مجموعتين متكافئتين، شريطة أن تعتمد على عدد كبير بما يكفي من الوحدات. يمكننا اختبار هذا الافتراض تجريبيًا من خلال بيانات خط الأساس من عينة التقييم لدينا، والتحقق من أنه لا توجد في الحقيقة فروقات منهجية في الخصائص الملحوظة بين مجموعتي المعالجة والمقارنة قبل بدء البرنامج. بعد ذلك، عقب إطلاق البرنامج، إذا رصدنا الفروقات في الحاصل بين مجموعتي المعالجة والمقارنة، فسنعلم أنه يمكن تفسير تلك الفروقات فقط بإدخال البرنامج، بما أن المجموعتين كانتا متطابقتين عند خط الأساس، قبل بدء البرنامج، وبما أنهما تتعرضان للعوامل البيئية الخارجية نفسها بمرور الزمن. بهذا المعنى، فإن مجموعة المقارنة تتحكم في جميع العوامل التي قد تفسر أيضًا الحصلة محل الاهتمام.

لتقدير أثر برنامج في إطار التعيين العشوائي، فإننا ببساطة نأخذ الفرق بين الحصلة من خلال المعالجة (الحصلة المتوسطة لمجموعة المعالجة المعينة عشوائيًا) وتقديرنا للواقع المضاد (الحصلة المتوسطة لمجموعة المقارنة المعينة عشوائيًا). يمكننا أن نثق بأن أثرنا المقدّر يشكّل الأثر الحقيقي للبرنامج، بما أننا استبعدنا جميع العوامل الملحوظة وغير الملحوظة التي كان يمكن أن تفسّر بخلاف ذلك الفرق في الحاصل بصورة معقولة. ناقشنا في المربعات ٢-٤ إلى ٦-٤ تطبيقات واقعية للتعيين العشوائي لتقدير أثر عدد من عمليات التدخل المختلفة حول العالم. وافترضنا في الشكل ١-٤ أن جميع الوحدات في المجتمع الإحصائي المؤهل ستُعَيّن إما في مجموعة المعالجة أو في مجموعة المقارنة. ولكن، في بعض الحالات، لا يلزم تضمين جميع الوحدات في التقييم. على سبيل المثال، إذا كان المجتمع الإحصائي للوحدات المؤهلة يضم مليون أم، وأدرت تقييم فعالية المكافآت المالية فيما يتعلق باحتمال تلقيح أطفالهن، فقد يكون كافيًا اختيار عينة عشوائية من ١٠٠٠ أم على سبيل المثال، وتعيين أولئك الأمهات الـ ١٠٠٠ في مجموعة المعالجة أو مجموعة المقارنة. يوضح الشكل ٢-٤ هذه العملية. بالمنطق نفسه الموضح سابقًا، يحافظ اختيار عينة عشوائية من المجتمع الإحصائي للوحدات المؤهلة لتكوين عينة التقييم على المجتمع الإحصائي للوحدات المؤهلة. ضمن مجموعة التقييم، يحافظ التعيين العشوائي للأفراد في مجموعتي المعالجة والمقارنة مرة أخرى على الخصائص. نناقش أخذ العينات أكثر في الفصل الخامس عشر.

الصلاحية الخارجية والداخلية

تضمن الخطوات المذكورة أعلاه للتعيين العشوائي للمعالجة الصلاحية الداخلية والخارجية لتقديرات الأثر (الشكل ٢-٤).

المربع ٢-٤: التعيين العشوائي بمثابة قاعدة لتخصيص الأموال للبرنامج: التحويلات النقدية المشروطة والتعليم في المكسيك

ال ١٨١ المتبقية بمثابة مجموعة مقارنة قبل دخول البرنامج في العام الثالث. استنادًا إلى التعيين العشوائية، وجد Schultz (٢٠٠٤) زيادة متوسطة في الالتحاق بنسبة ٣,٤ في المائة لجميع الطلاب في الصفوف ٨-١، وكانت الزيادة الكبرى في صفوف الفتيات اللاتي أكملن الصف السادس، بنسبة ١٤,٨ في المائة. السبب المحتمل هو أن الفتيات يملن إلى التسرب من المدرسة بمعدلات أعلى مع تقدمهن في السن، ولذلك أعطيت تحويلات أكبر قليلًا كي يبقين في المدرسة بعد مستويات الصف الابتدائي. بعد ذلك، استُكملت هذه الآثار قصيرة الأمد بالاستقراء للتنبؤ بأثر برنامج Progresas على مدى أطول على التعليم والدخل المكتسب مدى الحياة.

يوفر برنامج Progresas، الذي أصبح اسمه الآن "Prospera"، تحويلات نقدية للأمهات الفقيرة في ريف المكسيك، شريطة تسجيل أولادهن في المدرسة وخضوع الأطفال لفحوصات صحية منتظمة (انظر المربع ١-١ في الفصل الأول). تشكّل التحويلات النقدية للأطفال في الصفوف من ٣ إلى ٩ نسبة ٥٠٪ إلى ٧٥٪ تقريبًا من تكلفة التعليم، وهي مضمونة لمدة ثلاث سنوات. حُدّدت المجتمعات والأسر المستحقة للبرنامج استنادًا إلى مؤشر الفقر الذي أنشئ من خلال بيانات التعداد وجمع بيانات خط الأساس. ونظرًا لوجود حاجة إلى التدرج في البرنامج الاجتماعي واسع النطاق، اختير ثلثا المجتمعات المحلية تقريبًا (٣١٤ من أصل ٤٩٥) عشوائيًا لتلقي البرنامج في أول عامين، واستخدمت المجتمعات

المصدر: Schultz 2004. اختبارًا للذقة، جمع Schultz التعيين العشوائي مع طريقة الاختلاف في الاختلافات التي نوقشت في الفصل السابع.

المربع ٣-٤: التعيين العشوائي للمنح لتحسين فرص عمل الشباب في شمال أوغندا

التمويل عشوائيًا. طلبت الحكومة المركزية من حكومات المناطق تقديم أكثر من ضعف عدد المقترحات التي يمكنها تمويلها. وبعض فحص المقترحات، حُدّدت الحكومة قائمة بـ ٥٣٥ عرضًا مؤهلًا للبرنامج. بعد ذلك، عيّنت المقترحات عشوائيًا في مجموعة المعالجة أو مجموعة المقارنة، بواقع ٢٦٥ في مجموعة المعالجة و٢٧٠ في مجموعة المقارنة.

(تابع)

في عام ٢٠٠٥، شرعت حكومة أوغندا في برنامج يهدف إلى الحد من بطالة الشباب وتعزيز الاستقرار الاجتماعي في المنطقة الشمالية المتضررة من النزاع. وقد دعا برنامج فرص الشباب مجموعات من الشباب إلى تقديم مقترحات للمنح في أنشطة الأعمال والتدريب المهني. قُدّمت آلاف المقترحات، ولكن الحكومة كانت قادرة على تمويل بضع مئات فقط.

بالاستفادة من ارتفاع الطلب على البرنامج، عمل المقيّمون مع الحكومة لاختيار المجموعات التي تلقت

المربع ٣-٤: التعيين العشوائي للمنح لتحسين فرص عمل الشباب في شمال أوغندا (تابع)

المقارنة. وكانوا يكسبون أيضًا نسبة تزيد بـ ٣٨ في المائة ولديهم أسهم رأسمالية أكثر بنسبة ٥٧ في المائة. ولكن، لم يجد الباحثون أثرًا على التماسك الاجتماعي أو السلوك المعادي للمجتمع.

بلغ متوسط مبلغ المنحة في مجموعة المعالجة ٣٨٢ دولارًا أمريكيًا لكل شخص. بعد أربع سنوات من الصرف، كان الشباب في مجموعة المعالجة أكثر قابلية بمرتين لممارسة حرفة ماهرة مقارنة بالشباب في مجموعة

المربع ٤-٤: التعيين العشوائي لتدخلات المياه والصرف الصحي في ريف بوليفيا

أسماء المجتمعات عشوائيًا ووُضعت في قائمة. عُيِّنَت المجتمعات في مقدمة القائمة ضمن مجموعة المعالجة. كان كاتب بالعدل مستقل يراقب كل قرعة، ثم يسجل النتائج ويصادق عليها ليعطي مستوى آخر من الشرعية للعملية. وبالنسبة إلى المجتمعات التي استُبعدت من البرنامج، التزمت حكومات البلديات باستخدام نفس القائمة المرتبة عشوائيًا لتخصيص التمويل المستقبلي بعد الانتهاء من التقييم. وبهذه الطريقة، لن تُستبعد أي مجتمعات من التدخل بغرض التقييم فحسب، ولكن ستوجد مجموعة مقارنة طالما أن قيود الميزانية حدّت من عدد المشاريع في كل بلدية.

ابتداءً من عام ٢٠١٢، نفذت الحكومة البوليفية، بدعم من بنك التنمية للبلدان الأمريكية، تعيينًا عشوائيًا لتدخلات المياه والصرف الصحي للمجتمعات الريفية الصغيرة. وقد حدّد البرنامج ضمن البلديات الـ ٢٤ الأكثر حاجةً في البلاد أكثر من ٣٦٩ مجتمعًا مؤهلًا للتدخل. وبما أن الموارد كانت متاحة لتغطية ١٨٢ مجتمعًا فقط، استخدم البرنامج التعيين العشوائي لمنح كل مجتمع مؤهل فرصة متساوية للمشاركة. نظّم المشرفون على البرنامج، بالاشتراك مع حكومات البلديات، سلسلة أحداث عُقدت فيها قُرعات بحضور رؤساء المجتمعات والصحافة والمجتمع المدني. قُسمت المجتمعات أولاً وفقًا لحجم التعداد السكاني فيها. بعد ذلك، وضمن كل مجموعة، سُحبت

المصدر: Inter-American Development Bank Project No. BO-L1065, <http://www.iadb.org/en/projects/project-description-title,1303.html?id=BO-L1065>.
ملاحظة: راجع القرعة العامة للتعيين العشوائي على الرابط <https://vimeo.com/86744573>.

تعني *الصلاحية الداخلية* أن الأثر المقدر للبرنامج خالٍ من جميع العوامل المربكة الأخرى—أو بمعنى آخر تعني أن مجموعة المقارنة توفر تقديرًا دقيقًا للواقع المضاد، ومن ثم فإننا نقدّر الأثر الحقيقي للبرنامج. تذكر أن التعيين العشوائي ينتج مجموعة مقارنة مكافئة إحصائيًا لمجموعة المعالجة عند خط الأساس، قبل بدء البرنامج. وبمجرد بدء البرنامج، تتعرض مجموعة المقارنة لنفس مجموعة العوامل الخارجية في مجموعة المعالجة بمرور الزمن، والاستثناء الوحيد هو البرنامج. لذلك، إذا ظهرت أي فروقات في الحصائل بين مجموعتي المعالجة والمقارنة، فيمكن أن تُعزى إلى وجود البرنامج في مجموعة المعالجة. تُضمّن الصلاحية الداخلية لتقييم الأثر من خلال عملية *التعيين العشوائي للمعالجة*.

المربع ٤-٥: التعيين العشوائي لحماية مياه الينابيع في سبيل تحسين الصحة في كينيا

منطقتين في غرب كينيا. ونتيجةً لوجود عوائق مادية وإدارية، قررت منظمة ICS تنفيذ البرنامج على مراحل على مدار أربعة أعوام. وقد أتاح ذلك للمقيمين استخدام ينابيع لم تتلق المعالجة بصفاتها مجموعة المقارنة. ومن بين الينابيع الـ ٢٠٠ المؤهلة، اختير ١٠٠ ينبوع عشوائيًا لتلقي المعالجة في أول عامين. توصلت الدراسة إلى أن حماية الينابيع حدّت من تلوث المياه البرازية بنسبة ٦٦ في المائة وإسهال الأطفال بين مستخدمي الينابيع بنسبة ٢٥ في المائة.

وُثّق الرابط بين جودة الماء والآثار الصحية في البلدان النامية. ولكن القيمة الصحية لتحسين البنية التحتية حول مصادر المياه أقل وضوحًا. قاس كريمز وآخرون (٢٠١١) آثار برنامج يوفر تكنولوجيا لحماية الينابيع بغرض تحسين جودة المياه في كينيا، وعيّنوا عشوائيًا الينابيع التي ستلقى المعالجة.

تحصل نسبة ٤٣ بالمائة من الأسر في ريف كينيا الغربي على مياه الشرب من ينابيع طبيعية. تغلق تكنولوجيا حماية الينابيع مصادر الينابيع لتقليل التلوث. وبدءًا من عام ٢٠٠٥، نفّذت منظمة دعم الطفل الدولية (ICS) غير الحكومية برنامجًا لحماية الينابيع في

المصدر: Kremer and others 2011.

المربع ٤-٦: التعيين العشوائي للمعلومات حول مخاطر فيروس نقص المناعة البشرية للحد من حمل المراهقات في كينيا

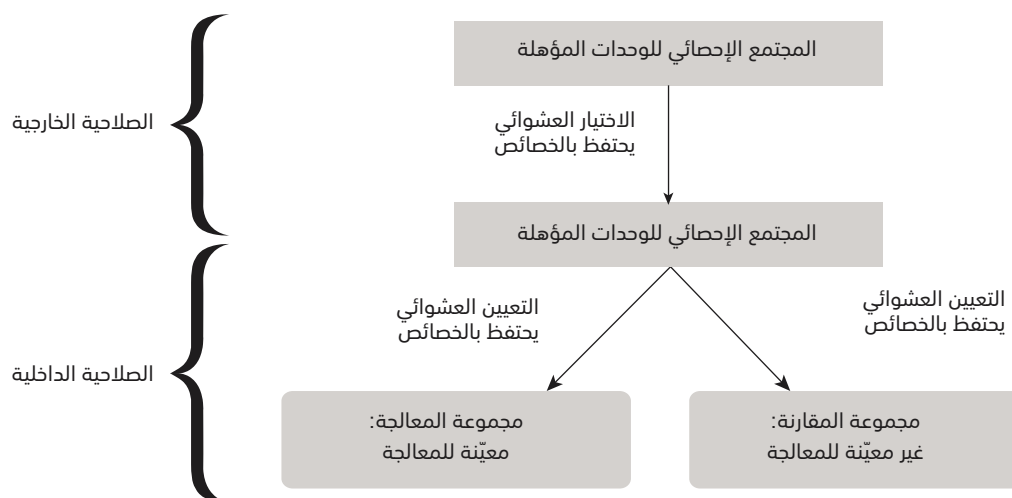
حسب الموقع، ونتائج الاختبارات، ونسبة النوع الاجتماعي للطلاب. عيّن ٧١ مدرسة عشوائيًا بعد ذلك للمعالجة الثانية، وقُسمت إلى طبقات للمشاركة في المعالجة الأولى. وقد أنتج ذلك مجموعتين من المدارس: المدارس المتلقية للمعالجة الأولى، والمدارس المتلقية للمعالجة الثانية، والمدارس غير المتلقية لأي من المعالجتين. وقد ضمن التعيين العشوائي للمدارس عدم وجود فرق منهجي في المعلومات التي تعرض لها الطلاب قبل بدء البرنامج. وبعد عامٍ من انتهاء البرنامج، توصل دوباس إلى أن حملة معلومات المخاطر النسبية أدت إلى انخفاض بنسبة ٢٨ بالمائة في احتمالية حمل الفتيات. على العكس من ذلك، لم يظهر أي تأثير على الحمل لدى المراهقات في المدارس التي تلقت المعالجة الأولى فقط.

اختبر دوباس (٢٠١١)، في تجربة عشوائية في غرب كينيا، فعالية علاجين تعليميين مختلفين لفيروس نقص المناعة البشرية/الإيدز في الحد من السلوك الجنسي غير الآمن بين المراهقين. تضمنت المعالجة الأولى تدريب معلمين على المنهاج الوطني لفيروس نقص المناعة البشرية/الإيدز، الذي كان يركز على تجنب المخاطر وتشجيع الامتناع عن ممارسة الجنس. كانت المعالجة الثانية، وهي حملة معلومات المخاطر النسبية، تهدف إلى الحد من الجنس بين الرجال المسنين والفتيات الأصغر سنًا عبر توفير معلومات حول معدلات فيروس نص المناعة البشرية مصنّفة حسب العمر والنوع الاجتماعي.

أُقيمت الدراسة في منطقتين ريفيتين في كينيا، وتضمنت العينة ٣٢٨ مدرسة ابتدائية. عيّن الباحثون عشوائيًا ١٦٣ مدرسة لتلقي المعالجة الأولى، مقسمة

المصدر: Dupas 2011.

الشكل ٢-٤ أخذ العينات عشوائيًا والتعيين العشوائي للمعالجة



المفهوم

الأساسي

يكون التقييم صالحًا داخليًا إذا كان يوفر تقديرًا دقيقًا للواقع المضاد من خلال مجموعة مقارنة صالحة.

المفهوم

الأساسي

يكون التقييم صالحًا خارجيًا إذا كانت عينة التقييم تمثل بدقة المجتمع الإحصائي للوحدات المؤهلة. يمكن تعميم نتائج التقييم حينها للمجتمع الإحصائي للوحدات الصالحة.

تعني *الصلاحية الخارجية* أن عينة التقييم تمثل بدقة المجتمع الإحصائي للوحدات المؤهلة. يمكن تعميم نتائج التقييم حينها للمجتمع الإحصائي للوحدات الصالحة. نستخدم أخذ العينات عشوائيًا للتأكد من أن عينة التقييم تعكس بدقة المجتمع الإحصائي للوحدات المؤهلة بحيث يمكن إجراء استكمال بالاستقراء للتأثيرات المحددة في عينة التقييم للمجتمع الإحصائي.

لاحظ أننا استخدمنا عملية عشوائية لغرضين اثنين: *الاختيار العشوائي* لعينة (للصلاحية الخارجية)، و*التعيين العشوائي* للمعالجة بمثابة طريقة لتقييم الأثر (للصلاحية الداخلية). يمكن أن ينتج تقييم الأثر تقديرات صالحة داخليًا للأثر من خلال التعيين العشوائي للمعالجة، ولكن، في حال إجراء التقييم على عينة غير عشوائية للمجتمع الإحصائي، يمكن ألا تُعَمِّم الآثار المقدّرة على المجتمع الإحصائي للوحدات المؤهلة. وعلى العكس من ذلك، إذا كان التقييم يستخدم عينة عشوائية للمجتمع الإحصائي للوحدات الصالحة، ولكن المعالجة لا تُعَيَّن بطريقة عشوائية، فإن العينة ستكون تمثيلية، ولكن مجموعة المقارنة قد لا تكون صالحة، ما قد يعرض الصلاحية الداخلية للخطر. قد تواجه البرامج في بعض السياقات قيودًا تتطلب مفاضلةً بين الصلاحية الداخلية والخارجية. على سبيل المثال البرنامج الذي نوقش سابقًا، الذي يستهدف الـ ٢٠ بالمائة من الأسر استنادًا إلى الدخل الأقل. إذا كان هذا البرنامج يضم جميع الأسر دون النسبة المئوية الـ ١٥، ولكنه يجري تقييم أثر للتعيين العشوائي بين عينة عشوائية للأسر في النسب المئوية

من الـ ١٥ إلى الـ ٢٥، فستكون لهذا التقييم صلاحية داخلية بفضل التعيين العشوائي: أي أننا سنعرف الأثر الحقيقي للمجموعة الفرعية للأسر في النسب المئوية من الـ ١٥ إلى الـ ٢٥. ولكن الصلاحية الخارجية لتقييم الأثر ستكون محدودة، وذلك لأنه لا يمكن استكمال النتائج بالاستقراء مباشرة للمجتمع الإحصائي من المستفيدين بأكمله: بالتحديد للأسر دون النسبة المئوية الـ ١٥.

متى يمكن استخدام التعيين العشوائي؟

يمكن استخدام التعيين العشوائي بمثابة قاعدة تخصيص برنامج في واحد من سيناريوهين محددتين:

١. عندما يكون المجتمع الإحصائي المؤهل أكبر من عدد الأماكن المتوفرة في البرنامج. عندما يتجاوز الطلب على برنامج ما العرض، يمكن استخدام القرعة لاختيار مجموعة المعالجة ضمن المجتمع الإحصائي المؤهل. وفي هذا السياق، تحصل كل وحدة في المجتمع الإحصائي على الفرصة نفسها (أو فرصة معروفة أكبر من الصفر وأصغر من الواحد) في أن تُختار للبرنامج. المجموعة التي تفوز بالقرعة هي مجموعة المعالجة، وباقي المجتمع الإحصائي الذي لا يخضع للبرنامج يمثل مجموعة المقارنة. طالما أن هناك عوائق تحول دون توسيع البرنامج ليشمل المجتمع الإحصائي بأكمله، يمكن الحفاظ على مجموعات المقارنة لقياس أثر البرنامج على المدى القصير وال المدى المتوسط وال المدى الطويل. وفي هذا السياق، لا تنشأ معضلة أخلاقية بسبب الاحتفاظ بمجموعة مقارنة إلى ما لانهاية، طالما أن مجموعة فرعية من المجتمع الإحصائي ستترك خارج البرنامج بسبب قيود السعة.

على سبيل المثال، لنفترض أن وزارة التعليم تريد توفير مكتبات مدرسية للمدارس العامة في جميع أنحاء البلاد، لكن وزارة المالية تخصص أموالاً تكفي لتغطية ثلثها فقط. إذا أرادت وزارة التعليم أن تحظى كل مدرسة عامة بفرصة متساوية للحصول على مكتبة، فستجري قرعة تمتلك فيها كل مدرسة فرصة الاختيار نفسها (١ من أصل ٣). تحصل المدارس التي تفوز بالقرعة على مكتبة جديدة وتشكّل مجموعة المعالجة، ولا يُعطى الثلثان المتبقيان من المدارس العامة في البلاد المكتبة ويشكلان مجموعة المقارنة. وما لم تُخصّص أموال إضافية لبرنامج المكتبة، تبقى مجموعة من المدارس دون تمويل للمكتبات، ويمكن استخدامها بمثابة مجموعة مقارنة لمعالجة الواقع المضاد.

٢. عندما يتطلب برنامج ما التنفيذ التدريجي حتى يشمل المجتمع الإحصائي بأكمله. عند تنفيذ برنامج على مراحل، فإن التوزيع العشوائي للترتيب الذي يتلقى المشاركون وفقه البرنامج يمنح كل وحدة مؤهلة الفرصة نفسها لتلقي المعالجة في المرحلة الأولى أو في مرحلة لاحقة من البرنامج. وطالما لم يتم إدخال المجموعة الأخيرة في البرنامج تدريجياً، فإنها تؤدي دور مجموعة مقارنة صالحة يمكن من خلالها تقدير الواقع المضاد للمجموعات التي تُدخّل على مراحل بالفعل. يمكن أن يسمح هذا الإعداد أيضاً للمقيّم باختيار آثار التعرض التفاضلي للمعالجة: أي أثر تلقي البرنامج لزمّن أطول أو أقصر.

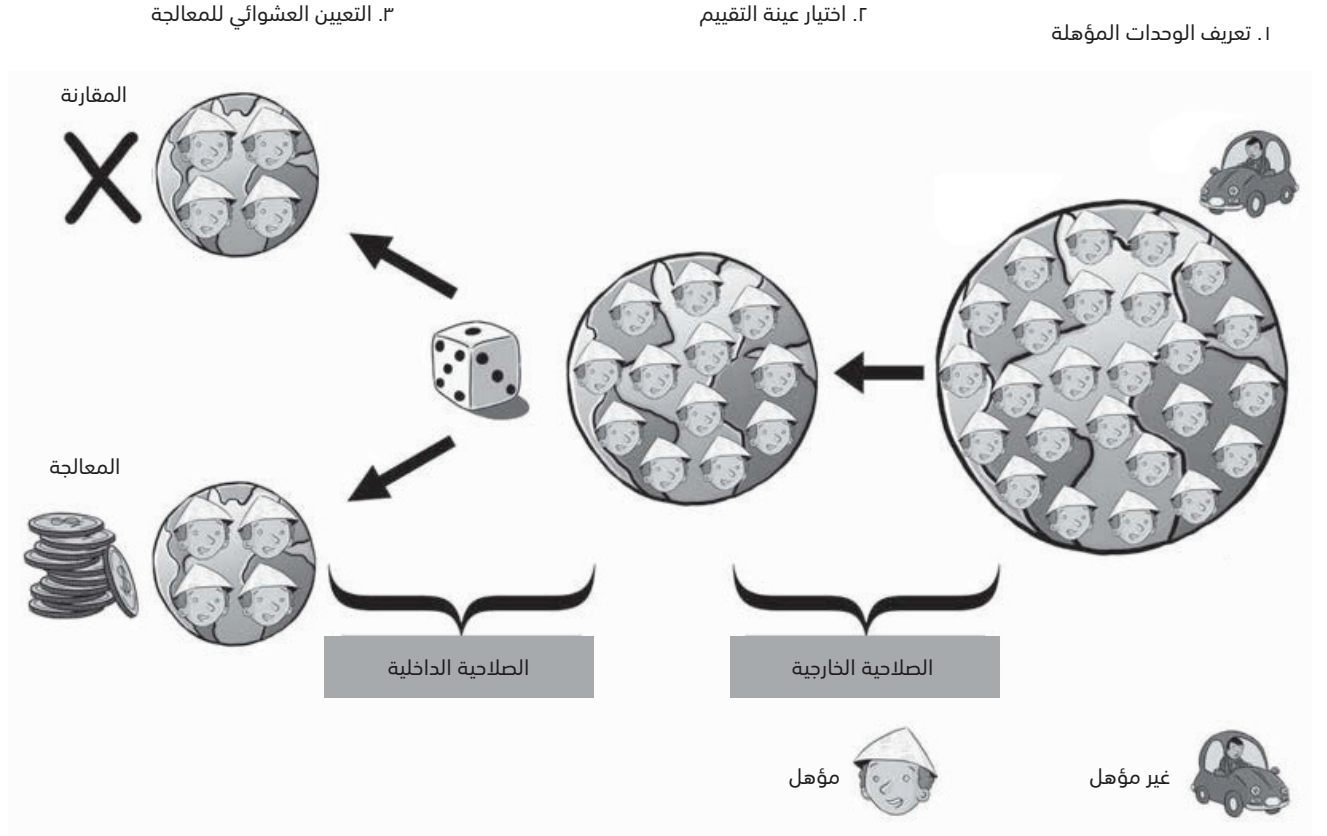
على سبيل المثال، افترض أن وزارة الصحة تريد تدريب الممرضات الـ ١٥٠٠٠ جميعاً في الريف على استخدام بروتوكول صحي جديد ولكنها بحاجة إلى ثلاث سنوات لتدريبهن جميعاً. في سياق تقييم الأثر، يمكن أن تعيّن الوزارة عشوائياً ثلث الممرضات لتلقي التدريب في العام الأول، وثلثاً آخر لتلقي التدريب في العام الثاني، وثلثاً لتلقي التدريب في العام الثالث: لتقييم أثر البرنامج التدريبي بعد عام واحد من تنفيذه، ستشكل مجموعة الممرضات المدربات في العام الأول مجموعة المعالجة، وستكون مجموعة الممرضات المعينة عشوائياً للتدريب في السنة الثالثة هي مجموعة المقارنة، لأنهم لم يتلقوا التدريب بعد.

كيف تعيّن المعالجة عشوائيًا؟

بعد أن ناقشنا ما يفعله التعيين العشوائي ولماذا ينتج مجموعة مقارنة جيدة، سننتقل إلى خطوات تعيين المعالجة عشوائيًا بنجاح. يوضح الشكل ٤-٣ هذه العملية.

الخطوة الأولى هي تحديد الوحدات المؤهلة للبرنامج. تذكر أنه اعتمادًا على برنامج محدّد، يمكن أن تكون الوحدة شخصًا أو مركزًا صحيًا أو مدرسة أو شركة أو قرية أو بلدية بأكملها. يتكون المجتمع الإحصائي للوحدات المؤهلة من أولئك الذين تهتم بمعرفة أثر برنامجك من أجلهم. على سبيل المثال، إذا كنت تنفذ برنامج تدريب لمعلمي المدارس الابتدائية في المناطق الريفية، فإن معلمي المدارس الابتدائية في المناطق الحضرية أو معلمي المدارس الثانوي لن ينتموا إلى الوحدات المؤهلة.

وبمجرد أن تحدد المجتمع الإحصائي للوحدات المؤهلة، فسيلزم مقارنة حجم المجموعة بعدد الملاحظات المطلوب للتقييم. يُحدّد حجم عينة التقييم من خلال حسابات القوة الإحصائية ويرتكز على أنواع الأسئلة التي تودّ أن تتم الإجابة عنها (انظر الفصل الخامس عشر). إذا كان المجتمع الإحصائي صغيرًا، فقد يتعين تضمين جميع الوحدات المؤهلة في التقييم. بدلاً من ذلك، إذا كان هناك وحدات معالجة أكثر من المطلوب للتقييم، فإن الخطوة ٢ هي اختيار عينة وحدات من المجتمع الإحصائي المطلوب تضمينه في عينة التقييم.



تجري هذه الخطوة الثانية بصورة رئيسية للحد من تكاليف جمع البيانات. إذا وُجد أنه يمكن استخدام البيانات من أنظمة المتابعة القائمة للتقييم، وأن تلك الأنظمة تشمل المجتمع الإحصائي للوحدات المؤهلة بأكمله، فقد لا تحتاج إلى سحب عينة تقييم عشوائية. ولكن، تخيل تقييماً يتضمن فيه المجتمع الإحصائي للوحدات المؤهلة عشرات آلاف المعلمين في كل مدرسة في البلاد، وأنت بحاجة إلى جمع معلومات مفصلة عن المعرفة والممارسات التربوية للمعلمين. إن مقابلة كل معلم في البلاد وتقييمه قد تكون باهظة الكلفة وغير ممكنة من الناحية اللوجستية. استناداً إلى حساباتك للقوة الإحصائية، يمكنك تحديد أنه للإجابة عن سؤال التقييم، يكفي أخذ عينة من ١٠٠٠ معلم موزعين على ٢٠٠ مدرسة. طالما أن عينة المعلمين تمثل المجتمع الإحصائي للمعلمين بأكمله، فإن أي نتائج يتوصل إليها التقييم ستكون صالحة خارجياً ويمكن تعميمها على بقية المعلمين في البلاد. وسيكون جمع بيانات عن هذه العينة المكونة من ١٠٠٠ معلم في ٢٠٠ مدرسة أقل تكلفة بكثير من جمع بيانات عن كل معلم في جميع المدارس في البلاد.

الخطوة ٣ هي صياغة مجموعات المعالجة والمقارنة من الوحدات في عينة التقييم من خلال التعيين العشوائي. في الحالات التي يتعين فيها إجراء تعيين عشوائي في محفل عام، وليكن على التلفاز، فقد تحتاج إلى استخدام تقنية بسيطة كرمي قطعة نقدية أو اختيار أسماء من داخل قبعة. تفترض الأمثلة التالية أو وحدة التوزيع العشوائي هو شخص فردي، ولكن المنطق نفسه ينطبق على التوزيع العشوائي للمزيد من وحدات الملاحظة العنقودية، كالمدارس أو الشركات أو المجتمعات:

١. إذا أردت تعيين ٥٠ بالمائة من الأفراد في مجموعة المعالجة و ٥٠ بالمائة في مجموعة المقارنة، فيمكنك إجراء قرعة بعملة نقدية لكل شخص. ويجب عليك أن تحدد مسبقاً أي وجه للعملة المعدنية يحدد الشخص في مجموعة المعالجة.

٢. إذا أردت تعيين ثلث عينة التقييم في مجموعة المعالجة، فيمكنك رمي حجر نرد لكل شخص. عليك أن تحدد قاعدةً أولاً. على سبيل المثال، يعني ظهور العدد ١ أو ٢ عند رمي حجر نرد التعيين في مجموعة المعالجة، بينما يعني ظهور الأعداد ٣ أو ٤ أو ٥ أو ٦ التعيين في مجموعة مقارنته. وهذا سترمي حجر نرد مرة واحدة لكل شخص في مجموعة التقييم وستعيّنه استناداً إلى العدد الذي يظهر.

٣. اكتب أسماء كل شخص في ورقة، بحيث تكون جميع الأوراق متطابقة في الحجم والشكل. اطوِ الأوراق بحيث لا يمكن رؤية الأسماء، واخبط الورق في قبعة أو وعاء آخر. وقبل أن تبدأ السحب، حدّد قاعدتك: أي عدد الأوراق التي ستسحبها وأن سحب اسم يعني تعيين الشخص الذي يظهر اسمه في مجموعة المعالجة. وبمجرد أن تتضح القاعدة، اطلب من شخص من الجمهور (شخص غير متحيز، كطفل) سحب أكبر عدد من الأوراق يكافئ العدد الذي تحتاج إليه من المشاركين في مجموعة المعالجة.

إذا احتجت إلى الكثير من الوحدات (ولنقل أكثر من ١٠٠)، فإن استخدام مقاربات بسيطة مثل هذه سيستغرق وقتاً طويلاً للغاية، وستعين عليك استخدام عملية مؤتمتة. لفعل ذلك، عليك أن تقرر أولاً قاعدة بشأن كيفية تعيين المشاركين استناداً إلى أرقام عشوائية. على سبيل المثال، إذا احتجت إلى تعيين ٤٠ من أصل ١٠٠ وحدة من عينة التقييم في مجموعة المعالجة، فقد تقرر تعيين الوحدات الـ ٤٠ ذات الأرقام العشوائية العليا في مجموعة المعالجة وتعيين الوحدات المتبقية في مجموعة المقارنة. لتنفيذ التعيين العشوائي، عليك تعيين رقم عشوائي في كل وحدة في عينة التقييم باستخدام مولد أرقام عشوائية لجداول البيانات، أو برنامج إحصائي اختصاصي (الشكل ٤-٤)، واستخدام القاعدة التي اخترتها سابقاً لتكوين مجموعتي المعالجة والمقارنة. من المهم تحديد القاعدة قبل توليد الأرقام العشوائية، وإلا فقد تميل إلى تحديد قاعدة استناداً إلى الأرقام العشوائية التي تراها، وهذا من شأنه إبطال التعيين العشوائي.

إن المنطق الكامن وراء العملية المؤتمتة لا يختلف عن التعيين العشوائي القائم على رمي القطعة النقدية أو اختيار الأسماء من قبعة: إنها آلية تحدد عشوائيًا ما إذا كانت كل وحدة ضمن مجموعة المعالجة أو مجموعة المقارنة.

وسواء أكنت تستخدم قرعة عامة، أو أرقامًا عشوائية يولدها الحاسوب، فمن المهم توثيق العملية لضمان شفافيتها. وهذا يعني أنه يجب تحديد قاعدة التعيين أولًا بصورة مسبقة وإبلاغ الجمهور عنها. ثانيًا، عليك الالتزام بالقاعدة بمجرد أن تسحب الأرقام العشوائية. ثالثًا، عليك أن تكون قادرًا على توضيح أن العملية كانت عشوائية حقًا. في حالي القُرعات ورمي حجر النرد، يمكنك تسجيل فيديو للعملية، ويتطلب التعيين القائم على الحاسوب للأرقام العشوائية توفير سجل لحساباتك، بحيث يمكن أن يكرر المدققون العملية.

الشكل ٤-٤ استخدام جدول بيانات للتعيين العشوائي للمعالجة

	A	B	C	D	E	F	G	H
1	Random number	Between 0 and 1.						
2	Goal	Assign 50% of evaluation sample to treatment group.						
3	Rule	If random number is above 0.5: assign person to treatment group; otherwise: assign person to comparison group.						
5	Unit identification	Name	Random number*	Final random number**	Assignment			
6	1001	Ahmed	0.7698674	0.479467635	0			
7	1002	Elisa	0.4054534	0.945729597	1			
8	1003	Anna	0.3584427	0.933658744	1			
9	1004	Jung	0.5010306	0.383305299	0			
10	1005	Tuya	0.8799600	0.102877439	0			
11	1006	Nilu	0.1764322	0.228446592	0			
12	1007	Roberto	0.0030776	0.444725231	0			
13	1008	Priya	0.7512858	0.817004226	1			
14	1009	Grace	0.1331390	0.955775449	1			
15	1010	Fathia	0.8735385	0.873459852	1			
16	1011	John	0.0089322	0.211028126	0			
17	1012	Alex	0.0762848	0.574082414	1			
18	1013	Nafula	0.5760701	0.151608805	0			
19	* type the formula =RAND(). Note that the random numbers in Column C are volatile: they change everytime you do a calculation.							
20	** Copy the numbers in column C and "Paste Special>Values" into Column D. Column D then gives the final random numbers.							
21	*** type the formula =IF(C(row number)>0.5,1,0)							

ما المستوى الذي تجري عنده التعيين العشوائي؟

يمكن إجراء التعيين العشوائي عند مستويات متعددة، بما فيها مستوى الفرد أو الأسرة أو الشركة أو المجتمع أو المنطقة. وعمومًا، يتأثر المستوى الذي تُعَيَّن عنده الوحدات عشوائيًا في مجموعتي المقارنة والمعالجة تأثيرًا كبيرًا بطريقة تنفيذ البرنامج ومكانه. على سبيل المثال، في حال تنفيذ برنامج صحي على مستوى العيادة الصحية، فستحدد أولاً عينة عشوائية للعيادات الصحية ثم تُعَيَّن بعضها عشوائيًا في مجموعة المعالجة والبعض الآخر في عينة المقارنة. وعندما يكون مستوى التعيين العشوائي أعلى أو أكثر تجميعًا، كما هو الحال على مستوى المناطق أو المقاطعات في بلد ما، فقد يصبح من الصعب إجراء تقييم أثر لأن عدد المناطق أو المقاطعات في معظم البلدان ليس كبيرًا بما فيه الكفاية كي تتمخض عنه مجموعتا معالجة ومقارنة متوازنتان. على سبيل المثال، إذا كان بلد ما يضم ست مقاطعات، فإن كلاً من مجموعتي المعالجة والمقارنة ستضم ثلاث مقاطعات فقط، وهذا لا يكفي لضمان توازن خصائص خط الأساس لمجموعتي المعالجة والمقارنة. علاوةً على ذلك، لكي يعطي التعيين العشوائي تقديرات أثر غير متحيزة، فمن المهم ضمان كون العوامل الخارجية ذات الإطار الزمني المحدد (مثل الطقس أو دورات الانتخابات المحلية) متماثلة في المتوسط ضمن مجموعتي المعالجة والمقارنة. عندما يزداد مستوى التعيين، يصبح من غير المحتمل بصورة متزايدة أن تتوازن هذه العوامل عبر مجموعتي المعالجة والمقارنة. على سبيل المثال، الهطول المطري عامل خارجي ذو إطار زمني محدد لأنه يختلف منهجيًا من عام لآخر. فعند التقييم في القطاع الزراعي، سنحتاج إلى ضمان أن فترات الجفاف تؤثر على مقاطعات المعالجة والمقارنة بصورة متساوية. وبوجود ثلاث مقاطعات في كل من مجموعتي المقارنة والمعالجة، سيكون من السهل فقدان هذا التوازن. من ناحية أخرى، لو خُفضت وحدة التعيين إلى مستوى المقاطعة الفرعية، كالبلدية، فمن المحتمل أكثر أن يتوازن الهطول المطري بين مجموعتي المعالجة والمقارنة بمرور الزمن.

وعلى صعيد آخر، كلما انخفض مستوى التعيين العشوائي أكثر—على سبيل المثال هبوطًا إلى مستوى الفرد أو الأسرة—ازدادت فرص تأثر مجموعة المقارنة دون قصد بالبرنامج. ثمة نوعان من المخاطر التي يجب وضعها في الحسبان عند اختيار مستوى التعيين، وهما التأثيرات غير المباشرة والامتثال غير المثالي. تحدث التأثيرات غير المباشرة عندما تؤثر مجموعة المعالجة بصورة مباشرة أو غير مباشرة على الحاصل في مجموعة المقارنة (أو العكس). ويحدث الامتثال غير المثالي عندما يشارك بعض أعضاء مجموعة المقارنة في البرنامج، أو عندما لا يشارك بعض أعضاء مجموعة المعالجة (راجع المناقشة الإضافية لهذين المفهومين في الفصل التاسع).




من خلال الدراسة المتأنيبة لمستوى التعيين العشوائي، يمكن تقليل خطر التأثيرات غير المباشرة والامتثال غير المثالي إلى الحد الأدنى. يمكن تعيين الأفراد في مجموعات أو تجمعات عنقودية، كالطلاب في المدارس أو الأسر في المجتمع، لتقليل تدفقات المعلومات والاتصالات بين الأفراد في مجموعتي المعالجة والمقارنة إلى الحد الأدنى. ولتقليل الامتثال غير المثالي، ينبغي اختيار مستوى التعيين أيضًا وفقًا لقدرة البرنامج للحفاظ على تمييز واضح بين مجموعتي المعالجة والمقارنة خلال التدخل. إذا كان البرنامج يضم أنشطة على مستوى المجتمع، فقد يصعب تجنب تعريض جميع الأفراد من المجتمع للبرنامج.

أحد الأمثلة الشهيرة للتأثيرات غير المباشرة توفير دواء لعلاج الديدان لدى الأطفال. إذا كانت أسر في مجموعة المعالجة تقع بالقرب من مجموعة المقارنة، فقد يتأثر الأطفال في أسر المقارنة إيجابيًا بالتأثيرات غير المباشرة للمعالجة لأن فرص إصابتهم بالديدان من جيرانهم ستنخفض (Kremer and Miguel 2004). لعزل أثر البرنامج، يتعين أن تقع أسر المعالجة والمقارنة بعيدًا عن بعضهما البعض بما يكفي لتجنب مثل هذه التأثيرات غير المباشرة. ولكن، مع ازدياد المسافة بين الأسر، ستزداد تكلفة تنفيذ البرنامج وإجراء المسوحات. بمثابة قاعدة جوهريّة، إذا كان من الممكن استبعاد التأثيرات غير المباشرة بصورة معقولة، فمن الأفضل إجراء التعيين العشوائي للمعالجة عند المستوى الأدنى الممكن لتنفيذ البرنامج، وهذا سيضمن كون عدد الوحدات في مجموعتي المعالجة والمقارنة أكبر ما يمكن.

تقدير الأثر بموجب التعيين العشوائي

بمجرد اختيارك عينة تقييم عشوائية وتعيينك المعالجة بطريقة عشوائية، سيكون من السهل جدًا تقدير أثر البرنامج. بعد تشغيل البرنامج لبعض الوقت، سيتعين قياس الحاصلات لمجموعتي المعالجة والمقارنة. أثر البرنامج هو ببساطة الفرق بين الحصلة المتوسطة (Y) لمجموعة المعالجة والحصلة المتوسطة (Y) لمجموعة المقارنة. على سبيل المثال، في المثال العام الوارد في الشكل ٤-٥، تساوي الحصلة المتوسطة لمجموعة المعالجة ١٠٠، وتساوي الحصلة المتوسطة لمجموعة المقارنة ٨٠، ومن ثم فإن أثر البرنامج يساوي ٢٠. نفترض حاليًا أن جميع الوحدات في مجموعة المعالجة تُعالج بصورة فاعلة ولا تُعالج أي وحدات من مجموعة المقارنة. في المثال الذي أوردناه عن برنامج تدريب المعلمين، يتلقى جميع المعلمين المعيّنين في مجموعة المعالجة التدريب ولا يتلقى أي من معلمي المقارنة هذا التدريب. ناقشنا في الفصل الخامس سيناريو (أكثر واقعية) عندما لا يكون الامتثال كاملاً؛ أي عندما تشارك نسبة أقل من ١٠٠ بالمائة من الوحدات في مجموعة المعالجة فعليًا في التدخل أو عند وصول بعض مجموعات المقارنة على البرنامج. وفي هذه الحالة، يبقى من الممكن الحصول على تقدير غير متحيز لأثر البرنامج من خلال التعيين العشوائي، على الرغم من أن تفسير النتائج سيتباين.

الشكل ٥-٤ تقدير الأثر بموجب التعيين العشوائي

المعالجة	المقارنة	تقييم
المتوسط (Y) لمجموعة المعالجة = ١٠٠	المتوسط (Y) لمجموعة المقارنة = ٨٠	الأثر $\Delta Y = ٢٠$
		

قائمة المراجعة: التعيين العشوائي

- التعيين العشوائي هي الطريقة الأقوى لتقدير الواقع المضاد، وهي تُعد المعيار الذهبي لتقييم الأثر. ولكن، ينبغي رغم ذلك التفكير في إجراء بعض الاختبارات الأساسية لتقدير صلاحية إستراتيجية التقييم ضمن السياق .
- ✓ هل خصائص خط الأساس متوازنة. قارن خصائص خط الأساس لمجموعة المعالجة ومجموعة المقارنة.^٦
- ✓ هل حدث أي عدم امتثال مع التعيين؟ تحقق مما إذا كانت جميع الوحدات المؤهلة قد تلقت المعالجة ومن أنه لم تتلقَّ أي وحدات غير مؤهلة المعالجة. في حال حدوث عدم امتثال، سيتعين عليك استخدام طريقة المتغير المساعد (راجع الفصل الخامس).
- ✓ هل أعداد الوحدات في مجموعتي المعالجة والمقارنة كبيرة بما يكفي؟ إذا لم تكن كذلك، فقد نحتاج إلى دمج طريقة التعيين العشوائي مع طريقة الاختلاف في الاختلافات (راجع الفصل السابع).
- ✓ هل ثمة أي سبب للاعتقاد بأن الحاصل لبعض الوحدات قد تعتمد بشكل ما على تعيين وحدات أخرى؟ هل يمكن أن يكون هناك أثر للمعالجة على الوحدات في مجموعة المقارنة (انظر الفصل التاسع)؟

تقدير أثر برنامج معونة التأمين الصحي (HISP): التعيين العشوائي

لنعد الآن إلى مثال برنامج معونة التأمين الصحي (HISP) ونتحقق من معنى التعيين العشوائي في هذا السياق. تذكر أنك تحاول تقدير أثر البرنامج من تجربة تضم ١٠٠ قرية معالجة.

بعد إجراء تقدير أثر باستخدام مقدّرات من المحتمل أن تكون متحيزة للواقع المضاد في الفصل الثالث (مع توصيات سياسية متضاربة)، قررت العودة إلى لوحة الرسم لإعادة التفكير في كيفية الحصول على تقدير أدقّ للواقع المضاد. بعد مداولات لاحقة مع فريقك للتقييم، تقتنع بأنه للحصول على تقدير صالح للواقع المضاد يلزم تحديد مجموعة من القرى التي تشبه قدر الإمكان قرى المعالجة الـ ١٠٠ من جميع النواحي، باستثناء أن إحدى المجموعتين شاركت في برنامج معونة التأمين الصحي (HISP) والأخرى لم تشارك. نظرًا إلى أن برنامج معونة التأمين الصحي (HISP) قد طُرح بمثابة برنامج تجريبي، وأن قرى المعالجة الـ ١٠٠ قد اختيرت عشوائيًا من بين جميع القرى الريفية في الدولة، تُلاحظ أن قرى المعالجة يجب أن يكون لها، في المتوسط، خصائص مماثلة للقرى الريفية غير المعالجة في البلد. لذلك، يمكن تقدير الواقع المضاد بطريقة صالحة بقياس النفقات الصحية للأسر المؤهلة في القرى الريفية التي لم تشارك في البرنامج.

ولحسن الحظ، في وقت استطلاعي خط الأساس والمتابعة، جمعت الدراسة الاستقصائية بيانات عن ١٠٠ قرية ريفية إضافية لم يُقدم إليها البرنامج. واختيرت القرى الـ ١٠٠ هذه عشوائيًا أيضًا من المجتمع الإحصائي للقرى الريفية في البلد. لذلك، تضمن الطريقة التي اختيرت بها مجموعتا القرى أن لهما خصائص متطابقة إحصائيًا، باستثناء أن قرى المعالجة الـ ١٠٠ قد تلقت برنامج معونة التأمين الصحي (HISP) وأن قرى المعالجة الـ ١٠٠ لم تتلقها. لقد حدث تعيين عشوائي للمعالجة.

ونظرًا للتعيين العشوائي للمعالجة، فإنك تثق تمامًا من أنه لا توجد عوامل خارجية بخلاف برنامج معونة التأمين الصحي (HISP) من شأنها أن تفسر أي اختلافات في الحاصل بين قرى المعالجة والمقارنة. للتحقق من صحة هذا الافتراض، عليك اختبار ما إذا كان للأسر المؤهلة في قرى المعالجة والمقارنة خصائص متشابهة عند خط الأساس، كما هو موضح في الجدول ٤-١.

تلاحظ أن الخصائص المتوسطة للأسر في قرى المعالجة والمقارنة متشابهة للغاية في الواقع. الفروق ذات الدلالة الإحصائية هي فقط فروق عدد سنوات التعليم لرب الأسرة والمسافة إلى المستشفى، وهذه الفروق صغيرة (٠,١٦ عامًا فقط أو أقل من ٦ بالمائة من العدد المتوسط

الجدول ١-٤ تقييم برنامج معونة التأمين الصحي (HISP): التوازن بين قرى المعالجة والمقارنة عند خط الأساس

خصائص الأسر	قرى المعالجة (n = ٢٩٦٤)	قرى المقارنة (n = ٢٦٦٤)	الفرق	t-stat
النفقات الصحية (المال الخاص لكل فرد سنوياً بالدولار)	١٤,٤٩	١٤,٥٧	-٠,٠٨	-٠,٧٣
عمر رب الأسرة (بالأعوام)	٤١,٦٦	٤٢,٢٩	-٠,٦٤	-١,٦٩
عمر الزوج (بالأعوام)	٣٦,٨٤	٣٦,٨٨	-٠,٠٤	-٠,١٢
عمر رب الأسرة (بالأعوام)	٢٠,٩٧	٢٠,٨١	-٠,١٦	-٢,٣٠
تعليم الزوج (بالأعوام)	٢٠,٧٠	٢٠,٦٧	-٠,٠٣	-٠,٤٣
رب الأسرة أنثى = ١	٠,٠٧	٠,٠٨	-٠,٠١	-٠,٥٨
الأصلي = ١	٠,٤٣	٠,٤٢	-٠,٠١	-٠,٦٩
عدد أفراد الأسرة	٥,٧٧	٥,٧١	-٠,٠٦	-١,١٢
لديها أرضية متسخة = ١	٠,٧٢	٠,٧٣	-٠,٠١	-١,٠٩
لديها حمام = ١	٠,٥٧	٠,٥٦	-٠,٠١	-١,٠٤
عدد هكتارات الأرض	١,٦٨	١,٧٢	-٠,٠٤	-٠,٥٧
المستشفى (كم)	١٠٩,٢٠	١٠٦,٢٩	-٢٠,٩١	-٢,٥٧

ملاحظة: مستوى الدلالة: ** = ١ بالمائة.

الجدول ٢-٤ تقييم برنامج معونة التأمين الصحي (HISP): التعيين العشوائي من خلال مقارنة المتوسطات

قرى المعالجة	قرى المقارنة	الفرق	t-stat
النفقات الصحية للأسر عند خط الأساس (دولار أمريكي)	١٤,٤٩	١٤,٥٧	-٠,٠٨
النفقات الصحية للأسر عند المتابعة (دولار أمريكي)	٧,٨٤	١٧,٩٨	-١٠,١٤

ملاحظة: مستوى الدلالة: ** = ١ بالمائة

لسنوات التعليم، و ٢,٩١ كيلومتر، أو أقل من ٣ بالمائة مع المسافة المتوسطة بين مجموعة المقارنة والمستشفى). حتى في حالة إجراء تجربة عشوائية على عينة كبيرة، يمكن توقع عدد صغير من الفروق بسبب الصدفة و خصائص الاختبار الإحصائي. في الواقع، من خلال استخدام مستويات دلالة قياسية تساوي ٥ بالمائة، يمكننا أن نتوقع أن تكون فروق في ٥ بالمائة تقريباً من الخصائص ذات دلالة إحصائية، على الرغم من أننا لن نتوقع أن يكون مقدار هذه الفروق كبيراً.

بعد إنشاء صلاحية مجموعة المقارنة، يمكنك الآن تقدير الواقع المضاد على أنه الإنفاق الصحي المتوسط للأسر المؤهلة في قرى المقارنة الـ ١٠٠. يوضح الجدول ٤-٢ متوسط النفقات الصحية للأسر المؤهلة في قرى المعالجة والمقارنة. تلاحظ أنه عند خط الأساس، لا يختلف متوسط النفقات الصحية للأسرة في مجموعتي المعالجة والمقارنة إحصائيًا، كما لك أن تتوقع بموجب التعيين العشوائي. نظرًا إلى أن لديك الآن مجموعة مقارنة صالحة، يمكنك إيجاد أثر برنامج معونة التأمين الصحي (HISP) ببساطة بأخذ الفرق بين النفقات الصحية المتوسطة المدفوعة من المال الخاص للأسر في قرى المعالجة وقرى المقارنة المعيّنة عشوائيًا في فترة المتابعة.

الأثر هو انخفاض مقداره ١٠,١٤ دولارات خلال عامين. يعطي تكرار هذه النتيجة من خلال تحليل انحدار خطي النتيجة نفسها كما هو موضح في الجدول ٤-٣. أخيرًا، تجري تحليل انحدار متعدد المتغيرات يتحكم في بعض الخصائص الملحوظة لأسر العينة، وتجد أن البرنامج قد خفض من نفقات الأسر المسجلة بمقدار ١٠,٠١ دولارات خلال عامين، وهذا يطابق تقريبًا نتيجة الانحدار الخطي.

مع التعيين العشوائي، يمكننا أن نؤكد أنه لا توجد عوامل تختلف منهجيًا بين مجموعات المعالجة ومجموعات المقارنة التي قد تفسر أيضًا الفرق في النفقات الصحية. بدأت كلتا مجموعتي القرى بخصائص متشابهة للغاية، وتعرضت للسياسات الوطنية والبرامج نفسها خلال عامي المعالجة. ومن ثم فإن السبب الوحيد المعقول الذي يجعل الأسر الفقيرة في مجتمعات المعالجة تتكبد نفقات أقل من الأسر في قرى المقارنة هو أن المجموعة الأولى تلقت برنامج التأمين الصحي والمجموعة الأخرى لم تحصل عليه.

الجدول ٣-٤ تقييم برنامج معونة التأمين الصحي (HISP): التعيين العشوائي من خلال تحليل الانحدار

الأثر المقدّر على النفقات الصحية الأسرية	الانحدار الخطي	الانحدار الخطي متعدد المتغيرات
-١٠,١٤**	-١٠,١٤**	-١٠,٠١**
(٠,٣٩)	(٠,٣٩)	(٠,٣٤)

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

السؤال ٣ الخاص ببرنامح HISP



- أ. لماذا يُشتق تقدير الأثر باستخدام الانحدار الخطي متعدد المتغيرات في الأساس دون تغيير عند التحكم في العوامل الأخرى، مقارنةً بالانحدار الخطي البسيط ومقارنة المتوسطات؟
- ب. استنادًا إلى الأثر المقدّر بطريقة التعيين العشوائي، هل ينبغي توسيع برنامح HISP على المستوى الوطني؟

مصادر إضافية

- للحصول على المواد المرفقة مع هذا الفصل وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (www.worldbank.org/ieinpractice).
- للحصول على مصادر إضافية حول تقييمات أثر التعيين العشوائي، اطلع على تقييم بنك التنمية للبلدان الأمريكية (www.iadb.org/evaluationhub).
- للحصول على نظرة عامة كاملة عن تقييمات أثر التعيين العشوائي، راجع الكتاب التالي والموقع الإلكتروني المصاحب:
Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized _ Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press (<http://runningres.com/>).
- للحصول على مناقشة مفصلة حول تحقيق التوازن بين مجموعات المعالجة والمقارنة من خلال التعيين العشوائي، راجع ما يلي:
Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4): 200–32.
- للاطلاع على سلوكيات التعيين العشوائي للتقييم في الكاميرون، راجع مجموعة أدوات التقييم لدى البنك الدولي، الوحدة ٣ (www.worldbank.org/health/impactevaluationtoolkit).


الملاحظات

١. يمكن الإشارة إلى التعيين العشوائي للمعالجة بصورة شائعة أيضًا بكلٍ من *التجارب العشوائية المضبوطة*، و*التقييمات العشوائية*، و*التقييمات التجريبية*، و*التجارب الاجتماعية*، بالإضافة إلى مصطلحات أخرى. وبتعبير، يتعين على التجربة تحديد الآثار من خلال التعيين العشوائي، ولكن المقيمين يستخدمون عادةً المصطلح تجربة فقط عندما يستخدم التقييم التعيين العشوائي.
٢. لاحظ أن هذا الاحتمال لا يعني بالضرورة فرصة بنسبة ٥٠-٥٠ للفوز بالقرعة. عمليًا، تمنح جميع تقييمات التعيين العشوائي كل وحدة مؤهلة احتمال اختيار يُحدّد بحيث يكون عدد الفائزين (المعالجات) مساويًا للعدد الكلي المتاح من الفوائد. على سبيل المثال، إذا كان لدى برنامح تمويل كافٍ لخدمة ١٠٠٠ مجتمع فقط من بين ١٠٠٠٠ مجتمع مؤهل، فسيُمنح كل مجتمع فرصة واحدة من كل ١٠ لاختياره للمعالجة.

- سُعْظَم القوة الإحصائية (وهو مفهوم يُناقش بتفصيل أكبر في الفصل الخامس عشر) عند تقسيم عينة التقييم بالتساوي بين مجموعتي المعالجة والمقارنة. في المثال هنا، من أجل حجم عينة إجمالي مؤلف من ٢٠٠٠ مجتمع، سُعْظَم القوة الإحصائية من خلال أخذ عينات من جميع مجتمعات المعالجة البالغ عددها ١٠٠٠ مجتمع إضافة إلى عينة فرعية من ١٠٠٠ مجتمع مقارنة، بدلاً من أخذ عينة عشوائية بسيطة من ٢٠ بالمائة من المجتمعات المؤهلة الأصلية البالغ عددها ١٠٠٠٠ مجتمع. (والتي من شأنها أن تنتج عينة تقييم لحوالي ٢٠٠ مجتمع معالجة و ١٨٠٠ مجتمع مقارنة).
٣. على سبيل المثال، تستخدم برامج الإسكان التي توفر منازل مدعومة القرعة على نحو متكرر لاختيار المشاركين في البرامج. وتستخدم العديد من المدارس المستقلة في الولايات المتحدة القرعة لاختيار المتقدمين المقبولين.
٤. وبالإضافة إلى إنشاء مجموعات لها الخصائص المتوسطة نفسها، ينشئ التعيين العشوائي أيضًا مجموعات لها توزيعات متشابهة.
٥. تتيح معظم البرامج لك مجموعة رقم أولي لجعل نتائج التعيين العشوائي شفافة وقابلة للتكرار.
٦. وكما ذكرنا، لأسباب إحصائية، لا يلزم أن تكون جميع الخصائص الملحوظة متشابهة في مجموعتي المعالجة والمقارنة حتى يكون التوزيع العشوائي ناجحًا. وحتى عندما تكون خصائص المجموعتين متساوية حقًا، يمكن أن نتوقع أن تظهر نسبة ٥ بالمائة من الخصائص بفارق ذي دلالة إحصائية عند استخدام مستوى ثقة يبلغ ٩٥ بالمائة للاختبار. تحظى المتغيرات التي يكون عندها الفرق بين مجموعتي المعالجة والمقارنة كبيرًا باهتمام خاص.

المراجع

- Bertrand, Marianne, Bruno Crépon, Alicia Marguerie, and Patrick Premand. 2016. "Impacts à Court et Moyen Terme sur les Jeunes des Travaux à Haute Intensité de Main d'oeuvre (THIMO): Résultats de l'évaluation d'impact de la composante THIMO du Projet Emploi Jeunes et Développement des compétence (PEJEDEC) en Côte d'Ivoire." Washington, DC: Banque Mondiale et Abidjan, BCP-Emploi.
- Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics* 129 (2): 697–752. doi: 10.1093/qje/qjt057.
- Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4): 200–232.
- Dupas, Pascaline. 2011. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 3 (1): 1–34.
- Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.

- 
- Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126: 145–205.
- Kremer, Michael, and Edward Miguel. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.
- Premand, Patrick, Oumar Barry, and Marc Smitz. 2016. "Transferts monétaires, valeur ajoutée de mesures d'accompagnement comportemental, et développement de la petite enfance au Niger. Rapport descriptif de l'évaluation d'impact à court terme du Projet Filets Sociaux." Washington, DC: Banque Mondiale.
- Schultz, Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74 (1): 199–250.

المتغيرات المساعدة

تقييم البرامج عندما لا يمثل الجميع لتعييناتهم

افتراضنا خلال مناقشة التعيين العشوائي في الفصل الرابع أن المشرف على البرنامج كانت لديه سلطة تعيين وحدات المعالجة والمقارنة، وأن أولئك المعيّنين في المعالجة يشاركون في البرنامج وأولئك المعيّنين في مجموعة المقارنة لا يشاركون فيه. بمعنى آخر، تتوافق الوحدات المعينة في مجموعتي المعالجة والمقارنة مع تعيينها. يتحقق الامتثال الكامل بصورة أكثر تكرارًا في البيئات المخبرية أو التجارب الطبية، حيث يمكن للباحث أن يتحقق بعناية أكبر أولاً من أن جميع الخاضعين للدراسة في مجموعة المعالجة يتلقون المعالجة، وثانيًا من أن أيًا من الخاضعين للدراسة في مجموعة المقارنة لن يتلقوها.^١ بتعميم أكبر، افترضنا في الفصل الرابع أنه يمكن للبرامج تحديد ماهية المشاركين المحتملين، واستثناء بعضهم، والتحقق من أن الآخرين يشاركون.

ولكن، في البرامج الاجتماعية الواقعية، ربما لا يكون من الواقعي التفكير في أن المشرف على البرنامج سيتمكن من ضمان الامتثال الكامل مع تعيين المجموعة. مع ذلك، تسمح العديد من البرامج للمشاركين المحتملين باختيار المشاركة، ومن ثم لا يمكنها استبعاد المشاركين المحتملين الذين يريدون المشاركة. بالإضافة إلى ذلك، لبعض البرامج ميزانية كبيرة كفاية لتوفير البرنامج للمجتمع الإحصائي المؤهل بأكمله على الفور، بحيث تعين عشوائيًا الأشخاص في مجموعتي المعالجة والمقارنة و تستبعد

المشاركين المحتملين لإضفاء الصفة الأخلاقية على التقييم. لذلك، نحتاج إلى طريقة بديلة لتقييم أثر هذه الأشكال من البرامج.

المفهوم الأساسي

تعتمد طريقة المتغير المساعد على مصدر خارجي ما للاختلاف من أجل تحديد حالة المعالجة. يؤثر المتغير المساعد على احتمال المشاركة في برامج ما، ولكنه خارج عن سيطرة المشارك ولا علاقة له بخصائص المشارك.

يمكن أن تساعدنا طريقة تُسمى *المتغيرات المساعدة* (IV) على تقييم البرامج ذات الامتثال غير المثالي، أو المشاركة الطوعية، أو التغطية الشاملة. وعمومًا، لتقدير الأثر، تعتمد طريقة المتغير المساعد (IV) على مصدر خارجي ما للاختلاف أو التباين (variation) من أجل تحديد حالة المعالجة. ولهذه الطريقة تطبيقات واسعة النطاق تتجاوز تقييم الأثر. يمكننا بديهيًا التفكير في المتغير المساعد باعتباره شيئًا خارجًا عن سيطرة الفرد ويؤثر على احتمال مشاركته في برنامج ما، ولكنه لا يرتبط بخصائصه.

نناقش في هذا الفصل كيف يمكن إيجاد هذا الاختلاف الخارجي، أو المتغير المساعد (IV)، من خلال قواعد تشغيل البرنامج الذي يخضع لسيطرة منفذي البرنامج أو فرق التقييم. لإنتاج تقديرات أثر صحيحة، يجب أن يحقق هذا المصدر الخارجي للاختلاف عددًا من الشروط، التي سنناقشها بالتفصيل في هذا الفصل. تبين أن التعيين العشوائي للمعالجة، كما يناقش الفصل الرابع، أداة مساعدة جيدة جدًا تحقق الشروط اللازمة. وسنستخدم طريقة المتغير المساعد في تطبيقين شائعين لتقييم الأثر. أولاً، سنستخدمها بمتابعة امتداد لطريقة التعيين العشوائي عندما لا تمثل جميع الوحدات لتعيينات المجموعات الخاصة بها. ثانيًا، سنستخدمها لتصميم ترويج عشوائي للمعالجة، وهو طريقة تقييم يمكن أن تفي بالغرض بالنسبة لبعض البرامج التي توفر اشتراكًا طوعيًا أو تغطية شاملة. يوضح المربع ٥-١ استخدامًا مبتكرًا لطريقة المتغير المساعد.

أنواع تقييمات الأثر

يقدّر تقييم الأثر دائمًا أثر البرنامج من خلال مقارنة نتائج مجموعة معالجة بتقدير الواقع المضاد الذي يُحصل عليه من مجموعة المقارنة. افترضنا في الفصل الرابع الامتثال الكامل للمعالجة؛ أي أن جميع الوحدات التي يُقدّم إليها البرنامج تشارك فيه فعليًا، ولا تتلقى أي وحدات مقارنة البرنامج. وفي هذا السيناريو، نقدّر متوسط تأثير المعالجة (ATE) للمجتمع الإحصائي.

عند تقييم برامج واقعية يمكن للمشاركين فيها أن يقرروا ما إذا كانوا يريدون المشاركة أم لا، يكون الامتثال الكامل أقل شيوعًا من بيانات من قبيل التجارب المخبرية. وتوفر البرامج عمليًا فرصة المعالجة لمجموعة محددة، أو أن بعض الوحدات تشارك ولا تشارك وحدات أخرى. وفي هذه الحالة، وفي غياب الامتثال الكامل، يمكن أن تقدّر تقييمات الأثر أثر تقديم برنامج أو أثر المشاركة في البرنامج.

المربع 0-1: استخدام المتغيرات المساعدة لتقدير أثر برنامج عالم سمس على الجاهزية للمدرسة

وهكذا، استخدم ليفين وكيرني (٢٠١٥) المسافة بين المنازل إلى أقرب برج تلفزيوني يرسل التردد الفائق بمثابة أداة للمشاركة في البرنامج. يقول الباحثان إنه بما أن الأبراج التلفزيونية شُيّدت في مواقع اختارتها الحكومة— قبل بث برنامج عالم سمس في الأصل— فلن يكون المتغير مرتبطًا بخصائص الأسر أو التغيرات في الدخل. توصل التقييم إلى نتائج إيجابية بشأن جاهزية الأطفال في سن ما قبل المدرسة للالتحاق بالمدارس. في المناطق التي كان يوجد فيها استقبال تلفزيوني للتردد الفائق عندما بدأ البرنامج، كان الأطفال أكثر عرضة للتقدم في المدرسة الابتدائية في السن المناسب. وكان هذا التأثير ملحوظًا للأطفال والأولاد من أصل أفريقي وغير لاتيني والأطفال في المناطق المحرومة اقتصاديًا.

سرعان ما اكتسب البرنامج التلفزيوني *عالم سمس* شعبية وإشادة من قبل النقاد بُعيد بثه أول مرة عام ١٩٦٩، وهو برنامج الهدف منه تحضير الأطفال في سن ما قبل المدرسة للمدرسة الابتدائية. ومنذ ذلك الحين، يشاهده ملايين الأطفال. وفي عام ٢٠١٥، سعى ليفين وكيرني إلى تقييم الآثار طويلة المدى للبرنامج في تقييم بأثر رجعي أجري في الولايات المتحدة. استخدم الباحثون منهج متغيرات مساعدة بالاستفادة من المحددات في تكنولوجيا البث التلفزيوني في السنوات الأولى للبرنامج. لم يكن بإمكان جميع الأسر متابعة البرنامج في سنواته القليلة الأولى. فقد كان يُبث فقط على قنوات التردد الفائق (UHF). وكان ثلثا سكان الولايات المتحدة الأمريكية فقط يعيشون في مناطق يمكن مشاهدة البرنامج فيها.

المصدر: Kearney and Levine 2015.

المفهوم الرئيسي

يُقدّر النية للعلاج (ITT) الفرق في الحاصل بين الوحدات المعينة في مجموعة المعالجة والوحدات المعينة في مجموعة المقارنة، بغض النظر عما إذا كانت الوحدات المعينة في مجموعة المعالجة تتلقى المعالجة بالفعل.

المفهوم الرئيسي

يُقدّر تأثير المعالجة على المعالجين (TOT) الفرق في الحاصل بين الوحدات التي تتلقى المعالجة بالفعل ومجموعة المقارنة.

في غياب الامتثال الكامل في مجموعة المعالجة، يُطلق على الأثر المقدّر اسم *النية للعلاج* (*intention-to-treat* أو ITT عند اختصارًا) عند مقارنة المجموعات التي قُدِّم إليها البرنامج عشوائيًا (في مجموعة المعالجة) أو لم يُقدِّم (في مجموعة المقارنة)—بغض النظر عما إذا كان هؤلاء في مجموعة المعالجة مسجلين بالفعل في البرنامج أم لا. ويعد تقدير النية للعلاج (ITT) هو المتوسط المرجح لحاصل المشاركين وغير المشاركين في مجموعة المعالجة بالمقارنة مع متوسط حصيلة مجموعة المقارنة. وتقدير النية للعلاج مهم للحالات التي نحاول فيها تحديد متوسط أثر تقديم برنامج ما، وتكون فيها المشاركة في البرنامج طوعية. وعلى النقيض من ذلك، قد نهتم أيضًا في معرفة أثر برنامج ما بالنسبة إلى مجموعة من الأفراد الذين يُقدِّم إليهم البرنامج ويشاركون فيه. يُسمى هذا الأثر المقدّر تأثير المعالجة على المعالجين (*treatment-on-the-treated* أو TOT عند اختصارًا). يكون تقدير النية للعلاج (ITT) وتأثير المعالجة على المعالجين (TOT) متماثلين عند وجود امتثال كامل. سنعود إلى الفرق بين تقدير النية للعلاج (ITT) وتأثير المعالجة على المعالجين (TOT) في الأقسام المستقبلية، ولكننا سنبدأ بمثال لتوضيح هذين المفهومين.

فكر في برنامج معونة التأمين الصحي (HISP)، الذي ناقشناه في فصول سابقة. نظرًا للاعتبارات التشغيلية، ولتقليل التأثيرات غير المباشرة إلى الحد الأدنى، فإن وحدة تعيين المعالجة التي اختارتها الحكومة هي القرية. يمكن للأسر في قرية المعالجة (القرى التي يُقدّم فيها برنامج التأمين الصحي) الاشتراك في خدمة التأمين الصحي طوعًا، في حين لا تستطيع الأسر في مجتمعات المقارنة ذلك. على الرغم من أن جميع الأسر في قرى المعالجة مؤهلة للاشتراك في برنامج التأمين الصحي، قد يقرر جزء من الأسر—لنقل ١٠ بالمائة—عدم الاشتراك (ربما لأنه لديها في الأصل تأمين من خلال وظائفها، أو لأنها بحالة صحية جيدة ولا تتوقع أنها بحاجة إلى الرعاية الصحية، أو لأي أسباب أخرى لا تعد ولا تحصى).

في هذا السيناريو، قررت نسبة ٩٠ بالمائة من الأسر في قرية المعالجة أن تشارك في البرنامج وتتلقى فعليًا الخدمات التي يوفرها البرنامج. نحصل على تقدير النية للعلاج بمقارنة متوسط حصيلة جميع الأسر التي قُدّم إليها البرنامج—أي لنسبة ١٠٠ بالمائة من الأسر في قرى المعالجة—مع متوسط الحصيلة في قرى المقارنة (حيث اشتركت القرى). على النقيض من ذلك، يمكن التفكير في تأثير المعالجة على المعالجين (TOT) على أنه الأثر المقدّر لـ ٩٠ بالمائة من الأسر في قرى المعالجة التي اشتركت في البرنامج. ومن المهم أن ننوه إلى أنه بما أن الأفراد الذين يشتركون في البرنامج عند تقديمه قد يختلفون عن الأفراد الذين يُقدّم إليهم البرنامج ولكنهم ينسحبون منه، فإن أثر تأثير المعالجة على المعالجين (TOT) لا يماثل بالضرورة الأثر الذي كنا سنحصل عليه لنسبة الـ ١٠ بالمائة من الأسر في قرى المعالجة التي لم تشارك لو أنها اشتركت. وعليه، لا يمكن استكمال تأثيرات المعالجة المكانية بالاستقراء مباشرة من مجموعة إلى أخرى.

الامتثال غير المثالي

كما ناقشنا، يُحبذ في البرامج الاجتماعية الواقعية أن يتوفر امتثال كامل مع معايير الاختيار في البرنامج (ومن ثم الالتزام بحالة المعالجة أو المقارنة)، وعادةً ما يسعى صانعو السياسات وفرق التقييم على حد سواء جاهدين إلى الاقتراب قدر الإمكان من ذلك الوضع المثالي. ولكن، في الواقع العملي، قد لا يحدث امتثال صارم بنسبة ١٠٠ في المائة لتعيينات المعالجة والمقارنة، على الرغم من أفضل الجهود التي يبذلها منفذ البرنامج وفريق التقييم. سندرس الآن الحالات المختلفة التي قد تحدث، وسنناقش الآثار المترتبة عن طرق التقييم التي يمكن استخدامها. نؤكد مقدمًا على أن الحل الأفضل للامتثال غير المثالي هو تجنبه في المقام الأول. وبهذا المعنى، يتعين على مديري البرامج وصانعي السياسات بذل قصارى جهدهم للحفاظ على الامتثال على أعلى مستوى ممكن في مجموعة المعالجة وأقل مستوى ممكن في مجموعة المقارنة.

هـب أنك تحاول تقييم برنامج تدريب المعلمين، حيث يوجد ٢٠٠٠ معلم مؤهل للمشاركة في تدريب تجريبي. عُيّن المعلمون عشوائيًا في إحدى مجموعتي المعالجة. يُعيّن ١٠٠٠ معلم عشوائيًا في مجموعة المعالجة، ويُعيّن ١٠٠٠ عشوائيًا في مجموعة المقارنة.

عندما يتلقى جميع المعلمين في مجموعة المعالجة التدريب، ولا يتلقى أي معلم التدريب في مجموعة المقارنة، تقدير متوسط تأثير المعالجة (ATE) من خلال أخذ الفرق في متوسط النتائج (لنقل درجات اختبار الطالب) بين المجموعتين. متوسط تأثير المعالجة (ATE) هذا هو متوسط أثر المعالجة على ١٠٠٠ معلم، علمًا أن جميع المعلمين المعيّنين في مجموعة المعالجة يحضرون الدورة فعليًا، بينما لا يحضر أي من المعلمين المعيّنين في مجموعة المقارنة.

تحدث حالة الامتثال غير المثالي عندما تختار بعض الوحدات المعينة في مجموعة المعالجة عدم الاشتراك أو تُترك دون معالجة. في مثال تدريب المعلمين، لا يحضر بعض المعلمين المعيّنين في مجموعة المعالجة فعليًا في اليوم الأول للدورة. وفي هذه الحالة، لا يمكننا حساب متوسط المعالجة لمجموعة المعلمين لأن بعض المعلمين لا يشتركون أبدًا؛ لذلك لا يمكننا أبدًا حساب حصائل المعالجة. ولكن يمكننا تقدير أثر البرنامج على المعلمين الذين يخضعون للمعالجة أو يوافقون بها. نريد تقدير أثر البرنامج على المعلمين الذين عُيّنوا في مجموعة المعالجة والمشاركين فعليًا. وهذا ما يُشار إليه باسم *تقدير تأثير المعالجة على المعلمين*. في مثال تدريب المعلمين، يوفر تقدير تأثير المعالجة على المعلمين الأثر للمعلمين المعيّنين في مجموعة المعالجة والذين يظهرون ويتلقون التدريب فعليًا.

الحالة الثانية للامتثال غير المثالي هي عندما يتمكن الأفراد المعيّنون في مجموعة المقارنة من المشاركة في البرنامج. هنا، لا يمكن تقدير الأثر مباشرة لمجموعة المعالجة بأكملها لأن بعض نظرائهم في مجموعة المقارنة لا يمكن ملاحظتها بدون معالجة. تم افتراض أن الوحدات المعالجة في مجموعة المقارنة تنتج تقديرًا للواقع المضاد لبعض الوحدات في مجموعة المعالجة، ولكنها تتلقى المعالجة، لذلك لا توجد طريقة لمعرفة أثر البرنامج على هذه المجموعة الفرعية من الأفراد. في مثال تدريب المعلمين، هب أن معظم المعلمين المتحفزين في مجموعة المقارنة يحضرون الدورة بطريقة ما. وفي هذه الحالة، لن يكون للمعلمين الأكثر تحفيزًا في مجموعة المعالجة نظراء في مجموعة المقارنة، ومن ثم لن يكون من الممكن تقدير أثر التدريب لتلك الشريحة من المعلمين المتحمسين. عند وجود عدم امتثال من أي جانب، يجب أن تفكر بعناية في نوع تأثير المعالجة الذي تقدره وكيفية تفسيره. ثمة خيار أول يقوم على حساب مقارنة مباشرة للمجموعة المعينة أصلًا للمعالجة مع المجموعة المعينة أصلًا للمقارنة؛ وهذا سينتج تقدير النية للعلاج. يقارن النية للعلاج الأشخاص الذين ننوي معالجتهم (الأشخاص المعيّنين في مجموعة المعالجة) مع الأشخاص الذين لا ننوي معالجتهم (الأشخاص المعيّنين في مجموعة المقارنة). إذا كان عدم الامتثال من جانب المعالجة فقط، فقد يكون هذا مقياسًا مثيرًا للاهتمام وملائمًا للأثر لأنه في أي حال، يمكن لمعظم صانعي السياسات ومديري البرنامج توفير برنامج فقط ولا يمكنهم فرض البرنامج على فئتهم السكانية المستهدفة.

في مثال تدريب المعلمين، قد تحتاج الحكومة إلى معرفة متوسط أثر البرنامج على جميع المعلمين المعيّنين، حتى لو لم يلتحق بعض المعلمين بالدورة. يُعزى ذلك إلى أنه حتى إذا وسعت الحكومة البرنامج، فمن المحتمل أن يكون هناك مدرسون لن يلتحقوا أبدًا. ولكن، إذا كان هناك عدم امتثال من جانب المقارنة، فإن تقدير النية للعلاج ليس سديدًا. في حالة تدريب المعلمين، بما أن مجموعة المقارنة للمعلمين تضم معلمين مدربين، فإن متوسط الحصيلة في مجموعة المقارنة قد تأثر بالمعالجة. لنفترض أن أثر تدريب الطلاب على الحصائل موجب. إذا كان غير الممثلين في مجموعة المقارنة هم المعلمون الأكثر تحفيزًا وكانوا يستفيدون أكثر من التدريب، فإن متوسط الحصيلة لمجموعة المقارنة سيكون متحيزًا إلى الأعلى (لأن المعلمين المتحمسين في مجموعة المقارنة والذين خضعوا للتدريب سيزيدون متوسط الحصيلة) وسيكون تقدير النية للعلاج متحيزًا للأسفل (نظرًا لأنه يمثل الفرق بين متوسط الحصائل في مجموعتي المعالجة والمقارنة).

في ظل ظروف عدم الامتثال هذه، ثمة خيار ثانٍ يقوم على تقدير ما يعرف باسم متوسط تأثير المعالجة المكاني (*local average treatment effect*) أو ما نشير إليه اختصارًا بـ (LATE). يحتاج تأثير المعالجة المكاني (LATE) إلى التفسير بتأني، فهو يمثل تأثيرات البرنامج فقط لمجموعة فرعية محددة من المجتمع الإحصائي. وبالتحديد، عند وجود عدم امتثال في مجموعة المعالجة ومجموعة المقارنة يمثل تأثير المعالجة المكاني الأثر على المجموعة الفرعية من الممثلين. في مثال تدريب المعلمين، إذا كان ثمة عدم امتثال في كل من مجموعة المعالجة والمقارنة، فإن تقدير تأثير المعالجة المكاني (LATE) يكون صالحًا فقط للمعلمين في مجموعة المعالجة الذين اشتركوا في البرنامج والذين ما كانوا ليشاركوا لو تم تعيينهم للمقارنة مجموعة.

سنشرح في بقية هذا الفصل كيفية تقدير تأثير المعالجة المكاني (LATE)، وكذلك كيفية تفسير النتائج على حدٍّ سواء. تنطبق مبادئ تقدير تأثير المعالجة المكاني (LATE) عند وجود عدم امتثال في مجموعة المعالجة أو مجموعة المقارنة أو كليهما في وقت واحد. تأثير المعالجة على المعلمين هو ببساطة تأثير المعالجة المكاني في الحالة الأكثر تحديدًا عند وجود عدم امتثال فقط في مجموعة المعالجة. لذلك، يركز ما تبقى من هذا الفصل على كيفية تقدير تأثير المعالجة المكاني (LATE).

التعيين العشوائي للبرامج والالتحاق النهائي

تخيّل أنك تقيّم أثر برنامج تدريب وظيفي على أجور الأفراد. يُعيّن البرنامج عشوائيًا على مستوى الأفراد. تُعيّن مجموعة المعالجة في البرنامج، بينما لا تُعيّن مجموعة المقارنة. ومن المرجح أن تعثر على ثلاثة أنواع من الأفراد في المجتمع الإحصائي:

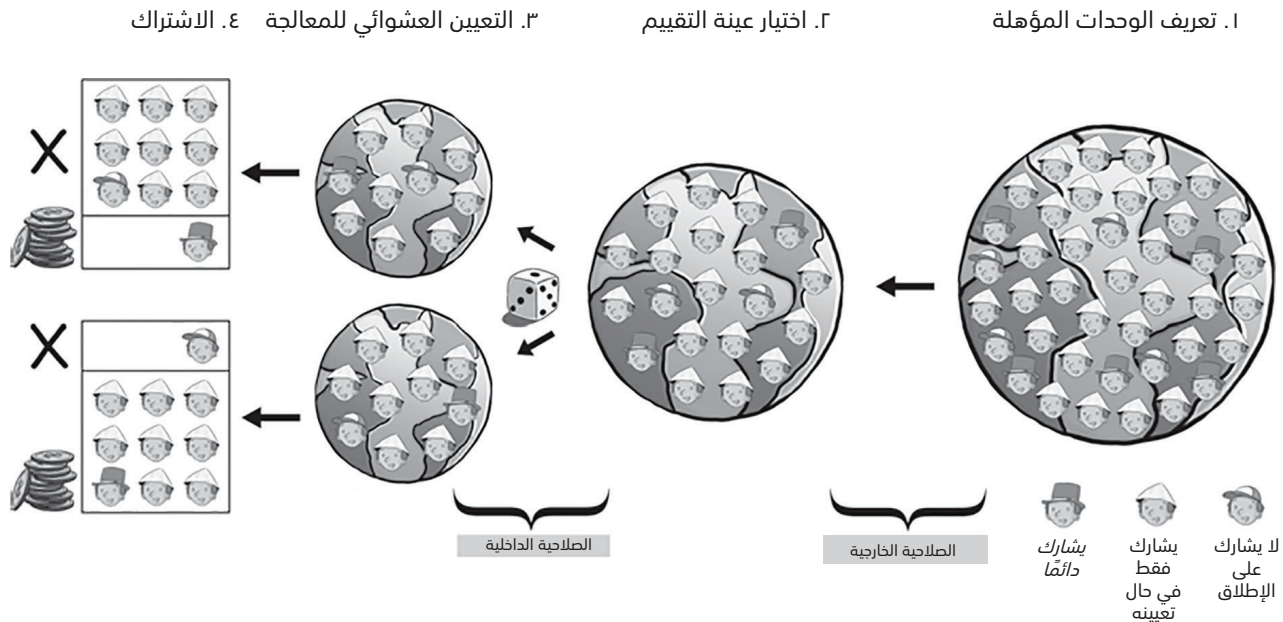
- **يشارك في حال تعيينه.** يُقصد بذلك الأفراد الذين يمثلون لتعيينهم. فإذا عُيّنوا في مجموعة المعالجة (عُيّنوا في البرنامج)، فإنهم يلتحقون به أو يشتركون فيه. وإذا عُيّنوا في مجموعة المقارنة (لم يُعيّنوا في البرنامج)، فإنهم لا يشاركون فيه.

- **لا يشارك على الإطلاق.** يُقصد بذلك الأفراد الذين لا يشاركون في البرنامج أو لا يلتحقون به على الإطلاق، حتى ولو عُيِّنوا في مجموعة المعالجة. إذا عُيِّن هؤلاء الأفراد في مجموعة المعالجة، فسيكونون غير ممثلين.
- **يشارك دائماً.** يُقصد بذلك الأفراد الذين سيجدون طريقة للاشتراك في البرنامج أو الالتحاق به، حتى لو عُيِّنوا في مجموعة المقارنة. إذا عُيِّن هؤلاء الأفراد في مجموعة المقارنة، فسيكونون غير ممثلين.

في سياق برنامج التدريب الوظيفي، قد تتضمن مجموعة **لا يشارك على الإطلاق** أشخاصاً غير متحفّزين لا يحضرون حتى لو عُيِّنوا في مكان في الدورة. في المقابل، فإن الأفراد في المجموعة **يشارك دائماً**، متحفزون لدرجة أنهم يعثرون على طريقة لدخول البرنامج حتى ولو عُيِّنوا في الأصل في مجموعة المقارنة. تضم مجموعة **يشارك في حال تعيينه** الأفراد الذين يلتحقون في الدورة إذا عُيِّنوا فيها، ولكنهم لا يسعون إلى الالتحاق إذا عُيِّنوا في مجموعة المقارنة.

يمثل الشكل ١-٥ التعيين العشوائي في البرنامج والاشتراك أو الالتحاق النهائي عند وجود المجموعات **يشارك في حال تعيينه**، **ولا يشارك على الإطلاق ويشارك دائماً**. هب أن المجتمع الإحصائي المستهدف يضم ٨٠ بالمائة من الأفراد ضمن النوع **يشارك في حال تعيينه**، و ١٠ بالمائة ضمن النوع **لا يشارك على الإطلاق**، و ١٠ بالمائة ضمن النوع **يشارك دائماً**. إذا أخذنا عينة عشوائية من المجتمع الإحصائي المخصص لعينة التقييم، فإن عينة التقييم ستضم أيضاً ٨٠ بالمائة من الأفراد ضمن النوع **يشارك في حال تعيينه**، و ١٠ بالمائة ضمن النوع **لا يشارك على الإطلاق**، و ١٠ بالمائة من النوع **يشارك دائماً**. ومن ثم إذا عيّنّا عشوائياً عينة التقييم في مجموعة معالجة ومجموعة مقارنة،

الشكل ١-٥ التعيين العشوائي مع الامتثال غير المثالي



فسيكون لدينا تقريبًا ٨٠ بالمائة من الأفراد ضمن النوع *يشارك في حال تعيينه*، و ١٠ بالمائة من الأفراد ضمن النوع *لا يشارك على الإطلاق*، و ١٠ بالمائة من الأفراد ضمن النوع *يشارك دائمًا* في كلتا المجموعتين. في المجموعة المعينة للمعالجة، سيشارك الأفراد ضمن النوعين *يشارك في حال تعيينه* و *يشارك دائمًا* وستبقى مجموعة الأفراد ضمن النوع *لا يشارك على الإطلاق* بعيدة. في مجموعة المقارنة، سيشارك الأفراد ضمن النوع *يشارك دائمًا*، بينما سيبقى الأفراد ضمن النوعين *يشارك في حال تعيينه* و *لا يشارك على الإطلاق* خارجًا. من المهم أن نتذكر أنه رغم علمنا بوجود هذه الأنواع الثلاثة من الأفراد موجودة في المجتمع الإحصائي، لا يمكننا بالضرورة تمييز نوع الفرد حتى نلاحظ سلوكيات معينة. في مجموعة المعالجة، ستمكن من التعرف على النوع *لا يشارك على الإطلاق* عندما لا يشارك، ولكننا لن نتمكن من تمييز النوع *يشارك في حال تعيينه* عن النوع *يشارك دائمًا* بما أن الأشخاص من كلا النوعين سيشاركون. في مجموعة المقارنة، ستمكن من التعرف على النوع *يشارك دائمًا* عندما يشترك، ولكننا لن نتمكن من تمييز النوع *يشارك في حال تعيينه* عن النوع *لا يشارك على الإطلاق* بما أن الأشخاص من كلا النوعين لن يشاركون.



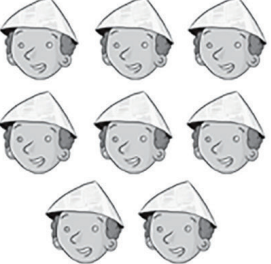
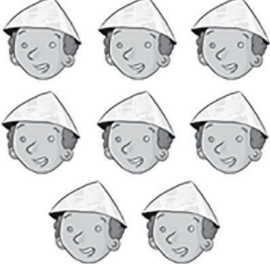
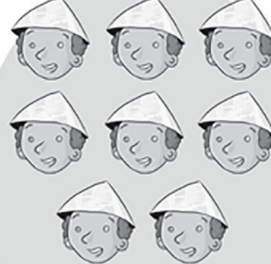


تقدير الأثر بموجب التعيين العشوائي مع الامتثال غير المثالي

بعد تحديد الفرق بين تعيين برنامج والاشتراك أو الالتحاق الفعلي، ننتقل إلى تقدير تأثير المعالجة المكاني (LATE) للبرنامج. يُجرى هذا التقييم في خطوتين اثنتين، وهما موضحتان في الشكل ٥-٢٠

لتقدير آثار برنامج خاضع للتعيين العشوائي وله امتثال غير مثالي، نقدّر أولاً أثر تقدير النية للعلاج. تذكر أن قيمته هي مجرد الاختلاف المباشر في مؤشر الحصة (Y) للمجموعة التي عُيِّنَت للمعالجة وأنه المؤشر نفسه للمجموعة التي لم نعيّنها للمعالجة. على سبيل المثال، إذا كان متوسط الأجر (Y) لمجموعة المعالجة الأولى هو ١١٠ دولارات أمريكية، وكان متوسط الأجر لمجموعة المقارنة ٧٠ دولارًا أمريكيًا، فإن تقدير النية للعلاج للأثر سيساوي ٤٠ دولارًا أمريكيًا (١١٠ - ٧٠ دولارًا ناقص ٧٠ دولارًا).

ثانيًا، نحتاج إلى استعادة تقدير تأثير المعالجة المكاني (LATE) لمجموعة النوع *يشارك في حال تعيينه* من تقدير النية للعلاج. ولإجراء ذلك، نحتاج إلى تحديد المصدر الذي أتى منه الفرق البالغ ٤٠ دولارًا. لنمض في ذلك عبر الإزالة. أولاً، نعلم أن الفرق لا يمكن أن يكون نتيجة للفرق بين الأشخاص الذين لا يشتركون (النوع *لا يشارك على الإطلاق*) في مجموعتي المعالجة والمقارنة. وذلك لأن الأشخاص ضمن النوع *لا يشارك على الإطلاق* لا يشاركون على الإطلاق في البرنامج، ولذا، بالنسبة إليهم، لا فرق فيما إذا كانوا في مجموعة المعالجة أو مجموعة المقارنة. ثانيًا، نعرف أن الفرق البالغ ٤٠ دولارًا أمريكيًا لا يمكن أن ينتج عن الاختلافات بين الأشخاص ضمن النوع *يشارك دائمًا* في مجموعتي المعالجة والمقارنة لأن الأشخاص ضمن النوع *يشارك دائمًا* يشترك دائمًا في البرنامج. وبالنسبة إليهم أيضًا، لا فرق بينما إذا كانوا في مجموعة المعالجة أو مجموعة المقارنة.

الشكل ٢-٥ تقدير متوسط تأثير المعالجة المكاني في ظل التعيين العشوائي مع الامتثال غير المثالي

آثار	المجموعة غير المعيّنة للمعالجة	المجموعة المعيّنة للمعالجة	
Δ النسبة المئوية للمشاركين = ٨٠٪ $\Delta Y = ITT = 40$ $LATE = 40/80\% = 50$	النسبة المئوية للمشاركين = ١٠٪ المتوسط Y لغير المعيّنين في المعالجة $v_0 =$	النسبة المئوية للمشاركين = ٩٠٪ المتوسط Y للمعّنين في المعالجة = ١١٠	
—			لا يشارك على الإطلاق
			يشارك فقط إذا عُيّن للمعالجة
—			يشارك دائمًا

ملاحظة: Δ = الأثر السببي، Y = الحصيلة. نحصل على تقدير النية للعلاج (ITT) بمقارنة حصائل المعيّنين في مجموعة المعالجة مع المعيّنين في مجموعة المقارنة، بغض النظر عن الاشتراك الفعلي. يوفر تقدير متوسط تأثير المعالجة المكاني (LATE) أثر البرنامج على الذين يشتركون في البرنامج فقط في حال تعيينهم في البرنامج (يشارك في حال تعيينه). لا يوفر تقدير متوسط تأثير المعالجة المكاني (LATE) أثر البرنامج على الذين لا يشتركون على الإطلاق (الأفراد ضمن النوع لا يشارك على الإطلاق) أو الذين يشتركون دائمًا (الأفراد ضمن النوع يشارك دائمًا).

ومن ثم فإن الفرق في الحصائل بين المجموعتين لا بد أن يأتي بالضرورة من أثر البرنامج فقط على المجموعة المتأثرة بتعيينها في المعالجة أو المقارنة: أي المجموعة يشارك في حال تعيينه. لذلك، إذا كان بإمكاننا تحديد الأفراد ضمن النوع يشارك في حال تعيينه في كلتا المجموعتين، فسيكون من السهل تقدير أثر البرنامج عليهم. وفي الواقع، على الرغم من معرفتنا بأن هذه الأنواع الثلاثة من الأفراد موجودة في المجتمع الإحصائي، لا يمكننا فصل الأفراد الفريدين استنادًا إلى ما إذا كانوا ضمن المجموعة يشارك في حال تعيينه، أو لا يشارك على الإطلاق، أو يشارك دائمًا. في المجموعة التي عُيّنَت للمعالجة، يمكننا تحديد الأفراد ضمن النوع لا يشارك على الإطلاق (لأنهم لم يشاركوا)، ولكن لا يمكننا التمييز بين المجموعة يشارك دائمًا والمجموعة يشارك في حال تعيينه (لأن الأشخاص من كلتا المجموعتين مشاركون). في المجموعة التي لم تُعَيّن للمعالجة، يمكننا تحديد المجموعة يشارك دائمًا (لأن المجموعة تشارك في البرنامج)، ولكن لا يمكننا التمييز بين المجموعة لا يشارك على الإطلاق ويشارك فقط في حال تعيينه.

ولكن، بمجرد أن نلاحظ أن ٩٠ في المائة من الوحدات في المجموعة التي تم تخصيص العلاج لها ، يمكننا أن نستنتج أن ١٠ في المائة من الوحدات في مجتمعنا الإحصائي يجب أن تكون ضمن النوع *لا يشارك على الإطلاق* (أي الجزء من الأفراد في المجموعة المعيّنة للمعالجة والذين لم يشتركوا). إضافةً إلى ذلك، إذا لاحظنا أن ١٠ في المائة من الوحدات في المجموعة غير المعيّنة في المعالجة يشتركون، فإننا نعرف أن ١٠ بالمائة من الأفراد ضمن النوع *يشارك دائماً* (مرة أخرى، نسبة الأفراد في مجموعتنا التي لم تُعيّن للمعالجة والذين اشتركوا). وهذا يترك ٨٠ بالمائة من هذه الوحدة في المجموعة *يشارك في حال تعيينه*. نعلم أن الأثر الكامل البالغ ٤٠ دولارًا أمريكيًا أتى من الفرق في الاشتراك لـ ٨٠ بالمائة من الوحدات في عينتنا الذين ينتمون إلى النوع *يشارك في حال تعيينه*. والآن، إذا كانت نسبة ٨٠ في المائة من الوحدات مسؤولة عن متوسط أثر قدره ٤٠ دولارًا أمريكيًا للمجموعة المعينة بأكملها للمعالجة، فإن الأثر على نسبة الـ ٨٠ في المائة هذه من المجموعة *يشارك في حال تعيينه* يجب أن يساوي ٨٠/٤٠، أو ٥٠ دولارًا أمريكيًا. بتعبير آخر، فإن أثر البرنامج للنوع *يشارك في حال تعيينه* يساوي ٥٠ دولارًا أمريكيًا، ولكن عند نشر الأثر عبر المجموعة المعيّنة للمعالجة بأكملها، فإن متوسط الأثر يخفّ بنسبة الـ ٢٠ بالمائة التي لم تتكون ممثلة للتعين العشوائي الأصلي.

تذكر أن إحدى المشكلات الأساسية المتصلة بالاختيار الذاتي في البرامج هي أنه لا يمكنك دائمًا معرفة سبب اختيار بعض الأشخاص للمشاركة وعدم اختيار البعض الآخر لها. عندما تجري تقييمًا تُعيّن فيه الوحدات عشوائيًا للبرنامج، ولكن المشاركة الفعلية تطوعية أو هناك طريقة كي تدخل الوحدات في مجموعة المقارنة في البرنامج، فعندها تكون لدينا إشكالية مماثلة: لن نفهم دائمًا العمليات السلوكية التي تحدد ما إذا كان الفرد يتصرف على أنه *لا يشارك على الإطلاق*، أو *يشارك دائماً*، أو *يشارك في حال تعيينه*. ولكن، شريطة ألا يكون عدم الامتثال كبيرًا جدًا، فإن التعيين العشوائي لا يزال يوفر أداة قوية لتقدير الأثر. ولكن الجانب السلبي للتعين العشوائي في حالة الامتثال غير المثالي هو أن تقدير التأثير هذا لم يعد صالحًا للمجتمع الإحصائي بأكمله. بدلًا من ذلك، ينبغي تفسير التقدير على أنه تقدير مكاني ينطبق فقط على مجموعة فرعية محددة ضمن المجتمع الإحصائي المستهدف، النوع *يشارك في حال تعيينه*. للتعين العشوائي لأي برنامج خاصيتان مهمتان اثنتان تسمحان لنا بتقدير متى يكون هناك أثر غير مثالي (انظر المربع ٢-٥):

١. ويمكن أن يكون بمثابة متنبئ بالاشتراك الفعلي في البرنامج إذا كان معظم الأشخاص يتصرفون على أنهم من النوع *يشارك في حال تعيينه*، بحيث يشاركون في المعالجة عند تعيينهم ولا يشاركون عند عدم تعيينهم في المعالجة.
٢. بما أن المجموعتين (المعالجة المعينة وغير المعينة) أنشئت من خلال عملية عشوائية، فإن خصائص الأفراد في المجموعتين لا ترتبط بأي شيء آخر - مثل القدرة أو الدافع - وهذا قد يؤثر أيضًا على النتائج (٧).

المربع ٥-٢: استخدام المتغيرات المساعدة للتعامل مع عدم الامتثال في برنامج قسيمة المدرسة في كولومبيا

كان هناك بعض حالات عدم الامتثال مع التعيين العشوائي. كانت نسبة ٩٠ بالمائة تقريباً من الفائزين في القرعة يستخدمون قسيمة أو نوعاً آخر من المنح الدراسية، وكانت نسبة ٢٤ بالمائة من الخاسرين في القرعة تتلقى منحة دراسية. باستخدام مصطلحاتنا السابقة، لا بد أن يكون المجتمع الإحصائي قد ضم ١٠ بالمائة من الأفراد من النوع *لا يشارك على الإطلاق*، و٢٤ بالمائة من الأفراد من النوع *يشارك دائماً*، و٦٦ بالمائة من الأفراد من النوع *يشارك في حال تعيينه*. ولذلك استخدم Angrist وآخرون (٢٠٠٢) أيضاً التعيين، أو حالة فوز طالب أو خسارته في القرعة، بمثابة متغير مساعد فعال عند المعالجة، أو تلقي منحة دراسية فعلياً. أخيراً، تمكّن الباحثون من حساب تحليل التكلفة والمنفعة لفهم أثر برنامج القسائم على كل من نفقات الأسرة والحكومة. واستنتجوا أن التكاليف الاجتماعية الإجمالية للبرنامج صغيرة، وتغوقها المنافع المتوقعة للمشاركين وأسرهم، ما يشير إلى أن برامج قائم على الطلب مثل برنامج PACES يمكن أن تكون وسيلة فعالة من حيث التكلفة لزيادة التحصيل التعليمي.

قدم برنامج توسيع تغطية المدرسة الثانوية (Programa de Ampliación de Cobertura de la Educación Secundaria، أو PACES)، في كولومبيا، قسائم لأكثر من ١٢٥٠٠ طالب تغطي ما يزيد قليلاً عن نصف تكلفة الالتحاق بالمدرسة الثانوية الخاصة. ونظراً إلى ميزانية برنامج PACES المحدودة، حُصّصت القسائم عبر قرعة. استفاد Angrist وآخرون (٢٠٠٢) من هذه المعالجة المعينة عشوائياً لتحديد تأثير برنامج القسائم على الحاصلات التعليمية والاجتماعية. توصل Angrist وآخرون (٢٠٠٢) إلى أن الفائزين بالقرعة كانوا أكثر عرضة بنسبة ١٠ في المائة لإكمال الصف الثامن وأحرزوا، في المتوسط، انحرافات معيارية أكبر بـ ٠,٢ في الاختبارات المعيارية بعد ثلاث سنوات من القرعة الأولية. وتوصلوا أيضاً إلى أن الآثار التعليمية كانت أكبر لدى الفتيات منها لدى الفتيان. نظر الباحثون بعد ذلك إلى أثر البرنامج على العديد من الحاصلات غير التعليمية، وتوصلوا إلى أن الفائزين بالقرعة كانوا أقل عرضة للزواج وكانوا يعملون لمدة أقل بـ ٢-١ ساعة في الأسبوع.

المصدر: Angrist and others 2002.

من الناحية الإحصائية، يؤدي التعيين العشوائي دور المتغير المساعد. وهو متغير يتنبأ بالاشتراك الفعلي للوحدات في برنامج ما، ولكنه لا يرتبط بالخصائص الأخرى للوحدات التي قد تكون مرتبطة بالحاصل. وفي حين لا يمكن لمسؤولي البرنامج السيطرة على جزء من قرار الأفراد بالاشتراك في برنامج ما، إلا أن جزءاً آخر من القرار يكون تحت سيطرتهم. وبالتحديد، فإن جزء القرار الذي يمكن ضبطه هو التعيين في مجموعتي المعالجة والمقارنة. طالما يتنبأ التعيين في مجموعات المعالجة والمقارنة بالاشتراك النهائي في البرنامج، يمكن استخدام التعيين العشوائي بمثابة أداة للتنبؤ بالاشتراك النهائي. يتيح لنا المتغير المساعد هذا استعادة تقديرات متوسط تأثير المعالجة المكاني من تقديرات تأثير النية للعلاج من أجل الوحدات من النوع *يشارك في حال تعيينه*.

يجب أن يحقق المتغير المساعد شرطين اثنين:

١. ينبغي ألا يرتبط المتغير المساعد بخصائص مجموعات المعالجة والمقارنة. ويتحقق ذلك عبر التعيين العشوائي للمعالجة بين الوحدات في عينة التقييم. تُعرّف هذه الحالة باسم الشرط الخارجي. من المهم ألا يؤثر المتغير المساعد مباشرة على الحصلة موضع الاهتمام. يجب أن تنتج التأثيرات فقط من خلال البرنامج الذي نهتم بتقييمه.

٢. يجب أن يؤثر المتغير المساعد على معدلات المشاركة في مجموعتي المعالجة والمقارنة بصورتين مختلفتين. نفكر عادةً في زيادة المشاركة في مجموعة المقارنة. ويمكن التحقق من ذلك عن طريق التحقق من أن المشاركة أعلى في مجموعة المعالجة مقارنة بمجموعة المقارنة. تُعرّف هذه الحالة باسم الملاءمة.

تفسير تقدير متوسط تأثير المعالجة المكاني

تبرز أهمية الفرق بين تقدير متوسط تأثير المعالجة وتقدير متوسط تأثير المعالجة المكاني بالتحديد عندما يتعلق الأمر بتفسير نتائج التقييم. لنفكر منهجيًا في كيفية تفسير تقدير متوسط تأثير المعالجة المكاني. أولاً، علينا تمييز الأفراد الذين يمثلون لبرنامج (النوع *يشارك في حال تعيينه*) مختلفون عن الأفراد الذين لا يمثلون (النوع *لا يشارك على الإطلاق ويشارك دائماً*). وفي مجموعة المعالجة بالتحديد، قد يكون غير الممثلين/غير المشاركين (*لا يشاركون على الإطلاق*) هم الذين يحصلون على القليل من التدخل. في مجموعة المقارنة، من المحتمل أن يكون غير الممثلين/المشاركين (النوع *يشارك دائماً*) هم مجموعة الأفراد الذين يتوقعون الاستفادة على الأكثر من المشاركة. في مثالنا عن تدريب المعلمين، قد يكون المعلمون المعينون للتدريب والذين يقررون عدم المشاركة (النوع *لا يشارك على الإطلاق*) هم أولئك الذين يشعرون بأنهم لا يحتاجون إلى التدريب، أو المعلمين الذين لديهم تكلفة زمنية محتملة أعلى (على سبيل المثال، لأن لديهم عملاً ثانياً أو أطفالاً عليهم العناية بهم)، أو المعلمين الذين يخضعون لإشراف متراخٍ بإمكانهم الإفلات من خلال عدم الحضور. من ناحية أخرى، قد يكون المعلمون المعينون في مجموعة المقارنة ولكنهم يشاركون في كل الأحوال (النوع *يشارك دائماً*) هم أولئك الذين يشعرون بحاجة مطلقة للتدريب، أو المعلمين الذين ليس لديهم أطفال ليعتنوا بهم، أو المعلمين الذين لديهم مدير صارم يصرّ على أن كل شخص بحاجة إلى التدريب. ثانياً، نعرف أن تقدير متوسط تأثير المعالجة المكاني يوفر الأثر لمجموعة فرعية محددة من المجتمع الإحصائي؛ حيث يأخذ في الحسبان فقط المجموعات الفرعية غير المتأثرة بأي نوع من أنواع عدم الامتثال. بمعنى آخر، يأخذ في الحسبان فقط النوع الذي *يشارك في حال تعيينه*. بما أن النوع *يشارك في حال تعيينه* مختلف عن النوعين *لا يشارك على الإطلاق ويشارك دائماً*، فإن الأثر الذي حصلنا عليه من خلال تقدير تأثير المعالجة المكاني (LATE) لا ينطبق على النوعين *لا يشارك على الإطلاق* أو *يشارك دائماً*. على سبيل المثال، لو كانت وزارة التعليم تنوي إجراء جولة ثانية من التدريب وفرض التدريب على المعلمين الذين *لا يشاركون على الإطلاق* والذين لم يتدربوا في الجولة الأولى،

فلن نعرف ما إذا كان سيكون لهؤلاء المعلمين تأثيرات أقل من تأثيرات المعلمين الذين شاركوا في الجولة الأولى أو أعلى منها أو تساويها. وعلى نحو مماثل، إذا وجد المعلمون الأكثر تحفيزًا ذاتيًا دائمًا طريقة للاشتراك في برنامج تدريب المعلمين على الرغم من تعيينهم في مجموعة المقارنة عشوائيًا، فإن متوسط تأثير المعالجة المكاني للممثلين في كل من مجموعتي المعالجة والمقارنة لا يعطينا معلومات حول أثر البرنامج للمعلمين المتحفزين للغاية (يشارك دائمًا). لا ينطبق تقدير متوسط تأثير المعالجة المكاني إلا على مجموعة فرعية محددة من المجتمع الإحصائي: تلك الأنواع التي لا تتأثر بعدم الامتثال - أي فقط النوع الممثل - ولا ينبغي استكمالها بالاستقراء لمجموعات فرعية أخرى من المجتمع الإحصائي.

الترويج العشوائي كمتغير مساعد

في القسم السابق، رأينا كيفية تقدير الأثر بناءً على التعيين العشوائي للمعالجة، حتى لو كان الامتثال لمجموعات المعالجة والمقارنة المحددة أصلاً غير مثالي. بعد ذلك، نقترح مقارنة شبيهة للغاية يمكن تطبيقها لتقييم البرامج ذات الأهلية الشاملة أو الاشتراك مفتوح أو التي لا يستطيع مدير البرنامج فيها التحكم في من يشارك ومن لا يشارك. يوقّر هذه المقاربة، الذي يُسمى الترويج العشوائي (المعروف أيضًا باسم تصميم التشجيع)، تشجيعًا إضافيًا لمجموعة عشوائية من الوحدات للاشتراك في البرنامج. يؤدي هذا الترويج العشوائي بمثابة متغير مساعد. ويؤدي دور مصدر خارجي للاختلاف يؤثر على احتمال تلقي المعالجة، ولكنه لا يرتبط بخلاف ذلك بخصائص المشاركين. تسمح برامج الاشتراك الطوعي عادةً للأفراد المهتمين بالبرنامج بأن يقرروا بأنفسهم الاشتراك والمشاركة. مرة أخرى، فكر في برنامج التدريب الوظيفي التي نوقشت سابقًا—ولكن التعيين العشوائي هذه المرة غير ممكن، ويمتلك أي فرد يرغب في الاشتراك في البرنامج حرية فعل ذلك. تماشيًا إلى حد كبير مع المثال السابق، سنتوقع مصادفة أنواع مختلفة من الأشخاص: مجموعة الممثلين، ومجموعة الذين لا يشاركون على الإطلاق ومجموعة الذين يشاركون دائمًا.

- يشاركون دائمًا. يُقصد بذلك الأفراد الذين يشاركون في البرنامج دائمًا.
- لا يشاركون على الإطلاق. يُقصد بذلك الأفراد الذين لا يشاركون على الإطلاق.
- الممثلون أو الذين يشاركون في حال الترويج لهم. وفي هذا السياق، لأي شخص يودّ الاشتراك في البرنامج الحرية في فعل ذلك. بيد أن بعض الأشخاص قد يكونون مهتمين بالمشاركة، ولكن بسبب جملة من الأسباب، قد لا تتوفر لديهم معلومات كافية أو حافز صحيح للاشتراك.

الممثلون هنا هم الذين يشاركون في حال الترويج لهم: وهم مجموعة من الأفراد الذين يشتركون في البرنامج فقط إذا مُنحوا حافزًا أو حائزًا أو ترويجيًا إضافيًا يشجعهم على الاشتراك. وبدون هذا الحافز الإضافي، سيبقى من يشارك في حال الترويج ببساطة خارج البرنامج.

بالعودة إلى مثال التدريب الوظيفي، إذا كانت الوكالة التي تنظم التدريب تحظى بتدريب جيد ولديها القدرات الكافية، فقد تعتمد سياسة "الباب المفتوح"، وتعالج كل شخص عاطل عن العمل يرغب في المشاركة. ولكن، من غير المحتمل أن يتقدم كل شخص عاطل عن العمل فعليًا للمشاركة أو أن يعرف أصلًا بأن البرنامج موجود. وقد يتردد بعض الأشخاص العاطلين عن العمل في الاشتراك لأنهم لا يعرفون سوى القليل عن محتوى التدريب ويجدون صعوبة في الحصول على معلومات إضافية. افترض الآن أن وكالة التدريب الوظيفي توظف عامل توعية مجتمعية للتجول في أنحاء البلدة وتشجيع مجموعة مختارة عشوائيًا من الأشخاص العاطلين عن العمل على الاشتراك في برنامج تدريب وظيفي. يترك هذا العامل على الأبواب حاملاً قائمة بأسماء الأشخاص العاطلين عن العمل المختارين، ويصف البرنامج التدريبي، ويعرض المساعدة على الشخص للاشتراك في البرنامج المعني. الزيارة شكلٌ من أشكال الترويج أو التشجيع على المشاركة في البرنامج. بطبيعة الحال، لا يمكن للعامل إجبار أي شخص على المشاركة. بالإضافة إلى ذلك، يمكن للأشخاص العاطلين عن العمل الذين لا يزورهم عامل التوعية للاشتراك أيضًا، على الرغم من أنه سيتعين عليهم الذهاب إلى الوكالة بأنفسهم لفعل ذلك. لدينا الآن إحدًا مجموعتان من الأشخاص العاطلين عن العمل: ويُقصد بذلك الأفراد الذين عُيِّنوا عشوائيًا وزارهم عامل التوعية، والأفراد الذين لم يتلقوا زيارة عشوائيًا. إذا كان جهد التوعية فعالاً، فينبغي أن يكون معدل الاشتراك بين الأشخاص العاطلين عن العمل الذين زارهم العامل أعلى منه بين الأشخاص العاطلين عن العمل الذين لم يزورهم العامل.

لنتحدث الآن حول الكيفية التي يمكننا من خلالها تقييم برنامج التدريب الوظيفي هذا. لا يمكننا أن نقارن فقط الأشخاص العاطلين عن العمل الذين يشتركون مع أولئك الذين لا يشتركون. السبب في ذلك هو أن العاطلين عن العمل الذين يشتركون هم على الأرجح مختلفون تمامًا عن أولئك الذين لا يشتركون بالطرق الملحوظة وغير الملحوظة: فقد يكونون ذوي مستوى تعليمي أعلى أو أدنى (يمكن ملاحظة ذلك بسهولة)، وربما يكونون أكثر حماسًا وحرصًا على العثور على وظيفة (من الصعب ملاحظة ذلك وقياسه).

ولكن، ثمة بعض الاختلاف الإضافي الذي يمكننا استغلاله للعثور على مجموعة مقارنة صالحة. فكر للحظة فيما إذا كان بإمكاننا مقارنة مجموعة أشخاص عُيِّنوا عشوائيًا لتلقي زيارة من عامل التوعية مع المجموعة التي لم تُرَزر. ونظرًا إلى أن المجموعات التي رُوِّج لها والمجموعات التي لم يُرَوج لها حُدِّدت عشوائيًا، فإن كلتا المجموعتين تضم تركيبين متطابقين من الأشخاص المتحفزين للغاية (يشارك دائمًا) الذين سيشاركون سواء طرق عامل التوعية أبوابهم أم لا. تضم كلتا المجموعتين أيضًا أشخاصًا غير متحفزين (لا يشارك على الإطلاق) لن يشتركوا في البرنامج على الرغم من كل جهود عامل التوعية. وأخيرًا، إذا كان عامل التوعية فعالاً في التحفيز على الاشتراك، فإن بعض الأشخاص (يشارك في حال الترويج لهم) سيشاركون في البرنامج إذا زارهم عامل التوعية، ولكنهم لن يشتركوا إذا لم يزورهم.

وبما أن عامل التوعية زار مجموعة من الأفراد المعيّنين عشوائيًا، يمكننا اشتقاق تقدير تأثير المعالجة المكاني (LATE) كما ناقشنا سابقًا. الفارق الوحيد هو أنه بدلاً من تعيين البرنامج عشوائيًا، فإننا نرّوِّج لهم عشوائيًا. طالما يظهر الأشخاص من النوع *يشارك في حال الترويج* (الأشخاص الذين يشتركون عندما نوعيهم ولا يشتركون عندما لا نوعيهم) بأعداد كافية، فإن لدينا اختلافًا إضافيًا بين المجموعة مع الترويج والمجموعة بدون ترويج يسمح لنا بتحديد أثر التدريب على النوع *يشارك في حال الترويج*. بدلاً من الامتثال لمجموعة المعالجة، يمثل الأشخاص من النوع *يشارك في حال الترويج* الآن للترويج.

ولكي تعمل هذه الإستراتيجية، نحتاج إلى أن تكون جهود التوعية أو الترويج فعالة في زيادة الاشتراك زيادةً كبيرة بين مجموعة الأفراد من النوع *يشارك في حال الترويج*. وفي الوقت نفسه، لا نريد أن تؤثر أنشطة الترويج بحد ذاتها على الحاصل النهائية للفائدة (مثل الأرباح)، نظرًا إلى أننا في النهاية مهتمون في الأساس بتقدير أثر برنامج التدريب وليس أثر إستراتيجية الترويج على الحاصل النهائية. على سبيل المثال، إذا عرض عاملو التوعية مبالغ كبيرة من المال للأشخاص العاطلين عن العمل لحثهم على المشاركة، فسيصعب تخمين ما إذا كانت أي تغييرات لاحقة في الدخل ناتجة عن التدريب أو نشاط الترويج بحد ذاته.

إن الترويج العشوائي إستراتيجية إبداعية ينتج عنها مكافئ مجموعة مقارنة لأغراض تقييم الأثر. ويمكن استخدامه عندما يكون لدى البرنامج اشتراك مفتوح ويكون من المجدي تنظيم حملة ترويجية تستهدف عينة عشوائية من المجتمع الإحصائي موضع الاهتمام. الترويج العشوائي مثال آخر عن المتغيرات المساعدة يسمح لنا بتقدير الأثر بطريقة غير متحيزة. ولكن من جديد، كما هو الحال مع التعيين العشوائي ذي الامتثال غير المثالي، فإن تقييمات الأثر التي تعتمد على الترويج العشوائي توفر تقديرًا لتأثير المعالجة المكاني (LATE): تقدير مكاني للأثر على مجموعة فرعية محددة من المجتمع الإحصائي، وهي المجموعة *يشارك في حال الترويج*. وكما في السابق، لا يمكن استكمال تقدير تأثير المعالجة المكاني (LATE) هذا بالاستقراء للمجتمع الإحصائي بأكمله، وذلك نظرًا إلى أن المجموعتين *يشارك دائمًا* ولا *يشارك على الإطلاق* من المحتمل أن تكونا مختلفتين تمامًا عن المجموعة *يشارك في حال الترويج*.

هل قلنا "ترويج"؟

يسعى الترويج العشوائي إلى زيادة الاشتراك في برنامج طوعي في عينة فرعية مختارة عشوائيًا من المجتمع الإحصائي. ويمكن أن يأخذ الترويج بحد ذاته عدة أشكال. على سبيل المثال، قد نختار بدء حملة إعلامية لتوعية الأفراد الذين لم يشتركوا لأنهم لم يكونوا يعرفون محتوى البرنامج أو يفهمونه بالكامل. بدلاً من ذلك، قد نختار تقديم حوافز للتسجيل، مثل تقديم هدايا أو جوائز صغيرة أو إتاحة وسائل النقل.

كما ناقشنا في حالة المتغير المساعد بصورة أعم، يجب تحقيق عدد من الشروط لكي ينتج مقارنة الترويج العشوائي تقديرًا صالحًا لأثر البرنامج:

١. يجب أن تكون المجموعتان المروّج وغير المروّج فيها متشابهتين. أي أن متوسط خصائص المجموعتين يجب أن تكون متكافئة إحصائيًا. ويتحقق ذلك عبر التعيين العشوائي لأنشطة التوعية أو الترويج بين الوحدات في عينة التقييم.
٢. ينبغي ألا يؤثر الترويج بحد ذاته مباشرةً على الحاصل موضع الاهتمام. وهذا يتطلب شديد الأهمية بحيث يمكننا القول إن التغيرات في الحاصل موضع الاهتمام ناتجة عن البرنامج بحد ذاته وليس عن الترويج.
٣. يجب أن تغيّر حملة الترويج معدلات الاشتراك تغييرًا كبيرًا في المجموعة المروّج فيها بالنسبة إلى المجموعة غير المروّج فيها. نفكر عادةً بزيادة الاشتراك عبر الترويج. ويمكن التحقق من ذلك عبر التأكد من أن معدلات الاشتراك أعلى في المجموعة التي تتلقى الترويج منها في المجموعات التي لا تتلقاه.

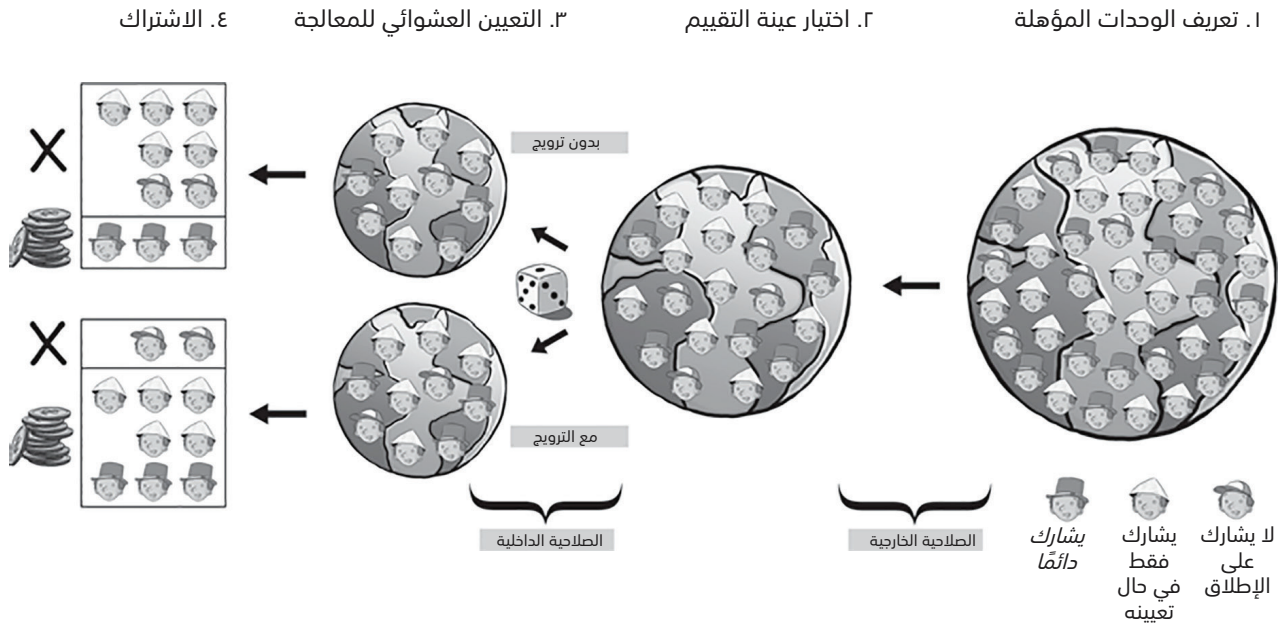
عملية الترويج العشوائي

يقدم الشكل ٥-٣ عملية الترويج العشوائي. وكما في الطرق السابقة، نبدأ بالمجتمع الإحصائي الذي يحتوي على الوحدات المؤهلة للبرنامج. وعلى عكس التعيين العشوائي، لا يمكننا بعد الآن اختيار من سيتلقى البرنامج ومن لن يتلقاه عشوائيًا لأن البرنامج طوعي تمامًا. ولكن، ستكون هناك ثلاثة أنواع من الوحدات ضمن المجتمع الإحصائي للوحدات المؤهلة:

- *يشاركون دائمًا*. يُقصد بذلك الذين يريدون دائمًا الاشتراك في البرنامج.
- *يشاركون في حال الترويج*. يُقصد بذلك الذين سيسجلون في البرنامج فقط في حال منحهم ترويجًا إضافيًا.
- *لا يشاركون على الإطلاق*. يُقصد بذلك الذين لا يريدون التسجيل في البرنامج على الإطلاق سواء أُمُنحوا الترويج أم لا.

من جديد، لاحظ أن كون الشخص من النوع *يشارك دائمًا* أو *يشارك في حال الترويج* أو *لا يشارك على الإطلاق* هو خاصية جوهرية للوحدات لا يمكن لفريق تقييم البرنامج قياسها بسهولة لأنها مرتبطة بعوامل مثل التحفيز والذكاء والمعلومات.

وبمجرد تحديد المجتمع الإحصائي المؤهل، فإن الخطوة التالية هي اختيار عينة عشوائية من المجتمع الإحصائي كي تكون جزءًا من التقييم.



هذه هي الوحدات التي سنجمع البيانات عنها. في بعض الحالات—على سبيل المثال، عندما تكون لدينا بيانات عن المجتمع الإحصائي الكامل ذي الوحدات المؤهلة—قد نقرر تضمين هذا المجتمع الإحصائي بأكمله في عينة التقييم. وبمجرد تعريف عينة التقييم، يعيّن الترويج العشوائي عينة التقييم عشوائيًا في مجموعة رُوّج فيها ومجموعة لم يُرَوّج فيها. وبما أننا نختار عشوائيًا أعضاء كل من المجموعة التي رُوّج فيها والمجموعة غير المروّج فيها، فستشترك كلتا المجموعتين في خصائص عينة التقييم الإجمالية، وستكون هذه الخصائص معادلة للمجتمع الإحصائي ذي الوحدات المؤهلة. لذلك، ستكون للمجموعة المروّج فيها والمجموعة غير المروّج فيها خصائص متشابهة.

بعد انتهاء حملة الترويج، يمكننا ملاحظة معدلات الاشتراك في كلتا المجموعتين. في المجموعة المروّج فيها، سيشارك الأفراد من النوع *يشارك دائمًا*. وعلى الرغم من أننا نعرف الوحدات من النوع *يشارك دائمًا* في المجموعة غير المروّج فيها، لن نتمكن من التمييز بين النوعين *لا يشارك على الإطلاق* و*يشارك في حال الترويج* في المجموعة. على النقيض من ذلك، في المجموعة المروّج فيها، سيشارك الأفراد من كلا النوعين *يشارك في حال الترويج* و*يشارك دائمًا*، في حال لن يشارك الأفراد من النوع *لا يشارك على الإطلاق*. إذًا، في المجموعة المروّج فيها، سنتمكن من تحديد المجموعة من النوع *لا يشارك دائمًا*، ولكننا لن نستطيع التمييز بين المجموعتين *يشارك في حال الترويج* و*يشارك دائمًا*.








تقدير الأثر بموجب الترويج العشوائي

تخيّل أنه من أجل مجموعة من ١٠ أفراد، يمكن أن ترفع حملة الترويج الاشتراك من ٣٠ بالمائة في المجموعة غير المروّج فيها (٣ يشارك دائمًا) إلى ٨٠ بالمائة في المجموعة المروّج لها (٣ يشارك دائمًا و٥٠ يشارك في حال الترويج). افترض أن الحصيلة المتوسطة لجميع الأفراد في المجموعة غير المروّج لها (١٠ أفراد) يساوي ٧٠، وأن متوسط الحصيلة لكل الأفراد في المجموعة المروّج لها (١٠ أفراد) يساوي ١١٠ (الشكل ٥-٤). إذن، ماذا سيكون أثر البرنامج؟

أولاً، لنحسب الفرق المباشر في الحاصل بين المجموعتين المروّج لها وغير المروّج لها، الذي يساوي ٤٠ (١١٠ ناقص ٧٠). نعرف أنه لا يأتي أي من هذا الفرق المساوي ٤٠ يأتي من النوع *لا يشارك على الإطلاق* لأنهم لا يشاركون في أي من المجموعتين. نعرف أيضاً أنه لا ينبغي أن يأتي أي من هذا الفرق المساوي ٤٠ من النوع *يشارك دائماً* لأن هذه المجموعة تشارك في كلتا المجموعتين. لذا فإن كل هذا الفرق البالغ ٤٠ لا بد أن يأتي من النوع *يشارك في حال الترويج*.

الخطوة التالية هي الحصول على تقدير تأثير المعالجة المكاني (LATE) على النوع *يشارك في حال الترويج*. نعلم أن الفرق الكلي بين المجموعتين المروّج فيها وغير المروّج فيها والبالغ ٤٠ يمكن أن يُعزى إلى النوع *يشارك في حال الترويج*، الذي يشكل ما يصل إلى ٥٠ بالمائة فقط من المجتمع الإحصائي. لتقييم متوسط تأثير البرنامج على أحد الممثلين، نقسم ٤٠ على النسبة المئوية للنوع *يشارك في حال الترويج* في المجتمع الإحصائي. على الرغم من أنه لا يمكننا تحديد النوع *يشارك في حال الترويج*، يمكننا استنتاج ما يجب أن يكون النسبة المئوية لهم من المجتمع الإحصائي: وهو الفرق في معدلات الاشتراك بالفئات للمجموعتين المروج لها وغير المروج لها (٥٠ بالمائة، أو ٠,٥). لذلك، يساوي تقدير متوسط تأثير المعالجة المكاني للبرنامج على المجموعة *يشارك في حال الترويج* $0,5/40 = 0,0125$. ونظراً إلى أن الترويج يُعَيّن عشوائياً، فإن المجموعتين المروج لها وغير المروج لها لهما خصائص متساوية. ومن ثم فإن الاختلافات التي نلاحظها في متوسط الحاصل بين المجموعتين لا بد أن تكون ناتجة عن الحقيقة بأن المجموعة المروّج فيها، النوع *يشارك في حال الترويج*، تشارك، بينما لا تشارك المجموعة غير المروّج فيها.

الشكل ٤-٥ تقدير متوسط تأثير المعالجة المكاني في ظل الترويج العشوائي

آثار	المجموعة غير المروّج لها	المجموعة المروّج لها	
Δ النسبة المئوية للمشاركين = ٥٠% $\Delta Y = 40$ LATE = 40/50% = 80	النسبة المئوية للمشاركين = ٣٠% المتوسط Y للمجموعة غير المروّج لها = ٧٠	النسبة المئوية للمشاركين = ٨٠% المتوسط Y للمجموعة المروّج لها = ١١٠	
—			لا يشارك على الإطلاق
			يشارك في حال الترويج
—			يشارك دائماً

ملاحظة: Δ = الأثر السببي، Y = الحصلة. الرموز التي تظهر مقابل الخلفية المظلمة تمثل المشاركين.

من جديد، لا ينبغي أن نستنتج مباشرة الآثار المقدرة للنوع *يشارك في حال الترويج* للمجموعات الأخرى، وذلك لأنها من المحتمل أن تكون مختلفة تمامًا عن المجموعتين من النوعين *لا يشارك على الإطلاق* و*يشارك دائمًا*. يقدم المربع ٣-٥ مثالاً عن ترويج عشوائي لمشروع في بوليفيا.

المربع ٣-٥: الترويج العشوائي للاستثمارات في البنية التحتية للتعليم في بوليفيا

الاتفاق كان أعلى بين المجتمعات المروّج فيها. استخدم نيومان وآخرون (٢٠٠٢) الترويج العشوائي بمثابة متغير مساعد. وقد توصلوا إلى أن الاستثمارات في التعليم قد نجحت في تحسين مقاييس جودة البنية التحتية، مثل الكهرباء، ومرافق الصرف الصحي، والكتب المدرسية لكل طالب، ونسب الطلاب إلى المعلمين. بيد أنهم اكتشفوا أثرًا بسيطًا على الحاصل التعليمي، باستثناء انخفاض نسبته ٢,٥ بالمائة تقريبًا في معدّل التسرّب من المدارس. ونتيجة لهذه النتائج، تركز وزارة التعليم وصندوق الاستثمار الاجتماعي حاليًا المزيد من الانتباه والموارد على "برمجيات" التعليم، بتمويل تحسينات البنية التحتية المادية فقط عندما تدمج التدخل.

أضفت بوليفيا عام ١٩٩١ الطابع المؤسسي على صندوق الاستثمار الاجتماعي (SIF) ووسعت نطاقه، حيث وفر التمويل للمجتمعات الريفية لإجراء استثمارات صغيرة في التعليم والصحة والبنية التحتية للمياه. وقد أنشأ البنك الدولي، الذي كان يساعد في تمويل صندوق الاستثمار الاجتماعي، تقدير أثر ضمن تصميم البرنامج.

وكجزء من تقييم الأثر لمكون التعليم، اختيرت مجتمعات في منطقة تشاكو عشوائيًا للترويج الفعال لتدخل صندوق الاستثمار الاجتماعي وتلقت زيارات إضافية وتشجيعًا للتقدّم من موظفي البرنامج. كان البرنامج مفتوحًا لجميع المجتمعات المؤهلة في المنطقة وكان مدفوعًا بالطلب، حيث إنه كان على المجتمعات التقدم للحصول على تمويل لمشروع معين. لم تلتحق جميع المجتمعات بالبرنامج، ولكن

المصدر: Newman and others 2002.

تقدير أثر برنامج معونة التأمين الصحي (HISP): الترويج العشوائي

لنحاول الآن استخدام طريقة الترويج العشوائي تقدير أثر برنامج معونة التأمين الصحي (HISP). افترض أن وزارة الصحة اتخذت قرارًا تنفيذيًا بإتاحة معونة التأمين الصحي على الفور لأي أسرة ترغب في التسجيل. تلاحظ أن هذا سيناريو مختلف عن حالة التعيين العشوائي التي درسناها حتى الآن. ولكنك تعرف واقعياً أن التوسيع الوطني سيكون متزايدًا بمرور الزمن، لذا تتوصل إلى اتفاق

لمحاولة تسريع الاشتراك في مجموعة فرعية عشوائية من القرى من خلال حملة ترويجية. تنجز في عينة فرعية عشوائية من القرى مجهودات ترويجية مكثفة تتضمن التواصل والتسويق الاجتماعي بهدف زيادة الوعي ببرنامج معونة التأمين الصحي (HISP). الأنشطة الترويجية مصممة بعناية لتجنب محتوى قد يشجع من غير قصد تغيرات في سلوكيات أخرى مرتبطة بالصحة، لأن من شأن هذا إبطال كون الترويج متغيرًا مساعدًا (IV). بدلًا من ذلك، يركز الترويج حصريًا على زيادة الاشتراك في برنامج معونة التأمين الصحي (HISP). بعد عامين من الترويج للبرنامج وتنفيذه، تجد أن ٤٩,٢٪ من الأسر في القرى المعيّنة عشوائيًا للترويج قد اشتركت في البرنامج، بينما سجّلت نسبة ٨,٤٪ فقط من الأسر في القرى غير المروّج فيها (الجدول ١-٥).

ونظرًا إلى أن القرى المروّج فيها وغير المروّج فيها عيّنت عشوائيًا، فأنت تعلم أن متوسط خصائص المجموعتين يجب أن يكونا هما نفسهما في حال عدم وجود ترويج. ويمكنك التحقق من ذلك الافتراض بمقارنة النفقات الصحية الأساسية (وأي خصائص أخرى) لمجمعي الإحصاء. وبعد عامين من تنفيذ البرنامج، تلاحظ أن متوسط الإنفاق الصحي في القرى المروّج فيها يبلغ ١٤,٩٧ دولارًا أمريكيًا، مقارنة بـ ١٨,٨٥ دولارًا أمريكيًا في المناطق غير المروّج فيها (بفارق سالب ٣,٨٧ دولارات أمريكية). ولكن، نظرًا إلى أن الفرق الوحيد بين القرى المروّج فيها والقرى غير المروّج فيها هو أن الاشتراك في البرنامج أعلى في القرى المروّج فيها (بفضل الترويج)، فإن هذا الفرق البالغ ٣,٨٧ دولارًا أمريكيًا في النفقات الصحية لا بد أن يرجع إلى النسبة الإضافية البالغة ٤٠,٧٨٪ من الأسر التي اشتركت في القرى المروّج فيها بسبب الترويج. لذلك، نحتاج إلى تعديل الفرق في الإنفاق الصحي لإيجاد أثر البرنامج على النوع *يشارك في حال الترويج*. لعمل ذلك، نقسم تقدير النية للعلاج - أي الفرق المباشر بين المجموعات المروّج فيها والمجموعات غير المروّج فيها - على النسبة المئوية للنوع *يشارك في حال الترويج*: $3,87 / 0,4078 = 9,49$ - دولارات أمريكية.

الجدول ١-٥ تقييم برنامج معونة التأمين الصحي (HISP): الترويج العشوائي من خلال مقارنة المتوسطات

القرى المروّج فيها	القرى غير المروّج فيها	الفرق	t-stat
النفقات عند خط الأساس (دولار أمريكي) ١٧,١٩	١٧,٢٤	٠,٠٥-	٠,٤٧-
النفقات عند المتابعة (دولار أمريكي) ١٤,٩٧	١٨,٨٥	٣,٨٧-	١٦,٤٣-
معدل الاشتراك في برنامج HISP ٤٩,٢٠٪	٨,٤٢٪	٤٠,٧٨٪	٤٩,٨٥

ملاحظة: مستوى الدلالة: ** = ١ بالمائة.

الجدول ٢-٥ تقييم برنامج معونة التأمين الصحي (HISP): الترويج العشوائي من خلال تحليل الانحدار

الانحدار الخطي متعدد المتغيرات	الانحدار الخطي	الأثر المقدّر على النفقات الصحية الأسرية (دولار أمريكي)
**٩,٧٤- (٠,٤٦)	**٩,٥٠- (٠,٥٢)	

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

ثم يقدّر زميلك، خبير الاقتصاد القياسي الذي يقترح استخدام الترويج العشوائي باعتباره متغيرًا مساعدًا، أثر البرنامج من خلال إجراء المربعات الصغرى على مرحلتين (راجع الدليل الفني عبر الإنترنت على الموقع الإلكتروني <http://www.worldbank.org/ieinpractice> للحصول على المزيد من التفاصيل حول المقارنة الاقتصادية القياسية لتقدير الآثار عبر المتغير المساعد). وتوصل إلى النتائج الموضحة في الجدول ٢-٥. هذا الأثر المقدّر صالح للأسر التي اشتركت في البرنامج بسبب الترويج وما كانت لتشارك لولا ذلك: بمعنى آخر النوع يشارك في حال الترويج.

السؤال ٤ الخاص ببرنامج HISP



- ما الشروط المطلوبة لقبول نتائج تقييم الترويج العشوائي لبرنامج HISP؟
- استنادًا إلى هذه النتائج، هل ينبغي توسيع برنامج معونة التأمين الصحي (HISP) على المستوى الوطني؟

محددات طريقة الترويج العشوائي

الترويج العشوائي هو إستراتيجية ناجعة لتقدير أثر البرامج الطوعية والبرامج ذات الأهلية الشاملة، لا سيما لأنه لا يتطلب استبعاد أي وحدات مؤهلة. ولكن، لهذه المقارنة بعض المحددات الجديرة بالملاحظة مقارنة بالتعيين العشوائي للمعالجة.

أولاً، يجب أن تكون إستراتيجية الترويج فعالة. إذا لم تؤدّ حملة الترويج إلى زيادة الاشتراك، فلن يظهر فرق بين المروج فيها وغير المروج فيها، ولن يكون هناك ما يمكن مقارنته. لذلك من الضروري تصميم حملة الترويج بعناية وتجريبها على نطاق واسع للتأكد من أنها ستكون فعالة. من الناحية الإيجابية، يمكن أن يساعد تصميم حملة الترويج مديري البرامج من خلال تعليمهم كيفية زيادة الاشتراك بعد انتهاء فترة التقييم.

ثانيًا، تقدّر طريقة الترويج العشوائي أثر البرنامج لمجموعة فرعية من المجتمع الإحصائي للوحدات المؤهلة فقط (تأثير المعالجة المكاني).

بالتحديد، يُقدّر متوسط الأثر المكاني للبرنامج من مجموعة الأفراد الذين اشتركوا في البرنامج فقط عند تشجيعهم على ذلك. ولكن، قد تكون للأفراد في هذه المجموعة خصائص مختلفة جدًا عن خصائص الأفراد الذين يشتركون دائمًا أو لا يشتركون على الإطلاق. لذلك، قد يختلف متوسط تأثير المعالجة للمجتمع الإحصائي بأكمله عن متوسط تأثير العلاج المقدر للأفراد الذين يشاركون فقط عند تشجيعهم. ولن يُقدّر تقييم الترويج العشوائي الآثار لمجموعة الأفراد الذين يشتركون في البرنامج بدون تشجيع. في بعض السياقات، قد تكون هذه المجموعة (النوع *يشارك دائمًا*) هي بدقة المجموعة التي صُمم البرنامج لإفادتها. وفي هذا السياق، سيسلط تصميم الترويج العشوائي الضوء على الآثار المتوقعة للمجتمعات الإحصائية الجديدة التي ستشارك من الترويج الإضافي، ولكن ليس على الآثار على المجتمعات الإحصائية التي تلتحق بالفعل بمفردها.

قائمة المراجعة: الترويج العشوائي كمتغير مساعد

يؤدي الترويج العشوائي إلى تقديرات صحيحة للواقع المضاد إذا زادت حملة الترويج بصورة كبيرة من قبول البرنامج دون التأثير مباشرة على الحاصل موضع الاهتمام.

✓ هل الخصائص الأساسية متوازنة بين الوحدات التي تلقت حملة الترويجي وتلك التي لم تتلقها؟ قارب الخصائص الأساسية للمجموعتين.

✓ هل تؤثر حملة الترويج بصورة كبيرة على الالتحاق في البرنامج؟ ينبغي ذلك. قارن معدلات الالتحاق في البرنامج في العينات الفرعية المروّج لها وغير المروّج لها.

✓ هل تؤثر حملة الترويج مباشرة على الحاصل؟ ينبغي ألا تؤثر. لا يمكن عادةً اختبار هذا مباشرة، لذلك يتعين عليك الاعتماد على النظرية والحس العام والمعرفة الجيدة لإعداد تقييم الأثر للحصول على التوجيه.

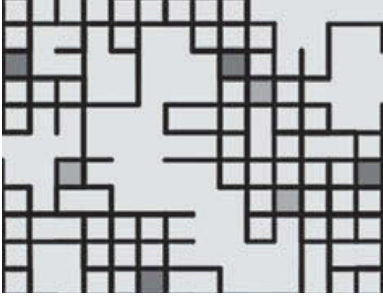
مصادر إضافية

- للحصول على المواد المرتبطة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للحصول على مصادر إضافية حول المتغير المساعد، اطلع على تقييم بنك التنمية للبلدان الأمريكية (<http://www.iadb.org/evaluationhub>)

١. في العلوم الطبية، يتلقى المرضى في مجموعة المقارنة عادةً دواءً وهميًا: أي شيء مثل قرص سكر لا ينبغي أن يكون له أي تأثير على الحصلة المقصودة. يجري ذلك لمراقبة أثر الدواء التجريبي بصورة إضافية، أي أن التغييرات المحتملة في السلوك والحاصل التي يمكن أن تحدث ببساطة من فعل تلقي المعالجة، حتى لو كانت المعالجة نفسها غير فعالة.
٢. تقابل هاتان الخطوتان تقنية الاقتصاد القياسي القائمة على المربعات الصغرى ذات المرحلتين، التي تنتج تقييمًا لمتوسط تأثير المعالجة المحلي.

المراجع

- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92 (5): 1535–58.
- Kearney, Melissa S., and Philip B. Levine. 2015. "Early Childhood Education by MOOC: Lessons from *Sesame Street*." NBER Working Paper 21229, National Bureau of Economic Research, Cambridge, MA.
- Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16 (2): 241–74.



تصميم انقطاع الانحدار

تقييم البرامج التي تستخدم مؤشر الأهلية

غالبًا ما تستخدم البرامج الاجتماعية مؤشرًا لتحديد المؤهل للاشتراك في البرنامج وغير المؤهل للاشتراك فيه. على سبيل المثال، تُوجّه برامج مكافحة الفقر عادةً إلى الأسر الفقيرة، التي تُحدّد من خلال درجة الفقر أو مؤشره. يمكن أن تستند درجة الفقر إلى صيغة تقيس مجموعة من الأصول الأسرية الأساسية بمثابة بديل (أو تقدير) للمتوسطات (مثل الدخل أو الاستهلاك أو القوة الشرائية).^١ تصنف الأسر ذات الدرجات المنخفضة على أنها فقيرة، وتُعدّ الأسر ذات الدرجات الأعلى أفضل حالاً نسبياً. تُحدّد برامج مكافحة الفقر عادةً عتبة أو درجة انقطاع تُعدّ الأسر تحتها فقيرة ومؤهلة للبرنامج. من الأمثلة على ذلك نظام كولومبيا لاختيار المستفيدين من الإنفاق الاجتماعي (انظر المربع ٦-١). ومن بين الأمثلة الأخرى درجات الاختبار (انظر المربع ٦-٣). يمكن منح القبول الجامعي لأصحاب الأداء الأعلى في اختبار معياري تُصنّف نتائجه من أصحاب الأداء الأدنى إلى الأعلى. إذا كان عدد الخانات الزمنية محدودًا، فسيُمنح الطلاب الذين حصلوا على درجة أعلى من حد معين القبول فقط (مثل أعلى ١٠ في المائة من الطلاب). في كلا المثالين، هناك مؤشر أهلية مستمر (درجة الفقر ودرجة الاختبار، على التوالي) يسمح بتصنيف المجتمع الإحصائي موضع الاهتمام، بالإضافة إلى الحد أو درجة الانقطاع التي تحدد المؤهلين وغير المؤهلين.

المربع ١-٦: استخدام تصميم انقطاع الانحدار لتقييم أثر خفض المصروفات الدراسية على معدلات الالتحاق في المدارس في كولومبيا.

أو مستوى تعليمه، مستمرة على طول درجة SISBEN عند خط الأساس، أي أنه ما من "قفزات" في الخصائص على طول درجة SISBEN. الثاني هو أن للأسر على جانبي درجات الانقطاع خصائص متشابهة، ما يكون مجموعات مقارنة ذات مصداقية. ثالثاً، كانت عينة كبيرة من الأسر متاحة. وأخيراً، احتفظت الحكومة بالمعادلة المستخدمة لحساب مؤشر SISBEN سرية، وذلك لحماية النتائج من التلاعب.

باستخدام طريقة تصميم انقطاع الانحدار، وجد الباحثون أنه كان للبرنامج أثر إيجابي كبير على معدلات الالتحاق بالمدارس. وبالتحديد، كان الالتحاق أعلى بنسبة ٣ في المائة لطلاب المدارس الابتدائية من الأسر التي تقل فيها درجة الانقطاع رقم ١، وأعلى بنسبة ٦ بالمائة لطلاب المدارس الثانوية من الأسر بين درجتي الانقطاع رقم ١ ورقم ٢. توفر هذه الدراسة دليلاً على فوائد تقليل التكاليف المباشرة للتعليم، لا سيما للطلاب المعرضين للخطر. ولكن مؤلفيها يدعون أيضاً إلى إجراء مزيد من الأبحاث حول مرونة الأسعار للإبلاغ على نحو أفضل عن تصميم برامج الدعم مثل هذا البرنامج.

استخدم أوركويولا، وليندن، وباريرا-أوسوريو (٢٠٠٧) تصميم انقطاع الانحدار (RDD) لتقييم أثر برنامج للحد من الرسوم المدرسية في كولومبيا (جراتويتاد) على معدلات الالتحاق في مدينة بوغوتا. يُوجّه البرنامج استناداً إلى مؤشر يُسمى SISBEN، وهو مؤشر فقر مستمر للفقر تُحدّد قيمته وفقاً لخصائص الأسرة، مثل الموقع، و مواد بناء المنزل، والخدمات المتوفرة هناك، والعوامل الديموغرافية، والصحة، والتعليم، والدخل، ومهن أفراد الأسرة. وقد حددت الحكومة درجتي انقطاع على طول مؤشر SISBEN: الأطفال للأسر التي تقل درجاتها عن درجة الانقطاع رقم ١ مؤهلون للحصول على التعليم المجاني من الصفوف الأولى إلى الحادي عشر، وأطفال للأسر التي تتراوح درجاتها بين درجتي الانقطاع رقم ١ و ٢ مؤهلون للحصول على دعم بنسبة ٥٠٪ على رسوم الصفين العاشر والحادي عشر، والأطفال من الأسر الحاصلة على درجات أعلى من الدرجة النهائية رقم ٢ غير مؤهلين للحصول على تعليم مجاني أو إعانات.

استخدم المؤلفون تصميم انقطاع الانحدار لأربعة أسباب: الأول هو أن خصائص الأسرة، مثل دخل رب الأسرة

المصدر: Barrera-Osorio, Linden, and Urquiola 2007.

تصميم انقطاع الانحدار طريقة تقييم للأثر يمكن استخدامها للبرامج التي لديها مؤشر أهلية مستمر مع حد أهلية معرّف بوضوح (درجة انقطاع) لتحديد من هو مؤهل ومن غير مؤهل. لتطبيق تصميم انقطاع الانحدار، يجب أن تتحقق الشروط الرئيسية التالية:

١. يجب أن يرتب المؤشر الأشخاص أو الوحدات بطريقة مستمرة أو "سلسلة". إن للمؤشرات مثل درجات الفقر، أو درجات الاختبار، أو العمر العديد من القيم التي يمكن ترتيبها من الصغيرة إلى الكبيرة، ومن ثم يمكن اعتبارها سلسلة. على النقيض من ذلك، لا يمكن تصنيف المتغيرات ذات الفئات المنفصلة أو "مستودعات البيانات" التي لها بعض القيم الممكنة فقط أو التي لا يمكن تصنيفها. ومن أمثلة الحالة الأخيرة حالة التوظيف (موظف أو عاطل عن العمل)، أو المستوى التعليمي الأعلى المحقق (ابتدائي، أو ثانوي، أو جامعي، أو دراسات عليا)، أو ملكية السيارة (نعم أو لا)، أو بلد الميلاد.

٢. يجب أن يحتوي المؤشر على درجة انقطاع محددة بوضوح: أي نقطة على المؤشر يُصنّف السكان فوقها أو تحتها على أنهم مؤهلون للبرنامج. على سبيل المثال، يمكن تصنيف الأسر التي تقل درجة مؤشر الفقر لها عن ٥٠ من ١٠٠ على أنها فقيرة، وقد يُصنّف الأفراد الذين يبلغون من العمر ٦٧ عامًا أو أكبر على أنهم مؤهلون للحصول على معاش تقاعدي، وقد يُصنّف الطلاب الحاصلون على ٩٠ درجة أو أكثر من ١٠٠ في الاختبار على أنهم مؤهلون للحصول على منحة دراسية. درجات الانقطاع في هذه الأمثلة هي ٥٠ و ٦٧ و ٩٠ على التوالي.

٣. يجب أن يكون الانقطاع فريدًا للبرنامج موضع الاهتمام، أي، يجب ألا تكون هناك برامج أخرى، تستخدم درجة الانقطاع نفسها، باستثناء البرنامج المراد تقييمه. على سبيل المثال، إذا كانت درجة الفقر الأقل من ٥٠ تؤهل الأسرة لتلقي التحويلات النقدية، والتأمين الصحي، والنقل العام المجاني، فلن نتمكن من استخدام طريقة تصميم انقطاع الانحدار (RDD) لتقدير أثر برنامج التحويلات النقدية في حد ذاته.

٤. لا يمكن لمندوبي الإحصاء، أو المستفيدين المحتملين، أو مشرفي البرنامج، أو السياسيين التلاعب بدرجة فرد أو وحدة بعينها. يعمل تصميم انقطاع الانحدار (RDD) على تقدير الأثر حول نقطة قطع الأهلية بأنه الفرق بين متوسط حصيلة الوحدات على الجانب المعالج من نقطة قطع الأهلية ومتوسط حصيلة الوحدات على الجانب غير المعالج (المقارن) من الانقطاع.

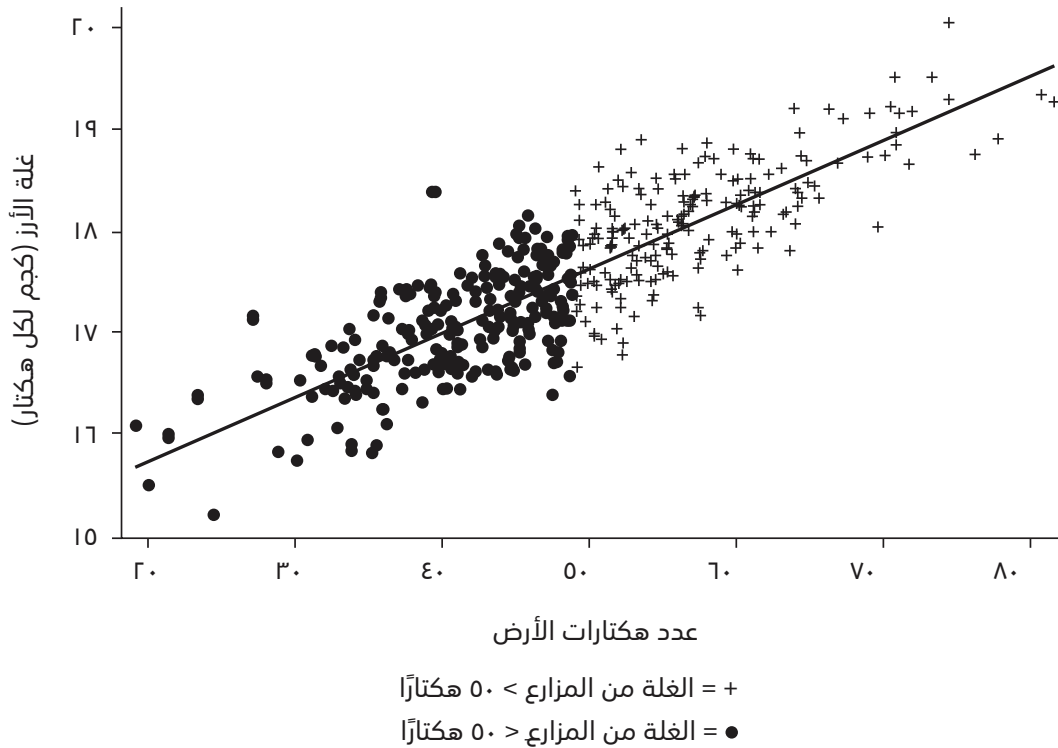
المفهوم الأساسي

تصميم انقطاع الانحدار هو طريقة مناسبة لتقييم الأثر للبرامج التي تستخدم مؤشرًا مستمرًا لتصنيف المشاركين المحتملين والذين لديهم نقطة انقطاع على طول المؤشر الذي يحدّد ما إذا كان المشاركون المحتملون مؤهلين لتلقي البرنامج أم لا.

تأمل برنامجًا زراعيًا يهدف إلى تحسين إجمالي غلال الرز عبر دعم شراء المزارعين للسماذ. يستهدف البرنامج المزارع الصغيرة ومتوسطة الحجم، التي يصنفها على أنها مزارع مساحة أراضيها أقل من ٥٠ هكتارًا. قبل بدء البرنامج، قد نتوقع أن تكون مخزجات المزارع الصغيرة أقل من المزارع الكبيرة، كما هو موضح في الشكل ٦-١، الذي يحدد حجم المزرعة وإنتاج الأرز. ودرجة الأهلية في هذه الحالة هي عدد هكتارات المزرعة، ويساوي الانقطاع ٥٠ هكتارًا. تحدّد قواعد البرنامج على أن المزارع التي تقل مساحتها عن ٥٠ هكتارًا مؤهلة لتلقي إعانات الأسمدة، أما المزارع التي تبلغ مساحتها ٥٠ هكتارًا أو أكثر فليست مؤهلة. وفي هذه الحالة، قد نتوقع أن نرى عددًا من المزارع التي مساحتها ٤٨ أو ٤٩ أو حتى ٤٩,٩ هكتارًا تشارك في البرنامج. لن تشارك مجموعة أخرى من المزارع التي مساحتها ٥٠ و ٥٠,١ و ٥٠,٢ هكتارًا في البرنامج لأنها تقع فقط في الجانب غير المؤهل من الانقطاع. من المحتمل أن تكون مجموعة المزارع التي مساحتها ٤٩,٩ هكتارًا مشابهة جدًا لمجموعة المزارع التي مساحتها

٥٠,١ هكتارًا من جميع النواحي، باستثناء أن إحدى المجموعتين تلقت إعانة الأسمدة والمجموعة الأخرى لم تحصل عليها. مع ابتعادنا عن نقطة قطع الأهلية، قد يزداد الفرق بين المزارع المؤهلة وغير المؤهلة. بيد أن حجم المزرعة مقياس جيد لمدى اختلافهما، ما يسمح لنا بالسيطرة على الكثير من هذه الاختلافات.

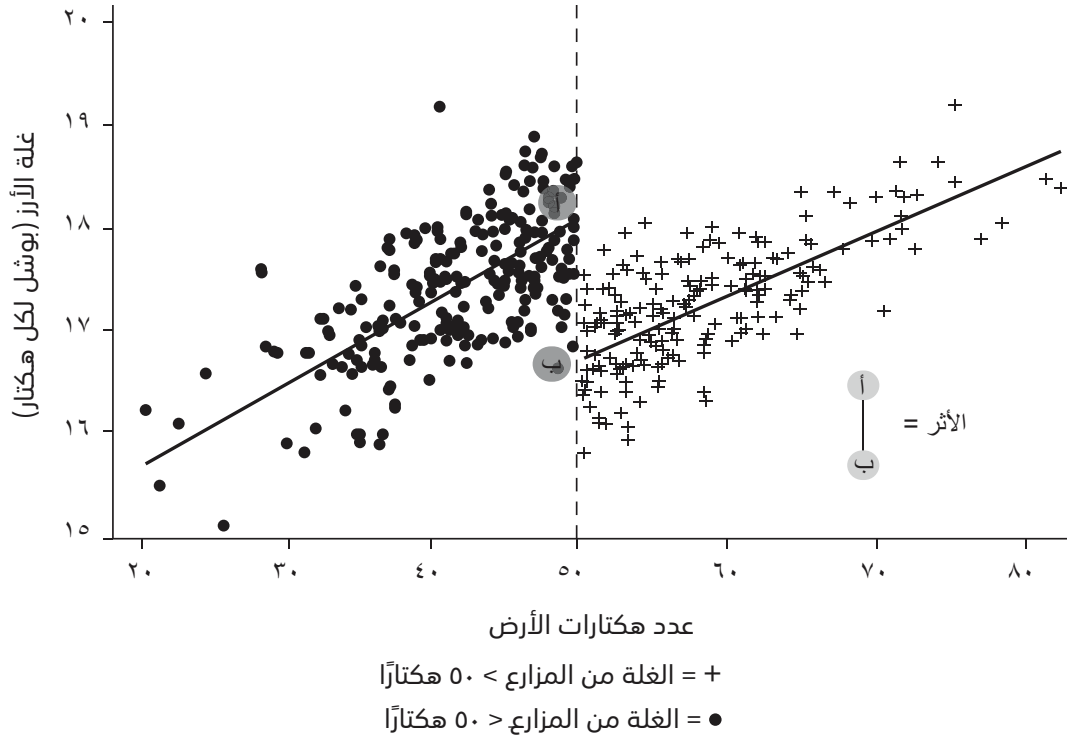
الشكل ٦-١. غلة الأرز، المزارع الصغيرة مقابل المزارع الكبيرة (خط الأساس)



وبمجرد بدء البرنامج وتقديم الإعانة في تكلفة الأسمدة للمزارع الصغيرة والمتوسطة، يمكن لتقييم الأثر استخدام تصميم انقطاع الانحدار (RDD) لتقييم أثره (الشكل ٦-٢). يحسب تصميم انقطاع الانحدار (RDD) الأثر على أنه الفرق في الحاصل، مثل غلال الأرز، بين الوحدات على جانبي انقطاع الأهلية، والذي يساوي في مثالنا حجم مزرعة يساوي ٥٠ هكتارًا. تشكل المزارع التي كانت كبيرة جدًا بحيث يتعذر تسجيلها في البرنامج مجموعة المقارنة، وتنتج تقديرًا لحصيلة الواقع المضاد لتلك المزارع في مجموعة المعالجة التي كانت صغيرة بما يكفي تحديدًا للتسجيل. ونظرًا إلى أن هاتين المجموعتين من المزارع كانتا متشابهتين جدًا في الأساس وتتعرضان لنفس المجموعة من العوامل الخارجية بمرور الزمن (مثل الطقس، وصدمة الأسعار، والسياسات الزراعية المحلية والوطنية)، لا بد أن يكون السبب المعقول الوحيد لاختلاف الحاصل هو البرنامج نفسه.

وبما أن مجموعة المقارنة تتكون من مزارع توجد فوق عتبة الأهلية مباشرة، فإن الأثر الذي يعطيه تصميم انقطاع الانحدار (RDD) صالح محليًا فقط—أي في الجوار المحيط بدرجة انقطاع الأهلية. وهكذا نحصل على تقدير لمتوسط تأثير المعالجة المكاني أو (LATE) (انظر الفصل الخامس). أثر برنامج إعانة الأسمدة صالح للمزارع الكبيرة ومتوسطة الحجم: أي تلك التي مساحتها أقل من ٥٠ هكتارًا.

الشكل ٢-٦. غلة الأرز، المزارع الصغيرة مقابل المزارع الكبيرة (متابعة)



لن يتمكن تقييم الأثر بالضرورة من تحديد أثر البرنامج على المزارع الصغرى مباشرة—ولنقل تلك التي مساحة أرضها ١٠ أو ٢٠ فدانًا—حيث يمكن أن يختلف أثر إعانة الأسمدة بطرق مهمة عن المزارع متوسطة الحجم التي مساحة أرضها ٤٨ أو ٤٩ هكتارًا. إحدى إيجابيات طريقة تصميم انقطاع الانحدار هي أنه بمجرد تطبيق قواعد أهلية البرنامج، لا يلزم ترك أي وحدات مؤهلة بدون معالجة لغرض تقييم الأثر. المفاضلة هي أن آثار الملاحظات بعيدة جدًا عن الانقطاع لن تكون معروفة. يقدم المربع ٢-٦ مثالاً على استخدام تصميم انقطاع الانحدار لتقييم برنامج شبكة الأمان الاجتماعي في جامايكا.

تصميم انقطاع الانحدار الضبابي

بمجرد أن نتحقق من عدم وجود دليل على التلاعب في مؤشر الأهلية، فربما نظل نواجه تحديًا إذا لم تحترم الوحدات تعيينها في مجموعات المعالجة أو المقارنة. بمعنى آخر، قد تختار بعض الوحدات المؤهلة للبرنامج على أساس مؤشر أهليتها عدم المشاركة، بينما قد تجد الوحدات الأخرى التي لم تتأهل للبرنامج على أساس مؤشر الأهلية الخاص بها طريقة للمشاركة على أي حال. عندما تمثل جميع الوحدات للمهمة التي تتوافق معها على أساس مؤشر أهليتها، فإننا نقول إن تصميم انقطاع الانحدار (RDD) "واضح"، بينما إذا كان هناك عدم امتثال على أي جانب من جانبي الانقطاع، فإننا

المربع ٢-٦: شبكات الأمان الاجتماعي المستندة إلى مؤشر الفقر في جامايكا

ومستويات تحفيز متشابهان، حيث تقدّم كلا نوعي الأسر للبرنامج. واستخدم الباحثان أيضًا درجة الأهلية للبرنامج في تحليل الانحدار للمساعدة على ضبط أي فروق بين المجموعتين.

وجد Levy و Ohls (٢٠١٠) أن برنامج PATH رفع من الالتحاق بالمدرسة لدى الأطفال الذين تتراوح أعمارهم بين ٦ و ١٧ وسطيًا بمقدار ٠,٥ يوم في الشهر، وهو ارتفاع كبير قياسًا إلى نسبة الالتحاق المرتفعة تمامًا البالغة ٨٥ بالمائة. علاوةً على ذلك، ازدادت الزيارات الصحية للأطفال في الأعمار بين ٠ و ٦ سنوات بمقدار ٣٨ بالمائة تقريبًا. وبينما لم يتمكن الباحثان من العثور على أي آثار طويلة المدى على التحصيل الدراسي أو حالة الرعاية الصحية، فقد توصلا إلى أن حجم الآثار التي وجدوها كان متسقًا على نطاق واسع مع برامج التحويلات النقدية المشروطة المنفذة في بلدان أخرى. وواحد من الجوانب المثيرة للاهتمام لهذا التقييم هو أنه جمع بيانات نوعية وكمية، باستخدام أنظمة معلومات، ومقابلات، ومجموعات تركيز، ومسوح للأسر المعيشية.

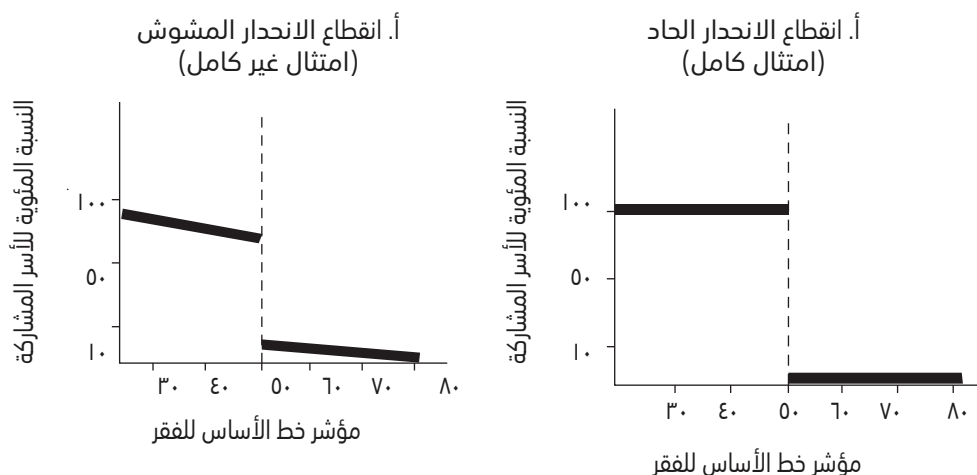
استُخدمت طريقة تصميم انقطاع الانحدار (RDD) لتقييم أثر مبادرة شبكة أمان اجتماعي في جامايكا. ففي عام ٢٠٠١، بدأت حكومة جامايكا برنامج تنمية من خلال الصحة والتعليم (PATH) لزيادة الاستثمارات في رأس المال البشري وتحسين توجيه منافع الرعاية الاجتماعية إلى الفقراء. قدّم البرنامج منحةً صحيةً وتعليميةً للأطفال في الأسر الفقيرة المؤهلة، شريطة الالتحاق بالدراسة وتلقي معالجات رعاية صحية منتظمة. بلغ متوسط الإعانة الشهرية لكل طفل ٦,٥٠ دولارات أمريكية تقريبًا، بالإضافة إلى إعفاء حكومي من الرسوم الصحية والتعليمية.

تمكّن Levy و Ohls (٢٠١٠) من خلال تحديد الأهلية للبرنامج من خلال صيغة تحديد درجات، من مقارنة الأسر تحت عتبة الأهلية مباشرة بالأسر فوقها مباشرة (على بُعد ما بين ٢ و ١٥ نقطة من الانقطاع). يبرر الباحثان استخدام طريقة تصميم انقطاع الانحدار (RDD) ببيانات أساسية توضح أن الأسر المعالجة وأسر المقارنة لديهما مستوي فقر متشابهان مقيسان بمتوسط الدرجات البديل،

المصدر: Levy and Ohls 2010.

نقول إن تصميم انقطاع الانحدار "ضبابي" (الشكل ٦-٣). إذا كان تصميم انقطاع الانحدار ضبابيًا، فيمكننا استخدام مقارنة المتغير المساعد لتصحيح عدم الامتثال (انظر الفصل الخامس). تذكر أنه في حالة التعيين العشوائي مع عدم الامتثال، استخدمنا التعيين العشوائي كمتغير مساعد لتصحيح عدم الامتثال. في حالة تصميم انقطاع الانحدار، يمكننا استخدام التعيين الأصلي استنادًا إلى مؤشر الأهلية بصفته المتغير المساعد. ولكنّ فعل ذلك فيه عيب؛ إذ إن تقدير تصميم انقطاع الانحدار سيصبح أكثر موضوعية، وذلك من حيث أنه لن يعود صالحًا لجميع الملاحظات القريبة من الانقطاع، وسيمثل بدلًا من ذلك الأثر للمجموعة الفرعية من المجتمع الإحصائي التي تقع بالقرب من نقطة الانقطاع والتي تشارك في البرنامج فقط بسبب معايير الأهلية.

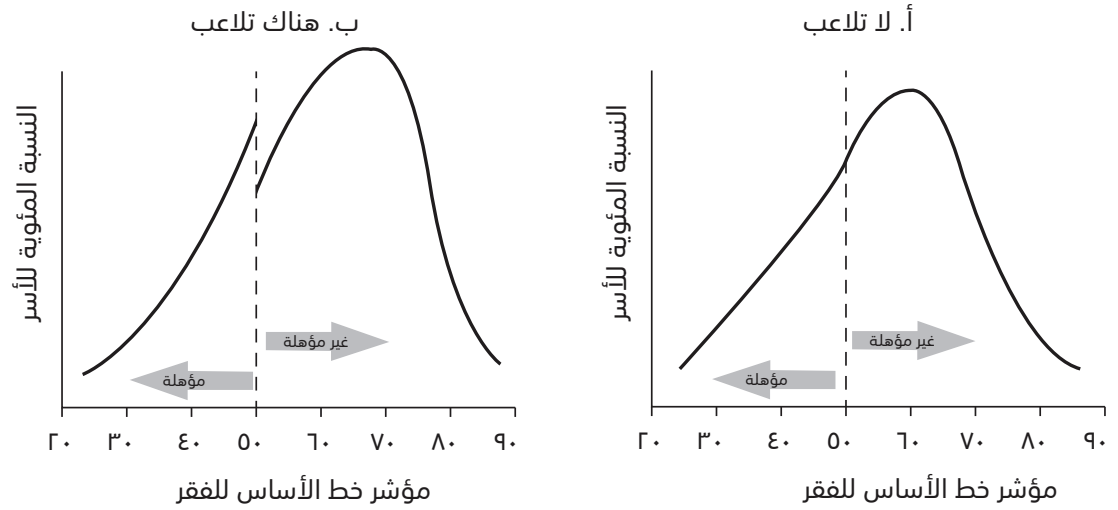
الشكل ٣-٦ الامتثال للتعيين



التحقق من صلاحية تصميم انقطاع الانحدار

لكي يعطي انقطاع الانحدار الضبابي متوسط تأثير معالجة مكاني غير متحيز عند نقطة الانقطاع، من المهم عدم التلاعب بمؤشر الأهلية حول نقطة الانقطاع بحيث يمكن للفرد تغيير حالة السيطرة.^٢ يمكن أن يأخذ التلاعب بمعايير الأهلية العديد من الصيغ. على سبيل المثال، يمكن أن يغير مندوبو الإحصاء الذين يجمعون البيانات المستخدمة لحساب درجة الأهلية إجابة أو إجابتين من المستجيبين. أو قد يكذب المستجيبون عمدًا على مندوبي الإحصاء إذا كانوا يعتقدون أن فعل ذلك من شأنه أن يؤهلهم للبرنامج. بالإضافة إلى ذلك، يمكن أن يزداد التلاعب بالدرجات سوءًا مع مرور الزمن، وذلك مع بدء مندوبي الإحصاء، والمستجيبين، والسياسيين في تعلّم "قواعد اللعبة". في مثال إعانة الأسمدة، من شأن التلاعب حول نقطة الانقطاع أن يحدث لو تمكّن أصحاب المزارع من تعديل سندات ملكية الأرض أو الإبلاغ الخاطئ عن أحجام مزارعهم. أو ربما يعثر مزارع لديه أرض مساحتها ٥٠,٣ هكتارًا على طريقة لبيع نصف هكتار كي يتأهل للبرنامج، إذا كانت الفوائد المتوقعة من إعانة الأسمدة تستحق فعل ذلك.

يوضح الشكل ٦-٤ علامة توضيحية للتلاعب. يوضح المقطع "أ" توزيع الأسر وفقًا لمؤشرها الأساسي عند عدم وجود تلاعب. كثافة الأسر حول الانقطاع (٥٠) مستمرة (أو سلسلة). توضح اللوحة (ب) حالة مختلفة: يبدو أن عددًا أكبر من الأسر "مجمعة" تحت الانقطاع مباشرة، بينما يمكن العثور على عدد قليل نسبيًا من الأسر فوق الانقطاع مباشرة. وبسبب غياب سبب مسبق للاعتقاد بأنه يجب أن تكون هناك إزاحة كبيرة في عدد الأسر حول الانقطاع مباشرة، فإن حدوث هذه الإزاحة في التوزيع حول الانقطاع هو دليل على أن الأسر قد تتلاعب في الدرجات للوصول إلى البرنامج. وهناك اختبار آخر للتلاعب، وهو مؤشر الأهلية بالنسبة إلى متغير الحصيلة عند خط الأساس ويتحقق من عدم وجود انقطاع أو "قفزة" حول خط الانقطاع مباشرة.



المربع ٣-٦: تأثير تصنيف الطلاب حسب درجات الاختبار على أداء المدرسة في كينيا

المعنيين في القسم منخفض الأداء. وعند حساب المتوسط، كانت درجات اختبار الخط النهائي في المدارس التي عينت الطلاب لأقسام ذات أصحاب الأداء الأعلى أو الأقل بمقدار ٠,١٤ انحرافًا معياريًا أعلى من درجات المدارس التي لم تستخدم هذه الطريقة واستخدمت بدلًا منها التعيين العشوائي لإنشاء مجموعات مكافئة من الطلاب. ولم تكن هذه النتائج مدفوعة فقط من قبل الطلاب في القسم عالي الأداء، بل أبدى الطلاب في القسم منخفض الأداء أيضًا تحسينات في درجات الاختبار. بالنسبة إلى الطلاب الموجودين حول درجة الانقطاع، توصل الباحثون إلى أنه لم يكن هناك فرق ذو دلالة إحصائية في درجات اختبار الخط النهائي. ترفض هذه النتائج الفرضية القائلة إن الطلاب يستفيدون مباشرة من وجود زملائهم الأكثر تفوقًا.

أجرى كريم، ودوباس، ودوفلو (٢٠١١) تجربة على ١٢١ مدرسة ابتدائية في غرب كينيا لاختبار ما إذا كان تعيين الطلاب في الفصول الدراسية استنادًا إلى الأداء يحسن الحاصلات التعليمية. في نصف المدارس، قُسم طلاب الصف الأول عشوائيًا إلى قسمين مختلفين اثنين للفصول الدراسية. في النصف الثاني من المدارس، عُيّن الطلاب إما في قسم عالي الأداء أو منخفض الأداء استنادًا إلى درجات اختباراتهم الأولية، باستخدام درجة الاختبار بمثابة نقطة انقطاع.

سمح تصميم انقطاع الانحدار (RDD) للباحثين باختبار ما إذا كان تكوين الطلاب في الفصل يؤثر مباشرة على درجات الاختبار. وقد قارنوا درجات اختبار خط النهاية للطلاب الذين كانوا حول الانقطاع مباشرة لمعرفة ما إذا كان أولئك المعينون في القسم عالي الأداء قد أبلوا بلاءً حسنًا أكثر من

المصدر: Duflo, Dupas, and Kremer 2011.



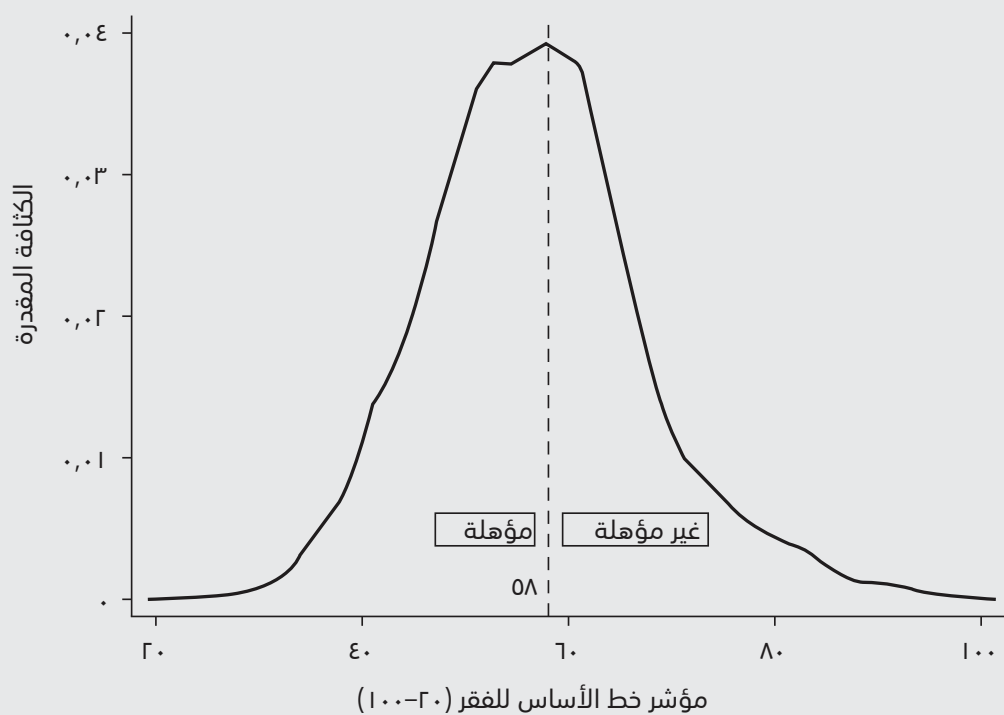
فكر الآن كيف يمكن تطبيق تصميم انقطاع الانحدار (RDD) على برنامج معونة التأمين الصحي (HISP). بعد إجراء المزيد من الاستقصاء حول تصميم معونة التأمين الصحي (HISP)، تجد أنه بالإضافة إلى اختيار القرى عشوائيًا، وجهت السلطات البرنامج نحو الأسر منخفضة الدخل باستخدام خط الفقر الوطني. يركز خط الفقر على مؤشر للفقر يعيّن لكل أسرة في البلاد درجة بين ٢٠ و ١٠٠ استنادًا إلى أصولها، وشروط سكنها، وبنيتها الاجتماعية السكانية. عيّن خط الفقر رسميًا على ٥٨. معنى ذلك أن جميع الأسر ذات الدرجة ٥٨ أو الأقل من ذلك تُصنّف بأنها فقيرة، وكل الأسر ذات الدرجة الأعلى من ٥٨ تُعدّ غير فقيرة. وحتى في قرى المعالجة، فإن الأسر الفقيرة فقط مؤهلة للاشتراك في برنامج معونة التأمين الصحي (HISP). تتضمن مجموعة بياناتك معلومات عن الأسر الفقيرة والأسر غير الفقيرة في قرى المعالجة.

قبل إنجاز تقديرات تصميم انقطاع الانحدار، تقرر التحقق مما إذا كان هناك أي أدلة على التلاعب بمؤشر الأهلية. وبمثابة خطوة أولى، تتحقق مما إذا كانت كثافة مؤشر الأهلية تثير أي مخاوف بشأن التلاعب بالمؤشر. وتمثّل بيانًا النسبة المئوية للأسر مقابل مؤشر خط الأساس للفقر (الشكل ٦-٥).^٣ لا يشير الشكل إلى أي "تجمّع" للأسر تحت الانقطاع البالغ ٥٨ مباشرة.

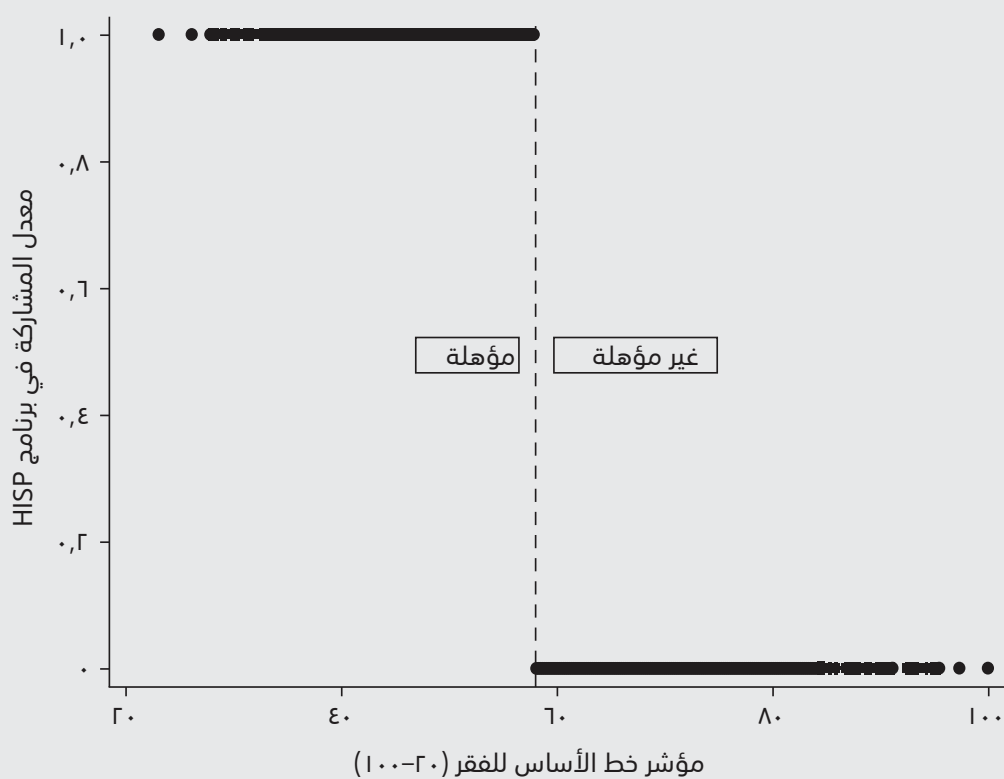
بعد ذلك، تتحقق مما إذا كانت الأسر قد احترمت تعيينها في مجموعتي المعالجة والمقارنة على أساس درجة الأهلية. تمثل بيانًا المشاركة في البرنامج بالنسبة إلى مؤشر خط الأساس للفقر (الشكل ٦-٦) وتجد أنه بعد عامين من بدء البرنامج التجريبي، سُمح فقط للأسر ذات الدرجة ٥٨ أو أقل (أي إلى يسار خط الفقر) بالاشتراك في برنامج HISP. بالإضافة إلى ذلك، فإن جميع الأسر المؤهلة قد اشتركت في برنامج HISP. بمعنى آخر، ستجد امتثالًا كاملاً وسيكون لديك تصميم انقطاع انحدار "حاد".

يمكنك الآن التقدم نحو تطبيق طريقة تصميم انقطاع الانحدار لحساب أثر البرنامج. باستخدام بيانات المتابعة، يمكنك مرة أخرى تمثيل العلاقة بين الدرجات في مؤشر الفقر والنفقات الصحية المتنبأ بها وإيجاد العلاقة الموضحة في الشكل ٦-٧.

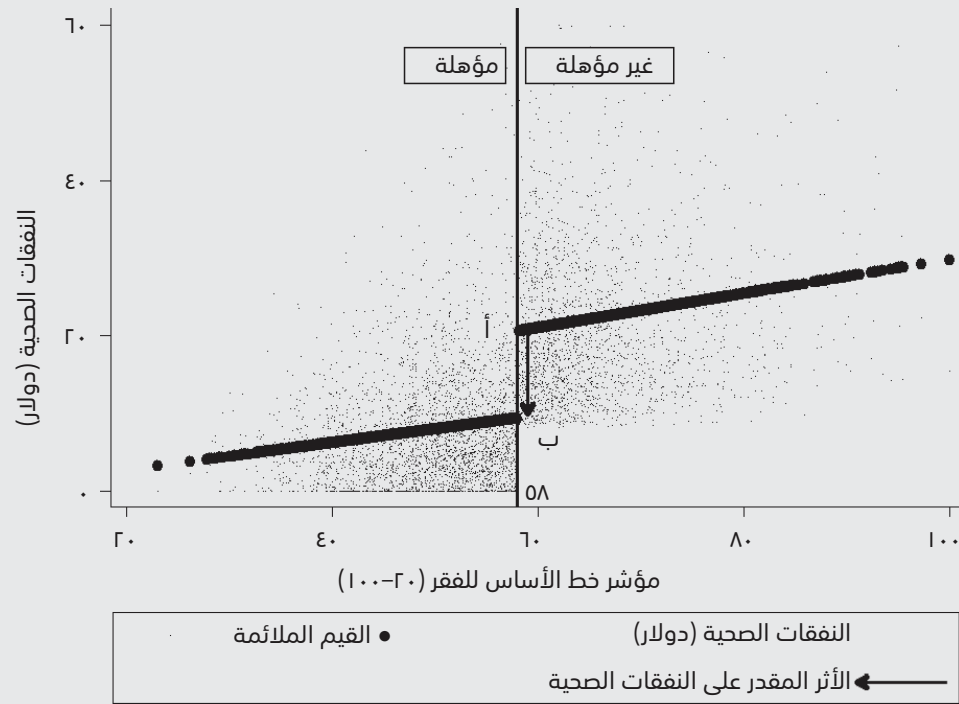
الشكل ٥-٦ برنامج HISP: كثافة الأسر، وفقًا لمؤشر خط الأساس للفقر



الشكل ٦-٦ المشاركة في برنامج HISP، حسب مؤشر خط الأساس للفقر



الشكل ٧-٦ مؤشر الفقر والنفقات الصحية، برنامج HISP، بعد عامين



الجدول ١-٦ تقييم برنامج معونة التأمين الصحي (HISP): تصميم انقطاع الانحدار من خلال تحليل الانحدار

الانحدار الخطي متعدد المتغيرات	
الأثر المقدر على النفقات الصحية - **٩,٠٣	القيم الملائمة
الأسرية (٠,٤٣)	

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

في العلاقة بين مؤشر الفقر والنفقات الصحية المتنبأ بها، تجد انفصالاً أو انقطاعاً واضحاً، عند خط الفقر (٥٨).

يعكس الانقطاع انخفاضاً في النفقات الصحية للأسر المؤهلة لتلقي البرنامج. ونظراً للتشابه الكبير بين الأسر على كلا جانبي درجة الانقطاع البالغة ٥٨، فإن التفسير المعقول لمستوى النفقات المختلف هو أن إحدى مجموعتي الأسر كانت مؤهلة للاشتراك في البرنامج، ولم تكن الأخرى مؤهلة. تقدّر هذا الفرق من خلال انحدار عبر النتائج الموضحة في الشكل ١-٦.

السؤال ٥ الخاص ببرنامح HISP



- أ. هل النتيجة الموضحة في الجدول ٦-١ صالحة لجميع الأسر المؤهلة؟
ب. بالمقارنة مع الأثر المقدر بطريقة التعيين العشوائي، ما الذي توضحه هذه النتيجة بشأن الأسر ذات مؤشر الفقر الأقل من ٥٨؟
ج. استنادًا إلى تقدير الأثر عبر تصميم انقطاع الانحدار، هل ينبغي توسيع برنامح HISP على المستوى الوطني؟

محددات طريقة تصميم انقطاع الانحدار وتفسيرها

يوفر تصميم انقطاع الانحدار تقديرات لمتوسط تأثير المعالجة المكاني (LATE) حول انقطاع الأهلية عند النقطة التي تكون فيها وحدات المعالجة والمقارنة أكثر شبهاً. وكلما اقتربت من الانقطاع، ازداد تشابه الوحدات على أي من جانبي الانقطاع. وفي الواقع، عندما تقترب كثيرًا من درجة الانقطاع، فستكون الوحدات على أي من جانبي الانقطاع متشابهة لدرجة أن مقارنتك ستكون بالجودة نفسها إذا ما اخترت مجموعتي المعالجة والمقارنة باستخدام التعيين العشوائي للمعالجة.

ونظرًا إلى أن طريقة تصميم انقطاع الانحدار تقدر أثر البرنامج حول درجة الانقطاع، أو مكانيًا، فلا يمكن تعميم التقدير بالضرورة ليشمل الوحدات ذات الدرجات الأبعد عن درجة الانقطاع: أي حيث قد لا يكون الأفراد المؤهلون وغير المؤهلين بدرجة التشابه نفسها. يمكن النظر إلى عدم قدرة طريقة تصميم انقطاع الانحدار على توفير تقدير لمتوسط تأثير المعالجة لجميع المشاركين في البرنامج على أنها نقطة قوة ومحدد في الوقت نفسه، اعتمادًا على مسألة التقييم موضع الاهتمام. إذا كان التقييم يسعى في المقام الأول إلى الإجابة عن السؤال: "هل يجب أن يكون البرنامج موجودًا أم لا؟"، فإن متوسط تأثير المعالجة لجميع السكان المؤهلين قد يكون هو المعيار الأكثر صلة، ومن الواضح أن طريقة تصميم انقطاع الانحدار لن تكون مثالية. ولكن، إذا كان سؤال السياسة موضع الاهتمام هو: "هل ينبغي قطع البرنامج أو توسيعه عند الهامش" —أي للمستفيدين (المحتملين) حول الانقطاع— فإن طريقة تصميم انقطاع الانحدار تنتج بدقة التقدير المكاني موضع الاهتمام لتوجيه هذا القرار السياسي المهم.

كما ذكرنا، قد يكون هناك تعقيد إضافي عندما يكون الامتثال على أي من جانبي الانقطاع غير مثالي. يحدث تصميم انقطاع الانحدار الضبابي عندما تتمكن وحدات غير مؤهلة استنادًا إلى درجة مؤشرها من الوصول إلى البرنامج، أو عندما تختار وحدات مؤهلة استنادًا إلى درجة مؤشرها عدم المشاركة في البرنامج.

وفي هذه الحالة، يمكننا استخدام منهجية متغير مساعد شبيهة بالمنهجية الموضحة في الفصل ٥: سيستخدم موقع الوحدات فوق درجة الانقطاع أو تحتها بمثابة متغير مساعد للمشاركة الملحوظة في البرنامج. وكما كان الحال في الأمثلة التي نوقشت في الفصل ٥، فإن القيام بذلك له عيب: يمكننا تقدير الأثر فقط لتلك الوحدات الحساسة لمعايير الأهلية—درجة الاشتراك في حال الأهلية، وليس *يشارك دائماً* أو لا يشارك أبداً.

إن حقيقة أن تقديرات طريقة تصميم انقطاع الانحدار تؤثر فقط حول درجة الانقطاع تسبب أيضاً تحديات من حيث القوة الإحصائية للتحليل. في بعض الأحيان، تُستخدم مجموعة محدودة فقط من الملاحظات التي تقع بالقرب من درجة الانقطاع في التحليل، ما يقلل عدد الملاحظات في تحليل تصميم انقطاع الانحدار بالنسبة للطرق التي تحلل جميع الوحدات في مجموعتي المعالجة والمقارنة. للحصول على قدرة إحصائية كافية عند تطبيق تصميم انقطاع الانحدار، سنحتاج إلى اختيار عرض نطاق حول درجة الانقطاع يتضمن عدداً كافياً من الملاحظات. عملياً، عليك أن تحاول استخدام عرض نطاق كبير قدر الإمكان، مع الحفاظ على التوازن في الخصائص المرصودة للمجتمع الإحصائي فوق درجة الانقطاع وتحتها. يمكنك بعدئذٍ تشغيل التقدير عدة مرات باستخدام عروض نطاق مختلفة للتحقق مما إذا كانت التقديرات حساسة لعرض النطاق المختار.

من المحاذير الأخرى عند استخدام طريقة تصميم انقطاع الانحدار أن التوصيف قد يكون حساساً للصيغة الوظيفية المستخدمة في تمثيل العلاقة بين درجة الأهلية والحصيلة موضع الاهتمام. في المثال المقدم في هذا الفصل، افترضنا أن العلاقة بين مؤشر الأهلية والحصيلة خطية. في الواقع، يمكن أن تكون العلاقة أعقد، بما في ذلك العلاقات غير الخطية والتفاعلات المتبادلة بين المتغيرات. إذا لم تأخذ في الحسبان هذه العلاقات المعقدة في التقدير، فقد يتم الخلط بينها وبين الانقطاع، ما يؤدي إلى تفسير غير صحيح للأثر المقدر لتصميم انقطاع الانحدار. عملياً، يمكنك تقدير أثر البرنامج باستخدام العديد من الصيغ الوظيفية (الخطية، ومن الدرجة الثانية، والمكعبة، ومن الدرجة الرابعة، وما شابه ذلك) لتقييم ما إذا كانت تقديرات الأثر في الواقع حساسة للصيغة الوظيفية.

وأخيراً، كما نوقش أعلاه، هناك بضعة شروط مهمة لقاعدة الأهلية والانقطاع. أولاً، يجب أن تكون فريدة بالنسبة إلى البرنامج ذي الصلة. ويمكن استخدام مؤشر للفقر يصنف الأسر أو الأفراد، على سبيل المثال، لتوجيه مجموعة متنوعة من البرامج الاجتماعية نحو الفقراء. وفي هذه الحالة، لن يكون من الممكن عزل أثر برنامج محدد لمكافحة الفقر عن جميع البرامج التي تستخدم معايير الاستهداف نفسها. ثانياً، ينبغي أن تكون قاعدة الأهلية والانقطاع مقاومة للتلاعب من قبل مندوبي الإحصاء، أو المستفيدين المحتملين، أو مشرفي البرامج، أو السياسيين. يتسبب التلاعب في مؤشر الأهلية في حدوث انقطاع في المؤشر يقوض الشرط الأساسي اللازم كي تعمل الطريقة: أي أن مؤشر الأهلية يجب أن يكون مستمرًا حول الانقطاع.

وحتى في وجود هذه الحدود، فإن تصميم انقطاع الانحدار هي طريقة تقييم أثر فعالة لتوليد تقديرات غير متحيزة لأثر برنامج بالقرب من انقطاع الأهلية. يستفيد تصميم انقطاع الانحدار من قواعد تعيين البرنامج، واستخدام مؤشرات الأهلية المستمرة، الشائعة بالفعل في العديد من البرامج الاجتماعية. عند تطبيق قواعد التوجيه على أساس المؤشر، لا يلزم استبعاد مجموعة من الأسر أو الأفراد المؤهلين من تلقي المعالجة من أجل التقييم لأنه يمكن استخدام تصميم انقطاع الانحدار بدلاً من ذلك.

قائمة المراجعة: تصميم انقطاع الانحدار

يتطلب تصميم انقطاع الانحدار أن يكون مؤشر الأهلية متواصلًا حول درجة الانقطاع وأن تكون الوحدات متشابهة في المنطقة المجاورة أعلى وأدنى درجة الانقطاع.

- ✓ هل المؤشر مستمر حول درجة الانقطاع في وقت خط الأساس؟
- ✓ هل هناك أي أدلة لعدم الامتثال مع القاعدة التي تحدد الأهلية للمعالجة؟ اختبر مما إذا كانت جميع الوحدات المؤهلة قد تلقت المعالجة ومن أنه لم تتلق أي وحدات غير مؤهلة المعالجة. إذا وجدت عدم توافق، فستحتاج إلى دمج تصميم انقطاع الانحدار مع منهجية متغير مساعد لتصحيح هذا "الانقطاع الضبابي".^٤
- ✓ هل ثمة أي أدلة عن أن درجات المؤشر قد تم التلاعب بها للتأثير على المؤهلين للبرنامج؟ اختبر ما إذا كان توزيع درجة المؤشر سلسًا عند نقطة الانقطاع. إذا وجدت أدلة عن "تجمّع" درجات المؤشرات إما فوق نقطة الانقطاع أو تحتها، فقد يشير ذلك إلى التلاعب.
- ✓ هل الانقطاع فريد بالنسبة إلى البرنامج الخاضع للتقييم، أم أن برامج أخرى تستخدم الانقطاع أيضًا؟

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- لمزيد من المعلومات حول تقييم برنامج حوالات نقدية باستخدام تصميم انقطاع الانحدار، راجع منشور المدونة على مدونة أثر التنمية للبنك الدولي (<http://blogs.worldbank.org/impactevaluations/>).
- لمراجعة المشاكل العملية في تنفيذ تصميم انقطاع الانحدار، انظر Guido Imbens وThomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

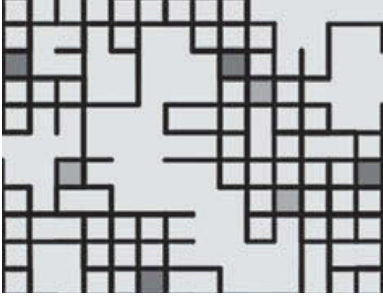
ملاحظات

١. يُسمى ذلك أحيانًا اختبار المتوسطات الممثلة.
٢. يُشار إلى مؤشر الأهلية المستمر أحيانًا بأنه متغير الفرض.
٣. ملاحظة تقنية: قُدّرت الكثافة باستخدام طريقة نواة Epanechnikov أحادية المتغير.
٤. وفي هذه الحالة، يمكنك استخدام الموقع إلى يسار نقطة الانقطاع أو يمينها بمثابة متغير مساعد للاشتراك الفعلي في البرنامج في المرحلة الأولى من تقدير المربعات الصغرى على مرحلتين.

المراجع

- Barrera-Osorio, Felipe, Leigh Linden, and Miguel Urquiola. 2007. "The Effects of User Fee Reductions on Enrollment: Evidence from a Quasi-Experiment." Columbia University and World Bank, Washington, DC.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101: 1739–74.
- Imbens, Guido, and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.
- Levy, Dan, and Jim Ohls. 2010. "Evaluation of Jamaica's PATH Conditional Cash Transfer Programme." *Journal of Development Effectiveness* 2 (4): 421–41.





الاختلاف في الاختلافات

تقييم البرنامج عندما تكون قاعدة التعيين أقل وضوحًا

تنتج الطرق الثلاث لتقييم الأثر التي نوقشت حتى الآن - التعيين العشوائي، والمتغيرات المساعدة (IV)، وتصميم انقطاع الانحدار (RDD) - جميعًا تقديرات للواقع المضاد من خلال قواعد التعيين الصريحة للبرنامج التي يعرفها فريق التقييم ويفهمها. لقد ناقشنا السبب في أن هذه الطرق توفر تقديرات ذات مصداقية للواقع المضاد من خلال افتراضات وشروط قليلة نسبيًا. يوفر النوعان التاليان من الطرق - الاختلاف في الاختلافات (DD) وطرق المطابقة - لفريق التقييم مجموعة إضافية من الأدوات التي يمكن تطبيقها عندما تكون قواعد تعيين البرنامج أقل وضوحًا أو عندما لا تكون أي من الطرق الثلاث الموضحة سابقًا ذات جدوى. من الشائع استخدام الاختلاف في الاختلافات والمطابقة في هذه الحالة، ولكن كليهما يتطلب عادةً افتراضات أقوى من طرق التعيين العشوائي، أو المتغير المساعد، أو تصميم درجة الانحدار. بديهيًا، إذا لم نكن نعرف قاعدة تعيين البرنامج، فلدينا مجهول إضافي في تقييمنا، الذي علينا وضع افتراضات عنه. وبما أن الافتراضات التي نضعها ليست صحيحة بالضرورة، فقد لا يوفر الاختلاف في الاختلافات أو المطابقة تقديرات موثوقة آثار البرنامج.

المفهوم الأساسي

يقارن الاختلاف في
الاختلافات التغيرات في
الحصائل بمرور الزمن
بين الوحدات المشتركة
في برنامج ما (مجموعة
المعالجة) والوحدات
غير المشتركة (مجموعة
المقارنة). وهذا يسمح لنا
بتصحيح أي فروق ثابتة
بمرور الزمن بين مجموعتي
المعالجة والمقارنة.

تقارن طريقة الاختلاف في الاختلافات التغيرات في الحصائل بمرور الزمن بين فئة سكانية مشتركة في برنامج ما (مجموعة المعالجة) وفئة سكانية غير مشتركة (مجموعة المقارنة). لتأخذ مثلاً برنامج إصلاح الطرقات الذي يُنفَّذ على مستوى المنطقة ولكن لا يمكن تعيينه عشوائياً بين المقاطعات، والذي لم يُعَيَّن أيضاً على أساس مؤشر بانقطاع محدد بوضوح يسمح بتصميم انقطاع الانحدار. بإمكان مجالس المنطقة اختيار الاشتراك في البرنامج أو عدم الاشتراك فيه. أحد أهداف البرنامج تحسين وصول السكان إلى سوق العمل، وأحد مؤشرات الحصيلة معدّل التوظيف. كما نوقش في الفصل ٣، فإن مجرد ملاحظة التغيّر قبل وبعد في معدلات التوظيف للمناطق التي تشترك في البرنامج لن ترصد الأثر السببي للبرنامج، وذلك لأن العديد من العوامل الأخرى قد تؤثر أيضاً على التوظيف بمرور الوقت. وفي الوقت نفسه، ستكون مقارنة المناطق التي اشتركت والتي لم تشترك في برنامج إصلاح الطرق إشكالية إذا كان ثمة أسباب غير ملحوظة لاشتراك بعض المناطق في البرنامج وعدم اشتراك مناطق أخرى (إشكالية تحيز الاختيار التي نوقشت في سيناريو المشتركين وغير المشتركين).

ولكن، ماذا لو دمجنا الطريقتين وقارنا التغيرات قبل وبعد في الحصائل لمجموعة مشتركة في البرنامج مع التغيرات قبل وبعد لمجموعة لم تشترك في البرنامج؟ يسيطر الفرق في الحصائل قبل وبعد بالنسبة إلى المجموعة المشتركة—الفرق الأول—على العوامل الثابتة بمرور الزمن في تلك المجموعة، بما أننا نقارن المجموعة بنفسها. يتبقى لدينا رغم ذلك عوامل تتغير بمرور الزمن (عوامل متغيرة بمرور الزمن) لهذه المجموعة. إحدى طرق رصد هذه العوامل المتغيرة بمرور الزمن هي قياس التغيّر قبل وبعد في الحصائل لمجموعة لم تشترك في البرنامج ولكنها تعرضت لمجموعة الشروط البيئية نفسها—الفرق الثاني. إذا "استبعدنا" الفرق الأول للعوامل الأخرى المتغيرة بمرور الزمن، التي تؤثر على الحصيلة موضع الاهتمام بطرح الفرق الثاني، فإننا بذلك نكون قد أزلنا مصدر التحيز الذي كان يقلقنا في المقارنات البسيطة قبل وبعد. تؤدي مقارنة الاختلاف في الاختلافات دوراً يتوافق مع اسمه. فهي تجمع بين تقديرين زائفين للواقع المضاد (المقارنات قبل وبعد، والمقارنات بين أولئك الذين يختارون الاشتراك وأولئك الذين يختارون عدم الاشتراك) لإنتاج تقدير أفضل للواقع المضاد. في مثال برنامج إصلاح الطرقات، قد تقارن طريقة الاختلاف في الاختلافات التغيرات في التوظيف قبل تنفيذ البرنامج وبعده للأفراد المشتركين في البرنامج مع التغيرات في التوظيف في المناطق التي لم تشترك في البرنامج.

من المهم التنويه إلى أن ما نقدّره هو الواقع المضاد للتغيّر في الحاصلات لمجموعة المعالجة: تقديرنا لهذا الواقع المضاد هو التغيّر في الحاصلات لمجموعة المقارنة. لا حاجة إلى أن تكون لمجموعتي المعالجة والمقارنة الشروط نفسها قبل التدخل. ولكن، لكي يكون الاختلاف في الاختلافات صالحًا، يجب أن تمثل مجموعة المقارنة بدقة التغيّر في الحاصلات، الذي كانت مجموعة المعالجة ستعايشه، في غياب البرنامج. لتطبيق الاختلاف في الاختلافات، من الضروري قياس الحاصلات في المجموعة التي تتلقّى البرنامج (مجموعة المعالجة) والمجموعة التي لا تتلقّى البرنامج (مجموعة المقارنة)، قبل البرنامج وبعده. نقدّم في المربع ٧-١ مثالاً استُخدمت فيه طريقة الاختلاف في الاختلافات لفهم أثر المبادرات الانتخابية على تنفيذ برنامج حوالات نقدية في البرازيل استنادًا إلى معدلات التسرب من المدارس. يوضح الشكل ٧-١ طريقة الاختلاف في الاختلافات لمثال إصلاح الطرقات. العام ٠ هو عام خط الأساس. في العام ١، تشترك مجموعة معالجة من المناطق في البرنامج، بينما لا تشترك مجموعة المقارنة.

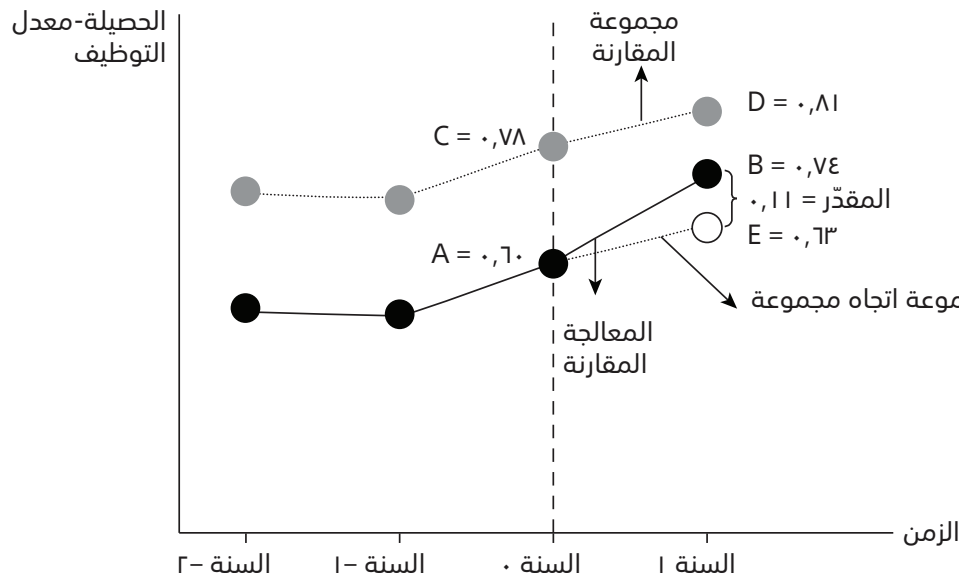
المربع ٧-١: استخدام الاختلاف في الاختلافات لفهم تأثير الحوافز الانتخابية على معدلات التسرب من المدارس في البرازيل

من المدارس في البلديات مع رؤساء البلديات في الفترة الأولى مقابل الفترة الثانية. وقد افترضوا أنه بما أن البرازيل كانت تعتمد فترتين بالحد الأقصى للسياسيين المحليين، فإن رؤساء البلديات كانت تشغلهم إعادة الانتخاب، ولذلك كانوا يتصرفون على نحو مختلف من رؤساء البلديات في الفترة الثانية، الذين لم يكن ذلك من شواغلهم. عمومًا، نجح البرنامج في خفض معدلات التسرب من المدرسة بمعدل ٨ في المائة للمستفيدين. توّصل الباحثون إلى أن أثر البرنامج كان أكبر بـ ٣٦ بالمائة في البلديات التي كان فيها رؤساء بلديات في فترة الولاية الأولى. وقد خلصوا إلى أن شواغل إعادة الانتخاب حفزت السياسيين المحليين على زيادة جهودهم في تنفيذ برنامج Bolsa Escola.

في دراسة تجريبية للمبادرات الانتخابية المحلية، تناول دي جانفري، وفينان، وسادوليت (٢٠١١) آثار التحويل النقدي المشروط (CCT) في البرازيل. كان برنامج Bolsa Escola يمنح الأمهات في الأسر الفقيرة راتبًا شهريًا شريطة التحاق أطفالهن في المدرسة. كان برنامج CCT برنامجًا فيدراليًا شبيهًا ببرنامج Oportunidades في المكسيك (انظر المربعين ١-١ و ٢-٤)، لكنه كان لامركزيًا على مستوى البلديات. كانت حكومات البلديات مسؤولة عن تحديد المستفيدين وتنفيذ البرنامج. قدّر دي جانفري، وفينان، وسادوليت، باستخدام طريقة الاختلاف في الاختلافات أثر البرنامج على معدلات التسرب من المدارس. وتوصلوا إلى تباين ملحوظ في أداء البرنامج عبر البلديات. لاستكشاف هذا التباين، قارن الباحثون التحسن في معدلات التسرب

المصدر: De Janvry, Finan, and Sadoulet 2011.

الشكل ١-٧ طريقة الاختلاف في الاختلافات



ملاحظة: يجب قراءة جميع الاختلافات بين النقاط بصفاتها فروقاً رأسية في الحاصل على المحور الرأسي.

ينتقل مستوى الحصيلة (معدل التوظيف) لمجموعة المعالجة من A ، قبل بدء البرنامج، إلى B بعد بدء البرنامج، في حين تنتقل الحصيلة لمجموعة المقارنة من C ، قبل بدء البرنامج، إلى D ، بعد بدء البرنامج.

ستتذكر تقيمي واقعنا المضاد: الفرق في الحاصل قبل التدخل وبعده لمجموعة المعالجة ($B - A$) والفرق في الحاصل بعد التدخل بين مجموعتي المعالجة والمقارنة ($B - D$). في الاختلاف في الاختلافات، يُتَحَصَّل على تقدير الواقع المضاد بحساب التغير في الحاصل لمجموعة المقارنة ($D - C$)، ثم طرحه من التغير في الحاصل لمجموعة المعالجة ($B - A$). إن استخدام التغير في الحاصل لمجموعة المقارنة بصفته تقدير الواقع المضاد للتغير في الحاصل بالنسبة إلى مجموعة المعالجة يشبه افتراض أنه لو لم تشارك المجموعة المشتركة في البرنامج، فإن حصيلتها كانت ستتطور بمرور الزمن على طول اتجاه المجموعة غير المشتركة نفسه: أي أن التغير في الحصيلة للمجموعة المشتركة كان سيكون من A إلى E ، كما هو موضح في الشكل ١-٧.

بإيجاز، يُحَسَّب أثر البرنامج ببساطة على أنه الفرق بين الفرقين:

$$\text{الاختلاف في الاختلافات} = (D - C) - (B - A) = (0.81 - 0.78) - (0.74 - 0.60) = 0.11$$

يمكن تقديم العلاقات المعروضة في الشكل ١-٧ أيضًا في جدول بسيط. يفصل الجدول ١-٧ مكونات تقديرات الاختلاف في الاختلافات. يضم الصف الأول حواصل مجموعة المعالجة قبل التدخّل (A) وبعد التدخّل (B). الفرق قبل وبعد لمجموعة المقارنة هو الفرق الأول (B - A). يضم الصف الثاني حواصل مجموعة المقارنة قبل التدخل (C) وبعد التدخل (D)، ومن ثم فإن الفرق الثاني هو (D - C). تحسب طريقة الاختلاف في الاختلافات تقدير الأثر كما يلي:

١. نحسب الفرق في الحصلة (Y) بين الموقفين السابق واللاحق لمجموعة المعالجة (B - A).

٢. نحسب الفرق في الحصلة (Y) بين الموقفين السابق واللاحق لمجموعة المقارنة (D - C).

٣. ثم نحسب الفرق بين الفرق في الحاصل لمجموعة المعالجة (B - A) والفرق لمجموعة المقارنة (D - C)، أو الاختلاف في الاختلافات $DD = (B - A) - (D - C)$. الاختلاف في الاختلافات هو تقدير الأثر لدينا.

يمكننا أيضًا حساب الاختلاف في الاختلافات بالعكس: نحسب أولاً الفرق في الحصلة بين مجموعة المعالجة ومجموعة المقارنة في الموقف اللاحق، ثم نحسب الفرق في الحصلة بين مجموعة المعالجة ومجموعة المقارنة في الموقف السابق، ونطرح الثاني من الأول.

$$\text{الاختلاف في الاختلافات} = (A - C) - (B - D) = DD = (0,78 - 0,60) - (0,81 - 0,74) = 0,11$$

الجدول ١-٧ طريقة حساب الاختلاف في الاختلافات

الفرق	قبل	بعد	
B - A	A	B	المعالجة/غير المشتركة
D - C	C	D	المقارنة/غير المشتركة
DD = (B - A) - (D - C)	A - C	B - D	

الفرق	قبل	بعد	
0,14	0,60	0,74	المعالجة/غير المشتركة
0,03	0,78	0,81	المقارنة/غير المشتركة
0,11 = 0,03 - 0,14 = DD	0,18-	0,07-	الفرق

كيف تكون طريقة الاختلاف في الاختلافات مفيدة؟

لفهم مدى فائدة الاختلاف في الاختلافات، لنبدأ بتقديرنا الزائف الثاني للواقع المضاد، الذي نوقش الفصل ٣، والذي يقارن وحدات مشتركة في برنامج مع وحدات غير مشتركة فيه. تذكر أن مصدر القلق الأساسي لهذه المقارنة كان أن مجموعتي الوحدات قد تكون لهما خصائص مختلفة، وأن هذه الخصائص - وليس البرنامج - هي التي تفسر الفرق في الحاصل بين المجموعتين. وكانت الاختلافات غير الملحوظة في الخصائص مثيرة للقلق بشكل خاص. وفق التعريف، من المستحيل أن تشمل الخصائص غير الملحوظة في التحليل.

تساعد طريقة الاختلاف في الاختلافات على حل هذه المشكلة إلى الحد الذي يمكن فيه افتراض أن العديد من خصائص الوحدات أو الأفراد ثابتة بمرور الزمن (أو غير متغيرة بمرور الزمن). ففكر، على سبيل المثال في الخصائص الملحوظة، مثل سنة ميلاد شخص، أو موقع منطقة بالقرب من المحيط، أو مناخ بلدة، أو مستوى تعليم والد. ومن المحتمل أن معظم هذه الأنواع من المتغيرات، على الرغم من ارتباطها بشكل معقول بالحاصل، لن تتغير على مدار التقييم. باستخدام منطق التفكير نفسه، قد نستنتج أن الكثير من الخصائص غير المرصودة للأفراد ثابتة تقريباً بمرور الزمن. ففكر على سبيل المثال في السمات الشخصية أو تاريخ الأسرة الصحي. قد يكون من المعقول ألا تتغير هذه الخصائص الجوهرية للأشخاص بمرور الزمن.

وبدلاً من مقارنة الحاصل بين مجموعتي المعالجة والمقارنة بعد التدخّل، تقارن طريقة الاختلاف في الاختلافات الاتجاهات بين مجموعتي المعالجة والمقارنة. الاتجاه بالنسبة إلى الفرد هو الفرق في حاصل ذلك الفرد قبل البرنامج وبعده. بطرح حالة الحاصل قبل من حالة الحاصل بعد، يمكننا إلغاء أثر كل الخصائص الفريدة بالنسبة للأفراد والتي لا تتغير بمرور الزمن. ومما يثير الاهتمام أننا لا نلغي (أو نسيطر على) أثر الخصائص الملحوظة غير المتغيرة بالنسبة إلى الزمن، بل أيضاً أثر الخصائص غير الملحوظة وغير المتغيرة بالنسبة إلى الزمن، مثل تلك المذكورة. يصف المربع ٧-٢ دراسة استخدمت طريقة الاختلاف في الاختلافات لتقدير أثر زيادة وجود الشرطة على حوادث سرقة السيارات في بوينس آيرس.

المفهوم الأساسي

بدلاً من مقارنة الحاصل بين مجموعتي المعالجة والمقارنة بعد التدخّل، تقارن طريقة الاختلاف في الاختلافات الاتجاهات بين مجموعتي المعالجة والمقارنة.

المربع ٧-٢: استخدام الاختلاف في الاختلافات لدراسة تأثيرات نشر أفراد الشرطة لمكافحة الجريمة في الأرجنتين

الجريمة. على النقيض من ذلك، لم تكن الزيادة في نشر قوى الشرطة في الأرجنتين ترتبط إطلاقًا بحدوث سرقات السيارات، ومن ثم فإن الدراسة لا تعاني من مشكلة السببية المتزامنة. تمكّن شارغروودسكي وديتيلا من استخدام طريقة الاختلاف في الاختلافات لتقدير أثر زيادة وجود الشرطة على حدوث سرقة السيارات.

وقد كشفت النتائج عن أثر رادع إيجابي لوجود الشرطة على حدوث الجرائم، بيد أن هذا الأثر كان مكانيًا. في المجمعات التي كانت تضم مباني لليهود تحصل على الحماية من الشرطة، انخفضت سرقات السيارات انخفاضًا كبيرًا بالمقارنة مع المجمعات الأخرى: بنسبة ٧٥ بالمائة. ولم يتوصل الباحثون إلى أي آثار على سرقات السيارات على بعد مجمع أو مجمعين من المباني المحمية.

درس ديتيلا (٢٠٠٥) ما إذا كانت الزيادة في تعداد قوى الشرطة قد قللت من الجرائم في الأرجنتين. ففي عام ١٩٩٤، دفع هجوم إرهابي على مركز يهودي كبير في بوينس آيرس الحكومة الأرجنتينية إلى زيادة حماية الشرطة للمجمعات التابعة لليهود في البلاد.

وسعيًا إلى فهم أثر وجود الشرطة على حدوث الجرائم، جمع شارغروودسكي وديتيلا بيانات بشأن عدد سرقات السيارات في كل مبنى ضخم في ثلاثة أحياء في بوينس آيرس قبل الهجوم الإرهابي وبعده. ثم جمعا هذه المعلومات مع البيانات الجغرافية لمواقع المؤسسات اليهودية في الأحياء. طرحت هذه الدراسة مقارنة مختلفة عن اندحارات الجرائم النموذجية. غالبًا ما تواجه الدراسات التي تُعنى بتأثير حفظ الأمن إشكالية داخلية النشأة، ذلك أن الحكومات تميل إلى زيادة وجود الشرطة في المناطق التي ترتفع فيها معدلات

المصدر: DiTella and Schargrodsky 2005.

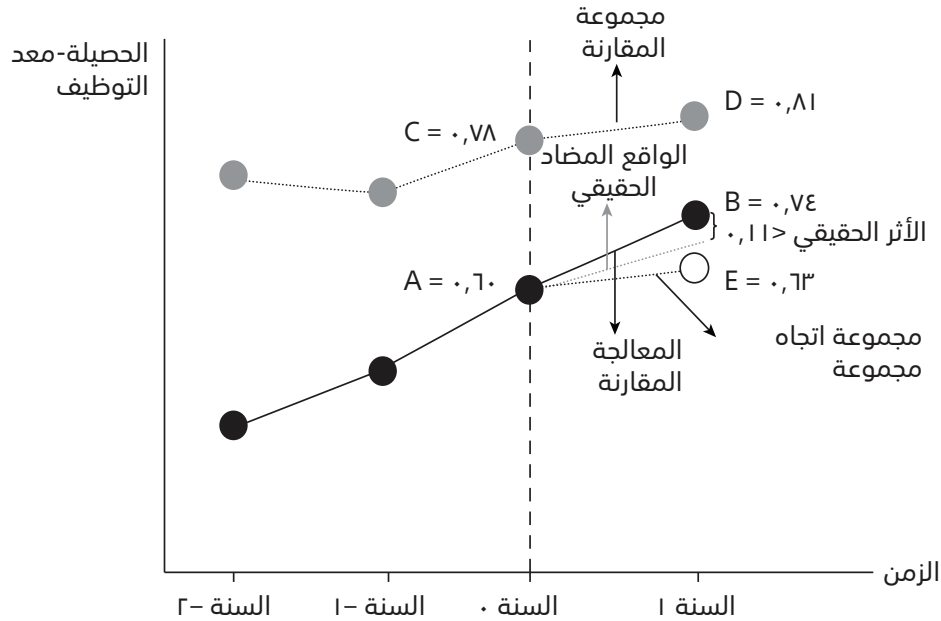
افتراض "الاتجاهات المتساوية" في الاختلاف في الاختلافات

على الرغم من أن الاختلاف في الاختلافات يسمح لنا بالتعامل مع الاختلافات الثابتة بالنسبة إلى الزمن بين مجموعتي المعالجة والمقارنة، إلا أنه لن يساعدنا في القضاء على الاختلافات المتغيرة بالنسبة إلى الزمن بين مجموعتي المعالجة والمقارنة. في مثال برنامج إصلاح الطرقات، إذا كانت مناطق المعالجة تستفيد أيضًا من إنشاء ميناء بحري جديد في وقت إصلاح الطرقات نفسه، فلن نتضمن من فصل أثر إصلاح الطرقات عن أثر إنشاء الميناء البحري باستخدام مقارنة الاختلاف في الاختلافات. ولكي توفر الطريقة تقديرًا صالحًا للواقع المضاد، يجب علينا افتراض عدم وجود فروقات متغيرة بالنسبة إلى الزمن من هذا القبيل بين مجموعتي المعالجة والمقارنة.

وثمة طريقة أخرى للتفكير في ذلك، ألا وهي أنه في غياب البرنامج، يتعين أن تسير الاختلافات في الحاصل بين مجموعتي المعالجة والمقارنة بالتوازي. أي أنه في غياب المعالجة، ينبغي أن تزداد الحاصل أو تنقص بالمعدل نفسه في كلتا المجموعتين؛ نطلب أن تُبدي الحاصل اتجاهات متساوية في غياب المعالجة. وبطبيعة الحال، ليس ثمة طريقة كي نثبت أن الاختلافات بين مجموعتي المعالجة والمقارنة كانت ستسير على التوازي في غياب البرنامج. السبب في ذلك هو أنه لا يمكننا ملاحظة ما كان سيحدث لمجموعة المقارنة في غياب البرنامج—بمعنى آخر، لا يمكننا ملاحظة الواقع المضاد.

وهكذا، عندما نستخدم طريقة الاختلاف في الاختلافات، علينا أن نفترض أنه في غياب البرنامج، كانت الحصيلة في مجموعة المعالجة ستسير بالتوازي مع الحصيلة في مجموعة المقارنة. يوضح الشكل ٧-٢ انتهاكاً للافتراض الأساسي. إذا كان اتجاه الحاصلين مختلفين لمجموعتي المعالجة والمقارنة، فإن أثر المعالجة المقدّر والمستحصل بطرق الاختلاف في الاختلافات لن يكون صالحاً، أو سيكون متحيزاً. يُعزى ذلك إلى أن اتجاه مجموعة المقارنة ليس تقديرًا صحيحًا لاتجاه الواقع المضاد الذي كان من الممكن أن يسود لمجموعة المعالجة في غياب البرنامج. وكما يوضح الشكل ٧-٢، إذا كانت الحاصلات الواقعية لمجموعة المقارنة تنمو بسرعة أبطأ من حاصلات مجموعة المعالجة في غياب البرنامج، فإن استخدام الاتجاه لمجموعة المقارنة بمثابة تقدير للواقع المضاد لمجموعة المعالجة يؤدي إلى تقدير متحيز لأثر البرنامج؛ وبتحديد أكبر، سنبالغ في تقدير أثر البرنامج.

الشكل ٧-٢ الاختلاف في الاختلافات عند اختلاف اتجاهات الحاصلات



اختبار صلاحية افتراضات "تساوي الاتجاهات" في الاختلاف في الاختلافات

يمكن تقييم افتراض تساوي الاتجاهات الكامن، على الرغم من أنه لا يمكن إثباته. اختبار الصلاحية الأول هو مقارنة التغيرات في حصائل مجموعتي المعالجة والمقارنة بصورة متكررة قبل تنفيذ البرنامج. يعني هذا، في برنامج إصلاح الطرقات، أننا سنقارن التغير في معدلات التوظيف بين مجموعتي المعالجة والمقارنة قبل بدء البرنامج: أي بين العام ٢ والعام ١، وبين العام ١ والعام ٠. إذا كانت الحصائل تسير بالتوازي قبل بدء البرنامج، فإننا نتحقق من أن الحصائل كانت ستواصل السير بالتوازي بعد التدخل. للتحقق من المساواة في اتجاهات ما قبل التدخل، نحتاج إلى عمليتي ملاحظة متسلسلتين اثنتين على الأقل لمجموعتي المعالجة والمقارنة قبل بدء البرنامج. وهذا يعني أن التقييم سيتطلب ثلاث عمليات ملاحظة متسلسلة: عمليتي ملاحظة قبل التدخل لتقييم اتجاهات ما قبل البرنامج، وعملية ملاحظة واحدة على الأقل بعد التدخل لتقييم الأثر باستخدام طريقة الاختلاف في الاختلافات.

ثمة طريقة أخرى لاختبار افتراض الاتجاهات المتساوية، ألا وهي إجراء ما يُسمى الاختبار الوهمي. بالنسبة إلى هذا الاختبار، تجري تقديرًا إضافيًا للاختلاف في الاختلافات باستخدام مجموعة معالجة "وهمية": أي مجموعة تعرف أنها لم تتأثر بالبرنامج. افترض، على سبيل المثال، أنك تقدر كيف تؤثر الدروس الخصوصية الإضافية لطلاب الصف السابع على احتمال ذهابهم إلى المدرسة، وأنت تختار طلاب الصف الثامن بمثابة مجموعة مقارنة. لاختبار ما إذا كان لدى الطلاب في الصفين السابع والثامن الاتجاهات نفسها في ما يتعلق بالالتحاق بالمدرسة، يمكنك اختبار ما إذا كان لدى الطلاب في الصفين الثامن والسادس الاتجاهات نفسها. تعلم أن طلاب الصف السادس لا يتأثرون بالبرنامج، لذلك إذا أجريت تقدير الاختلاف في الاختلافات باستخدام طلاب الصف الثامن بمثابة مجموعة مقارنة وطلاب الصف السادس بمثابة مجموعة معالجة وهمية، فعليك إيجاد الأثر الصفري. وإذا لم تفعل ذلك، فإن الأثر الذي تعثر عليه لا بد أن يأتي من بعض الاختلافات الكامنة في الاتجاهات بين طلاب الصف السادس وطلاب الصف الثامن. وهذا بدوره يثير الشك بشأن ما إذا كان يمكن افتراض أن طلاب الصف السابع والصف الثامن لديهم اتجاهات متساوية في غياب البرنامج.

ثمة طريقة ثالثة لاختبار افتراض الاتجاهات المتساوية، ألا وهي إجراء اختبار وهمي ليس من خلال مجموعة معالجة وهمية فحسب، بل من خلال حصيلة وهمية أيضًا. في مثال التعليم، قد تحتاج إلى اختبار صلاحية استخدام طلاب الصف الثامن بمثابة مجموعة مقارنة عبر تقدير أثر التعليم على حصيلة تعرف أنها لا تتأثر به، مثل عدد الأشقاء لدى الطلاب. إذا توصلت من خلال تقدير الاختلاف في الاختلافات إلى أثر للتعليم على عدد الأشقاء لدى هؤلاء الطلاب، فإنك تعلم حينها أن مجموعة المقارنة لديك معيبة.

ثمة طريقة رابعة لاختبار افتراض الاتجاهات المتساوية، ألا وهي إجراء تقدير الاختلاف في الاختلافات باستخدام مجموعات المقارنة المختلفة. في مثال اليوم، ستجري أولاً التقدير باستخدام طلاب الصف الثامن بمثابة مجموعة المقارنة، ثم ستجري تقديرًا آخر باستخدام طلاب الصف السادس بمثابة مجموعة المقارنة. إذا كانت كلتا المجموعتين مجموعة مقارنة صالحة، فستجد أن الأثر المقدر متماثل تقريبًا في كلتا عمليتي الحساب. نقدم في المربعين ٣-٧ و ٤-٧ مثالين لتقييم الاختلاف في الاختلافات الذي استخدم مزيجًا من هذه الطرق لاختبار افتراض الاتجاهات المتساوية.

المربع ٣-٧: اختبار افتراض الاتجاهات المتعادلة: خصصة المياه ووفيات الأطفال في الأرجنتين

أو المستويات التاريخية لوفيات الأطفال. وقد تحققوا من قوة نتائجهم من خلال إجراء اختبار وهمي باستخدام حصيلة وهمية: فقد ميزوا بين أسباب وفيات الأطفال المرتبطة بظروف المياه، مثل الأمراض المعدية والطفيلية، وبين تلك التي لا علاقة لها بظروف المياه مثل الحوادث والأمراض الخلقية. ثم اختبروا أثر خصصة خدمات المياه بشكل منفصل للمجموعتين الفرعيتين من أسباب الوفيات. وتوصلوا إلى أن خصصة خدمات المياه كانت مرتبطة بانخفاض الوفيات الناجمة عن الأمراض المعدية والطفيلية، ولكنها لا ترتبط بانخفاض الوفيات الناجمة عن أسباب مثل الحوادث والأمراض الخلقية. في النهاية، حدّد التقييم أن وفيات الأطفال انخفضت بنسبة ٨ بالمائة تقريبًا في المناطق التي طبقت الخصصة، وأن الأثر بلغ ذروته، حوالي ٢٦ بالمائة في المناطق الأفقر، حيث كان توسّع شبكة المياه هو الأكبر. سلّطت هذه الدراسة الضوء على عدد من مناقشات السياسات حول خصصة الخدمات العامة. استنتج الباحثون أن القطاع الخاص المنظم في الأرجنتين أثبت نجاحًا أكبر من القطاع العام في تحسين مؤشرات الوصول، والخدمة، والأهم من ذلك، ووفيات الأطفال.

استخدم شارغودسكي، وغيرتler، وغالياني (٢٠٠٥) طريقة الاختلاف في الاختلافات للإجابة عن سؤال سياسة مهم، وهو: هل تحسّن خصصة خدمات توفير المياه الحوائل الصحية وتساعد على التخفيف من حدة الفقر؟ خلال التسعينيات، شرعت الأرجنتين في إحدى حملات الخصصة الكبرى، حيث حولت شركات المياه المحلية إلى شركات خاصة منظمة. استغرقت عملية الخصصة أكثر من عقد، وبلغ عدد عمليات الخصصة ذروته بعد عام ١٩٩٥، لتشمل في النهاية حوالي ٣٠ بالمائة من بلديات البلاد و٦٠ بالمائة من السكان. استفاد التقييم من التباين في وضع الملكية بمرور الوقت لتحديد أثر الخصصة على معدل وفيات الأطفال دون سن الخامسة.

قبل عام ١٩٩٥، كانت معدلات وفيات الأطفال تتراجع بالتدريج نفسها في جميع أنحاء الأرجنتين. وبعد عام ١٩٩٥، تراجعت معدلات الوفيات بسرعة أكبر في البلديات التي خصصت خدمات المياه.

قال الباحثون إنه في هذا السياق، من المرجح أن يكون افتراض الاتجاهات المتساوية الذي يقوم عليه الاختلاف في الاختلافات صحيحًا. وبالتحديد، أوضحوا أنه لم تُلحَظ فروق في اتجاهات وفيات الأطفال بين بلديات المقارنة والمعالجة قبل بدء حركة التخصيص. وأوضحوا أيضًا أن قرار الخصصة لم يكن مرتبطًا بالصدمات الاقتصادية

المصدر: Galiani, Gertler, and Schargrodsky 2005.

المربع ٧-٤: اختبار افتراض الاتجاهات المتعادلة: تشييد المدارس في إندونيسيا

حللت دوفلو (٢٠٠١) آثار برنامج تشييد مدارس في إندونيسيا على المدينين المتوسط والطويل على حائل التعليم وسوق العمل. شرعت إندونيسيا عام ١٩٧٣ في برنامج إنشاء مدارس ابتدائية وشيدت أكثر من ٦١ ألف مدرسة ابتدائية. لاستهداف الطلاب الذين لم يسبق لهم الالتحاق بالمدرسة، عينت الحكومة عدد المدارس التي ستُشيد في كل منطقة بما يتناسب مع عدد الطلاب غير الملتحقين في المنطقة. سعت دوفلو إلى تقييم أثر البرنامج على التحصيل التعليمي والأجور. قيس التعرض للمعالجة من خلال عدد المدارس في المنطقة، وُحددت مجموعتا المعالجة والمقارنة حسب العمر عند إطلاق البرنامج. كانت مجموعة المعالجة تتكون من رجال ولدوا بعد عام ١٩٦٢، لأنهم سيكونون صغارًا في السن بما يكفي للاستفادة من المدارس الابتدائية الجديدة التي أنشئت عام ١٩٧٤. كانت مجموعة المعالجة تتكون من رجال ولدوا قبل عام ١٩٦٢، الذين سيكونون كبارًا في العمر بما يحول دون استفادتهم من البرنامج.

استخدمت دوفلو طريقة الاختلاف في الاختلافات لتقدير أثر البرنامج على متوسط التحصيل التعليمي والأجور، بمقارنة الاختلافات في الحاصل بين المناطق ذات التعرض المرتفع والمنخفض. لتوضيح أن هذه كانت طريقة تقييم صالحة، احتاجت أولاً إلى اختبار افتراض الاتجاهات المتساوية عبر

المناطق. لاختبار ذلك، استخدمت دوفلو اختبارًا وهميًا باستخدام مجموعة معالجة وهمية. وقارنت المجموعة التي تتراوح أعمارها بين ١٨ و٢٤ عامًا عام ١٩٧٤ مع المجموعة التي تتراوح أعمارها بين ١٢ و١٧ عامًا. وبما أن أفراد كلتا المجموعتين كانوا كبارًا في السن بما يحول دون استفادتهم من البرنامج الجديد، ينبغي ألا تكون التغيرات في تحصيلهم العلمي مختلفة منهجيًا عبر المناطق. كان التقدير من انحدار الاختلاف في الاختلافات هذا قريبًا من الصفر. وقد أوضحت هذه النتيجة إلى أن التحصيل التعليمي لم يكن يزداد بسرعة أكبر قبل بدء البرنامج في المناطق التي ستصبح في نهاية المطاف مناطق عالية التعرض مقارنة بالمناطق منخفضة التعرض. أوضح الاختبار الوهمي أيضًا أن استراتيجية التحديد التي تعتمد على العمر في وقت إنشاء المدرسة ستفي بالغرض.

توصل التقييم إلى نتائج إيجابية على التحصيل الدراسي والأجور للطلاب الذين كان لديهم تعرض عالٍ للبرنامج، أي أولئك الذين كانوا دون سن الثامنة حين شُيِّدت المدارس. بالنسبة إلى هؤلاء الطلاب، كانت كل مدرسة جديدة تُنشأ لكل ١٠٠٠ طفل ترتبط بزيادة قدرها ٠,١٢ إلى ٠,١٩ سنة في التحصيل العلمي وزيادة بنسبة ٣,٠ بالمائة إلى ٥,٤ بالمائة في الأجور. وقد رفع البرنامج أيضًا من احتمال إكمال الطفل للمرحلة الابتدائية بمقدار ١٢ بالمائة.

المصدر: Duflo 2001.



تقييم أثر برنامج معونة التأمين الصحي (HISP): استخدام الاختلاف في الاختلافات

يمكن استخدام الاختلاف في الاختلافات لتقييم برنامج معونة التأمين الصحي (HISP). في هذا السيناريو، لديك جولتان من البيانات حول مجموعتين من الأسر: مجموعة اشتركت في البرنامج، وأخرى لم تشترك. بتذكر حالة المجموعتين المشتركة وغير المشتركة، تدرك أنه لا يمكنك ببساطة مقارنة متوسط النفقات الصحية للمجموعتين بسبب تحيز الاختيار. ونظرًا إلى أن لديك بيانات على فترتين لكل أسرة في العينة، يمكنك استخدام هذه البيانات لحل بعض هذه التحديات من خلال مقارنة التغير في النفقات الصحية للمجموعتين، بافتراض أن التغير في النفقات الصحية للمجموعة غير المشتركة يؤثر على ما كان سيحدث لنفقات المجموعة المشتركة في غياب البرنامج (انظر الجدول ٧-٢). لاحظ أنه لا تهم طريقة حسابك للفرق المزدوج.

بعد ذلك، تقدّر التأثير باستخدام تحليل الانحدار (الجدول ٧-٣). باستخدام انحدار خطي بسيط لحساب تقدير الاختلاف في الاختلافات البسيط، تجد أن البرنامج قلّل من النفقات الصحية للأسر بمقدار ٨,١٦ دولارات. ثم تنقّح تحليلك بإضافة المزيد من متغيرات المقارنة. بمعنى آخر، تستخدم انحدارًا خطيًا متعدد المتغيرات يأخذ في الحسبان مجموعة من العوامل الأخرى، وتوجد الانخفاض نفسه في النفقات الصحية الأسرية.

الجدول ٧-٢ تقييم برنامج معونة التأمين الصحي (HISP): الاختلاف في الاختلافات مقارنة المتوسطات

الفرق	قبل (خط الأساس)	بعد (المتابعة)	
٦,٦٠-	١٤,٤٩	٧,٨٤	المشتركون
١,٥١	٢٠,٧٩	٢٢,٣٠	غير المشتركين
٨,١٦ = ١,٥١ - ٦,٦٠ = DD			الاختلاف في الاختلافات

ملاحظة: يقدم الجدول النفقات الأسرية المتوسطة (بالدولار) للأسر المشتركة وغير المشتركة، قبل إدخال برنامج HISP وبعده.

الجدول ٧-٣ تقييم برنامج معونة التأمين الصحي (HISP): الاختلاف في الاختلافات من خلال تحليل الانحدار

الاختلاف الخطي متعدد المتغيرات	الاختلاف الخطي	
٨,١٦- (٠,٣٢)	٨,١٦- (٠,٣٢)	الأثر المقدّر على النفقات الصحية الأسرية

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة.

السؤال ٦ الخاص ببرنامج HISP



- أ. ما الافتراضات الأساسية المطلوبة لقبول هذه النتيجة من الاختلاف في الاختلافات؟
ب. استنادًا إلى النتيجة من الاختلاف في الاختلافات، هل ينبغي توسيع برنامج HISP على المستوى الوطني؟

محددات طريقة الاختلاف في الاختلافات

حتى عندما تكون الاتجاهات متساوية قبل بدء التدخل ، فإن التحيز في تقدير الاختلاف في الاختلافات قد يستمر في الظهور ولا يُكتشف. يُعزى ذلك إلى أن الاختلاف في الاختلافات يُسند إلى التدخل أي اختلافات في الاتجاهات بين مجموعتي المعالجة والمقارنة التي تحدث من وقت بدء التدخل. في حال وجود أي عوامل أخرى تؤثر على الفرق في الاتجاهات بين المجموعتين ولم تؤخذ في الحسبان في الانحدار متعدد المتغيرات، فسيكون التقدير غير صالح أو متحيزًا. لنقل إنك تحاول تقدير أثر دعم الأسمدة على إنتاج الأرز، وتفعل ذلك بقياس إنتاج الأرز للمزارعين المدعومين (المعالجة) والمزارعين غير المدعومين (المقارنة) قبل توزيع الإعانات وبعده.

إذا كان هناك في السنة ١ جفاف يؤثر فقط على المزارعين المدعومين، فإن تقدير الاختلاف في الاختلافات سينتج تقديرًا غير صالح لأثر دعم الأسمدة. عمومًا، يؤثر أي عامل بشكل غير متناسب على إحدى المجموعتين، ويفعل ذلك في نفس الوقت الذي تتلقى فيه مجموعة المعالجة المعالجة - ولا يؤخذ في الحسبان في الانحدار - ولديه القدرة على إبطال أثر البرنامج أو التسبب في تحيزه. يفترض الاختلاف في الاختلافات عدم وجود عامل من هذا القبيل.

قائمة المراجعة: الاختلاف في الاختلافات

يفترض الاختلاف في الاختلافات أن اتجاهي الحاصل متشابهان في مجموعتي المقارنة والمعالجة قبل التدخل، وأن العوامل الوحيدة التي تفسر الاختلافات في الحاصل بين المجموعتين ثابتة بمرور الزمن، بغض النظر عن البرنامج نفسه.

✓ هل كانت الحاصل ستتحرك بالتوازي في مجموعتي المعالجة والمقارنة في غياب البرنامج؟ يمكن تقييم ذلك باستخدام العديد من اختبارات التزييف، مثل ما يلي: (١) هل تسير الحاصل في مجموعة المعالجة والمقارنة بالتوازي قبل التدخل؟ إذا توفرت جولتان من البيانات قبل بدء البرنامج، فاختبر لمعرفة ما إذا كان يظهر أي فرق في الاتجاهات بين المجموعتين. (٢) ماذا عن النتائج الزائفة التي ينبغي ألا تتأثر بالبرنامج؟ هل تتحرك بالتوازي قبل بدء التدخل وبعده في مجموعتي المعالجة والمقارنة؟

✓ أجر تحليل الاختلاف في الاختلافات باستخدام عدة مجموعات مقارنة معقولة. ينبغي أن تحصل على تقديرات متشابهة لأثر البرنامج.

✓ أجر تحليل الاختلاف في الاختلافات باستخدام مجموعتي المعالجة والمقارنة اللتين اخترتهما وحصيلة زائفة ينبغي ألا تتأثر بالبرنامج. وينبغي أن تتوصل إلى أثر صفري للبرنامج على الحاصل.

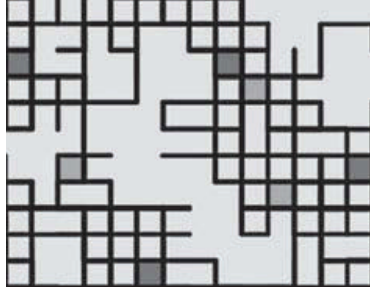
✓ أجر تحليل الاختلاف في الاختلافات باستخدام متغير الحصيلة المختار مع مجموعتين تعرف أنهما لم تتأثرا بالبرنامج. وينبغي أن تتوصل إلى أثر صفري للبرنامج.

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للمزيد من الافتراضات غير المصرح بها وراء الاختلاف في الاختلافات، انظر مدونة الأثر الإنمائي للبنك الدولي (<http://blogs.worldbank.org/impactevaluations>).

المراجع

- De Janvry, Alain, Frederico Finan, and Elisabeth Sadoulet. 2011. "Local Electoral Incentives and Decentralized Program Performance." *Review of Economics and Statistics* 94 (3): 672–85.
- DiTella, Rafael, and Ernesto Schargrotsky. 2005. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack." *American Economic Review* 94 (1): 115–33.
- Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795–813.
- Galiani, Sebastian, Paul Gertler, and Ernesto Schargrotsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113 (1): 83–120.



المطابقة

إنشاء مجموعة مقارنة صورية

المفهوم الأساسي

تستخدم المطابقة مجموعات بيانات وتقنيات إحصائية كبيرة لإنشاء أفضل مجموعة مقارنة ممكنة استنادًا إلى الخصائص الملحوظة.

تتكون الطريقة الموصوفة في هذا الفصل من مجموعة من التقنيات الإحصائية التي سنشير إليها مجتمعة باسم *المطابقة*. يمكن تطبيق طرق المطابقة في سياق أي قواعد تعيين برنامج تقريبًا، طالما توجد مجموعة لم تشارك في البرنامج. تستخدم المطابقة في الأساس تقنيات إحصائية لبناء مجموعة مقارنة صورية. بالنسبة إلى كل وحدة ممكنة قيد المعالجة، فإنها تحاول العثور على وحدة غير معالجة (أو مجموعة وحدات غير معالجة) لها أكثر الخصائص الممكنة تشابهًا. ادرس حالةً تحاول فيها تقييم أثر برنامج تدريب وظيفي على الدخل ويضم مجموعة بيانات، مثل سجلات الدخل والضرائب، ويضم أفرادًا مشتركين

في البرنامج وأفرادًا غير مشتركين. ليس للبرنامج الذي تحاول تقييمه أي قواعد تعيين واضحة (مثل التعيين العشوائي أو مؤشر الأهلية) تشرح سبب اشتراك بعض الأفراد في البرنامج وعدم اشتراك البعض الآخر فيه. في مثل هذا السياق، ستمكّنك طرق المطابقة من تحديد عدد الأفراد غير المشاركين الذين يبدوون أكثر شبهًا بالأفراد المعالجين، استنادًا إلى الخصائص المتاحة في مجموعة بياناتك. يصبح هؤلاء الأفراد غير المشتركين هم مجموعة المقارنة الذين تستخدمهم لتقدير الواقع المضاد.

الشكل ٨,١ المطابقة التامة لأربع خصائص

المجموعات غير المعالجة				المجموعات المعالجة			
العمر	النوع الاجتماعي	عدد أشهر عدم التوظيف	شهادة ثانوية	العمر	النوع الاجتماعي	عدد أشهر عدم التوظيف	شهادة ثانوية
٢٤	١	٨	١	١٩	١	٣	٠
٣٨	٠	١	٠	٣٥	١	١٢	١
٥٨	١	٧	١	٤١	٠	١٧	١
٢١	٠	٢	١	٢٣	١	٦	٠
٣٤	١	٢٠	٠	٥٥	٠	٢١	١
٤١	٠	١٧	١	٢٧	٠	٤	١
٤٦	٠	٩	٠	٢٤	١	٨	١
٤١	٠	١١	١	٤٦	٠	٣	٠
١٩	١	٣	٠	٣٣	٠	١٢	١
٢٧	٠	٤	٠	٤٠	١	٢	٠

يتطلب العثور على مطابق جيد لكل مشترك في البرنامج تقريب الخصائص التي تشرح قرار الفرد بالاشتراك في البرامج بأقرب شكل ممكن. ولسوء الحظ، فإن قول ذلك أسهل من فعله. إذا كانت قائمة الخصائص الملحوظة ذات الصلة كبيرة جدًا، أو إذا كانت كل خاصية تأخذ الكثير من القيم، فقد يصعب تحديد مطابق لكل من الوحدات في مجموعة المعالجة. عندما تزيد عدد الخصائص أو الأبعاد التي تريد مطابقة الوحدات التي اشتركت في البرنامج مقابلها، قد تواجه ما يسمى *لعنة الأبعاد*. فعلى سبيل المثال، إذا كنت تستخدم ثلاث خصائص مهمة فقط لتحديد مجموعة المقارنة المطابقة، مثل العمر والنوع الاجتماعي وما إذا كان الفرد حاصلًا على الشهادة الثانوية، فقد تجد مطابقات لجميع المشاركين المشتركين في البرنامج في مجموعة غير المشتركين، بيد أنك تخاطر بتجاهل الخصائص الأخرى التي قد تكون مهمة. ولكن، إذا زدت قائمة الخصائص - فرضًا، لتشمل عدد الأطفال، وعدد سنوات التعليم، وعدد أشهر عدم التوظيف، وعدد سنوات الخبرة، وما إلى ذلك - فقد لا تحتوي قاعدة بياناتك على تطابق جيد مع معظم المشاركين المشتركين في البرنامج ما لم تكن تحتوي على عدد كبير جدًا من الملاحظات. يوضح الشكل ٨-١ المطابقة استنادًا إلى أربع خصائص: العمر والنوع الاجتماعي وعدد أشهر عدم التوظيف والشهادة الثانوية.

مطابقة درجة الميل

لحسن الحظ، يمكن حل لعنة الأبعاد بسهولة كبيرة باستخدام طريقة تُسمى *مطابقة درجة الميل* (Rosenbaum and Rubin 1983). في هذه المقاربة، لا نحتاج بعد الآن إلى محاولة مطابقة كل مجموعة مشتركة بوحدة غير مشتركة لديها قيمة خصائص المراقبة الملحوظة نفسها بالضبط. بدلاً من ذلك، بالنسبة إلى كل وحدة في مجموعة المعالجة وفي مجموعة غير المشتركين، نحسب احتمال *اشتراك* هذه الوحدة في البرنامج (ما يُسمى درجة الميل) استنادًا إلى القيم الملحوظة لخصائصها (المتغيرات التفسيرية).

الدرجة عدد حقيقي بين ٠ و ١، وهذا يلخص أثر كل الخصائص الملحوظة على احتمالية الاشتراك في البرنامج. علينا أن نستخدم فقط الخصائص الأساسية الملحوظة لحساب درجة الميل. السبب في ذلك هو أن خصائص ما بعد المعالجة ربما تأثرت بالبرنامج نفسه، ومن شأن استخدام هذه الخصائص لتحديد مجموعة المقارنة المطابقة أن يتسبب في تحيز النتائج. عندما تؤثر المعالجة على الخصائص الفردية ونستخدم تلك الخصائص للمطابقة، فإننا نختار مجموعة مقارنة تبدو شبيهة بالمجموعة المعالجة بسبب المعالجة بحد ذاتها. وبدون المعالجة، ستبدو تلك الخصائص مختلفة. وهذا ينتهك المتطلبات الأساسية للتقدير الجيد للواقع المضاد: يجب أن تكون مجموعة المقارنة متشابهة في جميع الجوانب، باستثناء أن مجموعة المعالجة تتلقى المعالجة ومجموعة المقارنة لا تتلقاها.

وبمجرد حساب درجة الميل لجميع الوحدات، فيمكن مطابقة الوحدات في مجموعة المعالجة مع الوحدات في مجموعة غير المعالَجين التي لها درجة الميل الأقرب.^١ تصبح هذه الوحدات الأقرب هي مجموعة المقارنة المستخدمة لحساب تقدير للواقع المضاد. تحاول طريقة مطابقة درجة الميل تقليد تعيين العشوائي في مجموعتي المعالجة والمقارنة باختيار الوحدات التي لها ميول مشابهة لميول الوحدات في مجموعة المعالجة لمجموعة المقارنة. بما أن مطابقة درجة الميل ليست طريقة تعيين عشوائية ولكنها تحاول تقليد هذه الطريقة، فهي تنتمي إلى فئة الطرق شبه التجريبية.

يُنتج متوسط الفرق في الحاصل بين وحدتي المعالجة أو المقارنة ووحدات المقارنة المطابقة لهما الأثر المقدّر للبرنامج. باختصار، يُقدّر أثر البرنامج من خلال مقارنة متوسط حواصل مجموعة معالجة أو مجموعة مشتركة مع متوسط الحاصل بمتوسط الحاصل بين مجموعة فرعية مطابقة من الوحدات، حيث تستند المطابقة إلى الخصائص الملحوظة المتوفرة في البيانات المتاحة.

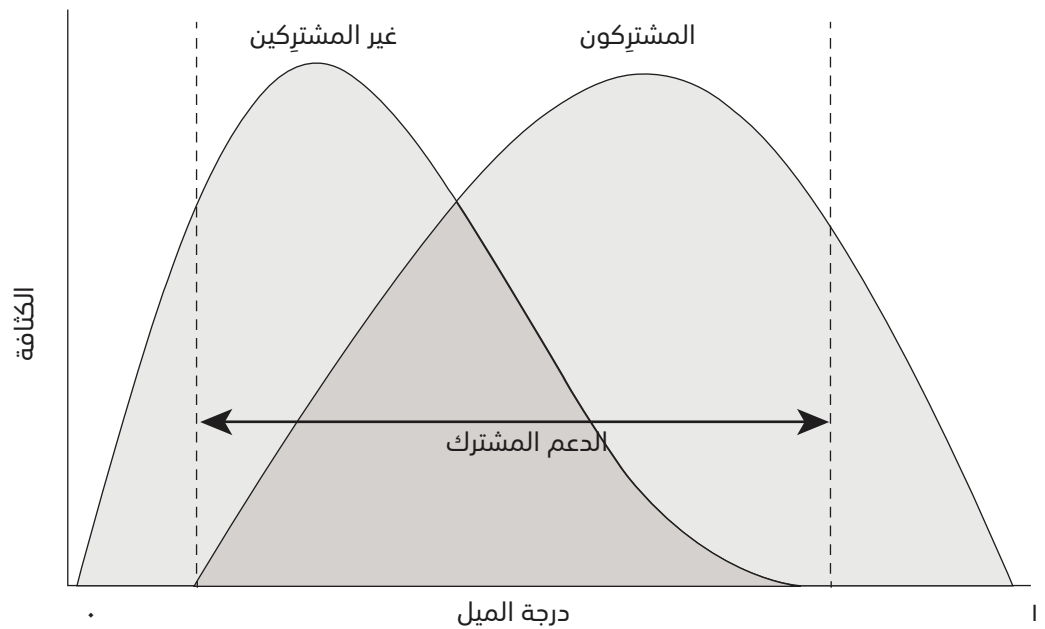
لكي تنتج مطابقة درجة الميل تقديرات لأثر البرنامج لجميع الملاحظات المعالجة، فينبغي مطابقة كل وحدة معالجة أو وحدة مشتركة مع وحدة غير مشتركة.^٢ ولكن، في حالة بعض الوحدات المشتركة عملياً، قد لا يكون لأي من الوحدات في مجموعة غير المشتركين درجات ميل متشابهة. من الناحية التقنية، قد يكون هناك افتقار إلى *الدعم المشترك*، أو انعدام التداخل، بين درجات الميل الخاصة بمجموعة المعالجة (أو المجموعة المشاركة) ودرجات الميل للمجموعة غير المشاركة.

يقدم الشكل ٨-٢ مثالاً عن الافتقار إلى الدعم المشترك. أولاً، نقدّر احتمال اشتراك كل وحدة في العينة في البرنامج استناداً إلى الخصائص الملحوظة لتلك الوحدة: أي درجة الميل. يوضح الشكل توزيع درجات الميل بصورة منفصلة للمشاركين وغير المشاركين. المشكلة هي أن هذه التوزيعات لا تتداخل بصورة تامة. في وسط التوزيع، من السهل نسبياً العثور على مطابقات لأن هناك كل من المشاركين وغير المشاركين بهذه المستويات من درجات الميل. ولكن، لا يمكن مطابقة المشاركين ذوي درجات الميل القريبة من ١ مع أي من غير المشاركين لأنه ما من غير مشتركين بهذا المستوى المرتفع من درجات الميل. بديهيًا، الوحدات ذات الاحتمالية العالية للاشتراك في البرنامج مختلفة جدًا عن الوحدات غير المشتركة لدرجة أننا لا نجد مطابقةً جيدًا لهم. وعلى نحوٍ مماثل، لا يمكن مطابقة غير المشاركين ذوي درجات الميل القريبة من ٠ مع أي من المشاركين لأنه ما من مشتركين بهذا المستوى المنخفض من درجات الميل. وهكذا يظهر الافتقار إلى الدعم المشترك في الأطراف أو نهايات توزيع درجات الميل.

في هذه الحالة، يقدر إجراء المطابقة متوسط تأثير المعالجة المكاني (LATE) للملاحظات على الدعم المشترك.

يلخص جالان ورافاليون (Jalan and Ravallion 2003) الخطوات التي ينبغي اتخاذها عند تطبيق مطابقة درجة الميل.^٣ أولاً، ستحتاج إلى مسوحات تمثيلية وقابلة للمقارنة بدرجة عالية يمكن فيها تحديد الوحدات المشتركة في البرنامج وتلك التي لم تشارك.

الشكل ٨-٢ مطابقة درجة الميل والدعم المشترك



ثانيًا، تجمّع العينتين وتقدّر احتمالية اشتراك كل فرد في البرنامج، استنادًا إلى الخصائص الفردية الملحوظة في المسح. تعطي هذه الخطوة درجة الميل. ثالثًا، تحصر العينة بالوحدات التي يظهر لها دعم مشترك في توزيع نقاط الميل. رابعًا، بالنسبة إلى كل مجموعة مشتركة، تحدّد مجموعة فرعية من الوحدات غير المشتركة التي لها درجات ميل مماثلة. خامسًا، تقارن الحاصلات لوحدات المعالجة والوحدات المشتركة ووحدات المقارنة أو الوحدات غير المشتركة المطابقة. الفرق في متوسط الحاصلات لهاتين المجموعتين الفرعيتين هو مقياس الأثر الذي يمكن أن يُسندَ إلى البرنامج لتلك الملاحظة المعالجة بالتحديد. سادسًا، يُنتج متوسط هذه الآثار الفردية تقديرًا لمتوسط تأثير المعالجة المكاني. عمليًا، تتضمن البرامج الإحصائية شائعة الاستخدام تعليمات مبرمجة مسبقًا تسير من الخطوات ٢ إلى ٦ تلقائيًا.

إجمالًا، من المهم تذكّر ثلاث قضايا حاسمة بشأن المطابقة. أولاً، يمكن أن تستخدم طرق المطابقة فقط الخصائص الملحوظة لإنشاء مجموعة مقارنة، وذلك لأنه لا يمكن أخذ الخصائص غير الملحوظة في الحسبان. إذا كانت هناك أي خصائص غير ملحوظة تؤثر على ما إذا كانت وحدة ما تشارك في البرامج وتؤثر على الحاصلات أيضًا، فإن تقديرات الأثر المُتحصّل عليها مع مجموعة المقارنة المطابقة ستكون متحيّزة. لكي تكون النتيجة المطابقة غير متحيّزة، فإنها تتطلب افتراضًا قويًا بأنه ما من فروق غير ملحوظة في مجموعات المعالجة والمقارنة التي ترتبط أيضًا بالحاصلات محل الاهتمام.

ثانيًا، يجب إجراء المطابقة باستخدام الخصائص غير المتأثرة بالبرنامج فقط. لا تقع معظم الخصائص المقيسة بعد بداية البرنامج في تلك الفئة. إذا كانت البيانات الأساسية (ما قبل التدخل) غير متاحة، وكانت البيانات الوحيدة تنتمي إلى ما بعد بدء التدخل، فإن الخصائص الوحيدة التي سنكون قادرين على استخدامها لبناء عينة مطابقة هي الخصائص التي لا تتأثر بالبرنامج (وهي قليلة عادةً)، مثل العمر والنوع الاجتماعي. وعلى الرغم من أننا نودّ مطابقة مجموعة أكثر ثراءً من الخصائص، بما في ذلك الحاصلات موضع الاهتمام، لا يمكننا فعل ذلك لأن تلك الخصائص قد تتأثر بالتدخل. ولا يوصى بالمطابقة استنادًا إلى خصائص ما بعد التدخل فحسب. إذا كانت البيانات الأساسية متوفرة، فيمكننا المطابقة استنادًا إلى مجموعة أكثر ثراءً من الخصائص، بما فيها الحاصلات موضع الاهتمام. ونظرًا إلى أن البيانات تُجمّع قبل التدخل، فلا يمكن أن تكون متغيرات ما قبل التدخل تلك قد تأثرت بالبرنامج. ولكن، إذا كانت البيانات الأساسية عن الحاصلات متوفرة، فينبغي ألا تستخدم طريقة المطابقة بمفردها. بل يجب عليك دمجها مع الاختلاف في الاختلافات للحد من خطر التحيز. يناقش القسم التالي هذا الإجراء.

ثالثًا، تكون نتائج تقدير طريقة المطابقة جيدة فقط بقدر جودة الخصائص المستخدمة للمطابقة. ورغم أنه من المهم توفير القدرة على استخدام عدد كبير من الخصائص، إلا أن الأكثر أهمية هو التمكن من المطابقة على أساس الخصائص التي تحدّد الاشتراك. وكلما كان لدينا فهم أفضل للمعايير المستخدمة لاختيار المشاركين، كانت قدرتنا على إنشاء مجموعة المقارنة المطابقة أفضل.

جمع طريقة المطابقة مع طرق أخرى

على الرغم من أن تقنية المطابقة تتطلب كمية كبيرة من البيانات وتنطوي على مخاطر تحيز كبيرة، فإنها تُستخدم لتقييم برامج التطوير في مجموعة واسعة من البيئات. وأكثر استخدامات المطابقة إقناعًا هي تلك التي تجمع المطابقة مع طرق أخرى وتلك التي تستخدم طريقة الوحدات الضابطة الاصطناعية. في هذا القسم، سنناقش طريقة الاختلاف في الاختلافات المتطابق وطريقة الوحدات الضابطة الاصطناعية.

الاختلاف في الاختلافات المتطابق

عندما تتوفر بيانات أساسية عن الحاصل، يمكن دمج المطابقة مع الاختلاف في الاختلافات للحد من خطورة التحيز في التقدير. ووفق ما وضعناه سابقًا، لا يمكن لمطابقة درجات الميل البسيطة أن تأخذ في الحسبان الخصائص غير الملحوظة التي قد تفسر سبب اختيار مجموعة الاشتراك في برنامج والتي قد تؤثر أيضًا على الحاصل. تعني المطابقة المدمجة بالاختلاف في الاختلافات على الأقل بأي خصائص غير ملحوظة وثابتة عبر الزمن بين المجموعتين. وهي تُنفَّذ على النحو التالي:

١. نفّذ المطابقة استنادًا إلى خصائص خط الأساس الملحوظة وغير الملحوظة (على النحو الموضح).
 ٢. بالنسبة إلى كل وحدة مشتركة، احسب التغير في الحاصل بين الفترتين قبل وبعد (الفرق الأول).
 ٣. بالنسبة إلى كل مجموعة مشتركة، احسب التغير في الحاصل بين الفترتين قبل وبعد لمقارنة هذه الوحدة المطابقة (الفرق الثاني).
 ٤. اطرح الفرق الثاني من الفرق الأول، أي طبق طريقة الاختلاف في الاختلافات.
 ٥. وأخيرًا، أوجد متوسط تلك الاختلافات المضاعفة.
- يقدم المربعان ٨-١ و ٨-٢ مثالين عن تقييمين استخدمتا طريقة الاختلاف في الاختلافات عمليًا.

المربع ٨-١: الاختلاف في الاختلافات المتطابق: تطوير الطرق الريفية والسوق المحلية في فيتنام

و٩٥ بلدية مقارنة. للحد من تحيُّز الاختيار المحتمل أكثر، استخدم الباحثان الاختلاف في الاختلافات لتقدير التغير في ظروف السوق المحلية. بعد البرنامج بعامين، أشارت النتائج إلى أن إعادة تأهيل الطرقات أدى إلى آثار إيجابية كبيرة بشأن وجود الأسواق المحلية وتوافر الخدمات وتكرارهما. تطورت أسواق جديدة في نسبة تزيد ١٠ بالمائة من بلديات المعالجة مقابل بلديات المقارنة. في بلديات المعالجة، كان من الأكثر شيوعاً أن تنتقل الأسر من الأنشطة الزراعية إلى أنشطة أكثر ارتباطاً بالخدمات، مثل الحياكة وتصفيف الشعر. ولكن النتائج كانت متباينة جداً عبر البلديات. في البلديات الفقيرة، تميل الآثار إلى أن تكون أعلى بسبب انخفاض مستويات التنمية الأولية للسوق. وقد استنتج الباحثان أن مشاريع تحسين الطرق الصغيرة يمكن أن تكون لها آثار أكبر إذا استهدفت مناطق ذات تنمية سوقية منخفضة في البداية.

استخدم مو، وفان دي والي (٢٠١١) مطابقة درجة الميل مع التباين المزدوج لتقدير أثر برنامج طرق ريفية على تنمية الأسواق المحلية. من عام ١٩٩٧ إلى عام ٢٠٠١، أعادت الحكومة الفيتنامية تأهيل ٥٠٠٠ كيلومتر من الطرق الريفية. اختيرت الطرق وفقاً لمعاري التكلفة والكثافة السكانية. وبما أن البلديات التي استفادت من الطرق المعاد تأهيلها لم تُختَر عشوائياً، استخدم الباحثون مطابقة درجة الميل لإنشاء مجموعة مقارنة. باستخدام بيانات من مسح خط الأساس، توصل الباحثان إلى مجموعة من العوامل على مستوى البلديات التي تؤثر على اختيار طريق في البلدية للبرنامج، مثل حجم الفئة السكانية، وحصة الأقليات العرقية، ومستويات المعيشة، وكثافة الطرق الموجودة، ووجود وسائل نقل الركاب. وقد قُدِّرَ درجات الميل استناداً إلى هذه الخصائص وحصرًا حجم العينة بمنطقة الدعم المشترك. تمخّض ذلك عن ٩٤ بلدية معالجة

المصدر: Mu and Van de Walle 2011.

المربع ٨-٢: الاختلاف في الاختلافات المتطابق: الأرضيات الإسمنتية وصحة الأطفال وسعادة الأمهات في المكسيك

استخدم الباحثون طريقة الاختلاف في الاختلافات جنباً إلى جنب مع المطابقة لمقارنة الأسر في كواويلا مع أسر مشابهة في ولاية

يوفر برنامج Firme Piso في المكسيك ٥٠ متراً مربعاً من الأرضيات الخزسانية للمنازل ذات الأرضيات الترابية (انظر المربع ١-٢). بدأ برنامج Piso Firme بصيغة برنامج محلي في ولاية كواويلا، بيد أنه اعتمد وطنياً. استفاد كثانيو وآخرون (٢٠٠٩) من التباين الجغرافي لتقييم أثر مجهود تحسين المساكن واسع النطاق هذا على حصائل الصحة والرفاهية.

(تابع)

المربع ٨,٢: الاختلاف في الاختلافات المتطابق: الأرضيات الإسمنتية وصحة الأطفال وسعادة الأمهات في المكسيك (تابع)

وانخفاض بنسبة ٤٩ بالمائة في الإسهال، وانخفاض بنسبة ٨١ بالمائة في الأنيميا، وتحسّن بنسبة ٣٦ بالمائة إلى ٩٦ بالمائة في التطور المعرفي للأطفال. جمع المؤلفون أيضًا بيانات حول رفاهية البالغين، ووجدوا أن الأرضيات الإسمنتية تجعل الأمهات أكثر سعادة، بزيادة نسبتها ٥٩٪ في الرضا الذاتي عن جودة الحياة، وانخفاض بنسبة ٦٩٪ في الرضا الذاتي عن جودة الحياة، وانخفاض بنسبة ٥٢٪ على مقياس تقدير الاكتئاب، وانخفاض بنسبة ٤٥ في المئة على مقياس تقدير الضغط النفسي الملحوظ.

اختتم كاتانيو وآخرون (٢٠٠٩) بتوضيح أن لمشروع Piso Firme أثرًا مطلقًا أكبر - على التنمية المعرفية للأطفال - بتكلفة أقل - من برنامج التحويلات النقدية المشروطة واسع النطاق في المكسيك بالإضافة إلى بالإضافة Oportunidades/Progresas، مشابهة في مجال الدعم الغذائي والتحفيز المعرفي في مرحلة الطفولة. منعت الأرضيات الإسمنتية أيضًا المزيد من حالات العدوى الطفيلية بالمقارنة مع معالجة التخلص من الديدان الشائعة. يقول المؤلفون إن برامج استبدال الأرضيات الترابية بأرضيات إسمنتية يمكن أن تحسّن تكاليف صحة الأطفال بفعالية في سياقات مشابهة.

دورانغو المجاورة، التي لم تكن قد طبقت البرنامج بعد في وقت المسح. لتحسين قابلية المقارنة بين مجموعتي المعالجة والمقارنة، حصر الباحثون عيّنهم على الأسر في المدن المجاورة التي تقع على جانبي الحدود بين الولايتين. ضمن هذه العينة، استخدموا تقنيات المطابقة لاختيار مجموعات المعالجة والمقارنة التي كانت الأكثر تشابهًا. وقد شملت خصائص ما قبل المعالجة التي استخدموها نسبة الأسر ذات الأرضيات الترابية، وعدد الأطفال الصغار، وعدد الأسر في كل مجتمع.

وبالإضافة إلى المطابقة، استخدم الباحثون متغيرات مساعدة لاستعادة متوسط تأثير المعالجة من تأثير النية للعلاج. بتقديم الأرضية الإسمنتية بمثابة متغير مساعد لوجود أرضيات إسمنتية بالفعل، وجدوا أن البرنامج أدى إلى انخفاض بنسبة ١٨,٢ بالمائة في وجود الطفيليات، و١٢,٤ بالمائة في انتشار الإسهال، و انخفاض بنسبة ١٩,٤ بالمائة في انتشار فقر الدم. علاوةً على ذلك، تمكنا من استخدام التباين في مقدار المساحة الإجمالية للأرضية المغطاة فعليًا بالإسمنت للتنبؤ في أن الاستبدال الكامل للأرضيات الترابية بأرضيات إسمنتية في الأسرة سيفضي إلى انخفاض بنسبة ٧٨ بالمائة في العدوى الطفيلية،

المصدر: Cattaneo and others 2009.

طريقة الوحدات الضابطة الاصطناعية

طريقة الوحدات الضابطة الاصطناعية هي طريقة مطابقة خاصة تسمح بتقدير الأثر في بيانات تتلقى فيها وحدة واحدة (مثل دولة أو شركة أو مستشفى) تدخلاً أو تتعرض لحدث ما. بدلاً من مقارنة هذه الوحدة المعالجة بمجموعة من الوحدات غير المعالجة، تستخدم هذه الطريقة معلومات حول خصائص الوحدة المعالجة والوحدات غير المعالجة لإنشاء وحدة مقارنة "اصطناعية" أو صناعية عن طريق موازنة كل وحدة غير معالجة بطريقة تجعل وحدة المقارنة الاصطناعية مماثلة للوحدة المعالجة بالشكل الأكبر. ويتطلب ذلك سلسلة طويلة من الملاحظات بمرور الوقت للخصائص الخاصة بكل من وحدة المعالجة والوحدات غير المعالجة. وهذا الجمع بين وحدات المقارنة في وحدة اصطناعية يوفر مقارنة أفضل للوحدة المعالجة عن أي وحدة غير معالجة بشكل فردي. يقدّم المربع ٨-٣ مثالاً عن تقييم يستخدم طريقة الوحدات الضابطة الاصطناعية.

المربع ٣-٨ طريقة الوحدات الضابطة الاصطناعية: الآثار الاقتصادية للصراع الإرهابي في إسبانيا

استخدم أبادي وغارديزابال (2003) طريقة الوحدات الضابطة الاصطناعية لدراسة الآثار الاقتصادية للصراع الإرهابي في إقليم الباسك. في بداية السبعينيات، كان إقليم الباسك واحدًا من أكثر الأقاليم ثراءً في إسبانيا، بيد أنه بحلول نهاية التسعينيات، وبعد 30 عامًا من النزاع، تراجع إلى المركز السادس في نصيب الفرد من الناتج المحلي الإجمالي. في بداية الإرهاب في أوائل السبعينيات، اختلف إقليم الباسك عن المناطق الإسبانية الأخرى في الخصائص

التي يعتقد أنها مرتبطة بإمكانية النمو الاقتصادي. لذلك، تعكس مقارنة نمو الناتج المحلي الإجمالي بين اقتصاد الباسك وبقية إسبانيا تأثير الإرهاب وتأثير هذه الاختلافات على النمو الاقتصادي قبل بدء الإرهاب. بمعنى آخر، يعطي مقارنة الاختلاف في الاختلافات نتائج متحيزة لأثر الإرهاب على النمو الاقتصادي في إقليم الباسك. للتعامل مع هذه الحالة، استخدم المؤلفان مجموعة من المناطق الإسبانية الأخرى لإنشاء منطقة مقارنة "اصطناعية".

المصدر: Abadie and Gardeazabal 2003.

تقدير أثر برنامج معونة التأمين الصحي (HISP): استخدام تقنيات المطابقة

بعد أن تعرفت على تقنيات المطابقة، قد تتساءل عما إذا كان بإمكانك استخدامها لتقدير أثر برنامج معونة التأمين الصحي (HISP). تقرر استخدام بعض تقنيات المطابقة لاختيار مجموعة من الأسر غير المشتركة التي تبدو مشابهة للأسر المشتركة استنادًا إلى الخصائص الأساسية الملحوظة. ولإجراء ذلك، تستخدم حزمة المطابقة بالبرامج الإحصائية لديك. تقارن الحزمة أولاً احتمال اشتراك أسرة في البرنامج استنادًا إلى القيم الملحوظة للخصائص (المتغيرات التفسيرية)، مثل عمر رب الأسرة وزوجه، ومستوى تعليمه، وما إذا كان رب الأسرة أنثى، وما إذا كان رب الأسرة من السكان الأصليين، وما إلى ذلك.

سننجز المطابقة باستخدام سيناريوهين اثنين. في السيناريو الأول، ثمة مجموعة كبيرة من المتغيرات للتنبؤ في الاشتراك، بما في ذلك الخصائص الاقتصادية والاجتماعية للأسرة. في السيناريو الثاني، ثمة معلومات قليلة للتنبؤ بالاشتراك (تعليم رب الأسرة وعمره فقط). كما هو موضح في الجدول ٨-١،

يكون احتمال مشاركة أسرة في البرنامج أصغر إذا كانت الأسرة أكبر سنًا، أو أعلى مستوى تعليميًا، أو ترأسها أنثى، أو لديها حمام، أو لديها مساحات أكبر من الأراضي. على العكس، فإن وجود عدد أكبر من أفراد الأسرة، ووجود أرضية ترابية، والعيش بعيدًا عن المستشفيات يزيد من قابلية اشتراك الأسرة في البرنامج. لذلك على العموم، يبدو أن الأسر الأكثر فقرًا والأقل تعليمًا هي الأكثر قابلية للاشتراك، وهو خبر سار لبرنامج موجّه نحو الفقراء.

وبما أن البرنامج قدّر احتمال اشتراك كل أسرة في البرنامج (درجة الميل)، عليك التحقق من توزيع درجة الميل للأسر المشتركة وأسر المقارنة المطابقة. يوضح الشكل ٨-٣ أن الدعم المشترك (عند استخدام مجموعة كاملة من المتغيرات التفسيرية) يمتد عبر التوزيع الكامل لدرجة الميل. وفي الواقع، لا يقع أي من الأسر المسجلة خارج منطقة الدعم المشترك. بمعنى آخر، يمكننا إيجاد أسرة مقارنة مطابقة لكل من الأسر المسجلة.

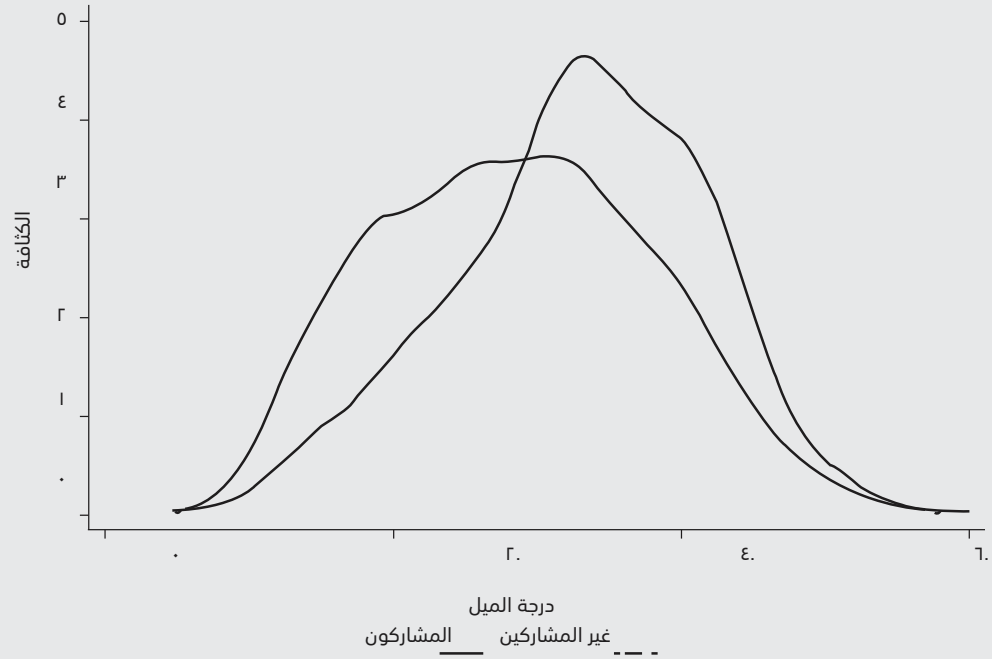
الجدول ٨-١ تقدير درجة الميل بناءً على الخصائص المرصودة لخط الأساس

المتغير التابع: المشترك = ١		مجموعة كاملة من المتغيرات	مجموعة محدودة من المتغيرات
المتغيرات التفسيرية خصائص خط الأساس		المعامل	المعامل
عمر رب الأسرة (أعوام)	**٠,٠١٣-	**٠,٠٢١-	
عمر الزوج (أعوام)	**٠,٠٠٨-	**٠,٠٤١-	
تعليم رب الأسرة (أعوام)	**٠,٠٢٢-		
تعليم الزوج (أعوام)	*٠,٠١٦-		
رب الأسرة أنثى = ١	٠,٠٢٠-		
الأصلي = ١	**٠,١٦١		
عدد أفراد الأسرة	**٠,١١٩		
أرضية ترابية = ١	**٠,٣٧٦		
حمام = ١	**٠,١٢٤-		
عدد هكتارات الأرض	**٠,٠٢٨-		
المسافة إلى المستشفى (كم)	**٠,٠٠٢		
ثابت	**٠,٤٩٧-	**٠,٠٠٤	

ملاحظة: الانحدار الاحتمالي. المتغير التابع يساوي ١ إذا اشتركت الأسرة في برنامج HISP ويساوي ٠ بخلاف ذلك. يمثل المعامل مساهمة كل متغير تفسيري مدرج في احتمال كون أسرة ما مشتركة في برنامج HISP.

مستوى الدلالة: * = ٥ بالمائة، ** = ١ بالمائة.

الشكل ٨,٣ المطابقة لبرنامج HISP:



تقرّر استخدام مطابقة المجاور الأقرب، أي أنك تطلب من البرنامج أن يحدد، لكل أسرة مشتركة، الأسرة غير المشتركة التي لديها أقرب درجة ميل إلى درجة ميل الأسرة المسجلة. يحصر البرنامج الآن العينة على الأسر في المجموعات المشتركة وغير المشتركة التي يمكنه أن يجد لها مطابقاً في المجموعة الأخرى. للحصول على الأثر المقدر باستخدام طريقة المطابقة، تحسب أولاً الأثر لكل أسرة فردياً (باستخدام أسرة المقارنة المطابقة لكل أسرة)، ثم متوسط تلك الآثار الفردية. يوضح الجدول ٨-٢ أن الأثر المقدر من تطبيق هذا الإجراء هو تخفيض مبلغ ٩,٩٥ دولارات أمريكية في النفقات الصحية للأسرة. أخيراً، يسمح البرنامج أيضاً بحساب الخطأ المعياري في الأثر المقدر باستخدام الانحدار الخطي (الجدول ٨-٣).^٤

تُدرك أن لديك أيضاً معلومات عن حائل خط الأساس في بيانات مسحك، لذلك قررت تنفيذ الاختلاف في الاختلافات المطابقة بالإضافة إلى استخدام المجموعة الكاملة من المتغيرات التفسيرية. أي أنك تحسب الفرق في النفقات الصحية الأسرية عند المتابعة بين الأسر المشتركة وأسر المقارنة المطابقة، وتحسب الفرق في النفقات الصحية الأسرية عند خط الأساس بين الأسر المشتركة وأسر المقارنة المطابقة، ثم تحسب الفرق بين هذين الفرقين. يوضح الجدول ٨-٤ نتيجة مقارنة الاختلاف في الاختلافات المطابق هذا.

الجدول ٢-٨ تقييم برنامج معونة التأمين الصحي (HISP): مطابقة خصائص خط الأساس ومقارنة المتوسطات

المشاركة	المطابقة المقارنة	الفرق
٧,٨٤	١٧,٧٩ (باستخدام مجموعة كاملة من المتغيرات التفسيرية)	٩,٩٥-
	١٩,٩ (استخدام مجموعة محدودة من المتغيرات التفسيرية)	١١,٣٥-

ملاحظة: يقارن هذا الجدول متوسط النفقات الصحية الأسرية للأسر المشتركة وأسر المقارنة المطابقة.

الجدول ٣-٨ تقييم برنامج معونة التأمين الصحي (HISP): مطابقة خصائص خط الأساس وتحليل الانحدار

الانحدار الخطي (المطابقة على مجموعة كاملة من المتغيرات التفسيرية)	الانحدار الخطي (المطابقة على مجموعة محدودة من المتغيرات التفسيرية)
٩,٩٥- ** (٠,٢٤)	١١,٣٥- ** (٠,٢٢)

الأثر المقدّر على النفقات الصحية الأسرية (دولار أمريكي)

ملاحظة: الأخطاء المعيارية بين قوسين. مستوى الدلالة: ** = ١ بالمائة..

الجدول ٤-٨ تقييم برنامج معونة التأمين الصحي (HISP): دمج الاختلاف في الاختلافات مع مطابقة خصائص خط الأساس

المقارنة	المطابقة باستخدام مجموعة كاملة من المتغيرات	الفرق
٧,٨٤	١٧,٧٩	٩,٩٥-
١٤,٤٩	١٥,٠٣	٠,٥٤
المتابعة خط الأساس		المطابقة الاختلاف في الاختلافات = ٩,٤١- ** (٠,١٩)

ملاحظة: الخطأ المعياري بين قوسين وهو محسوب باستخدام الانحدار الخطي. مستوى الدلالة: ** = ١ بالمائة.

السؤال ٧ الخاص ببرنامح HISP



- أ. ما الافتراضات الأساسية المطلوبة لقبول كل هذه النتائج استنادًا إلى طريقة المطابقة؟
- ب. لماذا تختلف نتائج طريقة المطابقة في حال استخدام مجموعة المتغيرات التفسيرية الكاملة مقابل مجموعة المتغيرات التفسيرية المحدودة؟
- ج. ماذا يحدث عندما تقارن نتيجة طريقة المطابقة مع نتيجة التعيين العشوائي؟ لماذا تعتقد أن النتائج مختلفة كثيرة ضمن مجموعة محدودة من المتغيرات التفسيرية؟ ما السبب في أن النتيجة أكثر تماثلًا عند المطابقة ضمن مجموعة كاملة من المتغيرات التفسيرية؟
- د. استنادًا إلى نتيجة طريقة المطابقة، هل ينبغي توسيع برنامج معونة التأمين الصحي (HISP) على المستوى الوطني؟

محددات طريقة المطابقة

على الرغم من إمكانية تطبيق إجراءات المطابقة في العديد من البيئات، بغض النظر عن قواعد تعيين البرنامج، إلا أنه ثمة أوجه قصور خطيرة متعددة. أولاً، تتطلب هذه الإجراءات مجموعات بيانات واسعة لعينات وحدات كبيرة، وحتى في حال إتاحتها، فقد يكون هناك افتقار إلى الدعم المشترك بين مجموعة المعالجة أو المجموعة المشاركة وبين مجموعة غير المشتركين. ثانيًا، يمكن إجراء المطابقة فقط استنادًا إلى الخصائص المرصودة، ووفق التعريف، لا يمكننا دمج الخصائص غير المرصودة في حساب درجة الميل. ولكي يحدّد إجراء المطابقة مجموعة مقارنة صالحة، علينا أن نتحقق من غياب أي فروق منهجية الخصائص غير المرصودة بين وحدات المعالجة ووحدات المقارنة المطابقة^٥ التي قد تؤثر على الحصيلة (Y). وبما أنه لا يمكننا إثبات عدم وجود مثل هذه الخصائص غير الملحوظة التي تؤثر على كل من المشاركة والحصائل، علينا افتراض عدم وجود أي منها. وعادةً ما يكون هذا افتراضًا قويًا جدًا. على الرغم من أن المطابقة تساعد على ضبط خصائص الخلفية المرصودة، لا يمكننا على الإطلاق استبعاد التحيز الناشئ عن الخصائص غير المرصودة. باختصار، افتراض عدم حدوث تحيّز اختيار ناشئ عن الخصائص غير الملحوظة قوي جدًا، والإشكالية الكبرى أنه يتعذر اختباره.

إن المطابقة وحدها أقل قوة على العموم من طرق التقييم الأخرى التي ناقشناها، لأنها تتطلب افتراضًا قويًا مفاده عدم وجود خصائص غير ملحوظة تؤثر في الوقت نفسه على المشاركة في البرنامج وحصائله. لا يتطلب التعيين العشوائي، والمتغير المساعد، وتصميم انقطاع الانحدار، من ناحية أخرى الافتراض غير القابل للاختبار بعدم وجود مثل هذه المتغيرات غير الملحوظة. وهي لا تتطلب أحيانًا مثل هذه العينات الكبيرة أو خصائص خلفية واسعة مثل مطابقة درجة الميل.

عادةً ما تُستخدم طرق المطابقة عمليًا عندما لا تكون خيارات تصميم التعيين العشوائي والمتغير المساعد وانقطاع الانحدار ممكنة. ينطوي ما يسمى المطابقة الوقائية (اللاحقة) على خطورة كبيرة عندما لا تتوفر أي بيانات خط أساس عن الحوائل موضع الاهتمام أو خصائص الخلفية. إذا كان التقييم يستخدم بيانات مسح جُمعت بعد بدء البرنامج (أي لاحقًا) لاستنتاج خصائص خلفية الأشخاص في خط الأساس، ثم يطابق المجموعة المعالجة بمجموعة مقارنة باستخدام تلك الخصائص المستنتجة، فقد يطابق بدون قصد استنادًا إلى الخصائص المتأثرة أيضًا بالبرنامج، في هذه الحالة، ستكون نتيجة التقدير غير صالحة أو متحيزة.

وعلى العكس من ذلك، عندما تكون بيانات خط الأساس متاحة، يمكن أن تكون المطابقة استنادًا إلى خصائص الخلفية الأساسية مفيدة جدًا عند دمجها بتقنيات أخرى، مثل الاختلاف في الاختلافات، ما يسمح بتصحيح الاختلافات بين المجموعات الثابتة بمرور الزمن. وتكون المطابقة أيضًا أكثر موثوقية عندما تكون قاعدة تعيين البرنامج والمتغيرات الأساسية معروفة، وفي هذه الحالة يمكن إجراء المطابقة على تلك المتغيرات. وبحلول هذا الوقت، قد تكون تقييمات الأثر هي الأفضل تصميمًا قبل البدء في تنفيذ البرنامج. وبمجرد بدء البرنامج، إذا لم تكن هناك طريقة كي يؤثر الشخص على كيفية تعيينه ولم يتم جمع بيانات أساسية، فسيكون هناك القليل من الخيارات الصارمة لتقييم الأثر، إن وجدت.

قائمة المراجعة: المطابقة

تعتمد المطابقة على افتراض مفاده أن الوحدات المشتركة وغير المشتركة متشابهة من حيث أي متغيرات غير ملحوظة يمكن أن تؤثر على احتمال المشاركة في البرنامج والحصيلة.

✓ هل تتحدد المشاركة في البرنامج بمتغيرات لا يمكن ملاحظتها؟ لا يمكن عادةً اختبار هذا مباشرة، لذلك سيتعين الاعتماد على الحس السليم والمعرفة الجيدة لإعداد تقييم الأثر.

✓ هل الخصائص الملحوظة متوازنة جيدًا بين المجموعات الفرعية المطابقة؟ قارن الخصائص الملحوظة لكل مجموعة معالجة مع مجموعة المقارنة المطابقة لها في خط الأساس.

✓ هل يمكن العثور على وحدة مقارنة مطابقة لكل وحدة معالجة؟ تحقق مما إذا كان هناك دعم مشترك كافٍ في توزيع درجات الميل. تشير مناطق الدعم المشترك الصغيرة إلى أن المشتركين وغير المشتركين مختلفون للغاية، وهذا يلقي بشكوك بشأن ما إذا كانت المطابقة طريقة ذات مصداقية.

مصادر إضافية


- للحصول على المواد المرفقة مع هذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- لمزيد من المعلومات عن المطابقة، انظر Rosenbaum, Paul. 2002. *Observational Studies*, second edition. Springer Series in Statistics. New York: Springer-Verlag.
- للمزيد حول تطبيق مطابقة درجة الميل، انظر Heinrich, Carolyn, Alessandro Maffioli, and Gonzalo Vásquez. 2010. "A Primer for Applying Propensity-Score Matching. Impact-Evaluation Guidelines." Technical Note IDB-TN-161, Inter-American Development Bank, Washington, DC.

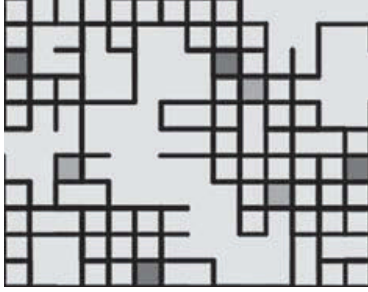
ملاحظات

1. ملاحظة تقنية: في الممارسة العملية، يتم استخدام العديد من التعريفات لما يشكل أقرب وحدة لإجراء المطابقة. يمكن تعريف وحدات المقارنة الأقرب استنادًا إلى طبقية درجة الميل - تحديد المجاورات الأقرب لوحدة المعالجة، بناءً على المسافة، ضمن نصف قطر معين - أو باستخدام تقنيات النواة. ويُعد التحقق من قوة نتائج المطابقة باستخدام خوارزميات مطابقة متعددة ممارسة جيدة. انظر Rosenbaum (2002) لمزيد من التفاصيل.
2. تركز المناقشة حول المطابقة في هذا الكتاب على المطابقة واحد لواحد. ولن يناقش الأنواع الأخرى للمطابقة، مثل المطابقة واحد إلى متعدد أو المطابقة مع الاستبدال/بدون استبدال. ولكن في جميع الأحوال، يبقى الإطار المفاهيمي الموصوف هنا ساريًا.
3. يمكن العثور على مراجعة مفصلة للمطابقة في Rosenbaum (2002).
4. ملاحظة تقنية: عندما لا تكون درجات ميل الوحدات المشتركة مشمولة بالكامل في منطقة الدعم المشترك، يجب تقدير الأخطاء المعيارية باستخدام السحب بدلاً من الانحدار الخطي.
5. بالنسبة إلى القراء ذوي الخلفية في مجال الاقتصاد القياسي، يعني هذا أن المشاركة مستقلة عن الحاصل، في ضوء خصائص الخلفية المستخدمة لإجراء المطابقة.

المراجع

- Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32.
- Cattaneo, Matias D., Sebastian Galiani, Paul J. Gertler, Sebastian Martinez, and Rocio Titiunik. 2009. "Housing, Health, and Happiness." *American Economic Journal: Economic Policy* 1 (1): 75–105.
- Heinrich, Carolyn, Alessandro Maffioli, and Gonzalo Vásquez. 2010. "A Primer for Applying Propensity-Score Matching. Impact-Evaluation Guidelines." Technical Note IDB-TN-161, Inter-American Development Bank, Washington, DC.

- 
- Jalan, Jyotsna, and Martin Ravallion. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching." *Journal of Business & Economic Statistics* 21 (1): 19–30.
- Mu, Ren, and Dominique Van de Walle. 2011. "Rural Roads and Local Market Development in Vietnam." *Journal of Development Studies* 47 (5): 709–34.
- Rosenbaum, Paul. 2002. *Observational Studies*, second edition. Springer Series in Statistics. New York: Springer-Verlag.
- Rosenbaum, Paul, and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies of Causal Effects." *Biometrika* 70 (1): 41–55.



معالجة التحديات المنهجية

تأثيرات المعالجة غير المتجانسة

لقد رأينا أن معظم طرق تقييم الأثر تنتج تقديرات صالحة للواقع المضاد في ظل افتراضات معيّنة فقط. ويمكن الخطر الرئيسي في تطبيق طريقة ما أن الافتراضات الكامنة لا تنطبق بصورة صحيحة، ما يؤدي إلى تقديرات متحيزة لأثر البرنامج. ولكن، ثمة أيضًا عدد من المخاطر الشائعة في معظم الأساليب التي ناقشناها. وسنناقش المخاطر الرئيسية في هذا الفصل.

ينشأ نوع واحد من المخاطر إذا كنت تقدّر أثر برنامج على مجموعة بأكملها، وكانت نتائجك تحجب بعض الاختلافات في الاستجابات للمعالجة بين مختلف المتلقين، أي عدم تجانس آثار المعالجة. تفترض معظم طرق تقييم الأثر أن البرامج تؤثر على الحاصل بطريقة بسيطة خطية لجميع وحدات المجتمع الإحصائي. وإذا كنت تعتقد أن مجتمعات إحصائية فرعية مختلفة قد تكون تعرضت لتأثير برنامج على نحو مختلف تمامًا، فقد يتعين عليك التفكير في الحصول على عينات مرتبة حسب كل مجتمع إحصائي فرعي. هب على سبيل المثال أنك مهتم بمعرفة أثر برنامج الوجبات المدرسية على الفتيات، ولكن ١٠ بالمائة من الطلاب فقط فتيات. في تلك الحالة، قد لا تحتوي حتى عينة عشوائية كبيرة من الطلاب على عدد كافٍ من الفتيات يسمح لك بتقدير أثر البرنامج على الفتيات. بالنسبة إلى تصميم عينة تقييمك، ستحتاج إلى تقسيم العينة إلى طبقات استنادًا إلى النوع الاجتماعي وتضمن عدد كبير بما فيه الكفاية من الفتيات للسماح لك بكشف حجم أثر معيّن.

الآثار السلوكية غير المقصودة

عند إجراء تقدير أثر، يمكنك أيضًا تحفيز استجابات سلوكية غير مقصودة من المجتمع الإحصائي الذي تدرسه، على النحو التالي:

- يحدث **تأثير هوثرون** عندما يتسبب مجرد مراقبتك للوحدات في تصّرف تلك الوحدات على نحو مختلف (انظر المربع ١-٩).
- يحدث **تأثير جون هنري** عندما تسعى وحدات المقارنة جاهدة للتعويض عن كونها غير مشمولة في المعالجة. (انظر المربع ١-٩).
- يمكن أن يؤدي **التوقع** إلى أثر سلوكي غير مقصود من نوع آخر. في الطرح العشوائي، يمكن أن تتوقع الوحدات في مجموعة المقارنة تلقي البرنامج في المستقبل والبدء في تغيير سلوكها قبل أن يصلها البرنامج بالفعل.

المربع ١-٩: القصة الشعبية عن تقييم الأثر: تأثير هوثرون وتأثير جون هنري

أصبحت لاحقًا موضوعًا مثيرًا للجدل وكانت إلى حد ما فاقدة للمصداقية، بقي مصطلح أثر هوثرون مستخدمًا. صاغ جاري ساريتسكي مصطلح **أثر جون هنري** عام ١٩٧٢ للإشارة إلى البطل الشعبي الأسطوري الأمريكي جون هنري، وهو "حفار فولاذ" كانت توكل إليه مهمة حفر فتحات مخصصة للمتفجرات في الصخور بحفّار فولاذي خلال إنشاء السكك الحديدية. ووفقًا للأسطورة، عندما عرف الرجل أنه كان يُقارن بالحفارة البخارية، أصبح يعمل بجد أكبر للتفوق على الآلة. وللأسف، مات نتيجة لذلك. ولكن المصطلح بقي قيد الاستخدام للإشارة إلى كيفية عمل وحدات المقارنة بجد أكثر للتعويض عن عدم منحها المعالجة.

يشير مصطلح **تأثير هوثرون** إلى التجارب التي أجريت من عام ١٩٢٤ إلى عام ١٩٣٢ في Hawthorne Works، وهو معمل تجهيزات كهربائية في ولاية إلينوي الأمريكية. اختبرت التجارب أثر تغير ظروف العمل (مثل زيادة شدة الضوء أو إنقاصها) على إنتاجية العمال، ووجدت أن أي تغييرات في ظروف العمل (زيادة الإضاءة أو إنقاصها، أو زيادة فترات الراحة أو إنقاصها، وما شابه ذلك) أدى إلى زيادة الإنتاجية. فسّر هذا على أنه تأثير الملاحظة: رأى العمال الذين كانوا جزءًا من التجربة أنفسهم مميزين، وزادت إنتاجيتهم بسبب هذا وليس بسبب التغيير في ظروف العمل. وعلى الرغم من أن التجارب الأصلية

المصادر: Landsberger 1958; Levitt and List 2009; Saretsky 1972.

- *تحيز الاستبدال* هو تأثير سلوكي آخر يؤثر على مجموعة المقارنة: ربما تتمكن المجموعات التي لم تُختَر لتلقي البرنامج من العثور على بدائل جيدة من خلال مبادرتها الشخصية.

إن الاستجابات السلوكية التي تؤثر على نحو غير متناسب على مجموعة المقارنة تمثل مشكلة لأنها قد تقوض الصلاحية الداخلية لنتائج التقييم، حتى إذا كنت تستخدم التعيين العشوائي كطريقة للتقييم. لا تمثل مجموعة المقارنة التي تعمل بجد أكثر للتعويض عن عدم تقديم المعالجة إليها، أو التي تغير سلوكها ترقبًا للبرنامج، الواقع المضاد تمثيلًا جيدًا.

إذا كان لديك سبب يدعوك للاعتقاد بأن هذه الاستجابات السلوكية غير المقصودة قد تكون موجودة، فإن البناء في مجموعات مقارنة إضافية لا تتأثر تمامًا بالتدخل هو خيارٌ أحيانًا - خيار يسمح لك باختبار هذه الاستجابات صراحةً. وقد يكون من الجيد جمع بيانات نوعية لفهم الاستجابات السلوكية على نحو أفضل.

الامتثال غير المثالي

الامتثال غير المثالي هو التعارض بين حالة المعالجة المعيّنة وحالة المعالجة الفعلية. ويحدث الامتثال غير المثالي عندما لا تتلقى بعض الوحدات المعيّنة في مجموعة الامتثال المعالجة، أو حين لا تشارك بعض الوحدات المعيّنة في مجموعة المقارنة في البرنامج. ناقشنا في الفصل ٥ الامتثال غير المثالي بالإشارة إلى التعيين العشوائي، ولكن يمكن أن يحدث الامتثال غير المثالي أيضًا في تصميم انقطاع الانحدار (كما يناقش الفصل ٦) وفي الاختلاف في الاختلافات (الفصل ٧). قبل أن تتمكن من تفسير تقديرات الأثر التي تنتجها أي طريقة، تحتاج إلى معرفة ما إذا كان الامتثال غير المثالي قد حدث في البرنامج. يمكن أن يحدث الامتثال غير المثالي بمجموعة من الطرق:

- لا يشارك جميع المشاركين المقصودين في البرنامج فعليًا. في بعض الأحيان، تختار الوحدات المعينة في برنامج ما عدم المشاركة.
- يُستثنى بعض المشاركين المقصودين من البرنامج بسبب أخطاء إشرافية أو تنفيذية.
- يُقدّم البرنامج بالخطأ لبعض وحدات مجموعة المقارنة وتشارك فيه.
- تتمكن بعض وحدات مجموعة المقارنة المشاركة في البرنامج، على الرغم من عدم تقديمه إليها.
- يُعيّن البرنامج استنادًا إلى مؤشر الأهلية المستمرة، ولكن لا يُفرض انقطاع الأهلية بصرامة.

• **الهجرة الاختيارية** هي هجرة تحدث استنادًا إلى حالة المعالجة. على سبيل المثال، يمكن أن يقارن التقييم الحاصل للبلديات المعالجة وغير المعالجة، ولكن يمكن أن يختار الأفراد الانتقال إلى بلدية أخرى إذا لم تعجبهم حالة المعالجة في بلديتهم.

عمومًا في ظل الامتثال غير المثالي، تنتج طرق تقييم الأثر القياسية تقديرات النية للعلاج (ITT). ولكن، يمكن استرداد متوسط تأثير المعالجة المكاني من تقديرات النية للعلاج باستخدام مقارنة المتغير المساعد. قدّمنا في الفصل ٥ هدف معالجة الامتثال غير المثالي في سياق التعيين العشوائي. وقد تمكّننا باستخدام تعديل النسبة المئوية للممثلين في عينة التقييم من استرداد متوسط تأثير المعالجة المكاني للممثلين من تقدير النية للعلاج. يمكن تمديد هذا "الإصلاح" ليشمل طرقًا أخرى من خلال تطبيق مقارنة متغير مساعد أعم. يضم المتغير المساعد مصدرًا خارجيًا للاختلاف يساعدك على توضيح أو تصحيح التحيز الذي قد ينبع من الامتثال غير المثالي. في حالة التعيين العشوائي بواسطة الامتثال غير المثالي، استخدمنا متغيرًا ١/٠ (ما يُدعى المتغير الصوري) يأخذ القيمة ١ إذا عُيّنَت الوحدة في الأصل في مجموعة المعالجة والقيمة ٠ إذا عُيّنَت الوحدة أصلًا في مجموعة المقارنة. في أثناء مرحلة التحليل، يُستخدم المتغير المساعد في سياق *انحدار ثنائي المراحل* يسمح لك بتحديد أثر المعالجة على الممثلين.

يمكن توسيع مقارنة المتغير المساعد في سياق طرق التقييم الأخرى:

- في سياق تصميم انقطاع الانحدار، المتغير المساعد الذي سنستخدمه متغير ١/٠ يشير إلى ما إذا كانت وحدة ما تقع في الجانب غير المؤهل أو الجانب المؤهل من درجة الانقطاع.
- وفي سياق الهجرة الانتقائية، سيكون أحد المتغيرات المساعدة لموضع الفرد بعد بداية البرنامج هو موقع الفرد قبل الإعلان عن البرنامج.

وعلى الرغم من إمكانية معالجة الامتثال غير المثالي باستخدام المتغيرات المساعدة، ثمة ثلاث نقاط من المهم تذكرها:

١. من وجهة نظر تقنية، من غير المحبذ مشاركة جزء كبير من مجموعة المقارنة في البرنامج. مع زيادة جزء مجموعة المقارنة التي تشترك في البرنامج، سينخفض جزء الممثلين في المجتمع الإحصائي، وسيكون متوسط تأثير المعالجة المحلي المقدّر باستخدام طريقة المتغير المساعد صالحًا فقط لجزء متقلص من السكان موضع الاهتمام. إذا تواصل ذلك لمدة أطول من اللازم، فقد تفقد النتائج دلالتها السياسية بالكامل، لأنها لن تعود تنطبق على جزء كافٍ من المجتمع الإحصائي موضع الاهتمام.

٢. وعلى العكس من ذلك، من غير المحبذ وجود جزء كبير غير مشترك من مجموعة المعالجة. مرة أخرى، إذا انخفض جزء مجموعة المعالجة الذي يشترك في البرنامج، فسينخفض جزء الممثلين في المجتمع الإحصائي. سيكون متوسط تأثير المعالجة المكاني مع طريقة المتغير المساعد صالحًا فقط لجزء متقلص من المجتمع الإحصائي موضع الاهتمام.

٣. كما ناقشنا في الفصل ٥، فإن طريقة المتغير المساعد تصلح فقط في ظروف محددة، وهي قطعًا ليست دلاً شاملاً.

التأثيرات غير المباشرة

التأثيرات غير المباشرة. هي مسألة أخرى قد تؤثر على التقييمات، سواء أكانت تستخدم التعيين العشوائي، أو تصميم انقطاع الانحدار، أو طريقة الاختلاف في الاختلافات. يحدث التأثير غير المباشر عندما يؤثر التدخل على غير المشاركين، وقد يكون إيجابياً أو سلبياً. هناك أربعة أنواع من التأثيرات غير المباشرة حسب أنجيلوتشي، ودي مارو (Angelucci and Di Maro 2015)

- **التأثيرات الخارجية.** هي التأثيرات التي تمتد من الأفراد المعالجين إلى الأفراد غير المعالجين. على سبيل المثال، فإن تلقيح الأطفال في قرية ضد الإنفلونزا يقلل من احتمالية إصابة السكان غير الملقحين في القرية نفسها بهذا المرض. هذا مثال عن تأثير خارجي إيجابي. وقد تكون التأثيرات الخارجية سلبية أيضاً. فعلى سبيل المثال، قد يتلف محصول أحد المزارعين عندما يضع جاره مبيدًا للأعشاب على قطعة أرضه وتنقل الرياح بعضًا من مبيد الأعشاب خارج نطاق أرضه.

- **التفاعلات الاجتماعية.** يمكن أن تنتج التأثيرات غير المباشرة عن التفاعلات المتبادلة الاقتصادية والاجتماعية بين المجتمعات الإحصائية المعالجة وغير المعالجة. ومن أمثلة ذلك مشاركة طالب يحصل على جهاز لوجي بمثابة جزء من برنامج تعزيز التعليم لهذا الجهاز مع طالب آخر لا يشارك في البرنامج.

- **تأثيرات توازن السياق.** تحدث هذه التأثيرات عندما يؤثر تدخل ما على المعايير السلوكية أو الاجتماعية ضمن سياق معين، مثل المكان الذي تُجرى فيه المعالجة. فعلى سبيل المثال، يمكن أن تؤثر زيادة حجم الموارد التي تتلقاها المراكز الصحية المعالجة حتى تتمكن من توسيع نطاق خدماتها على توقعات السكان بشأن ما ينبغي أن يكون عليه نطاق الخدمات المقدمة في جميع المراكز الصحية.

• **تأثيرات التوازن العام.** تحدث هذه التأثيرات عندما تؤثر التدخلات في العرض والطلب للسلع أو الخدمات، وبالتالي تغير سعر السوق لتلك السلع أو الخدمات. على سبيل المثال، قد يزيد برنامج يمنح النساء الفقيرات قسائم لاستخدام المرافق الخاصة من أجل الولادة من الطلب على الخدمات في المرافق الخاصة فجأة، ما يزيد من سعر الخدمة لجميع الأشخاص الآخرين. يقدم المربع ٩-٢ مثالاً عن تأثيرات غير مباشرة سلبية بسبب تأثيرات التوازن العام في سياق أحد برامج التدريب على الوظائف.

إذا كان غير المشارك الذي يعيش تأثيراً غير مباشر عضواً من مجموعة المقارنة، فإن التأثير غير المباشر ينتهك المتطلب الأساسي الذي مفاده أن حصيلة وحدة ما ينبغي ألا تتأثر بتعيين معين للمعالجة في الوحدات الأخرى. وإن افتراض قيمة معالجة الوحدة المستقرة (SUT- Stable Unit Treatment Value Assumption) (VA) أساسي لضمان إعطاء التعيين العشوائي تقديرًا غير متحيز للأثر. بديهيًا، إذا تأثرت مجموعة المقارنة بصورة غير مباشرة بالمعالجة التي تتلقاها مجموعة المعالجة (على سبيل المثال، يستعير طلاب المقارنة الأجهزة اللوحية من الطلاب المعالجين)، فعندئذٍ لا تمثل المقارنة بدقة الأثر الذي كان سيحدث لمجموعة المعالجة في غياب المعالجة (الواقع المضاد).

المربع ٩-٢ التأثيرات غير المباشرة السلبية نتيجة آثار التوازن العام: المساعدة في التوظيف وحاصل سوق العمل في فرنسا

عليها الباحث عن وظيفة الذي حصل على استشارة. ولاستقصاء هذا الافتراض، أجروا تجربة عشوائية تتضمن ٢٣٥ سوق عمل (كالمدن) في فرنسا. عُنيت أسواق العمل هذه عشوائيًا في واحدة من خمس مجموعات تتباين من حيث نسبة الباحثين عن وظيفة والذين سيعينون في معالجة الاستشارة (٠ بالمائة، و٢٥ بالمائة، و٥٠ بالمائة، و٧٥ بالمائة، و١٠٠ بالمائة). ضمن سوق العمل، عُنيت الباحثون المؤهلون عن وظائف في المعالجة باتباع هذا التناسب. وبعد ثمانية أشهر، وجد الباحثون أن الشباب العاطلين عن العمل الذين عُنيتوا في البرنامج كانوا أكثر قابلية بكثير للعثور على وظيفة مستقرة ممن لم يُعِينوا. ولكن، يبدو أن هذه المكاسب قد أتت على حساب عاملين مؤهلين لم يستفيدوا من البرنامج.

إن برامج المساعدة في التوظيف شائعة في الكثير من البلدان الصناعية. تتعاقد الحكومات مع كيان من طرف ثالث لمساعدة العاملين العاطلين عن العمل في البحث عن وظائف. وقد توصلت الكثير من الدراسات إلى أن هذه البرامج الاستشارية لها أثر إيجابي مهم وكبير على الباحثين عن وظائف.

درس كريون وآخرون (٢٠١٣) ما إذا كان لتقديم المساعدة الوظيفية للشباب المتعلمين الباحثين عن وظيفة في فرنسا آثار سلبية على الباحثين الآخرين عن وظيفة لم يدعمهم البرنامج. وقد افترضوا أنه قد تكون هناك آلية تأثير غير مباشر ناشطة: عندما لا يشهد سوق العمل الكثير من النمو، فإن مساعدة باحث عن وظيفة على العثور على وظيفة قد يكون ضارًا بباحث آخر عن وظيفة كان سيحصل على الوظيفة التي حصل

المصدر: Crépon and others 2013.

إذا لم يكن غير المشترك الذي يعايش تأثيرات غير مباشرة عضوًا في مجموعة المقارنة، فإن افتراض قيمة معالجة الوحدة المستقرة سيسري، وستبقى مجموعة المعالجة توفر تقديرًا جيدًا للواقع المضاد. ولكن، ما زلنا نحتاج إلى قياس التأثير غير المباشر لأنه يمثل أثرًا حقيقيًا للبرنامج. بمعنى آخر، فإن مقارنة حصائل مجموعتي المعالجة والمقارنة من شأنه أن يؤدي إلى تقديرات غير متحيزة لتأثير المعالجة على مجموعة المعالجة، ولكن هذا لن يراعي أثر البرنامج على المجموعات الأخرى.

يقدم كريمر وميغيل (Kremer and Miguel 2004)، اللذان درساً أثر تقديم علاج القضاء على الديدان للأطفال في المدارس الكينية (المربع ٣-٩)، مثالاً تقليدياً عن التأثيرات غير المباشرة الناتجة عن تأثيرات خارجية. الديدان المعوية طفيليات يمكن أن تنتقل من شخص إلى آخر من خلال ملامسة مادة برازية ملوثة. عندما يتلقى طفل دواء القضاء على الديدان، فسينخفض حمل الديدان لديه، وسينخفض أيضاً حمل الديدان للأشخاص الذين يقطنون نفس البيئة، وذلك لأنه لن يحدث بعد ذلك تلامس بينهم وبين ديدان الطفل. وهكذا، في مثال كينيا، عند تقديم الدواء إلى الأطفال في مدرسة ما، فإنه لن يفيد هؤلاء الأطفال فقط (فائدة مباشرة) بل سيفيد أيضاً الأطفال في المدارس المجاورة (فائدة غير مباشرة).

كما هو ممثل في الشكل ٩-١، فإن القضاء على الديدان في المجموعة أ من المدارس يقلل أيضاً من عدد الديدان التي تؤثر على الأطفال الذين لن يلتحقوا بالمدرسة أ. بالتحديد، يقلل عدد الديدان التي تؤثر على الأطفال الذين يلتحقون بمدارس مجموعة المقارنة ب، التي تقع بالقرب من مدارس المجموعة أ. ولكن مدارس المقارنة البعيدة عن مدارس المجموعة أ—التي تدعى مدارس المجموعة ج—لا تعايش مثل هذه التأثيرات غير المباشرة لأن الدواء الذي يُعطى في المجموعة أ لا يقتل أي ديدان تؤثر على الأطفال الذين يلتحقون بمدارس المجموعة ج. يناقش المربع ٣-٩ التقييم ونتائجه بمزيد من التفصيل.

تصميم تقدير أثر يراعي التأثيرات الثانوية

هَب أنك تصمم تقدير أثر لبرنامج تعتقد أنه ستحدث فيه تأثيرات غير مباشرة. كيف ينبغي التعامل مع ذلك؟ الشيء الأول الذي ينبغي فعله هو إدراك أنه يتعين توسيع هدف التقييم لديك. وفي حين أن التقييم القياسي يهدف إلى تقدير التأثير (أو التأثير السببي) لبرنامج ما على حصيلة موضع اهتمام للوحدات التي تتلقى المعالجة، فإن التقييم ذا التأثيرات غير المباشرة يجب أن يجيب عن سؤالين اثنين:

١. سؤال التقييم القياسي للأثر المباشر. ما الأثر (التأثير السببي) لبرنامج على حصيلة محل اهتمام للوحدات التي تتلقى المعالجة؟ هذا هو الأثر المباشر للبرنامج على الوحدات المعالجة.

المربع ٣-٩: العمل مع التأثيرات غير المباشرة: القضاء على الديدان، والعوامل الخارجية، والتعليم في كينيا

وتلقي نسبة مئوية صغيرة جدًا من وحدات مجموعة المقارنة للمعالجة)، تمكن الباحثان أيضًا من الاستفادة من مزايا عدم الامتثال لتحديد التأثيرات الخارجية أو التأثيرات غير المباشرة.

كان التأثير المباشر للتدخلات هو تقليل الإصابة المتوسطة إلى الشديدة بالديدان بنسبة ٢٦ نقطة مئوية للطلاب الذين تناولوا دواء الديدان. في حين انخفضت الإصابات المتوسطة إلى الشديدة بين الطلاب الذين كانوا يلتحقون بمدارس المعالجة ولكنهم لم يتناولوا الدواء بمقدار ١٢ نقطة مئوية عبر التأثير غير المباشر. وكانت هناك تأثيرات خارجية أيضًا بين المدارس.

ونظرًا إلى أن تكلفة معالجة الديدان منخفضة جدًا، وأن تأثيرات الصحة والتعليم مرتفعة نسبيًا، استنتج الباحثون أن القضاء على الديدان طريقة فعالة للتكاليف نسبيًا لتحسين معدلات المشاركة في المدارس. توضح الدراسة أيضًا أن الأمراض الاستوائية، كالديدان، قد تؤدي دورًا مهمًا في الحوائل التعليمية وتعزز الادعاءات بأن عبء المرض المرتفع في أفريقيا قد يساهم في انخفاض دخلها. وهكذا يجادل كريمر وميغيل بأن الدراسة تقدم حجة قوية للإعانات العامة لعلاج الأمراض ذات المنافع غير المباشرة المماثلة في البلدان النامية.

صُمم مشروع القضاء على أمراض الديدان في المدارس الابتدائية في بوسيا، كينيا، لاختبار مجموعة من جوانب معالجة الديدان والوقاية منها. وقد نُفذته منظمة دعم الأطفال الدولية الهولندية غير الربحية في إفريقيا، بالتعاون مع وزارة الصحة. شمل البرنامج ٧٥ مدرسة بلغ الالتحاق الكلي فيها ٣٠٠٠٠ طالب بين ٦ و١٨ عامًا. عولج الطلاب بدواء للديدان وفقًا لتوصيات منظمة الصحة العالمية وتلقوا أيضًا تعليمًا حول الوقاية من الديدان في شكل محاضرات صحية، ومخططات جدائية، وتدريب للمعلمين.

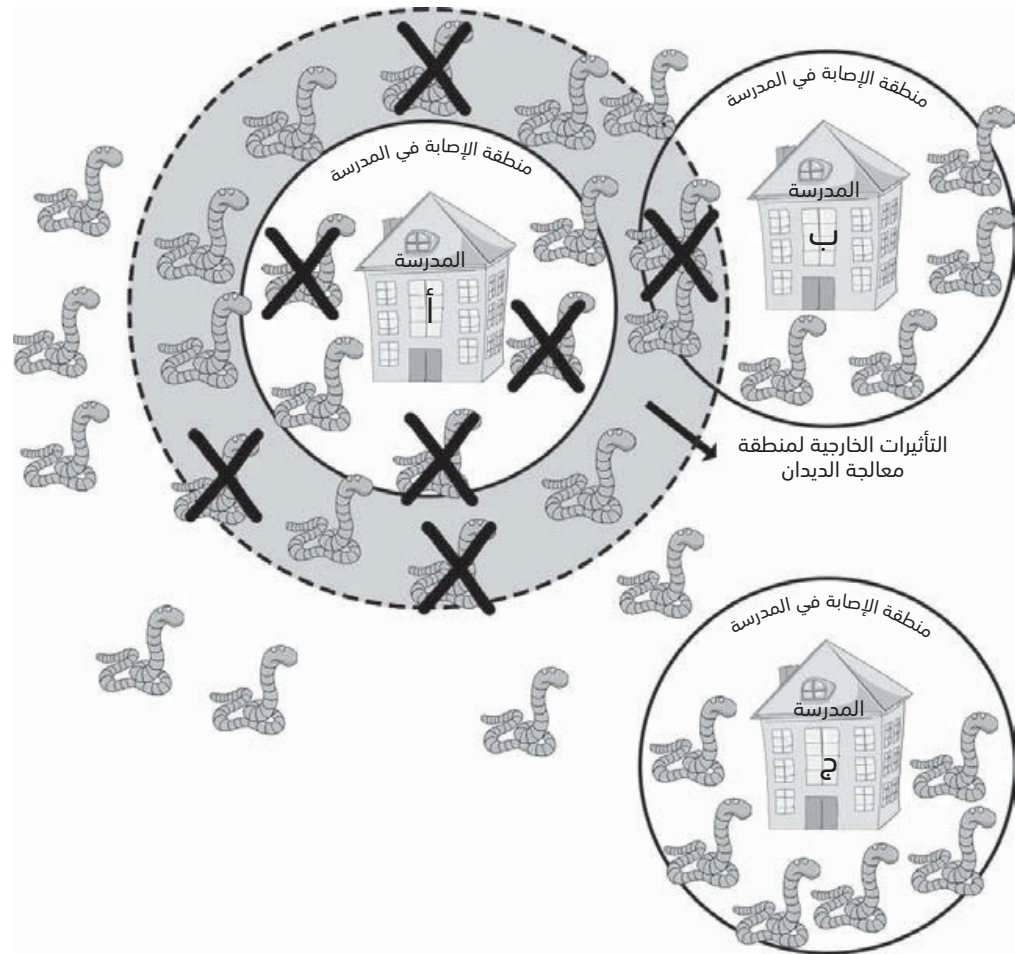
ونظرًا لمحددات إدارية ومالية، جرى الطرح على مراحل وفقًا للترتيب الأبجدي لأسماء المدارس، حيث تم البدء بمجموعة المدارس الأولى البالغة ٢٥ مدرسة عام ١٩٩٨، تلتها المجموعة الثانية عام ١٩٩٩، ثم الثالثة عام ٢٠٠١. ومن خلال التوزيع العشوائي على مستوى المدارس، تمكّن كريمر وميغيل (٢٠٠٤) من تقييم كل من أثر القضاء على الديدان على المدارس وتحديد التأثيرات غير المباشرة عبر المدارس باستخدام الاختلاف الخارجي قرب مدارس المقارنة على مدارس المعالجة. وعلى الرغم من أن الامتثال للتصميم العشوائي كان مرتفعًا نسبيًا (مع تلقي ٧٥ بالمائة من الطلاب المعيّنين للمعالجة لدواء الديدان،

المصدر: Kremer and Miguel 2004.

٢. *سؤال تقييم إضافي للأثر غير المباشر.* ما الأثر (التأثير السببي) لبرنامج على حصيلة محل اهتمام للوحدات التي لا تتلقى المعالجة؟ هذا هو الأثر غير المباشر للبرنامج على الوحدات غير المعالجة.

لتقدير الأثر غير المباشر على المجموعات المعالجة، سيتعين عليك اختيار مجموعة المقارنة بحيث لا تتأثر بالتأثيرات الثانوية. على سبيل المثال، قد تحتاج إلى أن تكون قرى المعالجة والمقارنة أو العيادات أو الأسر واقعة على مسافة بعيدة بما فيه الكفاية من بعضها البعض بحيث تكون التأثيرات غير المباشرة غير محتملة.

الشكل ١٠-٩ مثال تقليدي على التأثيرات غير المباشرة: العوامل الخارجية الإيجابية للقضاء على الديدان لدى أطفال المدارس



لتقدير الأثر غير المباشر على المجموعات غير المعالجة، عليك تحديد مجموعة مقارنة إضافية لكل مجموعة غير معالجة قد تتأثر بالتأثيرات غير المباشرة. على سبيل المثال، قد يجري العاملون الصحيون المجتمعيون زيارات منزلية لتوفير المعلومات للآباء حول فوائد التنوع الغذائي المحسن لأطفالهم. لنفترض أن العاملين الصحيين المجتمعيين يزورون فقط بعض الأسر في أي قرية معينة. قد تهتم في التأثيرات غير المباشرة على الأطفال في الأسر التي لم تتم زيارتها، وفي تلك الحالة ستحتاج إلى إيجاد مجموعة مقارنة لهؤلاء الأطفال. وفي الوقت نفسه، قد يكون من الممكن أن يؤثر التدخّل أيضًا على التنوع الغذائي للبالغين. إذا كان هذا الأثر غير المباشر موضع اهتمام للتقييم، فستكون مجموعة مقارنة مطلوبة أيضًا بين البالغين. ومع زيادة عدد قنوات التأثيرات غير المباشرة المحتملة، يمكن أن يبدأ التصميم في أن يصبح معقدًا نوعًا ما بسرعة.

تطرح التقييمات مع التأثيرات غير المباشرة تحديات محددة. أولاً، عندما تكون التأثيرات غير المباشرة محتملة، من المهم فهم آلية التأثيرات غير المباشرة: البيولوجية أو الاجتماعية أو البيئية أو ما شابهها. وإذا لم نكن نعرف ما هي آلية التأثير غير المباشر، فلن نتمكن من اختيار مجموعات مقارنة بدقة تتأثر بالتأثيرات غير المباشرة أو لا تتأثر بها. ثانيًا، يتطلب التقييم مع التأثيرات غير المباشرة جمع بيانات أكثر شمولاً من التقييم الذي لا يمثل ذلك فيه شاغلًا: هناك مجموعة مقارنة إضافية (في المثال السابق، القرى المجاورة). وقد تحتاج أيضًا إلى جمع بيانات عن الوحدات الإضافية (في المثال السابق، البالغون في الأسر المستهدفة بزيارات التغذية للأطفال). يتناول المربع ٩-٤ كيف عالج الباحثون التأثيرات غير المباشرة في تقييم برنامج تحويلات نقدية مشروطة في المكسيك.

المربع ٩-٤: تقييم التأثيرات غير المباشرة: التحويلات النقدية المشروطة والتأثيرات غير المباشرة في المكسيك

قرى المقارنة توفر واقعًا مضادًا صالغًا للأسر المؤهلة في قرى المعالجة، لغرض تقدير التأثيرات غير المباشرة داخل القرية للأسر غير المؤهلة. توصل الباحثان إلى دليل على تأثيرات غير مباشرة موجبة للاستهلاك. زاد استهلاك أغذية البالغين بنسبة ١٠ بالمائة تقريبًا في الشهر لدى الأسر غير المؤهلة في قرى المعالجة. وكان ذلك يوازي تقريبًا نصف متوسط الزيادة في استهلاك الأغذية بين الأسر المؤهلة. وقد دعمت النتيجة أيضًا فرضية مشاركة المخاطر في القرى. كانت الأسر غير المؤهلة في قرى المعالجة تتلقى قروضًا وتحويلات أكبر من الأسرة والأصدقاء بالمقارنة مع الأسر غير المؤهلة في قرى المقارنة. وهذا يعني أن التأثير غير المباشر يحدث خلال أسواق التأمين والائتمان المحلية. استنادًا إلى هذه النتائج، استنتج أنجيلوتشي ودي جيورجي أن التقييمات السابقة لبرنامج Progresha قللت من تأثير البرنامج بنسبة ١٢ بالمائة لأنها لم تكن تأخذ في الحسبان التأثيرات غير المباشرة على الأسر غير المؤهلة داخل قرى المعالجة.

درس أنجيلوتشي ودي جيورجي (٢٠٠٩) التأثيرات غير المباشرة في برنامج Progresha في المكسيك، الذي كان يقدم تحويلات نقدية شرطية للأسر (انظر المربعين ١-١ و ٢-٤). سعى الباحثان إلى استكشاف ما إذا كانت هناك مشاركة للمخاطر ضمن القرى. فإذا شاركت الأسر المخاطر، فقد تكون الأسر المؤهلة تحوّل جزءًا من التحويلات النقدية إلى أسر غير مؤهلة عبر القروض أو الهدايا.

نُفذ برنامج Progresha على مراحل على مدى عامين، حيث اختيرت ٣٢٠ قرية عشوائيًا لتلقي التحويلات النقدية في عام ١٩٩٨، و١٨٦ في عام ١٩٩٩. وهكذا، بين عامي ١٩٩٨ و ١٩٩٩، كانت هناك ٣٢٠ قرية معالجة و ١٨٦ قرية مقارنة. ضمن قرى المعالجة، حُدّدت أهلية الأسرة لعمليات التحويل وفق برنامج Progresha استنادًا إلى حالة الفقر، وجمعت بيانات التعداد لكلتا المجموعتين. وقد أنشأ ذلك أربع مجموعات فرعية ضمن العينة: فئات سكانية مؤهلة وغير مؤهلة ضمن قرى المعالجة والمقارنة. وعلى فرض أن البرنامج لم يؤثر على قرى المقارنة تأثيرًا غير مباشر، فقد كانت الأسر المؤهلة في

المصدر: Angelucci and De Giorgi 2009.

تَحَيِّزُ *التناقص* مسألة شائعة أخرى قد تؤثر على التقييمات، سواء أكانت تستخدم التعيين العشوائي، أو تصميم انقطاع الانحدار، أو طرق الاختلاف في الاختلافات. يحدث التناقص عندما تحتفي أجزاء من العينة بمرور الزمن، ولا يستطيع الباحثون العثور على جميع الأعضاء الأوليين لمجموعتي المعالجة والمقارنة في مسوحات المتابعة أو البيانات. على سبيل المثال، من بين ٢٥٠٠ أسرة شملها المسح في خط الأساس، يمكن للباحثين العثور على ٢٣٠٠ أسرة فقط في مسح متابعة بعد ذلك بعامين. وإذا عاد الباحثون وحاولوا إعادة مسح المجموعة نفسها، على سبيل المثال بعد ١٠ سنوات، فقد يتمكنون من العثور على عدد أقل حتى من الأسر الأصلية.

قد يحدث التناقص لأسباب عديدة. على سبيل المثال، قد ينتقل أعضاء في الأسر، أو حتى أسر بأكملها، إلى قرية أو مدينة أو منطقة أو حتى دولة أخرى. في مثال حديث لمتابعة طويلة المدى لتدخل إنمائي في مرحلة الطفولة المبكرة في جامايكا، في مسح المتابعة لمدة ٢٢ عامًا، هاجرت نسبة ١٨ بالمائة من العينة إلى الخارج (انظر المربع ٩-٥). وفي حالات أخرى، ربما لا يعود المستجيبون يرغبون في الإجابة عن مسح إضافي. أو ربما تمنع النزاعات أو انعدام الأمان في المنطقة فريق البحث من إجراء مسح في بعض المواقع التي كان يتضمنها خط الأساس.

يمكن أن يكون التناقص إشكاليًا لسببين. أولاً، ربما لا تعود عينة المتابعة تمثل بعد الآن المجتمع الإحصائي موضع الاهتمام. تذكر أنه عندما نختار العينة في وقت التعيين العشوائي، فإننا نختارها بحيث تمثل بدقة المجتمع الإحصائي موضع الاهتمام. بمعنى آخر، نختار عينة ذات صلاحية خارجية للمجتمع الإحصائي موضع الاهتمام. إذا كان مسح المتابعة أو جمع البيانات مشوبًا بتناقص كبير، فقد يقلقنا ذلك من أن عينة المتابعة قد تمثل فقط مجموعة فرعية محددة من المجتمع الإحصائي موضع الاهتمام. على سبيل المثال، إذا كان الأشخاص الأكثر تعليمًا في العينة الأصلية هم أيضًا الذين يهاجرون، فإن مسح المتابعة الذي نجريه يفوت أولئك المتعلمين، ولا يمثل بعد الآن بدقة المجتمع الإحصائي موضع الاهتمام، الذي يتضمن أولئك الأشخاص المتعلمين.

ثانيًا، قد لا تكون عينة المتابعة متوازنة بين مجموعتي المعالجة والمقارنة. افترض أنك تحاول تقييم برنامج يسعى إلى تعزيز تعليم الفتيات، وأن الفتيات المتعلّمات هن الأكثر قابلية للانتقال إلى المدينة بحثًا عن عمل. ومن ثم قد يُظهر مسح المتابعة الذي تجريه تناقصًا مرتفعًا على نحوٍ غير متناسب في مجموعة المعالجة، مقارنة بمجموعة المقارنة. وقد يؤثر ذلك على الصلاحية الداخلية للبرنامج: بمقارنة وحدتي المعالجة والمقارنة اللتين تجدهما عند المتابعة، لن يعود بإمكانك تقدير أثر البرنامج بدقة.

المربع ٩-٥: التناقص في الدراسات مع المتابعة طويلة المدى: تنمية الطفولة المبكرة والهجرة في جامايكا

ولكن، عند التفكير في مجموعة فرعية لأطفال أصبحوا عمالاً مهاجرين، كانت هناك إشارات للتناقص الانتقائي. من بين العمال المهاجرين الـ ٢٣، تسرّب تسعة من العينة، وكان جزء أكبر بكثير منهم ينتمي إلى مجموعة المعالجة. أوضح ذلك أن المعالجة كانت مرتبطة بالهجرة. بما أن العمال المهاجرين كانوا يكسبون أموالاً أكثر ممن بقوا في جامايكا، صعب ذلك من تقدير الآثار.

ولمعالجة التحيز المحتمل نتيجة التناقص بين العمال المهاجرين، استخدم الباحثون تقنيات اقتصادية قياسية. وقد تنبؤوا بأرباح المهاجرين الذين كانوا قد تسربوا من العينة من خلال انحدار المبيعات الصغرى العادية (OLS) باستخدام حالة المعالجة، والنوع الاجتماعي، والهجرة بمثابة محددات. توصل الباحثون باستخدام هذه التنبؤات في تقدير الأثر إلى أن البرنامج أفضى إلى نتائج مذهلة. فقد رفع التدخل في مرحلة الطفولة المبكرة الأرباح بنسبة ٢٥ بالمائة لعينة المعالجة. وكان هذا الأثر كبيراً بما يكفي لكي تلحق مجموعة معالجة الأطفال متوقفي النمو مع مجموعة المقارنة للأطفال متوقفي النمو بعد ٢٠ عامًا.

يمكن أن يكون التناقص إشكاليًا بالتحديد حين تمر سنوات كثيرة بين مسح خط الأساس ومسح المتابعة. عام ١٩٨٦، بدأ فريق في جامعة ويست لامديز دراسة لقياس الحاصل طويلة المدى لعملية تدخل في مرحلة الطفولة المبكرة في جامايكا. عام ٢٠٠٨، أجريت دراسة متابعة عندما كان عمر المشاركين الأصليين ٢٢ عامًا. وكان من الصعب على الدراسة الأصلية تتبع جميع المشاركين.

كان التدخل برنامجًا لمدة عامين يوفر تحفيًا نفسيًا اجتماعيًا ومكملات غذائية للأطفال الصغار الذين يعانون من توقف النمو في كنجستون، جامايكا. وُزّع ما مجموعه ١٢٩ طفلًا توزيعًا عشوائيًا على واحد من ثلاثة أذرع علاجية أو مجموعة مقارنة. أجرى الباحثون أيضًا مسحًا على ٨٤ طفلًا يعاني من توقف النمو لمجموعة مقارنة ثانية. في المتابعة، تمكن الباحثون من إعادة مسح حوالي ٨٠ بالمائة من المشاركين. ولم يكن هناك دليل على التناقص الانتقائي في العينة بأكملها، ما يعني أنه لم تكون هناك فروق ذات دلالة في خصائص خط الأساس لأولئك الذين كان من الممكن إجراء مسح لهم في عمر ٢٢ عامًا بالمقارنة مع أولئك الذين تعذر إجراء مسح لهم.

المصدر: Gertler and others 2014; Grantham-McGregor and others 1991.

إذا توصلت إلى وجود تناقص خلال مسح متابعة، فيمكن أن يساعدك الاختبارات التالية على تقدير مدى المشكلة. أولاً، تحقق مما إذا كانت خصائص خط الأساس للوحدات التي تسربت من العينة مساوية إحصائيًا لخصائص خط الأساس للوحدات التي أعيد مسحها بنجاح. وطالما أن خصائص خط الأساس لكلتا المجموعتين ليست مختلفة إحصائيًا، ينبغي أن تظل عينتك الجديدة تمثل المجتمع الإحصائي موضع الاهتمام.

ثانيًا، تحقق مما إذا كان معدل التناقص في مجموعة المعالجة مماثلًا لمعدل التناقص في مجموعة المقارنة. فإذا كان معدل التناقص مختلفين كثيرًا، فثمة مخاوف بأن عينتك لم تعد صالحة، وقد تحتاج إلى استخدام تقنيات إحصائية مختلفة لمحاولة تصحيح ذلك. إحدى الطرق الشائعة *ترجيح الاحتمال العكسي*، وهي طريقة تعيد ترجيح البيانات إحصائيًا (في هذه الحالة، بيانات المتابعة) لتصحيح غياب جزء من المستجيبين الأصليين. تعيد هذه الطريقة ترجيح عينة المتابعة بحيث تبدو مشابهة لعينة خط الأساس^١.

توقيت التأثيرات واستمرارها

قد تتكون قنوات النقل المحتملة بين المدخلات، والأنشطة، والمخرجات، والحصائل على الفور، أو بعد وقت قصير، أو بتأخير زمني كبير، وهي ترتبط ارتباطًا وثيقًا في العادة بتغيرات في السلوك البشري. أكد الفصل ٢ على مدى أهمية التفكير في هذه القنوات والتخطيط لها قبل بدء التدخل، وتطوير سلسلة سببية واضحة للبرنامج الخاضع للتقييم. للتبسيط، كنا نهمل مسائل التوقيت. بيد أنه من المهم التفكير في الجوانب المرتبطة بالتوقيت عند تصميم التقييم.

أولًا، لا تصبح البرامج بالضرورة فعالة فعالية كاملة بعد بدئها على الفور (King and Behrman 2009). وقد يحتاج مديرو البرنامج إلى وقت لتشغيل البرنامج، وقد لا يستفيد المستفيدون على الفور لأن التغييرات السلوكية تستغرق وقتًا، وقد لا تعدّل المؤسسات سلوكها على الفور أيضًا. من ناحية ثانية، بمجرد أن تغير المؤسسات والمستفيدون سلوكيات محددة، فقد تتواصل هذه السلوكيات حتى إذا انقطع البرنامج. فمثلًا، قد يحافظ البرنامج الذي يحفز الأسر على فرز القمامة وإعادة تدويرها وتوفير الطاقة على فاعلته بعد إزالة الحوافز، إذا تمكن من تغيير المعايير المنزلية حول كيفية التعامل مع القمامة والطاقة. عندما تصمم تقييمًا، عليك أن تكون حذرًا (وواقعيًا) جدًّا في تقييم الزمن الذي سيستغرقه البرنامج لبلوغ فاعليته الكاملة. وقد يلزم إجراء العديد من مسوحات المتابعة لقياس أثر البرنامج بمرور الزمن، أو حتى بعد قطع البرنامج. يوضح المربع ٩-٦ تقييمًا توضحت فيه بعض الآثار فقط بعد قطع التدخل الأولي.

المربع ٩-٦: تقييم التأثيرات طويلة المدى: المعونات واعتماد الناموسيات المعالجة بمبيدات حشرية في كينيا

أن الأس. التي كانت تتلقى المعونة التي خفضت سعر تلك الشبكات من ٣,٨٠ دولارات إلى ٠,٧٥ دولار كانت أكثر قابلية بنسبة ٦٠٪ لشرائها. وعندما قُدمت الشبكات بالمجان، ازداد معدل اعتمادها إلى ٩٨ بالمائة. وعلى المدى الطويل، تحوّلت معدلات الاعتماد الأعلى إلى رغبة أكبر في الدفع، وذلك لأن الأسر رأت فوائد اقتناء ناموسية معالجة بمبيدات حشرية. كانت الأسر التي تلقت إحدى المعلومات الكبيرة في المرحلة الأولى أكثر قابلية بثلاثة أضعاف لشراء شبكة ثانية في المرحلة الثانية بسعر يزيد عن الضعف.

توضح نتائج هذه الدراسة حدوث أثر تعلّم في عمليات التدخل الخاصة بالناموسيات المعالجة بمبيدات حشرية. ويشير ذلك إلى أهمية التفكير في آثار عمليات التدخل على المدى الطويل، وكذلك كشف استمرار التأثيرات.

صممت دوباس (٢٠١٤) تقييماً للأثر بغرض قياس الآثار قصيرة المدى وطويلة المدى على الطلب على الناموسيات المعالجة بمبيدات حشرية في بوسيا، في كينيا. حددت Dupas، باستخدام تجربة تسعير على مرتلتين، الأسر في العديد من مستويات معونة نوع جديد من الناموسيات المعالجة بمبيدات حشرية. وبعد عام واحد، مُنحت جميع الأسر في مجموعة فرعية من القرى الفرصة لشراء الشبكة نفسها. وهذا سمح للباحثين بقياس استعداد الأسر للدفع مقابل الحصول على ناموسيات معالجة بمبيدات حشرية وكيف تغيّر هذا الاستعداد اعتماداً على المعونة المعطاة في المرحلة الأولى من البرنامج.

إجمالاً، أشارت النتائج إلى أن المعونة المقدمة مرة واحدة أحدثت أثراً إيجابياً جذاً على اعتماد الناموسيات المعالجة بمبيدات حشرية والاستعداد للدفع على المدى الطويل. في الطور الأول للمعالجة، توصلت دوباس إلى

المصدر: Dupas 2014.

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).

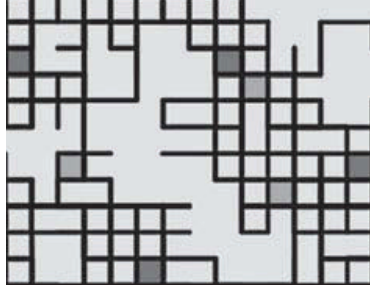
ملاحظة

١. ثمة طريقة إحصائية أكثر تقدماً تقود على تقدير "الحدود الحادة" لتأثيرات المعالجة (انظر Lee 2009).

المراجع

Angelucci, Manuela, and Giacomo De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption." *American Economic Review* 99 (1): 486–508.

- Angelucci, Manuela, and Vincenzo Di Maro. 2015. "Programme Evaluation and Spillover Effects." *Journal of Development Effectiveness*. doi: 10.1080/19439342.2015.1033441.
- Crépon, Bruno, Esther Dufl o, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* 128 (2): 531–80.
- Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187): 998–1001.
- Grantham-McGregor, Sally, Christine Powell, Susan Walker, and John Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation and Development of Stunted Children: The Jamaican Study." *Lancet* 338: 1–5.
- King, Elizabeth M., and Jere R. Behrman. 2009. "Timing and Duration of Exposure in Evaluations of Social Programs." *World Bank Research Observer* 24 (1): 55–82.
- Kremer, Michael, and Edward Miguel. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.
- Landsberger, Henry A. 1958. *Hawthorne Revisited*. Ithaca, NY: Cornell University Press.
- Lee, David. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–102.
- Levitt, Steven D., and John A. List. 2009. "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments." Working Paper 15016, National Bureau of Economic Research, Cambridge, MA.
- Saretsky, Gary. 1972. "The OEO P.C. Experiment and the John Henry Effect." *Phi Delta Kappan* 53: 579–81.



تقييم البرامج متعددة الأوجه

تقييم البرامج التي تجمع بين العديد من خيارات المعالجة

ناقشنا حتى الآن برامج تتضمن فقط معالجة من نوع واحد. في الواقع، تنشأ العديد من المسائل وثيقة الصلة بالسياسات في سياق البرامج متعددة الأوجه: أي البرامج التي تجمع بين العديد من خيارات العلاج.^١ وقد لا يهتم صانعو السياسات بمعرفة ما إذا كان البرنامج يعمل أم لا فحسب، ولكن أيضًا ما إذا كان البرنامج يعمل على نحو أفضل من برنامج آخر أو بتكلفة أقل. على سبيل المثال، إذا أردنا زيادة الالتحاق بالمدرسة، فهل من الأكثر فعالية تنفيذ تدخلات في جانب الطلب (مثل التحويلات النقدية للأسر) أو عمليات التدخل في جانب العرض (مثل زيادة حوافز المعلمين)؟ إذا أدخلنا عمليتي التدخل معًا، فهل تعملان على نحو أفضل من كل منهما منفردة؟ بمعنى آخر، هل هما مكملتان لبعضهما البعض؟ بدلاً من ذلك، إذا كانت الأولوية فعالية التكلفة للبرنامج، فقد تحتاج إلى تحديد المستوى الأمثل للخدمات التي ينبغي أن يحققه البرنامج. على سبيل المثال، ما المدى الأمثل لبرنامج تدريب مهني؟ هل لبرنامج يمتد ستة أشهر أثر أكبر على عثور المتدربين على وظائف من برنامج يمتد ثلاثة أشهر؟ وإذا كان الأمر كذلك، فهل هذا الفرق يكفي لتبرير الموارد الإضافية اللازمة لبرنامج يمتد ستة أشهر؟ وأخيرًا، قد يهتم صانعو السياسات في كيفية تعديل برنامج حالي لجعله أكثر فعالية، وقد يحتاجون إلى اختبار مجموعة من الآليات لإيجاد الآلية التي تعمل على النحو الأفضل.

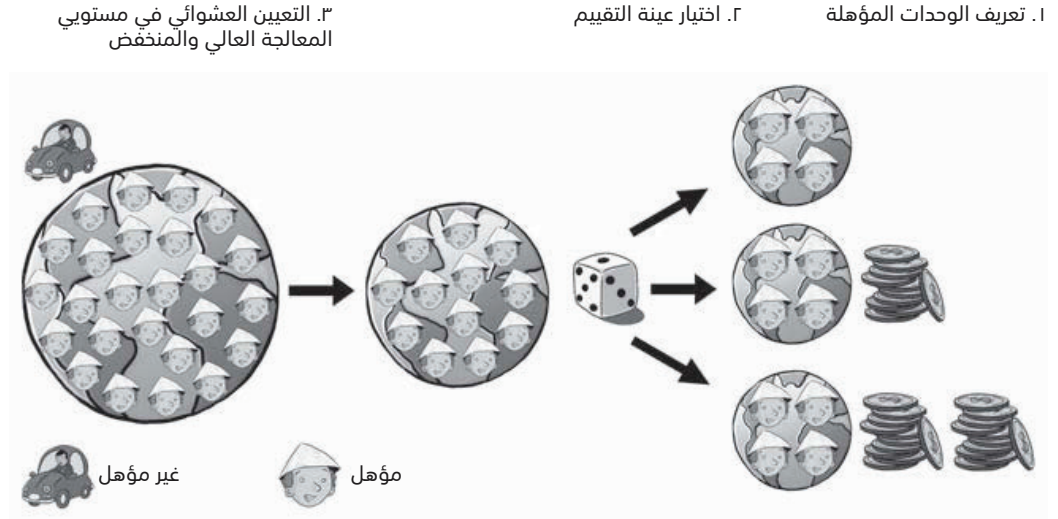
بخلاف مجرد تقدير أثر التدخل على حصيلة موضع اهتمام، يمكن أن تساعد تقييمات الأثر على الإجابة عن أسئلة أوسع، مثل الأسئلة التالية:

- ما أثر إحدى المعالجات مقارنة بأثر معالجة أخرى؟ على سبيل المثال، ما أثر برنامج يوفر تدريباً على التربية مقابل برنامج تغذية على التطور المعرفي للأطفال؟
 - هل الأثر المتضافر لمعالجة أولى ومعالجة ثانية أكبر من مجموع الأثرين الفرديين؟ على سبيل المثال، هل الأثر الكلي للتربية للتدخل التربوي والتدخل التغذوي أكبر من مجموع تأثيري التدخلين الفرديين أم أصغر منه أو يساويه؟
 - ما الأثر الإضافي لمعالجة أعلى كثافة مقارنة بمعالجة أقل كثافة؟ على سبيل المثال، ما الأثر على التطور المعرفي للأطفال متوقفي النمو إذا زارهم عامل اجتماعي في المنزل كل أسبوعين، بالمقارنة مع زيارتهم مرة واحدة في الشهر؟
- يقدم هذا الفصل أمثلة عن كيفية تصميم تقييمات الأثر لنوعين من البرامج متعددة الأوجه: برامج ذات مستويات متعددة للمعالجة نفسها، وأخرى ذات معالجات متعددة. أولاً، نناقش كيفية تصميم تقدير أثر لبرنامج ذي مستويات معالجة متعددة. ثم ننتقل إلى كيفية تفصيل الأنواع المختلفة لأثر برنامج ذي معالجات متعددة. تفترض المناقشة أننا نستخدم طريقة تعيين عشوائية، ولكن يمكن تعميمها على طرق أخرى.

تقييم برامج بمستويات معالجة متباينة

من السهل نسبياً تصميم تقدير أثر برنامج ذي مستويات معالجة متعددة. تخيل أنك تحاول تقدير أثر برنامج ذي مستويين للمعالجة: مرتفع (كالزيارات كل أسبوعين)، ومنخفض (كالزيارات الشهرية). تحتاج إلى تقدير أثر كلا الخيارين، وتحتاج أيضاً إلى معرفة مقدار تأثير الزيارات الإضافية على الحاصل. لفعل ذلك، يمكنك إجراء قرعة لتحديد من يتلقى المستوى العالي من المعالجة، ومن يتلقى المستوى المنخفض من المعالجة، ومن يُعَيَّن في مجموعة المقارنة. يوضح الشكل ١٠-١ هذه العملية.

الشكل ١٠-١ خطوات التعيين العشوائي لمستويين من المعالجة



كما في التعيين العشوائي القياسي، تقوم الخطوة ١ على تعريف المجتمع الإحصائي للوحدات المؤهلة لبرنامجك. وتقوم الخطوة ٢ على تحديد عينة عشوائية من الوحدات التي ينبغي تضمينها في التقييم، عينة التقييم. بمجرد أن تكون لديك عينة التقييم، تعيّن بعد ذلك في الخطوة ٣ الوحدات في المجموعة التي تتلقى المعالجة عالية المستوى، أو المجموعة التي تتلقى المعالجة منخفضة المستوى، أو مجموعة المقارنة. ونتيجة التعيين العشوائي في مستويات معالجة متعددة، ستكون قد أنشأت ثلاث مجموعات متميزة:

- المجموعة أ التي تشكل مجموعة المقارنة.
- المجموعة ب التي تتلقى المعالجة منخفضة المستوى.
- المجموعة ج التي تتلقى المعالجة عالية المستوى.

تضمن طريقة التعيين العشوائي، عند تطبيقها على النحو الصحيح، تشابه المجموعات الثلاث. لذلك، يمكنك تقدير تأثير المستوى العالي للمعالجة بمقارنة متوسط الحصلة للمجموعة ج مع متوسط الحصلة للمجموعة أ. ويمكنك أيضًا تقدير أثر المستوى المنخفض للمعالجة من خلال مقارنة متوسط الحصلة للمجموعة ب مع متوسط الحصلة للمجموعة أ. وأخيرًا، يمكنك تقييم ما إذا كان للمعالجة عالية المستوى أثر أكبر من المعالجة منخفضة المستوى من خلال مقارنة متوسط حصليتي المجموعتين "ب" و"ج".

يتبع تقدير أثر برنامج فيه أكثر من مستويي معالجة المنطق نفسه. إذا كانت هناك ثلاثة

مستويات للمعالجة، فستنشئ عملية التعيين العشوائي ثلاث مجموعات معالجة مختلفة، إضافةً إلى مجموعة مقارنة. عمومًا، بوجود مستويات معالجة مختلفة عددها n ، ستكون هناك مجموعات معالجة عددها t ، بالإضافة إلى مجموعة مقارنة. يقدم المربع ١٠-١ والمربع ١٠-٢ مثالين عن تقييمات الأثر التي تختبر أشكالًا مختلفة الشدة أو خيارات معالجة متعددة.

المفهوم الأساسي

المفهوم الأساسي عند تقييم برامج فيها العدد n من مستويات المعالجة المختلفة، ينبغي أن يكون هناك العدد n من مجموعات المعالجة زائد مجموعة مقارنة واحدة.

المربع ١٠-١ اختبار كثافة البرنامج لزيادة الالتزام بالمعالجة المضادة للفيروسات العكوسة

بينما كانت قيمة الفاصل من ٣/٢ إلى ١ تقابل مجموعة المقارنة.

توصل الباحثون إلى أن الرسائل الأسبوعية زادت من نسبة المرضى الذين يحققون التزامًا بنسبة ٩٠ بالمائة بالعلاج بالفيروسات العكوسة بنسبة ١٦-١٣ بالمائة تقريبًا، بالمقارنة مع غياب الرسائل. وكانت هذه الرسائل الأسبوعية أيضًا فعالة في الحد من تكرار انقطاعات المعالجة، التي تبين أنها سبب مهم لفشل مقاوم المعالجة في بيئات محدودة الموارد. وعلى عكس التوقعات، لم تكن إضافة عبارات تشجيعية في الرسائل الطويلة أكثر فعالية من الرسائل القصيرة ولا من عدم إرسال رسائل.

توصل الباحثون أيضًا إلى أنه بينما حسنت الرسائل الأسبوعية الالتزام، لم تحسنه الرسائل اليومية، ولكنهم لم يتمكنوا من إدراك السبب في أن الرسائل الأسبوعية كانت الأكثر فعالية. قد يفسر التعود، أو تناقص الاستجابة لحافز متكرر، هذه النتيجة، أو ربما كان المرضى يجدون ببساطة أن الرسائل اليومية تطفلية.

استخدم بوب إيتشيس وآخرون (٢٠١١) تصميمًا شاملاً متعدد المستويات لتقدير أثر استخدام رسائل التذكير القصيرة (SMS) على التزام مرضى فيروس نقص المناعة البشرية/الإيدز بالعلاج بمضادات الفيروسات العكوسة (القهقرية) في عيادة ريفية في كينيا. غيّرت الدراسة شدة المعالجة على طول بعدين اثنين: مدى تواتر إرسال الرسائل إلى المرضى (يوميًا أو أسبوعيًا)، وطول الرسالة (قصيرة أو طويلة). كانت الرسائل القصيرة تتضمن تذكيرًا ("هذا تذكير لك")، بينما كانت الرسائل الطويلة تتضمن تذكيرًا وعبرة تشجيعية ("هذا تذكير لك. تحلّ بالقوة والشجاعة، فنحن نهتم لأمرك"). وُزِعَ ما مجموعه ٥٣١ مريضًا توزيعًا عشوائيًا على واحدة من أربع مجموعات علاجية أو مجموعة المقارنة. كانت مجموعات المعالجة رسائل قصيرة أسبوعية، أو رسائل طويلة أسبوعية، أو رسائل قصيرة يومية، أو رسائل يومية طويلة.

وُزِعَ ثلث العينة على مجموعة المقارنة، ووُزِعَ الثلثان المتبقيان بالتساوي في كل من مجموعتين التدخل الأربع. أنشئت سلسلة من أعداد عشوائية بين ٠ و ١٠. وكانت أربعة فواصل بين ٠ و ٣/٢ تقابل مجموعات التدخل الأربع،

الجدول ١٠-١-١: ملخص تصميم البرنامج

المجموعة	نوع الرسالة	تكرار الرسالة	عدد المرضى
١	تذكير فقط	أسبوعي	٧٣
٢	تذكير + تشجيع	أسبوعي	٧٤
٣	تذكير فقط	يومي	٧٠
٤	تذكير + تشجيع	يومي	٧٢
٥	لا شيء (مجموعة المقارنة)	لا شيء	١٣٩

المصدر: Pop-Eleches and others 2011.

المربع ٢-١٠: اختبار بدائل البرنامج لمتابعة الفساد في إندونيسيا

والمساحين عينات أساسية للطرق الجديدة، وقدروا "تكلفة" المواد المستخدمة، ثم قارنوا حساباتهم بالميزانيات المبلغ عنها.

توصل أولكن إلى أن زيادة عمليات التدقيق الحكومية (من احتمال خضوع للتدقيق نسبته ٤ بالمائة إلى احتمال خضوع للتدقيق نسبته ١٠٠ بالمائة) قللت من النفقات بحوالي ٨ نقاط مئوية (من ٢٤ بالمائة). كان لزيادة مشاركة المجتمع في المتابعة أثر على العمالة المفقودة ولكن ليس على النفقات المفقودة. وكانت نماذج التعليقات فعالة فقط عندما وُزعت على أطفال في المدرسة ليقدموها لأسرهم وليس عندما قدمها رؤساء القرى.

في إندونيسيا، استخدم أولكن (٢٠٠٧) تصميمًا شاملاً لاختبار طرق مختلفة للسيطرة على الفساد، بدءًا من مقارنة الإنفاذ من أعلى إلى أسفل ووصولاً إلى زيادة المتابعة المجتمعية على مستوى القاعدة. وقد استخدم منهجية تعيين عشوائية في أكثر من ٦٠٠ قرية كانت تبني طرقًا بمثابة جزء من مشروع لتحسين البنية التحتية على مستوى البلاد.

كانت إحدى المعالجات المتعددة تتضمن اختيار بعض القرى عشوائيًا لإبلاغها بأن مشروع الإنشاء فيها سيخضع للتدقيق من قبل وكيل حكومي. بعد ذلك، لاختبار مشاركة المجتمع في المتابعة، طبق الباحثون عمليتي تدخل. أرسلوا دعوات لحضور اجتماعات مساءلة مجتمعية، وقدموا نماذج تعليقات يمكن تقديمها مباشرة. لقياس مستويات الفساد، أخذ فريق مستقل من المهندسين

المصدر: Olken 2007.

تقييم التدخلات المتعددة

بالإضافة إلى مقارنة مستويات متعددة للمعالجة، قد تحتاج إلى مقارنة خيارات معالجة مختلفة بالكامل. في الواقع، يفضل صانعو السياسات عادة مقارنة المزايا النسبية لعمليات تدخل مختلفة، بدلاً من معرفة أثر عملية تدخل واحدة فحسب.

تخيل أنك تريد تقدير أثر برنامج يشمل عمليتي تدخل بما يتعلق بالالتحاق بالمدارس، وهاتان العمليتان هما: التحويلات النقدية إلى أسر الطلاب والمشروطة بالالتحاق في المدرسة والتوصيل المجاني بالحافلات إلى المدرسة. أولاً، قد تحتاج إلى معرفة أثر كل عملية تدخل على حدة. هذه الحالة مطابقة تقريبًا للحالة التي تختبر فيها مستويات مختلفة من المعالجة لعملية تدخل واحدة: فبدلاً من التعيين العشوائي للوحدات في مستويات عالية ومنخفضة في مجموعتي معالجة ومقارنة، يمكنك تعيينها عشوائيًا في مجموعة التحويلات النقدية، ومجموعة نقل مجاني بالحافلة، ومجموعة النقل ومجموعة المقارنة. عمومًا، بوجود عمليات تدخل مختلفة عددها n ، ستكون هناك مجموعات معالجة عددها n ، بالإضافة إلى مجموعة مقارنة.

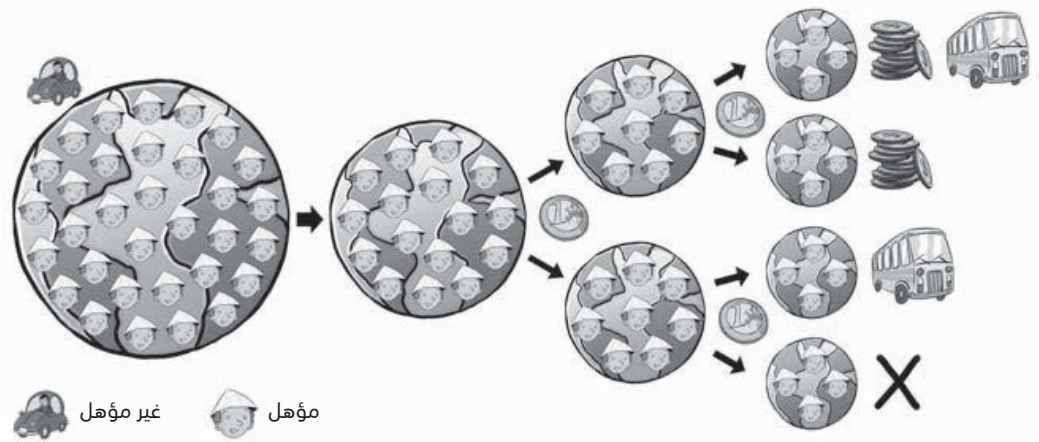
وبصرف النظر عن الرغبة في معرفة أثر كل عملية تدخل على حدة، قد تحتاج أيضًا إلى معرفة ما إذا كان دمجها أفضل من جمع التأثيرات الفردية فحسب. إذا ما نظرنا إلى البرنامج من منظور المشاركين، فإنه يتوفر في ثلاث صيغ مختلفة: التحويلات النقدية المشروطة فقط، أو النقل المجاني بالحافلات فقط، أو مزيج من التحويلات النقدية المشروطة والنقل المجاني بالحافلات.

يشبه التعيين العشوائي لبرنامج بعلميتي تدخل إلى حد كبير العملية الخاصة ببرنامج فيه عملية تدخل واحدة. والفرق الرئيسي هو الحاجة إلى إجراء العديد من القُرعات المستقلة بدلاً من قرعة واحدة. وينتج عن هذا تصميم متقاطع يُسمى أحياناً *التصميم الشامل*. يوضح الشكل ١٠-٢ هذه العملية. كما هو الحال من قبل، تقوم الخطوة ١ على تعريف المجتمع الإحصائي للوحدات المؤهلة لبرنامجك. والخطوة الثانية هي اختيار عينة من الوحدات العشوائية المؤهلة من المجتمع الإحصائي لصياغة عينة التقييم. وبمجرد حصولك على عينة التقييم، تقوم الخطوة ٣ على التعيين العشوائي لوحدات من عينة التقييم في مجموعة معالجة ومجموعة مقارنة. في الخطوة ٤، تستخدم قرعة ثانية لتعيين مجموعة جزئية من مجموعة المعالجة لتلقي عملية التدخل الثانية. أخيراً، في الخطوة ٥، تجري قرعة أخرى لتصميم مجموعة فرعية عن مجموعة المقارنة البدائية المقرر أن تتلقى عملية التدخل الثانية، بينما تبقى المجموعة الفرعية الأخرى للمقارنة الصرفة. ونتيجة للتعيين العشوائي في المعالجتين، ستكون قد أنشأت أربع مجموعات كما يوضح الشكل ١٠-٣.





- المجموعة أ تتلقى كلا نوعي التدخل (التحويلات النقدية والنقل بالحافلات).
 - المجموعة ب تتلقى التدخل ١ وليس التدخل ٢ (التحويلات النقدية فقط).
 - المجموعة ج لا تتلقى التدخل ٢ ولكنها تتلقى التدخل ١ (النقل بالحافلات فقط).
 - المجموعة د لا تتلقى التدخل ١ ولا التدخل ٢ وتشكل مجموعة مقارنة صرفة.
- تضمن طريقة التعيين العشوائي، عند تطبيقها على النحو الصحيح، تشابه المجموعات الأربع. لذلك، يمكنك تقدير أثر عملية التدخل الأولى بمقارنة الحصيلة (مثل معدل الحضور في المدرسة) للمجموعة ب مع حصيلة مجموعة المقارنة الصرفة، المجموعة د. ويمكنك أيضاً تقدير أثر عملية التدخل الثانية بمقارنة حصيلة المجموعة ج بحصيلة مجموعة المقارنة الصرفة، المجموعة د. بالإضافة إلى ذلك، يتيح هذا التصميم أيضاً إمكانية مقارنة الأثر المتزايد لتلقي عملية التدخل الثانية عندما تتلقى وحدة بالفعل عملية التدخل الأولى. تعطي مقارنة حصيلتي المجموعتين أ و ب أثر عملية التدخل الثانية للوحدات التي تلقت بالفعل عملية التدخل الأولى. تعطي مقارنة حصيلتي المجموعتين أ و ج أثر عملية التدخل الأولى للوحدات التي تلقت بالفعل عملية التدخل الثانية.

الشكل ٢-١٠ خطوات التعيين العشوائي لعمليتي تدخل

١. تعريف الوحدات المؤهلة ٢. اختيار عينة التقييم ٣. التعيين العشوائي للمعالجة الأولى ٤ و ٥. التعيين العشوائي في المعالجة الثانية



الشكل ٣-١٠ التصميم الشامل لبرنامج فيه عمليتا تدخل

عملية التدخل ١			
المعالجة	مجموعة	المعالجة	مجموعة
المجموعة ج	المجموعة أ		
		مجموعة	
المجموعة د	المجموعة ب		
			

لقد استخدم الوصف السابق مثال التعيين العشوائي لتفسير الكيفية التي يمكن بها تصميم تقدير أثر لبرنامج يتكون من عمليتي تدخل مختلفتين. عندما يضم برنامج أكثر من عمليتي تدخل اثنتين، يمكن زيادة عدد القرعات، ويمكن تقسيم التقييم أكثر من ذلك لإنشاء مجموعات تتلقى مجموعات مدمجة من عمليات التدخل. يمكن تنفيذ تصاميم من خلال معالجات متعددة ومستويات معالجة متعددة. وحتى إذا ازداد عدد المجموعات، فإن النظرية الأساسية الكامنة وراء التصميم تبقى نفسها، كما وُصف سابقًا.

جدير بالذكر أن تقييم أكثر من عملية تدخل واحدة أو اثنتين سي طرح تحديات عملية لكل من التقييم وتشغيل البرنامج، وذلك لأن تعقيد التصميم سيزداد أسياً بالنسبة إلى عدد أذرع المعالجة. لتقييم عملية تدخل واحدة، نحتاج فقط إلى مجموعتين اثنتين: مجموعة معالجة واحدة، ومجموعة مقارنة واحدة. لتقييم عمليتي تدخل واحدة، نحتاج إلى أربع مجموعات: ثلاث مجموعات معالجة، ومجموعة مقارنة واحدة. إذا أردت تقييم ثلاث عمليات تدخل، بما في ذلك جميع المجموعات المدمجة من عمليات التدخل الثلاث فستحتاج إلى $2 \times 2 \times 2$

= ٨ مجموعات في التقييم. عمومًا، لإجراء عملية تدخل تضم جميع المجموعات المدمجة من عمليات التدخل التي عددها n ، فستحتاج إلى 2^n من المجموعات. بالإضافة إلى ذلك، لكي تستطيع تمييز الاختلافات في الحاصل بين مجموعات مختلفة، يجب أن تضم كل مجموعة عددًا كافيًا من وحدات الملاحظة لضمان قدرة إحصائية كافية. عمليًا، يمكن أن يتطلب كشف الاختلافات بين أذرع تدخل مختلفة عينات أكبر مما هو عليه الحال عند مقارنة معالجة بمقارنة صرفة. فإذا نجح ذراعا المعالجة في إحداث تغييرات في الحاصل المرغوب فيها، فستكون هناك حاجة إلى عينات أكبر لاكتشاف الاختلافات الطفيفة المحتملة بين المجموعتين.^٣

وأخيرًا، يمكن اعتماد تصاميم شاملة أيضًا في تصاميم التقييم التي تجمع عدة طرق تقييم. تحدد القواعد التشغيلية التي توجه تعيين كل معالجة مجموعة الطرق التي يجب استخدامها. على سبيل المثال، قد تُخصص المعالجة الأولى استنادًا إلى درجة أهلية، ولكن الثانية تُخصص بأسلوب عشوائي. وفي تلك الحالة، يمكن أن يستخدم التصميم تصميم انقطاع انحدار لعملية التدخل الأولى وطريقة تعيين عشوائي لعملية التدخل الثانية.

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للمزيد من المعلومات حول تصميم تقييم الأثر بعدة خيارات معالجة، انظر Banerjee, Abhijit, and Esther Duflo.. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1: 151–78.

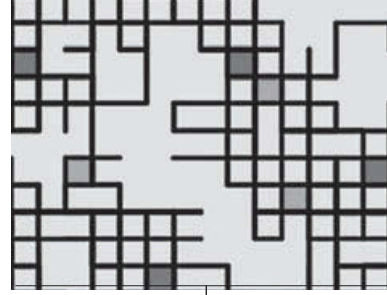
المفهوم الأساسي

لكي يقيم تقدير أثر المجموعات المدمجة بين n من المعالجات المختلفة، ستحتاج ما مجموعه 2^n من مجموعات المعالجة والمقارنة.

١. انظر (Banerjee and Duflo 2009) للاطلاع على مناقشة مفصلة.
٢. لاحظ أنه يمكن عملياً دمج القرعات الثلاث المنفصلة في قرعة واحدة تبقى تحقق النتيجة نفسها.
٣. لاختبار أثر عمليات التدخل المتعددة أيضاً تأثيراً أكثر دقة: فنظراً لأننا نزيد عدد عمليات التدخل أو مستويات المعالجة التي نختبرها بعضها مقابل البعض الآخر، فإننا نزيد من احتمال أننا سنجد أثراً في واحد من الاختبارات على الأقل، حتى ولو لم يكن هناك أثر. بمعنى آخر، من المحتمل أكثر أن نعثر على نتيجة إيجابية زائفة. لمنع ذلك، يجب تعديل الاختبارات الإحصائية لتراعي اختبار عدة فرضيات. يُشار إلى النتائج الإيجابية الزائفة أيضاً باسم الأخطاء من النوع الثاني. انظر الفصل ١٥ للاطلاع على مزيد من المعلومات عن الأخطاء من النوع الثاني والمراجع حول اختبار عدة فرضيات.

المراجع

- Banerjee, Abhijit, and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1: 151–78.
- Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.
- Pop-Eleches, Cristian, Harsha Thirumurthy, James Habyarimana, Joshua Zivin, Markus Goldstein, Damien de Walque, Leslie MacKeen, Jessica Haber, Sylvester Kimaiyo, John Sidle, Duncan Ngare, and David Bangsberg. 2011. "Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in a Resource-Limited Setting: A Randomized Controlled Trial of Text Message Reminders." *AIDS* 25 (6): 825–34.



الجزء ٣

كيفية تنفيذ تقييم الأثر

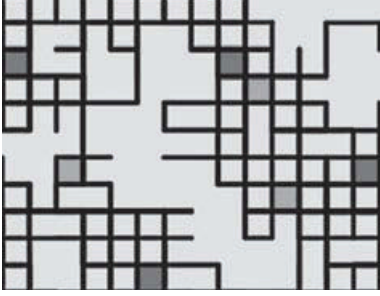
يركز الجزء ٣ من هذا الكتاب على كيفية تنفيذ تقييم الأثر: كيفية اختيار طريقة تقييم الأثر المتوافقة مع القواعد التشغيلية للبرنامج؛ كيفية إدارة تقييم الأثر، بما في ذلك ضمان إنشاء شراكة قوية بين فرق البحث والسياسات وإدارة الوقت والميزانية للتقييم؛ وكيفية التأكد من أن التقييم أخلاقي وذو مصداقية، مع اتباع مبادئ العمل مع الموضوعات البشرية والعلوم المفتوحة؛ وكيفية استخدام تقييم الأثر لتوجيه السياسة.

يحدد الفصل ١١ كيفية استخدام قواعد تشغيل البرنامج، أي الموارد المتاحة للبرنامج ومعايير اختيار المستخدمين وتوقيت التنفيذ، بمثابة أساس لاختيار طريقة تقييم الأثر. تم وضع إطار عمل بسيط لتحديد أي من أساليب تقييم الأثر المعروضة في الجزء الثاني أكثر ملاءمة لبرنامج معين، بناءً على قواعده التشغيلية. يناقش الفصل كذلك كيف أن الطريقة المفضلة هي الطريقة التي تتطلب الافتراضات الأضعف وأقل متطلبات من البيانات في سياق القواعد التشغيلية.

يناقش الفصل ١٢ العلاقة بين فرق السياسة والبحث وأدوار كل منهما. ويراجع الفرق بين الاستقلالية وعدم الانحياز، ويسلّط الضوء على الجوانب التي قد تكون حساسة عند تنفيذ تقييم الأثر. ويقدم الفصل إرشادات حول كيفية إدارة توقعات الأطراف المعنية، ويبرز بعض المخاطر المشتركة التي ينطوي عليها إجراء تقييمات الأثر، ويقدم اقتراحات حول كيفية التعامل مع هذه المخاطر. ويختتم الفصل بنظرة عامة على كيفية إدارة أنشطة تقييم الأثر، بما في ذلك تشكيل فريق التقييم، وتوقيت إجراء التقييم، وتحديد الميزانية، وجمع الأموال.

يقدم الفصل ١٣ لمحة عامة عن أخلاقيات وعلم تقييم الأثر، متضمنًا أهمية عدم حرمان المستفيدين المؤهلين من الفوائد من أجل التقييم، وكيفية تطبيق المبادئ الأساسية للبحث الأخلاقي الذي يشمل موضوعات البشر، ودور مجالس المراجعة المؤسسية التي تعتمد الأبحاث التي تشمل أشخاصًا وتتابعها، وأهمية ممارسة العلم المفتوح، بما في ذلك تسجيل التقييمات وإتاحة البيانات للجمهور لمزيد من البحث ولتكرار النتائج.

يقدم الفصل ١٤ نظرة ثاقبة حول كيفية استخدام تقييمات الأثر في توجيه السياسة، بما في ذلك نصائح حول كيفية جعل النتائج ذات صلة، ومناقشة أنواع المنتجات التي يمكن لتقييمات الأثر، بل وينبغي لها، أن تقدمها، وإرشادات حول كيفية استخلاص الاستنتاجات ونشرها لتعزيز أثر السياسة.



اختيار أسلوب تقييم الأثر

تحديد الأسلوب الذي يجب استخدامه لبرنامج معين

إن مفتاح تحديد الأثر السببي لبرنامج ما هو إيجاد مجموعة مقارنة صالحة لتقدير الواقع المضاد والإجابة على سؤال السياسة محل الاهتمام. ناقشنا، في الجزء ٢، عددًا من الطرق منها التعيين العشوائي والمتغيرات المساعدة وتصميم انقطاع الانحدار والاختلاف في الاختلافات والمطابقة. في هذا الفصل، نأخذ في الاعتبار السؤال عن الطريقة التي تختارها لبرنامج معين ترغب في تقييمه.

أولاً، نوضح أن القواعد التشغيلية للبرنامج توفر إرشادات واضحة عن كيفية العثور على مجموعات المقارنة، وبالتالي عن الطريقة الأكثر ملاءمة لسياق سياستك. المبدأ الشامل هو أنه إذا كانت القواعد التشغيلية لبرنامج ما محددة جيداً، فيمكنها حينئذٍ المساعدة في تحديد الطريقة الأنسب لتقييم هذا البرنامج المحدد.

ثانياً، الأساليب المقدمة في الجزء ٢ لها متطلبات مختلفة من البيانات وتعتمد على افتراضات أساسية مختلفة. تتطلب بعض الطرق افتراضات أقوى من غيرها لتقدير تغيرات الحصائل الناتجة عن التدخل بدقة. بشكل عام، نحن نفضل الطريقة التي تتطلب أضعف الافتراضات وأقل متطلبات من البيانات في سياق القواعد التشغيلية.

أخيرًا، نناقش كيفية اختيار وحدة التدخل. على سبيل المثال، هل سيتم تعيين البرنامج على المستوى الفردي أم على مستوى أعلى، مثل المجتمعات أو المقاطعات؟ بشكل عام، نفضل اختيار أصغر وحدة تدخل ممكنة ضمن القيود التشغيلية.

كيف يمكن لقواعد تشغيل البرنامج المساعدة على اختيار أسلوب تقييم الأثر

المفهوم الأساسي

تحدد القواعد التشغيلية لبرنامج أسلوب تقييم الأثر الأنسب لتقييم ذلك البرنامج، وليس العكس.

تتمثل إحدى الرسائل الرئيسية لهذا الكتاب في أنه يمكننا استخدام قواعد تشغيل البرنامج للعثور على مجموعات مقارنة صالحة، إلى الحد الذي تكون فيه القواعد التشغيلية للبرنامج محددة جيدًا. في الواقع، توفر قواعد تشغيل البرنامج دليلًا للطريقة الأنسب لتقييم هذا البرنامج بعينه. إن قواعد تشغيل البرنامج هي التي يمكنها أن تقود أسلوب التقييم، بل ويجب أن تقوده، وليس العكس. لا ينبغي أن يغير التقييم بشكل جذري العناصر الرئيسية لقواعد تعيين البرنامج المحددة جيدًا من أجل تصميم أكثر اتساقًا للتقييم.

إن القواعد التشغيلية الأكثر صلة بتصميم التقييم هي تلك التي تحدد الأشخاص المؤهلين للبرنامج وكيفية اختيارهم للمشاركة. تتكون مجموعات المقارنة من المؤهلين ولكن لا يمكن دمجهم في لحظة معينة (على سبيل المثال، عندما تكون هناك قيود على الموارد مع وجود طلب زائد)، أو أولئك الذين يقترحون من درجة أهلية المشاركة في البرنامج.

مبادئ قواعد تعيين البرامج المحددة جيدًا

المفهوم الأساسي

عند تصميم تقييمات الأثر السابقة، يمكننا دائمًا إيجاد مجموعات مقارنة صالحة إذا كانت القواعد التشغيلية لاختيار المستفيدين منصفًا وتتسم بالشفافية وخاضعة للمساءلة:

عند تصميم تقييمات الأثر السابقة، يمكننا دائمًا إيجاد مجموعات مقارنة صالحة إذا كانت القواعد التشغيلية لاختيار المستفيدين منصفًا وتتسم بالشفافية وخاضعة للمساءلة:

- **ترتيب قواعد تعيين البرنامج العادلة** أو تعطي الأولوية للأهلية بناءً على مؤشر متفق عليه للحاجة، أو تنص على أن كل فرد يحصل على مزايا البرنامج أو على الأقل لديه فرصة متساوية في الحصول على المزايا.
- **علنية** شفافية قواعد تعيين البرنامج، بحيث يمكن للأطراف الخارجية الموافقة عليها ضمنيًا ويمكنها متابعة ما إذا تم اتباعها فعليًا. يجب أن تكون قواعد الشفافية كمية ويمكن ملاحظتها بسهولة.
- **تعدد قواعد المساءلة** مسؤولية مسؤولي البرنامج، وتنفيذها هو أساس الأداء الوظيفي أو مكافآت هؤلاء المسؤولين.

تتسم القواعد التشغيلية للأهلية بالشفافية وتخضع للمساءلة عندما تستخدم البرامج معايير قابلة للقياس يمكن التحقق منها من قبل أطراف خارجية وعندما تُعلن هذه المعايير. يضمن الإنصاف والشفافية والمساءلة أن معايير الأهلية قابلة للتحقق من الناحية الكمية وأن يتم تنفيذها بالفعل كما هو مصمم. على هذا النحو، تعمل مبادئ الحوكمة الرشيدة هذه على تحسين احتمالية أن يفيد البرنامج المجتمع الإحصائي المستهدف، كما أنها أيضًا مفتاح التقييم الناجح. إذا لم تكن القواعد قابلة للقياس والتحقق من الناحية الكمية، فسيواجه فريق التقييم صعوبة في التأكد من أن التعيين لمجموعات المعالجة والمقارنة يحدث كما هو مُصمم، أو على الأقل، في توثيق كيف حدث ذلك بالفعل. إذا لم يتمكن أعضاء فريق التقييم بالفعل من التحقق من التعيين، فلن يتمكنوا من تحليل البيانات بشكل صحيح لحساب الآثار. يعد فهم قواعد تعيين البرنامج أمرًا بالغ الأهمية لاختيار طريقة مناسبة لتقييم الأثر.

عندما تنتهك القواعد التشغيلية أيًا من هذه المبادئ الثلاثة للحوكمة الرشيدة، نواجه تحديات في كل من إنشاء برنامج جيد التصميم وإجراء التقييم. من الصعب العثور على مجموعات مقارنة صالحة إذا كانت القواعد التي تحدد أهلية المنتفعين واختيارهم ليست عادلة ولا تتسم بالشفافية وخاضعة للمساءلة. في هذه الحالة، قد يتطلب تصميم تقييم الأثر توضيحات وتعديلات في طريقة عمل البرنامج. لكن، إذا كانت القواعد محددة جيدًا، فيمكن اختيار طريقة تقييم الأثر بناءً على قواعد تعيين البرنامج الحالية، التي نناقشها الآن بمزيد من التفصيل.

قواعد التشغيل الرئيسية

تحكم قواعد التشغيل عادةً ماهية مزايا البرنامج، وكيفية تمويل هذه المزايا وتوزيعها، وكيف يختار البرنامج المستفيدين. تعد القواعد التي تحكم البرامج واختيار المستفيدين مفتاح العثور على مجموعات مقارنة صالحة، حيث تغطي القواعد التي تحكم اختيار المستفيدين الأهلية وقواعد التخصيص/التوزيع في حالة الموارد المحدودة والتنفيذ التدريجي للمستفيدين بمرور الوقت. وبشكل أكثر تحديدًا، فإن القواعد الأساسية التي تمثل خريطة طريق للعثور على مجموعات المقارنة تجيب على ثلاثة أسئلة تشغيلية أساسية تتعلق بالموارد المتاحة للبرنامج ومعايير الأهلية وتوقيت التنفيذ:

١. *الموارد المتاحة.* هل يمتلك البرنامج موارد كافية للوصول للنطاق المطلوب والوصول إلى التغطية الكاملة لجميع المستفيدين المؤهلين؟ لا تمتلك الحكومات والمنظمات غير الحكومية موارد كافية دائمًا لتقديم خدمات البرنامج لكل شخص مؤهل ويتقدم للحصول على مزايا. في هذه الحالة، يجب على الحكومة أن تقرر أي من المتقدمين المؤهلين سيحصل على مزايا البرنامج وأيهم لن ينضم إليه. في كثير من الأحيان، تقتصر البرامج على مناطق جغرافية محددة، أو على عدد محدود من المجتمعات، على الرغم من أنه قد يكون هناك مستفيدون مؤهلون في مناطق أو مجتمعات أخرى.

٢. **معايير الأهلية.** من المؤهل للحصول على مزايا البرنامج؟ هل يعتمد تعيين البرنامج على نقطة انقطاع الأهلية، أم أنها متاحة للجميع؟ عادة ما تُقدم الرعاية الصحية الأولية والمدارس العامة على مستوى عام. تستخدم العديد من البرامج قواعد الأهلية التشغيلية التي تعتمد على الترتيب المستمر مع نقطة الانقطاع. على سبيل المثال، تحدد برامج المعاشات التقاعدية حدًا للسن بحيث يصبح الأفراد الذين يتجاوزونه مؤهلين لهذه البرامج. غالبًا ما تصنف برامج التحويلات النقدية الأسر بناءً على حالة الفقر المُقدرة، وتعتبر الأسر الأقل من حد نقطة الانقطاع المحدد مسبقًا مؤهلة.

٣. **توقيت التنفيذ.** هل يتم تسجيل المستفيدين المحتملين في البرنامج دفعة واحدة أم على مراحل مع مرور الوقت؟ في كثير من الأحيان، تمنع القيود الإدارية والمتعلقة بالموارد الحكومات والمنظمات غير الحكومية من تقديم المزايا إلى المجتمع الإحصائي المؤهل بأكمله على الفور. حيث يجب أن يطرحوا برامجهم مع مرور الوقت، وبالتالي يجب أن يقرروا من الذي يستفيد أولاً ومن يتم دمجه لاحقًا. إن أحد الأساليب الشائعة هي إدخال البرنامج على مراحل زمنية وجغرافية وكذلك إدماج جميع المستفيدين المؤهلين في قرية أو منطقة واحدة قبل الانتقال إلى المرحلة التالية.

اشتقاق مجموعات المقارنة من القواعد التشغيلية

عند تصميم تقييم الأثر السابق، فإن الإجابة على الأسئلة التشغيلية الثلاثة تحدد إلى حد كبير طريقة تقييم الأثر المناسبة لبرنامج معين. يحدد الجدول ١١-١ مجموعات المقارنة المحتملة لقواعد محددة لتشغيل البرنامج والأسئلة التشغيلية الأساسية الثلاثة المتعلقة بالموارد المتاحة وقواعد الأهلية وتوقيت التنفيذ. يتم تقسيم الأعمدة بناءً على ما إذا كان البرنامج يحتوي على موارد كافية لتغطية جميع المستفيدين المؤهلين المحتملين في النهاية (الموارد المتاحة)، ويتم تقسيمها أيضًا إلى البرامج ذات نقطة انقطاع وتصنيف أهلية مستمر وتلك التي لا تتميز بذلك (معايير الأهلية). يتم تقسيم الصفوف إلى التنفيذ المرحلي مقابل التنفيذ الفوري للبرنامج (توقيت التنفيذ). تسرد كل خلية المصادر المحتملة لمجموعات المقارنة الصالحة، بالإضافة إلى الفصل ذي الصلة الذي نوقشت فيه في الجزء ٢. يتم تمييز كل خلية بمؤشر: يشير الحرف الأولي إلى صف الجدول (أ، ب)، ويشير الرقم التالي إلى العمود (١ - ٤). على سبيل المثال، تشير الخلية أ١ إلى الخلية الموجودة في الصف الأول والعمود الأول من الجدول. على سبيل المثال، تحدد الخلية أ١ طرق التقييم الأكثر ملاءمة للبرامج ذات الموارد المحدودة، والتي لها معايير الأهلية، ويتم تنفيذها على مراحل بمرور الوقت.

معايير الأهلية.	زيادة الطلب على البرنامج (الموارد محدودة)		لا يوجد طلب زائد على البرنامج (الموارد كاملة)	
	(١)	(٢)	(٣)	(٤)
الترتيب المستمر للأهلية ونقطة الانقطاع	لا يوجد ترتيب مستمر للأهلية ونقطة الانقطاع	لا يوجد ترتيب مستمر للأهلية ونقطة الانقطاع	لا يوجد ترتيب مستمر للأهلية ونقطة الانقطاع	لا يوجد ترتيب مستمر للأهلية ونقطة الانقطاع
(أ) التنفيذ على مراحل بمرور الوقت	الخلية أ١ التعيين العشوائي (الفصل ٤) تصميم انقطاع الانحدار (الفصل ٦)	الخلية أ٢ التعيين العشوائي (الفصل ٤) المتغيرات المساعدة (الترويج العشوائي) (الفصل ٥) الاختلاف في الاختلافات (الفصل ٧) المطابقة (الفصل ٨)	الخلية أ٣ التعيين العشوائي للمراحل (الفصل ٤) تصميم انقطاع الانحدار (الفصل ٦)	الخلية أ٤ التعيين العشوائي للمراحل (الفصل ٤) المتغيرات المساعدة (الترويج العشوائي حتى الالتحاق الكبير) (الفصل ٥) الاختلاف في الاختلافات (الفصل ٧) الاختلاف في الاختلافات مع المطابقة (الفصل ٨)
	الخلية ب١ التعيين العشوائي (الفصل ٤) تصميم انقطاع الانحدار (الفصل ٦)	الخلية ب٢ التعيين العشوائي (الفصل ٤) المتغيرات المساعدة (الترويج العشوائي) (الفصل ٥) الاختلاف في الاختلافات (الفصل ٧) الاختلاف في الاختلافات مع المطابقة (الفصل ٨)	الخلية ب٣ تصميم انقطاع الانحدار (الفصل ٦)	الخلية ب٤ إذا كان أقل من الالتحاق الكامل: المتغيرات المساعدة (الترويج العشوائي) (الفصل ٥) الاختلاف في الاختلافات (الفصل ٧) الاختلاف في الاختلافات مع المطابقة (الفصل ٨)
(ب) التنفيذ الفوري				

توقيت التنفيذ

تحتاج معظم البرامج إلى التدرج في مراحل عبر الوقت إما بسبب قيود التمويل أو القيود اللوجستية والإدارية. تغطي هذه المجموعة أو الفئة الصف الأول من المخطط (الخلايا أ١ وأ٢ وأ٣ وأ٤) في هذه الحالة، إن القاعدة التشغيلية العادلة التي تتسم بالشفافية وتخضع للمساءلة هي إعطاء كل وحدة مؤهلة فرصة متساوية للحصول على البرنامج الأول والثاني والثالث وهكذا، مما يعني ضمناً بدء التشغيل العشوائي للبرنامج بمرور الوقت. في الحالات التي تكون فيها الموارد محدودة، أي حيث لن يكون هناك أبداً موارد كافية لتحقيق التوسع الكامل (الخلايا أ١ وأ٢ وب١ وب٢)، قد يظهر الطلب الزائد على هذه الموارد بسرعة كبيرة. ثم قد يكون إجراء قرعة لتحديد من سيشارك في البرنامج مقارنة قابلة للتطبيق لتوزيع المزايا بين الوحدات المؤهلة على قدم المساواة. في هذه الحالة، تحصل كل وحدة مؤهلة على فرصة متساوية للاستفادة من البرنامج. تعد القرعة مثلاً على القاعدة التشغيلية العادلة التي تتسم بالشفافية وتخضع للمساءلة لتوزيع مزايا البرنامج بين الوحدات المؤهلة.

تشتمل فئة أخرى من البرامج على تلك البرامج التي يتم تنفيذها على مراحل على مدار الوقت والتي يمكن للمسؤولين ترتيب المستفيدين المحتملين لها حسب الحاجة (الخليتان أ و ٣). إذا كانت المعايير المستخدمة لتحديد الأولوية من المستفيدين كمية ومتاحة ولديها حد نهائي للأهلية، فيمكن للبرنامج استخدام تصميم انقطاع الانحدار. تتكون الفئة العريضة الأخرى من البرامج التي لديها القدرة الإدارية ليتم تنفيذها على الفور: أي الخلايا الموجودة في الصف السفلي من المخطط. عندما يكون للبرنامج موارد محدودة وغير قادر على ترتيب المستفيدين (الخلية ب ٢)، يمكن استخدام التعيين العشوائي على أساس زيادة الطلب. إذا كان البرنامج يحتوي على موارد كافية للوصول للنطاق الكامل ولا توجد معايير للأهلية (الخلية ب ٤)، فإن الحل الوحيد هو استخدام المتغيرات المساعدة (الترويج العشوائي)، بافتراض القليل من الالتحاق الكامل للبرنامج. إذا كان البرنامج يستطيع تصنيف المستفيدين ويعتمد على معايير الأهلية، فيمكن استخدام تصميم انقطاع الانحدار.

تحديد الأولويات من المستفيدين

تتعلق الأسئلة التشغيلية الثلاثة الرئيسية بالأمر الحاسم المتعلق بكيفية اختيار المستفيدين، وهو أمر بالغ الأهمية لإيجاد مجموعات مقارنة صالحة. توجد مجموعات المقارنة أحياناً بين المجتمعات الإحصائية غير المؤهلة، وفي كثير من الأحيان بين المجتمعات الإحصائية المؤهلة ولكن يتم دمجهم في البرنامج لاحقاً. تعتمد كيفية تحديد الأولوية من المستفيدين جزئياً على أهداف البرنامج. هل هو برنامج معاشات لكبار السن، أم برنامج يستهدف الفقراء للتخفيف من حدة الفقر، أم برنامج تطعيم متاح للجميع؟

لتحديد الأولوية من المستفيدين على أساس الحاجة، يجب على البرنامج أن يجد مؤشراً يمكن قياسه والتحقق منه. من الناحية العملية، تعتمد جدوى تحديد الأولويات إلى حد كبير على قدرة الحكومة على قياس الاحتياجات وترتيبها. إذا تمكنت الحكومة من تصنيف المستفيدين بدقة على أساس الحاجة النسبية، فقد تشعر بأنها ملزمة أخلاقياً بطرح البرنامج حسب الحاجة. ومع ذلك، لا يتطلب الترتيب على أساس الحاجة مقياساً كمياً فحسب، بل يتطلب أيضاً القدرة والموارد لقياس هذا المؤشر لكل وحدة يمكن أن تشارك في البرنامج.

تستخدم بعض البرامج معايير اختيار يمكن استخدامها من حيث المبدأ لتصنيف الحاجة النسبية وتحديد الأهلية. على سبيل المثال، تسعى العديد من البرامج للوصول إلى الأفراد الفقراء. ومع ذلك، غالباً ما يكون من الصعب قياس مؤشرات الفقر الدقيقة التي تُصنف الأسر بصورة موثوقة، كما أن جمعها مكلف. إن جمع بيانات الدخل أو الاستهلاك عن جميع المستفيدين المحتملين لتصنيفهم حسب مستوى الفقر عملية معقدة ومكلفة، وسيكون من الصعب التحقق منها أيضاً. بدلاً من ذلك، تستخدم العديد من البرامج نوعاً من اختبار الوسائل البديلة لتقدير مستويات الفقر. وتعد هذه فهارس لمقاييس بسيطة يمكن ملاحظتها مثل الأصول والخصائص الاجتماعية الديموغرافية (Grosch and others 2008). يمكن أن تساعد اختبارات الوسائل البديلة في تحديد ما إذا كانت الأسرة أعلى من مستوى الانقطاع الإجمالي أم أقل منه، ولكنها قد تكون أقل دقة في توفير تصنيف مفصل للحالة الاجتماعية والاقتصادية أو للحاجة.

وبدلاً من مواجهة تكلفة وتعقيد ترتيب المستفيدين الفرديين المحتملين، تختار العديد من البرامج الترتيب بناءً على مستوى أعلى من التجميع، على سبيل المثال بناءً على مستوى المجتمع. إن تحديد تعيين البرنامج على مستوى مُجمَع له فوائد تشغيلية واضحة، ولكن غالباً ما يكون من الصعب إيجاد مؤشرات على ترتيب احتياجات على مستوى إجمالي أكثر.

يجب استخدام معايير أخرى لتحديد كيفية تسلسل طرح البرنامج في الحالات التي يتعذر فيها على البرنامج توزيع المزايا بناءً على الحاجة بشكل موثوق، لأن مؤشر الترتيب، الذي يمكن قياسه وقابل للتحقق، إما غير متوفر أو مكلف للغاية وعرضة للخطأ. وتعد المساواة أحد المعايير التي تتوافق مع الحوكمة الرشيدة. تتمثل قاعدة المساواة في منح كل شخص مؤهل فرصة متساوية للالتحاق أولاً، وبالتالي تعيين مكان عشوائياً في تسلسل المستفيدين المحتملين. عملياً، ونظراً للتحديات في ترتيب الاحتياجات، أصبح التعيين العشوائي لمزايا البرنامج هو قاعدة شائعة الاستخدام لتعيين البرامج. إنها قاعدة توزيع عادلة ومنصفة. كما أنها تنتج تصميم تقييم عشوائياً يمكن أن يوفر صلاحية داخلية جيدة إذا تم تنفيذه جيداً، ويمكن أن تعتمد على افتراضات أضعف مقارنة بالطرق الأخرى، التي تتم مناقشتها في القسم التالي.

مقارنة بين طرق تقييم الأثر

بعد تقييم ما أسلوب تقييم الأثر المناسب لقواعد تشغيل برنامج محددة، يمكن لفريق التقييم اختيار الطريقة التي بها أضعف افتراضات وأقل متطلبات للبيانات. يقدم الجدول ٢-١١ مقارنة بين طرق تقييم الأثر البديلة من حيث متطلبات البيانات لتنفيذها، والافتراضات الأساسية الضرورية لتفسير نتائجها على أنها آثار سببية للتدخل. يمثل كل صف طريقة منفصلة. يصف العمودان الأولان الطرق والوحدات الموجودة في مجموعة المقارنة. ويوضح العمودان الأخيران الافتراضات اللازمة لتفسير النتائج على أنها سببية والبيانات اللازمة لتنفيذ هذه الطرق.

تتطلب جميع الطرق افتراضات؛ أي، لكي نتمكن من تفسير النتائج على أنها سببية، يجب أن نؤمن بأن الحقائق صحيحة ولا يمكننا دائماً التحقق منها بشكل كامل تجريبياً. وعلى وجه الخصوص، لكل طريقة، يتمثل أحد الافتراضات الرئيسية في أن وسيلة مجموعة المقارنة التي تعتمد عليها الطريقة هي تقدير صحيح للواقع المضاد. في كل فصل من الفصول المتعلقة بالطرق في الجزء ٢، ناقشنا بعض الاعتبارات الخاصة بكيفية اختبار ما إذا كانت الطريقة صالحة في سياق معين. تعتمد بعض الطرق على افتراضات أقوى من غيرها.

الجدول ٢-١١ مقارنة طرق تقييم الأثر

البيانات المطلوبة	الافتراضات الرئيسية	من في مجموعة المقارنة؟	الوصف	المنهجية
بيانات حصائل المتابعة لمجموعات المعالجة مجموعات المقارنة؛ حصائل خط الأساس وخصائص أخرى لمجموعات العلاج والمقارنة للتحقق من التوازن.	ينتج التعيين العشوائي بشكل فعال مجموعتين متطابقتين إحصائيًا فيما يتعلق بالخصائص المرصودة وغير المرصودة (عند خط الأساس وحتى خط النهاية).	الوحدات المؤهلة التي تم تعيينها عشوائيًا لمجموعة المقارنة.	الوحدات المؤهلة التي تم تعيينها عشوائيًا لمجموعة المعالجة أو فرصة متساوية في اختيارها. يميل إلى إنشاء تقديرات أكثر صالحة داخليًا في ظل أضعف الافتراضات.	التعيين العشوائي
بيانات حصائل المتابعة لجميع الوحدات؛ بيانات عن المشاركة الفعالة في البرنامج؛ بيانات عن حصائل خط الأساس وغيرها من الخصائص.	تؤثر الأداة المساعدة على المشاركة في البرنامج ولكنها لا تؤثر بشكل مباشر على الحصائل (أي أن المساعدة تؤثر على الحصائل فقط من خلال تغيير احتمالية المشاركة في البرنامج).	الوحدات "المتثلة" التي تتأثر مشاركتها في البرنامج بالمساعدة (يمكنها المشاركة إذا تعرضت للمساعدة، ولا يمكنها المشاركة إذا لم تتعرض للمساعدة).	مساعدة عشوائية (مثل حملة ترويج) تحدث تغييرات في المشاركة في البرنامج الذي يجري تقييمه. تستخدم هذه الطريقة التغيير في الحصائل الناتج عن التغيير في معدلات المشاركة لتقدير آثار البرنامج.	المتغير المساعد (خاصة الترويج العشوائي)
بيانات حصائل المتابعة؛ مؤشر الترتيب ونقطة انقطاع الأهلية؛ بيانات عن حصائل خط الأساس وغيرها من الخصائص.	لتحديد تأثيرات البرنامج غير المتحيزة بالنسبة للمجتمع الإحصائي القريب من نقطة الانقطاع، فإن الوحدات الموجودة أسفل نقطة الانقطاع وفوقها مباشرة متطابقة إحصائيًا. لتحديد تأثيرات البرنامج غير المتحيزة لجميع المجتمع الإحصائي، يجب أن يمثل المجتمع الإحصائي القريب من نقطة الانقطاع جميع المجتمع الإحصائي.	الوحدات التي تقترب من الحد ولكنها غير مؤهلة للحصول على البرنامج.	يتم تصنيف الوحدات بناءً على معايير محددة كمية ومستمرة، مثل مؤشر الفقر. هناك حد يحدد ما إذا كانت الوحدة مؤهلة للمشاركة في البرنامج أم لا. حصائل المشاركين على جانب واحد من هذا الحد مقارنة بحصائل غير المشاركين على الجانب الآخر من هذا الحد.	تصميم انقطاع الانحدار

(تابع)

الجدول ٢-١١ (تابع)

البيانات المطلوبة	الافتراضات الرئيسية	من في مجموعة المقارنة؟	الوصف	المنهجية
بيانات خط الأساس والمتابعة حول الحصائل والخصائص الأخرى لكل من المشاركين وغير المشاركين.	إذا لم يكن البرنامج موجوداً، فإن الحصائل لمجموعات المشاركين وغير المشاركين كانت ستتم بالتوازي بمرور الوقت.	الوحدات التي لم تشارك في البرنامج (لأي سبب) والتي تم جمع البيانات الخاصة بها قبل البرنامج وبعده.	يتم استخدام التغيير في الحصائل عبر الوقت في مجموعة من غير المشاركين لتقدير ما كان يمكن أن يتغير في الحصائل لمجموعة من المشاركين في حالة عدم وجود برنامج.	الاختلاف في الاختلافات
بيانات حصائل المتابعة للمشاركين وغير المشاركين؛ بيانات عن المشاركة الفعالة في البرنامج؛ خصائص خط الأساس لأداء المطابقة.	لا توجد خاصية تؤثر على المشاركة في البرنامج تتجاوز الخصائص الملاحظة المستخدمة للمطابقة.	لكل مشارك، الوحدة غير المشاركة التي يُتوقع أن يكون لها نفس احتمالية المشاركة في البرنامج بناءً على الخصائص المرصودة.	لكل مشارك في البرنامج، تبحث الطريقة عن الوحدة "الأكثر تشابهاً" في مجموعة غير المشاركين (أقرب تطابق يعتمد على الخصائص المرصودة).	المطابقة خاصة مطابقة درجة الميل)

المصدر: مقتبس من الموقع الإلكتروني لمختبر عبد اللطيف جميل لمكافحة الفقر.

عندما تتساوى كل الأمور الأخرى، تكون الطريقة الأفضل هي الطريقة الأنسب للسياق التشغيلي والتي تتطلب أضعف الافتراضات وأقل البيانات. توضح هذه المعايير سبب استقرار الباحثين على التعيين العشوائي باعتباره المعيار الذهبي، وسبب كونه الطريقة المفضلة غالبًا. يتلاءم التعيين العشوائي مع العديد من السياقات التشغيلية، ويميل إلى إنشاء تقديرات أثر صالحة داخليًا في ظل أضعف الافتراضات. عند تنفيذه بشكل صحيح، فإنه ينشئ قابلية للمقارنة بين مجموعات المعالجة والمقارنة في الخصائص المرصودة وغير المرصودة. بالإضافة إلى ذلك، يميل التعيين العشوائي إلى طلب عينات أصغر من العينات اللازمة لتنفيذ الطرق شبه التجريبية (انظر المناقشة في الفصل ١٥). نظرًا لأن التعيين العشوائي أمر بديهي إلى حد ما، فإن هذه الطريقة تجعل توصيل النتائج إلى صانعي السياسات أمرًا سهلًا وبسيطًا.

المفهوم الأساسي

الطريقة المفضلة لتقييم الأثر هي الطريقة الأنسب للسياق التشغيلي، والتي تتطلب أضعف الافتراضات وأقل قدر من متطلبات البيانات.

قد تكون الطرق شبه التجريبية أكثر ملاءمة في بعض السياقات التشغيلية، ولكنها تتطلب المزيد من الافتراضات حتى تتمكن مجموعة المقارنة من تقديم تقدير صالح للواقع المضاد. على سبيل المثال، يعتمد الاختلاف في الاختلافات على الافتراض بأن تغييرات حائل مجموعة المقارنة توفر تقديرًا صحيحًا لتغيرات الواقع المضاد في حائل مجموعة المعالجة. وليس من الممكن دائمًا اختبار هذا الافتراض، أن حائل مجموعات المعالجة والمقارنة تنمو بالتوازي عبر الوقت، بدون موجات متعددة من البيانات قبل التدخل. يعتمد انقطاع الانحدار على قابلية مقارنة الوحدات التي تكون أقل بقليل من حد الأهلية وتلك الأعلى بقليل. وتتميز المطابقة، من بين الطرق كلها، بأقوى الافتراضات لأنها تفترض بشكل أساسي أي خصائص غير ملحوظة بين المشاركين في البرنامج وغير المشاركين. بشكل عام، كلما كانت الافتراضات أقوى، زادت مخاطر عدم التمسك بها أثناء الممارسة.

خطة احتياطية لتقييمك

في بعض الأحيان، لا تسير الأمور كما هو مخطط لها بالضبط حتى مع أفضل تصميم لتقييم الأثر وأفضل النوايا. في أحد برامج التدريب الوظيفي، على سبيل المثال، خطت الوكالة المنفذة لاختيار المشاركين بشكل عشوائي من مجموعة المتقدمين، بناءً على الزيادة المتوقعة في الاشتراك في البرنامج. ونظرًا لارتفاع معدل البطالة بين المجتمع الإحصائي المستهدف، كان من المتوقع أن تكون مجموعة المتقدمين لبرنامج التدريب الوظيفي أكبر بكثير من عدد الأماكن المتاحة. ولكن للأسف، لم يكن الإعلان عن البرنامج فعالاً كما هو متوقع، وفي النهاية، كان عدد المتقدمين أقل بقليل من عدد أماكن التدريب المتاحة. بدون الاشتراك الزائد الذي يمكن من خلاله الحصول على مجموعة مقارنة، وبدون وجود خطة احتياطية، كان لابد من إسقاط المحاولة الأولية لتقييم البرنامج بالكامل. يعد هذا النوع من المواقف شائعًا، وكذلك التغييرات غير المتوقعة في السياق التشغيلي أو السياسي للبرنامج. لذلك، من المفيد أن يكون لديك خطة احتياطية تستخدمها في حالة عدم نجاح المنهجية المختارة أولاً.

يعد التخطيط لاستخدام عدة طرق لتقييم الأثر ممارسة جيدة أيضًا من وجهة النظر المنهجية. إذا كانت لديك شكوك حول ما إذا كان هناك تحيز متبقي في إحدى طرقك، فستتمكن من التحقق من النتائج مقابل الطريقة الأخرى. عندما يتم تنفيذ برنامج في طرح عشوائي، سيتم في النهاية دمج مجموعة المقارنة في البرنامج. مما يحد من الوقت الذي تكون خلاله مجموعة المقارنة متاحة للتقييم. ومع ذلك، إذا تم تنفيذ تصميم الترويج العشوائي بالإضافة إلى تصميم التعيين العشوائي، فستوفر مجموعة مقارنة طوال مدة البرنامج بالكامل. قبل دمج المجموعة النهائية للطرح، ستوجد مجموعتان بديلان للمقارنة (من التعيين العشوائي والترويج العشوائي)، على الرغم من بقاء مجموعة مقارنة الترويج العشوائي فقط على المدى الطويل.

إيجاد أصغر وحدة تدخل ممكنة

بشكل عام، تحدد قواعد التشغيل أيضًا المستوى الذي يتم فيه تعيين التدخل، والذي يتعلق بطريقة تنفيذ البرنامج. على سبيل المثال، إذا تم تنفيذ برنامج صحي على مستوى المنطقة، فإن جميع القرى في المنطقة إما ستحصل البرنامج (كمجموعة) أو لا تحصل عليه. يمكن تنفيذ بعض البرامج بنجاعة على مستوى الفرد أو الأسرة، بينما يحتاج البعض الآخر إلى التنفيذ على مستوى المجتمع أو المستوى الإداري الأعلى. حتى إذا كان من الممكن تعيين البرنامج وتنفيذه على المستوى الفردي، فقد يفضل فريق البحث التقييمي مستوى أعلى من التجميع من أجل التخفيف من التداعيات المحتملة، أي التأثيرات غير المباشرة من المشاركة على الوحدات غير المشاركة (انظر المناقشة في الفصل ٩). يمكن أن يمثل تنفيذ التدخل على مستوى عالي مشكلة بالنسبة للتقييم لسببين رئيسيين. السبب الأول، تتطلب تقييمات التدخلات التي تم تعيينها وتنفيذها على مستويات أعلى، مثل المجتمع المحلي أو المنطقة الإدارية، أحجام عينات أكبر وستكون أكثر تكلفة مقارنةً بتقييمات التدخلات على مستوى أقل، على مستوى الفرد أو الأسرة مثلاً. يعد مستوى التدخل مهمًا لأنه يحدد وحدة التعيين لمجموعات المعالجة والمقارنة، ما يؤثر على حجم عينة التقييم وتكلفتها. بالنسبة للتدخلات التي يتم تنفيذها على مستويات أعلى، هناك حاجة إلى عينة كبيرة لتكون قادرة على اكتشاف الأثر الحقيقي للبرنامج. ستتم مناقشة الهدف الكامن وراء ذلك في الفصل ١٥، والذي يستعرض كيفية تحديد حجم العينة المطلوبة للتقييم ويناقش كيف يؤدي التنفيذ على المستويات الأعلى إلى إنشاء مجموعات تزيد من حجم العينة المطلوبة.

السبب الثاني، يصعب في المستويات الأعلى من التدخل العثور على عدد كافٍ من الوحدات لتضمينها في التقييم. ومع ذلك، فإن التعيين العشوائي يولد فقط مجموعات معالجة ومقارنة قابلة للمقارنة إذا تم إجراؤه بين عدد كافٍ من الوحدات. على سبيل المثال، إذا كان مستوى التجميع على مستوى المقاطعة وكانت بالدولة ست مقاطعات فقط، فمن غير المرجح أن يحقق التوزيع العشوائي التوازن بين مجموعتي المعالجة والمقارنة. في هذه الحالة، نفترض أن تصميم التقييم يخصص ثلاث مقاطعات لمجموعة المعالجة وثلاث مقاطعات لمجموعة المقارنة. من المستبعد جدًا أن تكون المقاطعات في مجموعة المعالجة ماثلة لمجموعة المقارنة، حتى إذا كان عدد الأسر داخل كل مقاطعة كبيرًا. هذا لأن أساس موازنة مجموعات المعالجة والمقارنة هو عدد الوحدات المخصصة لهما، وليس عدد الأفراد أو الأسر في العينة. لذلك، فإن القيام بالتعيين العشوائي على مستويات عالية من التنفيذ يتسبب في وجود مخاطر على الصلاحية الداخلية إذا كان عدد الوحدات غير كافٍ.

لتجنب المخاطر المرتبطة بتنفيذ التدخل على مستوى جغرافي أو إداري عالٍ، يحتاج فريق التقييم ومديرو البرنامج إلى العمل معًا للعثور على أصغر وحدة تدخل تكون مجدية من الناحية التشغيلية. هناك العديد من العوامل المختلفة التي تحدد أصغر وحدة تدخل ممكنة:

- وفورات الحجم والتعقيد الإداري في تنفيذ البرنامج
 - القدرة الإدارية على توزيع المزايا على مستوى الفرد أو الأسرة
 - المخاوف المحتملة بشأن التوترات المحتملة
 - المخاوف المحتملة بشأن التأثيرات غير المباشرة لمجموعة المقارنة وتداخلها.
- تعتمد أصغر وحدة تدخل ممكنة عادةً على وفورات الحجم والتعقيد الإداري لتنفيذ البرنامج. على سبيل المثال، قد يتطلب برنامج التأمين الصحي وجود مكتب محلي للمستفيدين لتقديم المطالبات والدفع لمقدمي الخدمات. يجب توزيع التكاليف الثابتة للمكتب على عدد كبير من المستفيدين، لذلك قد يكون من غير الفعال طرح البرنامج على المستوى الفردي، وسيكون أكثر نجاعة عند طرحه على مستوى المجتمع المحلي. ومع ذلك، في المواقف ذات أنواع التدخلات الجديدة وغير المختبرة، قد يكون من المفيد استيعاب أوجه القصور على المدى القصير وطرح البرنامج داخل المناطق الإدارية، وذلك لضمان مصداقية التقييم بشكل أفضل وخفض تكاليف جمع البيانات.

يجادل بعض مديري البرامج بأن البرامج المدارة محليًا، مثل برامج التأمين الصحي، لا تملك القدرات الإدارية لتنفيذ البرامج على المستوى الفردي. وهم قلقون من أن إنشاء أنظمة لتقديم مزايا مختلفة إلى مختلف المستفيدين داخل الوحدات الإدارية المحلية سيمثل عبئًا، وأنه قد يكون من الصعب ضمان تنفيذ تعيين مجموعات المعالجة والمقارنة على النحو المصمم. تمثل القلق الأخير تهديدًا خطيرًا لتقييم الأثر، حيث قد لا يتمكن مديرو البرنامج من تنفيذ البرنامج بشكل متنسق مع تصميم التقييم. في هذه الحالة، قد يكون التنفيذ على مستوى أعلى أو تبسيط تصميم تقييم الأثر أمرًا ضروريًا.

تفضل الحكومات أحيانًا تنفيذ البرامج على مستويات أكثر اتساعًا، مثل المجتمع المحلي، لأنها تقلق بشأن التوترات المحتملة عندما يلاحظ أعضاء مجموعة المقارنة جيرانهم في مجموعة المعالجة يتمتعون بالمزايا. تم تنفيذ العديد من البرامج بنجاح على مستوى الفرد أو الأسرة داخل المجتمعات دون نشوب توترات، لا سيما عندما تم توزيع المزايا بطريقة منصفة وتتسم بالشفافية وتخضع للمساءلة. ومع ذلك، فإن خطر نشوء التوترات يجب أن يؤخذ في الاعتبار في سياق تقييم الأثر المحدد.

أخيرًا، عندما يتم تعيين برنامج وتنفيذه على مستوى منخفض جدًا، مثل الأسرة أو المستوى الفردي، فإن تداخل مجموعة المقارنة قد يضر بالصلاحيات الداخلية للتقييم. على سبيل المثال، لنفترض أنك تقيم فعالية توفير مياه الصنبور على صحة الأسرة. إذا قمت بتثبيت الحنفيات لأسرة دون جارتها، فقد تشارك الأسرة المعالجة في استخدام الحنفية مع جار للمقارنة؛ عندئذٍ لن تمثل الأسرة المجاورة مقارنة حقيقية، حيث إنها ستستفيد من التأثير غير المباشر. يوضح المربع ١-١١ الآثار المترتبة على اختيار مستوى تنفيذ التدخل في سياق برامج التحويلات النقدية. من الناحية العملية، يحتاج مديرو البرامج إلى اختيار أصغر وحدة تدخل ممكنة والتي (١) تسمح بعدد كبير كافٍ من الوحدات للتقييم، (٢) تخفف من المخاطر على الصحة الداخلية، (٣) تلائم السياق التشغيلي.

المربع ١-١١ برامج التحويلات النقدية والمستوى الأدنى للتدخل

للتسجيل في البرنامج في ربيع عام ١٩٩٨، وأُتيحت الفرصة نفسها لجميع الأسر المؤهلة في مجتمعات المقارنة بعد ١٨ شهرًا، أي في شتاء ١٩٩٩. ومع ذلك، وجد فريق التقييم ارتباطًا جوهريًا في الحاصل بين الأسر داخل المجتمعات. لذلك، لتوليد قدرة إحصائية كافية للتقييم، احتاجوا إلى عدد أكبر من الأسر في العينة مما كان سيحتاجون إليه إذا كانوا قادرين على تعيين أسر فردية لمجموعات المعالجة والمقارنة. وبالتالي، أدت استحالة تنفيذ البرنامج على مستوى الأسرة إلى زيادة متطلبات حجم العينة وزيادة تكلفة التقييم. تنطبق قيود مماثلة على العديد من البرامج في قطاع التنمية البشرية.

تستخدم غالبية التحويلات النقدية المشروطة المجتمعات كوحدة أو مستوى من التدخل، لأسباب إدارية وتصميم البرنامج، وكذلك بدافع القلق بشأن التداعيات والتوترات المحتملة في المجتمع إذا تم تعيين المعالجة على مستوى أدنى.

على سبيل المثال، اعتمد تقييم برنامج التحويلات النقدية المشروطة في المكسيك، Progreso/Oportunidades، على إطلاق البرنامج على مستوى المجتمع المحلي في المناطق الريفية لتعيين المجتمعات بشكل عشوائي لمجموعات المعالجة والمقارنة. أُتيحت الفرصة لجميع الأسر المؤهلة في مجتمعات المعالجة

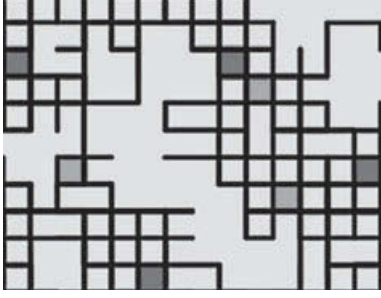
المصادر: Behrman and Hoddinott 2001; Skoufias and McClafferty 2001.

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).

المراجع

- Behrman, Jere R., and John Hoddinott. 2001. "An Evaluation of the Impact of PROGRESA on Preschool Child Height." Discussion Paper No. 104, Food Consumption and Nutrition Division, International Food Policy Research Institute, Washington, DC.
- Grosh, M. E., C. Del Ninno, E. Tesliuc, and A. Ouerghi. 2008. *For Protection and Promotion: The Design and Implementation of Effective Safety Nets*. Washington, DC: World Bank.
- Skoufias, Emmanuel, and Bonnie McClafferty. 2001. "Is Progreso Working? Summary of the Results of an Evaluation by IFPRI." International Food Policy Research Institute, Washington, DC.



إدارة تقييم الأثر

إدارة فريق التقييم، والوقت، والميزانية

يعد التقييم شراكة بين فريق السياسات وفريق الأبحاث. تعتمد كل مجموعة على الأخرى لنجاح التقييم. ويشكلون معًا فريق التقييم. تستند الشراكة إلى فهم الأدوار والمسؤوليات الخاصة بالفريقين، والالتزام المشترك بالتقييم، والاعتراف بما يحفز عمل الأفراد في التقييم. تعتبر الشراكة الفعالة أمرًا بالغ الأهمية لضمان المصداقية الفنية وتأثير السياسة العامة للتقييم.

يحدد هذا الفصل عناصر الشراكة الفعالة، بما في ذلك أدوار ومسؤوليات كل فريق. يستكشف كيف تعمل الشراكة في مراحل مختلفة من عملية التقييم والمراجعات والنماذج البديلة للتعاون. يتناول هذا الفصل أيضًا مسائل عملية تتعلق بالتوقيت وأعمال الميزانية.

فريق البحث: وظيفة البحث ووظيفة البيانات

يعد فريق البحث مسؤولاً عن الجودة الفنية والنزاهة العلمية لأعمال التقييم. تشمل مسؤولياته تصميم البحث وجودة البيانات والتحليل. تتكون فرق البحث عادةً من الأشخاص التاليين:

- **يعمل الباحث الرئيسي** مع صانعي السياسات ومنفذي البرامج على تحديد الأهداف الرئيسية وأسئلة السياسة والمؤشرات واحتياجات التقييم من المعلومات (غالبًا باستخدام نظرية التغيير كما تصورها سلسلة النتائج)، وتحديد منهجية تقييم الأثر، وتطوير خطة التقييم، وتحديد فريق البحث، وتسجيل تقييم الأثر، والحصول على الموافقات من مجلس المراجعة المؤسسية (IRB)، وإعداد خطة تقييم تشمل خطة تحليل أكثر تفصيلاً، وقيادة تحليل النتائج، والمشاركة مع فريق السياسات لنشر النتائج. يجب أن يكون الباحث الرئيسي قادرًا على العمل بفعالية مع فريق التقييم الكامل، متضمنًا المنظمة المسؤولة عن جمع البيانات والأعضاء الآخرين في فريق البحث وصانعي السياسات أو منفذي البرامج الذين يستخدمون البيانات ونتائج التقييم. قد يعمل عدد من الباحثين مع المحقق الرئيسي أو كمحققين رئيسيين مشاركين لقيادة أو دعم العمل التحليلي المحدد على العناصر، مثل أخذ العينات أو التقدير النوعي أو تحليل فعالية التكلفة.
- **يعمل مدير التقييم أو المنسق الميداني** مباشرة مع المحقق الرئيسي في التنفيذ اليومي للتقييم. ويشمل هذا العمل مع منفذي البرامج وصانعي السياسات في فريق السياسات والإشراف على العمل الميداني عند جمع البيانات الأولية. ويعد هذا الشخص مهمًا بشكل خاص في الحالات التي لا يكون فيها المحقق الرئيسي متمركزًا محليًا، حيث يتم تطبيق تقييم استباقي يحتاج إلى تنسيقه بدقة مع تنفيذ البرنامج، أو حيث يتم جمع البيانات الأولية.
- **يوجه خبير أخذ العينات** العمل على حسابات القوة الإحصائية وأخذ العينات. بالنسبة لنوع تقييم الأثر الكمي الذي تم تناوله في هذا الكتاب، يجب أن يكون خبير أخذ العينات قادرًا على إجراء حسابات القوة الإحصائية لتحديد أحجام العينات المناسبة للمؤشرات المحددة؛ تحديد العينة، ومراجعة نتائج العينة الفعلية مقابل العينة المصممة، وتقديم المشورة بشأن الآثار المترتبة على التحليل بما يتماشى مع خطة التحليل المسبق. غالبًا ما يقوم المحقق الرئيسي بالوظائف بشكل مباشر أو مع خبير أخذ العينات.

- يعد فريق جمع البيانات مسؤولاً عن تطوير أدوات جمع البيانات والأدلة المصاحبة ودفتر الرموز؛ جمع البيانات ورقمنتها وتنظيفها، وتقديم مجموعة بيانات نظيفة وموثقة، عندما يكون جمع البيانات الأولية مطلوباً. يناقش الفصل ١٦ مصادر البيانات والجوانب المختلفة لجمع البيانات.

فريق البحث: وظيفة السياسة ووظيفة إدارة البرنامج

يتكون فريق السياسات من صانعي السياسات ومنفذي البرامج:

- يحدد صانعو السياسات جدول أعمال البحث، ويحددون سؤال الدراسة الأساسي الذي يتعين تناوله، ويضمنون توفر الموارد الكافية للعمل، ويطبّقون النتائج على السياسة. يحتاجون في بداية التقييم إلى توضيح أهداف كل من البرنامج والتقييم، وكذلك نظرية التغيير والمؤشرات الرئيسية ذات الصلة، بما في ذلك الحد الأدنى لحجم التأثير ذي الصلة بالسياسة لمؤشرات الحاصل محل الاهتمام، على النحو المبين في الفصل ٢. يتمتع فريق السياسات بمعرفة حوار السياسات، والاتصالات مع الأطراف المعنية الرئيسية لضمان أن التقييم مصمم ليكون وثيق الصلة بالسياسات قدر الإمكان، ولضمان مشاركة الأطراف المعنية المناسبة ومتخذي القرار في النقاط الرئيسية في عملية التقييم.
- يتعاون منفذي البرنامج مع فريق البحث لمواءمة تصميم التقييم وتنفيذ البرنامج. ويشمل ذلك التحقق من أن تصميم التقييم يستند إلى معلومات دقيقة حول تشغيل البرنامج، والالتزام بتنفيذ البرنامج كما هو مخطط له، في حالة التقييمات السابقة. يقوم عادةً منفذو البرنامج التابعون لفريق السياسة أيضًا بإدارة ميزانية التقييم وغالبًا ما يشاركون في مساعدة فريق البحث في الإشراف على العمل الميداني لجمع البيانات.

من يهتم بالتقييم ولماذا؟

المفهوم الأساسي

تعد الشراكة الفعالة بين فريق السياسة وفريق البحث أمرًا بالغ الأهمية لضمان المصداقية الفنية وأثر السياسة العامة للتقييم.

من وجهة نظر فريق السياسة، يكون الاهتمام الأساسي عادةً هو معرفة ما إذا كان البرنامج أو الإصلاح فعالاً أم لا، وتكلفة تحقيق النتائج، وبالتالي السماح للفريق باتخاذ قرارات السياسة على أساس الأدلة المقدمة. ويهتم منفذو البرنامج المحليون بضمان تقدير جهودهم وحصولهم على التقدير ورؤية عملهم، والذي غالبًا ما يتجاوز حدود مسؤولياتهم اليومية. تتمثل إحدى الطرق الجيدة لتقدير هذه المساهمات في ضمان مشاركة الفرق المحلية بنشاط في أنشطة التقييم على نطاق أكبر. يمكن القيام بذلك من خلال عقد ورش عمل مشتركة، وكذلك عن طريق إصدار منشورات مشتركة، وضمان التدريب وبناء القدرات، وإشراك الباحثين المحليين الذين هم في وضع جيد للمساهمة بشكل جوهري ويمكن أن يكونوا بمثابة عامل ربط مهم بين فرق البحث وفرق السياسات.

للتقييمات قيمة من حيث الصالح العام وذلك عندما تطرح مسألة مهمة تتجاوز الاهتمام المباشر لفريق السياسة. وغالبًا ما يكون هذا الجانب من الاهتمامات الأساسية للباحثين الذين يستكشفون المسائل الأساسية المتعلقة بنظرية التغيير. على سبيل المثال، النتائج المتعلقة بكيفية تصرف الناس في ظل ظروف معينة أو كيفية عمل قنوات الإرسال، ما يسمح بتحقيق الآثار، ويمكن أن تسمح باستخلاص دروس عامة وتطبيقها في بيئات مختلفة. تساهم تقييمات الأثر بسرعة في إنشاء قاعدة أدلة عالمية بشأن أداء مجموعة من إصلاحات البرامج والسياسات، وتشكل مستودعات للمعرفة وثيقة الصلة بتصميم البرامج والسياسات. غالبًا ما يهتم المانحون ومعاهد السياسات بهذه القيمة الأوسع للمصلحة العامة ويقدمون بشكل متزايد الدعم المالي لإجراء التقييمات التي تساهم في قاعدة الأدلة هذه.

سيكون الباحثون أيضًا ملتزمين جدًا باستخدام منهجية تقييم قوية يمكن الدفاع عنها وسيرغبون في التأكد من مشاركتهم في تصميم تقييم الأثر، وفي تحليل البيانات، وفي إنتاج البحوث الأولية التي تلبى المتطلبات العلمية للنشر في المجلات الأكاديمية. تواجه فرق البحث متعددة التخصصات تحديًا إضافيًا يتمثل في ضمان وجود فهم مشترك بين أعضاء الفريق. قد يكون للتخصصات المختلفة، من قبيل الطب والاقتصاد، مناهج مختلفة لتسجيل التجارب أو إشراك الموضوعات أو الإبلاغ عن النتائج أو نشر النتائج مثلًا. من الأفضل توضيح هذه التوقعات المختلفة وفهمها في بداية التقييم. بغض النظر عن البروتوكولات المختلفة، من المتوقع أن تتبع فرق البحث المعايير العلمية والمبادئ الأخلاقية المقبولة عمومًا، التي تتم مناقشتها في الفصل ١٣.

يمكن أن تتسبب الاهتمامات المختلفة لفريق السياسة وفريق البحث في نشوب توترات تحتاج إلى إدراكها وإدارتها. يميل الباحثون إلى تقدير الدقة الفنية في تصميم التقييم أكثر من الجدوى التشغيلية لتنفيذ البرنامج. قد تكون الفرق أيضًا مهتمة بأسئلة تقييم مختلفة إلى حد ما. أخيرًا، قد لا يكون أي من الفريقين مهتمًا بنشر نتائج متباينة أو سلبية، لأن هذا قد يكون له تأثير سيئ على أداء البرنامج لفريق السياسة وربما يكون أقل اهتمامًا أكاديميًا لفريق البحث. قد يهتم فريق السياسة أيضًا بكونه انتقائيًا بشأن النتائج التي يتم إصدارها، في حين أن فريق البحث سيقدر بشدة القوة على نشر النطاق الكامل للنتائج. بالنسبة لفريق التقييم ككل، يعد تعزيز ثقافة الشفافية واحترام الأدلة أمرًا بالغ الأهمية. ينبغي مكافأة صانعي السياسات ومديري البرامج على التزامهم بوضع السياسات القائمة على الأدلة. حتى عندما تكون النتائج غير مواتية، يجب أن يُنسب الفضل إليهم لدعمهم للشفافية. وبالمثل، ينبغي تشجيع فريق البحث على الإبلاغ عن النتائج ونشرها، بغض النظر عن الاستنتاجات.

شراكة فريق البحث وفريق السياسة أثناء التقييم

تعتمد الجودة التقنية للتقييم وأثر سياسته على الشراكة النشطة بين فريق البحث وفريق السياسة في كل مرحلة من مراحل التقييم: التصميم والتنفيذ والتحليل والنشر. يلخص المربع ١٢-١ بعض المبادئ التوجيهية. مرحلة *التصميم*. أولاً، يحتاج صانعو السياسات إلى هيكلية الأسئلة البحثية الأساسية ونقلها بوضوح، وكذلك نظرية التغيير المصاحبة والمؤشرات الأساسية محل الاهتمام، والتأكد من أن فريق البحث يفهم هذه العناصر جيداً ويحترمها. لضمان ملاءمة السياسة، يحتاج فريق السياسة أيضاً إلى أخذ زمام المبادرة في هيكلية إستراتيجية المشاركة التي تضمن استشارة الأطراف المعنية المهمة وإبلاغها بتصميم التقييم وتنفيذه ونتائجه. وبالنسبة للباحثين، يحتاجون إلى أن يوضحوا لفريق السياسة الشروط اللازمة لإجراء تقييمات جيدة للأثر. ففي حالة التقييمات السابقة، يتضمن ذلك أولاً التحقق مع منفذي البرنامج وصانعي السياسات في فريق السياسة من أن عمليات البرنامج جيدة بما يكفي لضمان أن البرنامج الذي يتم تقييمه لن يتغير كثيراً أثناء التقييم، وبالتالي ستكون النتائج متصلة بأغراض السياسة. غالباً ما تكون "النقطة المثالية" لإجراء تقييم الأثر هي النقطة التي يتم فيها اختبار البرنامج بما يكفي لإثبات أنه يعمل بالطريقة المقصودة، والتي يمكن الاستدلال عليها من خلال إجراء تقييم صيرورة جيد، ولكن ليس على مستوى موسع، وبالتالي يتيح المزيد من الخيارات لبناء الوقائع المضادة المناسبة.

المربع ١٢-١: المبادئ التوجيهية للتعاون بين فرق السياسات والتقييم

- المشاركة المبكرة لزيادة خيارات تصميم التقييم وضمان شراكة فعالة بين فرق السياسة وفرق التقييم.
- وجود خطة واضحة لتقييم الأثر في البداية.
- فهم أدوار ومسؤوليات ودوافع الأطراف المعنية المختلفة ومنحها دوراً في التقييم.
- الاستمرار في المشاركة طوال التقييم لضمان التوافق المناسب بين التقييم والتدخل الذي يتم تقييمه.
- الاعتراف بالمخاطر والمزايا وإدارتهما، مع توضيح ما يمكن أن تفعله تقييمات الأثر وما لا يمكنها القيام به.
- تقدير الشفافية وضمان الموضوعية والاستعداد لاحترام النتائج، سواء كانت جيدة أو سيئة.

ثانيًا، يحتاج فريق البحث إلى فهم قواعد عمل البرنامج بوضوح: أي الموارد المتاحة ومعايير الأهلية لاختبار المستفيدين وتوقيت التنفيذ. ينبغي لفريق السياسة أن ينقل بوضوح قواعد التشغيل الثلاث هذه إلى فريق البحث، لأنها أساسية للوقوف على الخيارات المنهجية المتاحة في التقييم، على النحو الموضح بالتفصيل في الفصل ١١.

ثالثًا، يجب على فريق البحث إعداد خطة تقييم الأثر التي تحتوي على كل من الجوانب التشغيلية والبحثية، ومشاركتها مع صانعي السياسات لضمان تركيز التقييم على الأسئلة محل الاهتمام؛ أن عناصر التعاون مع فريق السياسات موضحة؛ وأن فريق التقييم واضح ومباشر بشأن الأسئلة المطروحة وطبيعة النتائج وتوقيتها (انظر المربع ١٢-٢). كما يعد اعتبار المخاطر وإستراتيجيات التخفيف المقترحة أمرًا مفيدًا. أخيرًا، يجب على فريق البحث الحصول على الموافقة الأخلاقية من مجلس المراجعة المؤسسي وتسجيل التقييم في سجل التجارب (انظر الفصل ١٣). يجب أن ينتج عن هذا الحوار، أثناء مرحلة التصميم، التزامًا واضحًا ومشتريًا بخطة التقييم بين أعضاء فرق السياسة والبحث، وكذلك توقعات واقعية ومسؤوليات متفق عليها بشكل متبادل. يوفر هذا الحوار فرصة لفريق البحث لتوضيح كل من قيمة تقييم الأثر، لا سيما تحديد السببية وإمكانية تعميم النتائج، وقيودها؛ مثل عدم تقديم رؤى حول سبب الحصول على نتائج معينة أو المفاضلات بين أحجام العينات وحسابات القوة الإحصائية أو الوقت الذي يستغرقه إنتاج بعض النتائج. يوفر هذا الحوار أيضًا الفرصة لفريق السياسات لتحديد الأسئلة ذات الأولوية ولضمان توافق التقييم جيدًا مع أسئلة السياسة محط الاهتمام.

مرحلة *التنفيذ*. تحتاج فرق السياسة والبحث إلى العمل معًا لضمان استمرار التنفيذ بسلاسة واستكشاف الأخطاء وإصلاحها. على سبيل المثال، في التجارب العشوائية المضبوطة، تحتاج الفرق إلى الاتفاق على أفضل طريقة عملية للاختيار العشوائي. بالإضافة إلى ذلك، خلال هذه المرحلة، يكون للتنسيق أهمية خاصة لضمان التوافق بين تصميم التقييم وتنفيذ البرنامج.

مرحلة *التحليل*. يجب أن يتوافق التحليل الذي يُجرى مع ما تم تحديده في خطة التقييم وفي خطة التحليل المسبق الأكثر تفصيلًا. يجب على فريق البحث تقديم النتائج ومناقشتها مع فريق السياسات في المراحل الرئيسية. ويجب في البداية، عند خط الأساس، أن يشمل ذلك مراجعة جودة البيانات التي تم جمعها والالتزام بخطة التقييم. سيساعد ذلك على ضمان أن تظل خطة التقييم المتصورة في مرحلة التصميم قابلة للتطبيق ويسمح بإجراء أي تعديلات ضرورية. كما تعد أيضًا فرصة ممتازة لمراجعة المنتجات التي سيتم تسليمها في كل مرحلة من مراحل التحليل ومعرفة ما إذا كان إنتاج تلك النتائج يسير على الطريق الصحيح فيما يتعلق باحتياجات فريق السياسة لاتخاذ القرار. بمجرد أن ينتهي فريق التقييم من تحليل الأثر، يجب تقديم النتائج الأولية ومشاركتها مع فريق السياسات لضمان الإجابة على أي سؤال والتحضير لمرحلة النشر.

المربع ١٢-٢: مخطط عام لخطة تقييم الأثر

١. المقدمة
٢. وصف التدخل
٣. أهداف التقييم
- ١-٣ الفرضيات، نظرية التغيير، سلسلة النتائج
- ٢-٣ أسئلة السياسات
- ٣-٣ مؤشرات الحاصل المحتمل
- ٤-٣ المخاطر
٤. تصميم التقييم
٥. تحديد العينات والبيانات
- ١-٥ إستراتيجية أخذ العينات
- ٢-٥ حسابات القوة الإحصائية
٦. نظرة عامة على خطة التحليل المسبق
٧. خطة جمع البيانات
- ١-٧ مسح خط الأساس
- ٢-٧ مسح (مسوحات) المتابعة
٨. المنتجات التي سيتم تسليمها
- ١-٨ تقرير خط الأساس
- ٢-٨ تقرير تقييم الأثر
- ٣-٨ ملخص السياسات
- ٤-٨ مجموعات البيانات الموثقة بالكامل وبروتوكولات التصميم والتحليل
٩. خطة النشر
١٠. البروتوكولات الأخلاقية الخاصة بحماية البشر
- ١-١٠ ضمان الموافقة المستنيرة
- ٢-١٠ الحصول على الموافقة من مجلس المراجعة المؤسسية (IRB)
١١. الجدول الزمني
١٢. الميزانية والتمويل
١٣. تكوين فريق التقييم وأدواره

مرحلة النشر. في هذه المرحلة، يحتاج فريق السياسات إلى التأكد من وصول نتائج التقييم إلى الأشخاص المناسبين في الوقت المناسب وبصيغة مناسبة. كما أنها المرحلة التي يتم فيها ضمان أن جميع البيانات الناتجة عن التقييم موثقة بشكل مناسب. غالبًا ما تستخدم الفرق إستراتيجيات ووسائل متعددة لنشر النتائج، مع مراعاة مختلف الفئات المستهدفة، على النحو الموضح في الفصل ١٤.

يعد التقييم توازنًا بين الخبرة الفنية والاستقلالية التي يساهم بها فريق البحث وملاءمة السياسة والتوجيه الإستراتيجي والتنسيق التشغيلي الذي يساهم به صانعو السياسات ومنفذو البرامج في فريق السياسة. يمكن استخدام مجموعة من النماذج لإنشاء هذه الشراكة بين فرق البحث وفرق السياسات وتنفيذها.

سيعتمد اختيار الطريقة على سياق وأهداف تقييم الأثر، وكذلك على اعتبار مجموعة من المخاطر. فمن ناحية، قد يقدم فريق بحث مستقل تمامًا مع تعاون محدود مع فريق السياسة تقييماً للأثر غير ذا صلة بأسئلة السياسة محل الاهتمام أو يستخدم منهجاً مقيداً بالتفاعلات غير الكافية مع منفذي البرنامج. ومن الناحية الأخرى، قد يؤدي فريق بحث مندمج تمامًا مع فريق السياسة إلى إنشاء مخاطر تضارب المصالح أو يؤدي إلى فرض رقابة على بعض النتائج إذا لم يتم تطبيق مبادئ العلوم المفتوحة (انظر الفصل ١٣). بالإضافة إلى ذلك، يمكن أن يكون للتقييمات، في كثير من الأحيان، غايات متعددة، منها بناء قدرات التقييم داخل الوكالات الحكومية وتوعية مشغلي البرامج بواقع مشاريعهم كما يتم تنفيذها في الميدان. قد تحدد جزئياً هذه الغايات الأوسع نطاقاً النموذج الذي سيتم اختياره.

بشكل عام، إن أكثر ما يهم جودة تقييم الأثر هو ما إذا كانت مقارنة الشراكة سينتج تقديرات غير متحيزة عن آثار البرنامج. طالما يتم احترام مبادئ أخلاقيات البحث والعلوم المفتوحة، يميل الحياد والموضوعية إلى أن يكونا أكثر أهمية لجودة تقييم الأثر من الاستقلال الوظيفي لفرق البحث وفرق السياسات. ومن الناحية العملية، غالباً ما يكون التعاون الوثيق بين فرق البحث وفرق السياسة مطلوباً لضمان وضع إستراتيجية تقييم الأثر الأعلى جودة.

نموذج الاستعانة بمصادر خارجية

بالنسبة إلى منفذي البرامج المشغولين الذين يديرون عمليات معقدة، فإن تكليف فريق خارجي بمسؤولية تصميم وتنفيذ تقييم الأثر غالباً ما يكون أمراً جذاباً. ويمكن أن تتخذ نماذج الاستعانة بمصادر خارجية أشكالاً مختلفة. يقوم مديرو البرامج في بعض الأحيان بالاستعانة بمصادر خارجية لتصميم تقييم الأثر وكذلك تنفيذ المسوح المختلفة (عادةً، مسح خط الأساس ومسح المتابعة)، وإسنادها إلى كيان واحد في عقد واسع النطاق. وفي حالات أخرى، يقوم مديرو البرامج في البداية بالاستعانة بمصادر خارجية للتصميم، ويتابعونه بعقود للمراحل المختلفة لجمع البيانات وتحليلها.

يتسبب الاستعانة بمصادر خارجية في الفصل بين تصميم تقييم الأثر وتنفيذه، ما قد يجعل تقييم الأثر أكثر استقلالية. ومع ذلك، فإن الاستعانة بمصادر خارجية بشكل كامل لتقييم الأثر يمكن أن يكون مصحوبًا بمخاطر كبيرة. حيث يمكن أن يحد إنشاء هذا النوع من العلاقات التعاقدية من التعاون بين تنفيذ البرنامج وفرق البحث المتعاقد معها. في بعض الحالات، يتم منح الفريق المتعاقد مجموعة من معلومات البرنامج المحددة مسبقًا، مع هامش ضئيل لمناقشة خطط التصميم والتنفيذ أو نطاق تشكيل البحث. وفي حالات أخرى، قد لا يتم تحديد قواعد البرنامج وطرق التنفيذ اللازمة لتصميم تقييم جيد للأثر. في مثل هذه الحالات، يكون للفريق المتعاقد المسؤول عن تقييم الأثر تأثير محدود في ضمان تحديد هذه العناصر.

وفي حالات أخرى أيضًا، قد يكون البرنامج قد تم تصميمه بالفعل أو قد يكون قد بدأ تنفيذه، مما قد يقيد بشدة الخيارات المنهجية للتقييم. غالبًا ما يُطلب من الفريق المتعاقد أن يتكيف مع التغييرات في تنفيذ البرنامج بأثر رجعي، دون المشاركة عن كثب أو إبلاغه أثناء التنفيذ. يمكن أن تؤدي هذه المواقف إلى تصميمات تقييم دون المستوى الأمثل أو إلى نشوب تحديات أثناء التنفيذ، حيث قد يكون لدى الفريق المتعاقد دوافع مختلفة عن الباحثين وصانعي السياسات الذين قادوا تصميم التقييم.

وأخيرًا، قد يمثل اختيار الفريق المتعاقد معه والإشراف عليه تحديًا لوحدة تنفيذ البرنامج. يجب مقدمًا مراعاة قواعد التعاقد بعناية وذلك لضمان نجاعة الاستعانة بمصادر خارجية وعدم وجود تضارب في المصالح. قد تحد بعض القواعد من إمكانية قيام فريق، متعاقد معه للمساهمة في تصميم تقييم الأثر، عن تقديم عطاءات لاحقة بشأن تنفيذه. وللتخفيف من هذه المخاطر، من الأفضل بشكل عام لفريق السياسة أن يكون لديه بالفعل تصميم لتقييم الأثر، بما في ذلك إستراتيجية التعريف، ومؤشرات الحاصلات الأساسية، وحسابات القوة الإحصائية الأولية، وأحجام العينة التقريبية. سيساعد هذا في توجيه عمليات الاتفاق والتعاقد، حيث إن هذه العناصر تؤثر بشدة على ميزانيات التقييم. يجب على فريق السياسات أيضًا وضع آليات لضمان الإشراف الفني القوي على تصميم تقييم الأثر وتنفيذه. يمكن أن يكون ذلك من خلال لجنة إشراف أو من خلال مراجعة فنية وعلمية منتظمة لمنتجات تقييم الأثر. وتشير تدابير التخفيف هذه مجتمعة إلى أن النموذج الأكثر فعالية عادة ليس نموذج الاستعانة بمصادر خارجية كاملة.

نموذج الشراكة

لا يقتصر التعاون بين فرق البحث وفرق السياسات بالضرورة على العلاقات التعاقدية فقط، حيث يمكن إقامة شراكات مفيدة بشكل متبادل عندما يهتم الباحثون بإجراء بحث حول مسألة تتعلق بالسياسات وعندما يسعى صانعو السياسات ومنفذو البرامج إلى ضمان إجراء تقييم جيد للأثر في مشروعهم. يتحلى الباحثون بحوافز تدفعهم لمعالجة المسائل الجديدة التي ستضيف إلى قاعدة الأدلة العالمية، ولدفع تقييم الأثر والمساهمة في ظهوره على نطاق أوسع. قد يكون فريق البحث قادرًا على الاستفادة من بعض التمويل المشترك لتقييم الأثر إذا كانت أهداف الممولين متوافقة بشكل وثيق مع تركيز البحث في التقييم.

يتمثل نوع آخر من النماذج المتكاملة التي أصبحت أكثر بروزًا، لا سيما في المؤسسات الكبرى ومنها البنك الدولي وبنك التنمية للبلدان الأمريكية، في استخدام القوة البحثية الداخلية لتقييم الأثر لدعم فرق السياسات والبرامج.

تنطوي مقارنة الشراكة على بعض المخاطر. ففي بعض الأحيان، قد يسعى الباحثون إلى دمج عناصر بحثية جديدة في تقييم الأثر والتي قد لا تتماشى تمامًا مع أهداف السياسة الحالية على المستوى المحلي، على الرغم من أنها يمكن أن تضيف قيمة أكثر على الصعيد العالمي. وقد لا يُقدر دائمًا صانعو السياسات ومنفذو البرامج الدقة العلمية اللازمة لإجراء تقييمات الأثر الصارمة، وقد يكون لديهم درجة تساهل أكبر من فريق البحث فيما يتعلق بالمخاطر المحتملة على تقييم الأثر.

للتخفيف من هذه المخاطر، يجب أن تكون أهداف فريق البحث وفرق السياسات متسقة بشكل وثيق. على سبيل المثال، يمكن أن تعمل فرق البحث وفرق السياسة معًا مقدمًا على خطة تقدير شاملة وتضع إستراتيجية مفصلة بالإضافة إلى أدوار ومسؤوليات الفرق المعنية (انظر المربع ١٢-٢). كما تعد خطة تقييم الأثر موضوعًا لتسليط الضوء على القواعد التشغيلية الرئيسية، فضلًا عن المخاطر التشغيلية المحتملة لتنفيذ تقييم الأثر.

يعد الالتزام المتبادل بتقييم الأثر كما ورد في خطة التقييم الواضحة أمرًا ضروريًا لكي تعمل الشراكة بسلاسة، حتى في غياب وجود علاقة تعاقدية. ومن الممارسات الجيدة لهذا الالتزام المتبادل أن يتخذ شكل اتفاقية مكتوبة، في شكل شروط مرجعية أو مذكرة تفاهم مثلًا، لتحديد الأدوار والمسؤوليات ونواتج تقييم الأثر. يمكن أيضًا تضمين هذه الجوانب في خطة تقييم الأثر.

النموذج المتكامل تمامًا

يتم تنفيذ بعض تقييمات الأثر من خلال نموذج متكامل تمامًا حيث يكون فريق البحث هو نفسه فريق تنفيذ البرنامج. يتم اتباع هذه المقاربة أحيانًا في تجارب الفعالية، حيث يتم اختبار تدخلات جديدة من أجل إثبات المفهوم. في هذه الحالة، يفضل الباحثون عمومًا الحفاظ على السيطرة على التنفيذ لضمان تنفيذ البرنامج بأكبر قدر ممكن من تصميمه الأصلي. في حين أن مثل هذه التقييمات للأثر تمثل أفضل طريقة لاختبار النظريات الأساسية وتحديد ما إذا كان تدخل معين يمكن أن يعمل في ظروف مثالية، فإن الخطر يكمن في أن النتائج قد تكون لها صلاحية خارجية محدودة.

يعرض المربع ٣-١٢ بعض الأمثلة على النماذج المختلفة التي يمكن أن تستخدمها فرق البحث والسياسة للتعاون.

المربع ٣-١٢: أمثلة على نماذج فريق البحث والسياسة معًا

وغيرها من الشركاء المنفذين. منذ أن تم وضع تصور للتقييم لأول مرة، يعمل الباحثون المنتسبون إلى IPA من شبكة عالمية من الجامعات مع المديرين القطريين في المكاتب الميدانية ذات الصلة لوضع خطة تصميم وتنفيذ للتقييم. يعد المديرين القطريين مسؤولين عن قيادة علاقات الشركاء والتوفيق بين المستثمرين الرئيسيين في فريق البحث وشركاء البرنامج في فريق السياسات لوضع اقتراح للتقييم. بمجرد الموافقة على الاقتراح، يقومون بتعيين طاقم عمل لإدارة المشروع لقيادة عملية جمع البيانات في الواقع، وجميعهم موجودون في المكتب الميداني ل IPA. إن التنسيق بين الباحثين ومنفذي البرنامج وثيق بشكل عام، وفي بعض الحالات، تكون مكاتب IPA مسؤولة أيضًا عن تنفيذ التدخل الذي يتم تقييمه.

نماذج الشراكة في البنك الدولي

في العقد الماضي، قام البنك الدولي بسرعة بتوسيع نطاق استخدام تقييمات الأثر السابقة لتقييم آثار بعض مشاريع التنمية التي يمولها. تقدم عدة مجموعات، منها SIEF و DIME (Development Impact Evaluation)، و GIL (Gender Strategic Impact Evaluation Fund)، و Innovation Lab، التمويل والدعم الفني لتقييم الأثر. عندما يتم وضع مشروع مبتكر أو عالي المخاطر بشكل خاص، يتم إعداد أنشطة تقييم الأثر، إما مدمجة في المشروع وتديرها نظيراتها من الحكومات، أو كأنشطة مستقلة يديرها البنك الدولي. وفي كلتا الحالتين، يتم إنشاء فريق تقييم، يتألف من فريق بحث، يشمل مزيًا من الخبراء الفنيين والأكاديميين، وفريق السياسات، متضمنًا عادة صانعي السياسات، ومنفذي البرامج، وقادة الفريق التشغيلي للمشروع.

(تابع)

عمليات التقييم بالاستعانة بمصادر خارجية في مؤسسة تحدي الألفية (Millennium Challenge Corporation)

تأسست مؤسسة تحدي الألفية (MCC)، وهي وكالة مساعدات أمريكية، في عام ٢٠٠٤، وتولي تركيزًا قويًا على المساءلة والنتائج. فهي تتطلب أن يكون لكل برنامج من برامجها التنموية خطة شاملة للمتابعة والتقييم، مع التركيز على التقييمات غير المنحازة والمستقلة. أدى هذا التركيز إلى قيام MCC بتطوير نموذج يتم فيه الاستعانة بمصادر خارجية لتصميم وتنفيذ التقييمات بالكامل وإسنادها لباحثين خارجيين. أدى الفصل بين فريق البرنامج والباحثين الخارجيين، في السنوات الأولى من عمليات MCC، إلى مشاكل في بعض الأحيان. ففي هندوراس، على سبيل المثال، صمم الباحثون تجربة عشوائية مضبوطة لبرنامج تدريب المزارعين. ومع ذلك، ونظرًا لأن عقد التنفيذ كان قائمًا على الأداء، كان للمنفذ حافزًا قويًا لإيجاد مزارعين ذوي أداء عالٍ للمشاركة في البرنامج، ولم يتم تعيين المزارعين المؤهلين بشكل عشوائي في البرنامج، مما أدى إلى إبطال صلاحية تصميم التقييم. مع إصدار التقييمات الخمسة الأولى لبرنامج تدريب المزارعين، فكرت MCC في تجارب مثل هذه وخلصت إلى أن التعاون بين المنفذين والمقيمين هو أمر بالغ الأهمية في جميع مراحل التصميم والتنفيذ. قامت المنظمة بتعديل نموذجها ليكون أكثر انتقائية عند تطبيق تقييمات الأثر من أجل تحقيق التوازن بين المساءلة والتعلم.

التكامل في منظمة Innovations for Poverty Action

في منظمة Innovations for Poverty Action (IPA)، وهي منظمة غير ربحية مقرها الولايات المتحدة، يعمل الباحثون وفرق السياسات معًا منذ بداية تصميم التقييم، وغالبًا منذ وقت نشأة البرنامج. يعتمد نموذج منظمة IPA على شبكة واسعة من المكاتب الميدانية التي يتمتع العديد منها بعلاقات قائمة مع الوكالات الحكومية

المربع ١٢-٣: أمثلة على نماذج تعاون فرق البحث والسياسة معًا (تابع)

تم التعاقد عليها جزئيًا مع المدرسة الوطنية للإحصاء (ENSEA) وتم إجراؤها جزئيًا داخليًا بواسطة فريق مخصص لجمع البيانات. قام البنك الدولي بتمويل أنشطة الرقابة الفنية والبحث، وكذلك قاد فريق التقييم. وساهم مختبر J-PAL عن طريق الأكاديميين المنتسبين. وقد أثبت هذا النموذج فاعليته في ضمان الدقة العلمية والملاءمة العالمية، وكذلك التوافق مع أولويات صانعي السياسات. كما يتطلب إدارة دقيقة للشراكات وتنسيقًا فعالًا بين مختلف الأطراف المعنية في فريق التقييم.

على سبيل المثال، في ساحل العاج، قامت مبادرة مشتركة بين البنك الدولي ومختبر عبد اللطيف جميل لمكافحة الفقر (J-PAL) والحكومة بتقييم مشروع توظيف الشباب وتنمية المهارات. تم تشكيل فريق تقييم، يتضمن فريق بحث يتألف من قائد فريق البنك الدولي وأكاديميين دوليين وخبراء محليين وفريق سياسات يضم متخصصين من وحدة تنفيذ المشروع، والوزارة التابعة، وموظفي البنك الدولي. حدد فريق التقييم المجالات ذات الأولوية لتقييم الأثر. ووضعت تجربة عشوائية مضبوطة سابقة. قامت الحكومة بصياغة الأسئلة الرئيسية ومولت جمع البيانات، والتي

المصادر: Bertrand and others 2016; IPA 2014; Sturdy, Aquino, and Molyneux 2014.

اختيار شريك فريق البحث

يتعين على صانعي السياسات ومنفذي البرامج أن يقرروا مع من سيشاركون في التقييم. وتتمثل المسائل الرئيسية في ما إذا كان فريق البحث، أو أجزاء منه، يمكن أن يكون فريقًا محليًا، وما نوع المساعدة الخارجية التي ستكون مطلوبة. تختلف القوة البحثية اختلافًا كبيرًا من بلد إلى آخر. غالبًا ما يتم التعاقد مع المؤسسات الدولية عند الحاجة إلى مهارات معينة، ويمكن أيضًا الشراكة مع المؤسسات المحلية. تقوم المؤسسات المحلية بوظائف جمع البيانات بوجه عام، نظرًا لمعرفتها العميقة بالسياق المحلي والبيئة المحلية. كما يوجد دافع عالمي قوي لضمان المشاركة الكاملة للباحثين المحليين في تقييم الأثر.

كلما زادت قدرة التقييم، أصبح من الشائع أن تقوم الحكومات والشركات الخاصة والمؤسسات متعددة الأطراف بتنفيذ تقييمات الأثر بالشراكة مع فرق البحث المحلية. يمكن لإشراك الباحثين المحليين أن يضيف قيمة كبيرة إلى تقييم الأثر، بالنظر إلى معرفتهم بالسياق المحلي. في بعض البلدان، لا يمنح التفويض بالبحث سوى للفرق التي تضم باحثين محليين. وبشكل عام، الأمر متروك لمدير التقييم لتقدير القوة المحلية وتحديد من سيكون مسؤولًا عن جوانب جهود التقييم. يمكن لشبكات تقييم الأثر الدولية التي تضم أكاديميين (مثل J-PAL أو IPA)، أو مجموعات البحث الخاصة، أو مجموعات تقييم الأثر في المؤسسات الدولية (مثل DIME و SIEF في البنك الدولي، أو SPD أو RES في بنك التنمية للبلدان الأمريكية) أن تساعد فرق السياسة على الاتصال بالباحثين الدوليين ذوي الخبرة الفنية للتعاون في تقييم الأثر.^١

هناك سؤال آخر وهو "هل يجب العمل مع وكالة خاصة أم عامة؟". يمكن أن تكون المجموعات أو المؤسسات البحثية الخاصة أكثر موثوقية في تقديم النتائج في الوقت المناسب، ولكن غالبًا ما تكون المؤسسات الخاصة أقل قابلية بشكل مفهوم لإدراج عناصر في التقييم، ما يجعل الجهد أكثر تكلفة بمجرد توقيع العقد. يمكن لفريق البحث أيضًا الاعتماد على المؤسسات البحثية والجامعات، حيث يمكن أن تضمن سمعتهم وخبراتهم الفنية أن تقبل الأطراف المعنوية نتائج التقييم على نطاق واسع. ومع ذلك، تفتقر هذه المؤسسات أحيانًا إلى الخبرة التشغيلية أو القوة على أداء بعض جوانب التقييم، مثل جمع البيانات. قد تحتاج هذه الجوانب إلى التعاقد داخليًا مع شريك آخر. قد يكون بناء القدرات في القطاع العام غايةً أيضًا ويمكن إدراجها في الشروط المرجعية لتقييم الأثر. أيًا كان الجمع بين النظراء الذي يتم إعداده في النهاية، فإن المراجعة السليمة لأنشطة التقييم السابقة للمتعاونين المحتملين ضرورية لاتخاذ خيار مستنير.

يجب تقييم قدرة وتوافر فريق بحث داخلي للقيام بأنشطة تقييم الأثر في ضوء الأنشطة الأخرى المسؤولين عنها، خاصة عند العمل مع وكالة عامة ذات مسؤوليات متعددة. يعد الوعي بعبء العمل مهمًا لتقييم كيفية تأثيره على جودة التقييم الذي يتم إجراؤه وكذلك على تكلفة فرصة التقييم فيما يتعلق بالجهود الأخرى التي تكون الوكالة الحكومية مسؤولة عنها.

كيفية تحديد وقت التقييم

ناقش الجزء ١ مزايا التقييمات السابقة المصممة أثناء إعداد البرنامج. يسمح التخطيط المسبق باختيارات أكثر في إنشاء مجموعات المقارنة، ويسهل جمع البيانات الأساسية ويساعد الأطراف المعنية على التوصل إلى توافق في الآراء حول أهداف البرنامج والمسائل محل الاهتمام.

على الرغم من أهمية التخطيط لعمليات التقييم مبكرًا في مرحلة تصميم المشروع، يجب أن يتم توقيت تنفيذها في "النقطة المثالية" المذكورة سابقًا بمجرد أن يكتمل البرنامج بما يكفي ليكون مستقرًا، ولكن قبل أن يتم توسيعه. وغالبًا ما تكون المشاريع التجريبية أو الإصلاحات الناشئة عرضة للمراجعة، سواء من حيث محتواها أو فيما يتعلق بكيفية تنفيذها وتوقيتها ومكانها والقائمين عليها. قد يحتاج مقدمو البرامج إلى وقت للتعليم وتطبيق قواعد تشغيلية جديدة باستمرار. ومن المهم تطبيق التقييمات على البرامج بعد أن تكون مستقرة وذلك لأن التقييمات تتطلب قواعد واضحة لتشغيل البرنامج لتوليد الوقائع المضادة المناسبة.

هناك مسألة رئيسية أخرى تتعلق بمدى الوقت المطلوب قبل قياس النتائج. ويتحقق التوازن الصحيح طبقًا للسياق المحدد: "إذا قمنا بالتقييم مبكرًا جدًا، فهناك خطر اكتشاف أثر جزئي فقط أو عدم وجود أثر على الإطلاق؛ وإذا كان التقييم متأخرًا جدًا، فهناك خطر في أن يفقد البرنامج دعم المانحين والجمهور أو ربما تم توسيع نطاق برنامج سيء التصميم" (King and Behrman 2009, 56).^٢ هناك مجموعة من العوامل التي يجب ترجيحها لتحديد موعد جمع بيانات المتابعة:

دورة البرنامج، بما في ذلك مدة البرنامج ووقت التنفيذ والتأخيرات المحتملة. يجب ربط تقييم الأثر بدورة تنفيذ البرنامج؛ لا يمكن أن يقود التقييم البرنامج قيد التقييم. تخضع التقييمات بحكم طبيعتها للإطار الزمني للبرنامج؛ حيث يجب أن تتماشى مع المدة المتوقعة للبرنامج. كما يجب أن تتكيف مع تأخيرات التنفيذ المحتملة عندما تكون البرامج بطيئة في تخصيص الفوائد أو عندما تتأخر بسبب عوامل خارجية.^٣ بشكل عام، على الرغم من ضرورة إدراج توقيت التقييم في المشروع منذ البداية، إلا أنه يتحلى المقيّمون بالمرونة وإجراء تعديلات أثناء تنفيذ المشروع. بالإضافة إلى ذلك، ينبغي وضع بند مخصص لتتبع التدخلات، باستخدام نظام متابعة قوي بحيث تكون جهود التقييم على علم بالوتيرة الفعلية للتدخل.

الوقت المتوقع الذي يحتاجه البرنامج للتأثير على الحاصل، وكذلك طبيعة الحاصل محل الاهتمام. عند حساب توقيت جمع بيانات المتابعة، يجب اعتبار مقدار الوقت المطلوب لظهور النتائج بعد تنفيذ البرنامج. تساعد سلسلة نتائج البرنامج في تحديد مؤشرات الحاصل والوقت المناسب لقياسها. تهدف بعض البرامج (مثل برامج دعم الدخل) إلى تقديم منافع قصيرة الأجل، بينما تهدف برامج أخرى (مثل برامج التعليم الأساسي) إلى تحقيق مكاسب طويلة الأجل. علاوة على ذلك، يستغرق ظهور بعض النتائج بطبيعتها (مثل التغيرات في متوسط العمر المتوقع أو الخصوبة من الإصلاح الصحي) وقتًا أكثر من غيرها (مثل الأرباح من برنامج التدريب).

على سبيل المثال، أثناء تقييم "صندوق الاستثمار الاجتماعي البوليفي"، الذي اعتمد على البيانات الأساسية التي تم جمعها في عام ١٩٩٣، لم يتم جمع بيانات المتابعة حتى عام ١٩٩٨ استنادًا إلى الوقت اللازم لتنفيذ التدخلات (مشاريع المياه والصرف الصحي، والعيادات الصحية، والمدارس) ولتظهر التأثيرات على صحة المجتمع الإحصائي المستفيد وتعليمه (Newman and others 2002). كانت هناك حاجة إلى فترة زمنية مماثلة لتقييم "مشروع التعليم الابتدائي في باكستان" الذي استخدم تصميمًا تجريبيًا يقوم بمسوحات أساسية ومتابعة لتقدير أثر المدارس المجتمعية على حائل الطلاب، بما في ذلك التحصيل الأكاديمي (King, Orazem, and Paterno 2008). ومع ذلك، غالبًا ما يتم جمع بيانات المتابعة في وقت أبكر مما هو موصى به، نظرًا لضغوط الحصول على النتائج في الوقت المناسب أو قيود الميزانية ودورة المشروع (McEwan 2014).

وبالتالي، فإن توقيت جمع بيانات المتابعة يعتمد على البرنامج قيد الدراسة، وكذلك على مؤشرات الحائل محل الاهتمام.

يمكن جمع بيانات المتابعة أكثر من مرة، بحيث يمكن النظر في النتائج قصيرة المدى ومتوسطة المدى ومقارنتها بينما لا تزال مجموعة المعالجة تحت وطأة التدخل. قد لا تقف بيانات المتابعة على الأثر الكامل للبرنامج إذا تم قياس المؤشرات في وقت مبكر جدًا. ومع ذلك، من المفيد جدًا توثيق الآثار قصيرة المدى التي يمكن أن توفر معلومات حول الحائل المتوقعة على المدى الطويل، وذلك لاستخلاص نتائج مبكرة لتقييم الأثر والتي يمكن أن تنشط الحوار بين فرق البحث وفرق السياسة وتحافظ على الاتصال مع عينة التقييم لتقليل تناقص العينة بمرور الوقت.

غالبًا ما تنتج مسوحات المتابعة، التي تقيس النتائج طويلة الأجل بعد تنفيذ البرنامج، أكثر الأدلة إقناعًا فيما يتعلق بفعالية البرنامج. على سبيل المثال، كانت النتائج الإيجابية من تقييمات الأثر طويلة الأجل لبرامج الطفولة المبكرة في الولايات المتحدة (Currie 2001; Currie and Thomas 1995, 2000) وجامايكا (Grantham-McGregor and others 1994; Gertler and others 2014) مؤثرة في إثبات أهمية الاستثمار في تدخلات الطفولة المبكرة. أحيانًا تكون الآثار طويلة المدى واضحة للبرنامج، ولكن حتى تصميم تقييم الأثر القوي قد لا يصمد أمام اختبار الزمن. على سبيل المثال، قد تبدأ الوحدات في المجموعة الضابطة في الاستفادة من التأثيرات غير المباشرة من مستخدمي البرنامج.

يمكن للفرق جمع بيانات المتابعة أكثر من مرة، بحيث يمكن النظر في النتائج قصيرة ومتوسطة وطويلة المدى ومقارنتها.

دورات وضع السياسات. عند تحديد توقيت التقييم يجب أيضًا اعتبار الوقت الذي نحتاج فيه إلى معلومات معينة لاتخاذ قرارات متعلقة بالسياسة ويجب أن يزامن التقييم وأنشطة جمع البيانات مع نقاط اتخاذ القرارات الرئيسية. يجب أن يتم توقيت إنتاج النتائج لإعداد الميزانيات أو توسيع البرنامج أو غيرها من قرارات السياسة.

تشكل الموازنة إحدى الخطوات الأخيرة لتفعيل تصميم التقييم. نستعرض، في هذا القسم، بعض بيانات تكلفة تقييم الأثر الحالية، ونناقش كيفية إعداد ميزانية لتقييم ما، ونقترح بعض الخيارات للتمويل.

مراجعة بيانات التكلفة

يوفر الجدولان ١-١٢ و ٢-١٢ أسس مقارنة مفيدة بشأن التكاليف المرتبطة بإجراء تقييمات الأثر الصارمة. وهما يحتويان على بيانات تكلفة تقييم الأثر لعدد من المشاريع التي يدعمها صندوق تقييم الأثر الإستراتيجي (SIEF) الذي يديره البنك الدولي. تأتي العينة في الجدول ١-١٢ من مراجعة شاملة للبرامج التي تدعمها مجموعات أبحاث تنمية الطفولة المبكرة والتعليم داخل SIEF. تم اختيار العينة في الجدول ٢-١٢ على أساس توافر إحصاءات الميزانية الحالية من مجموعة تقييمات الأثر التي يمولها SIEF.^٤

المفهوم الأساسي

تشكل تقييمات الأثر
بشكل عام نسبة صغيرة
فقط من ميزانيات البرامج
الإجمالية. بالإضافة إلى
ذلك، يجب مقارنة تكلفة
إجراء تقييم الأثر بتكاليف
عدم إجراء تقييم صحيح،
وبالتالي من المحتمل
تشغيل برنامج غير فعال.

تتراوح التكاليف المباشرة لأنشطة التقييم التي تمت مراجعتها في العينات المدرجة في الجدولين ١-١٢ و ٢-١٢ بين ١٣٠٠٠ دولار أمريكي و ٢,٧٨ مليون دولار أمريكي، بمتوسط تكلفة يبلغ حوالي مليون دولار أمريكي. على الرغم من أن هذه التكاليف تتباين بشكل كبير وقد تبدو مرتفعة من حيث القيمة المطلقة، إلا أن تقييمات الأثر لا تشكل عمومًا سوى نسبة صغيرة من ميزانيات البرامج الإجمالية. بالإضافة إلى ذلك، يجب مقارنة تكلفة إجراء تقييم الأثر بتكاليف عدم إجراء تقييم صحيح، وبالتالي من المحتمل تشغيل برنامج غير فعال. تسمح التقييمات للباحثين وصانعي السياسات بتحديد البرامج أو ميزات البرنامج التي تعمل والتي لا تعمل، والإستراتيجيات التي قد تكون الأكثر فعالية ونجاعة في تحقيق غايات البرنامج. ومن هذا المنطلق، فإن الموارد اللازمة لتنفيذ تقييم الأثر تشكل استثمارًا صغيرًا نسبيًا ولكنه مهم.

يُفصّل الجدول ٢-١٢ تكاليف عينة تقييمات الأثر التي يدعمها SIEF. تشمل التكاليف الإجمالية للتقييم وقت موظفي البنك الدولي، والاستشاريين الوطنيين والدوليين، والسفر، وجمع البيانات، وأنشطة التوزيع.^٥ كما هو الحال في جميع التقييمات تقريبًا التي لا يمكن استخدام البيانات الحالية لها، فإن أعلى تكلفة في التقييم هي جمع بيانات جديدة، حيث تمثل ٦٣ في المائة من إجمالي تكلفة التقييم، في المتوسط، كما هو موضح في الجدول.

تعكس هذه الأرقام أحجام وأنواع مختلفة من التقييمات. وبوجه عام، تعد التكلفة النسبية لتقييم برنامج تجريبي أعلى من التكلفة النسبية لتقييم برنامج وطني أو شامل. بالإضافة إلى ذلك، فإن بعض التقييمات لا تتطلب سوى مسح متابعة واحد فقط أو قد تكون قادرة على استخدام مصادر البيانات الموجودة، في حين أن البعض الآخر قد يحتاج إلى القيام بجولات متعددة من جمع البيانات. تعتمد تكاليف جمع البيانات إلى حد كبير على رواتب الفريق المحلي وتكلفة الوصول إلى المجتمع الإحصائي في عينة التقييم وطول الوقت في الميدان. لمعرفة المزيد حول كيفية تقدير تكلفة المسح في سياق معين، يوصى بأن يقوم فريق التقييم أولاً بالاتصال بوكالة الإحصاء الوطنية والبحث عن المعلومات من الفرق التي قامت بعمل المسح في الدولة.

الجدول ١٢-٢ التكاليف المفصلة لمجموعة مختارة من تقييومات الأثر المدعومة من البنك الدولي

الجدول ١٢-١ تكلفة تقييومات الأثر لمجموعة مختارة من المشروعات المدعومة من البنك الدولي

تقييم الأثر	الدولة	إجمالي تكلفة تقييم الأثر (دولار أمريكي)	إجمالي تكلفة البرنامج (دولار أمريكي)	تقييم الأثر في شكل نسبة مئوية من إجمالي تكلفة البرنامج
مشروع شبكة الأمان	بوركينافاسو	٧٥٠,٠٠٠	٣٨,٨٠٠,٠٠٠	١,٩
توظيف وتنمية مهارات المهاجرين	الصين	٢٢٠,٠٠٠	٥٠,٠٠٠,٠٠٠	٠,٤
مشروع شبكة الأمان الاجتماعي	كولومبيا	١٣٠,٠٠٠	٨٦,٤٠٠,٠٠٠	٠,٢
التغذية المتكاملة/شبكة الأمان الاجتماعي في مجال العمل (تجريبي)	جيبوتي	٤٨٠,٠٠٠	٥,٠٠٠,٠٠٠	٨,٨
برنامج استثمار القطاعات الاجتماعية	جمهورية الدومينيكان	٦٠٠,٠٠٠	١٩,٤٠٠,٠٠٠	٣,١
الحوافز القائمة على الأداء للمعلمين	غينيا	٢,٠٥٥,٠٠٠	٣٩,٦٧٠,٠٠٠	٤,٩
الحماية الاجتماعية	جاميكا	٨٠٠,٠٠٠	٤٠,٠٠٠,٠٠٠	٢,٠
معالجة سوء التغذية المزمن	مدغشقر	٦٥١,٠٠٠	١٠,٠٠٠,٠٠٠	٦,١
مراكز رعاية الأطفال المجتمعية (تجريبي)	ملاوي	٩٥٥,٠٠٠	١,٥٠٠,٠٠٠	٣٨,٩
المعلومات والتحويلات النقدية غير المشروطة	نيبال	٩٨٤,٠٠٠	٤٠,٠٠٠,٠٠٠	٢,٤
المساعدة الفنية لشبكة الأمان الاجتماعي	باكستان	٢,٠٠٠,٠٠٠	٦٠,٠٠٠,٠٠٠	٣,٣
مشروع الحماية الاجتماعية	بنما	١,٠٠٠,٠٠٠	٢٤,٠٠٠,٠٠٠	٤,٢
معايير معيشة المجتمع الأول	رواندا	١,٠٠٠,٠٠٠	١١,٠٠٠,٠٠٠	٩,١
المعلومات من أجل المساواة والتدخلات الحافزة للمعلمين	تنزانيا	٧١٢,٠٠٠	٤١٦,٠٠٠,٠٠٠	٠,٢
التدخلات في حجم الفصل وجودة المعلم	أوغندا	٦٣٩,٠٠٠	١٠٠,٠٠٠,٠٠٠	٠,٦
الصندوق الاجتماعي للتنمية	جمهورية اليمن	٢,٠٠٠,٠٠٠	١٥,٠٠٠,٠٠٠	١٣,٣
المتوسط		٩٣٦,٠٠٠	٥٩,٧٩٨,٠٠٠	٦,٢

المصدر: المصدر: عينة من تقييومات الأثر المدعومة من قبل مجموعات أبحاث تنمية الطفولة المبكرة والتعليم لصندوق تقييم الأثر الإستراتيجي للبنك الدولي.

ملاحظة: IE = تقييم الأثر

أ. التكلفة الإجمالية للبرنامج لا تشمل التكاليف المرتبطة بتقييم الأثر.

الجدول ٢-١٢ التكاليف المفصلة لمجموعة مختارة من تقييمات الأثر المدعومة من البنك الدولي

تقييم الأثر	الدولة	إجمالي التكلفة (دولار أمريكي)	حجم العينة	جمع البيانات (نسبة مئوية)	العاملون والاستشاريون (نسبة مئوية)	السفر (نسبة مئوية)	النشر وورش العمل (نسبة مئوية)	أخرى (نسبة مئوية)
بناء القدرات الأبوية للمساعدة في تغذية الطفل وصحته	بنجلاديش	٦٥٥,٠٠٠	٢,٠٧٤ أسرة	٢٧	٤٨	٥	٠	٢٠
سد فجوة التعلم المبكر لأطفال الغبر	بلغاريا	٧٠٢,٠٠٠	٦,٠٠٠ أسرة	٧٤	٢١	٤	١	٠
عنصر تنمية الطفولة المبكرة والتغذية في مشروع شبكة الأمان في بوركينا فاسو	بوركينا فاسو	٧٥٠,٠٠٠	٤,٧٢٥ أسرة ٢,٩٧٨ مدرسة	٥٥	٢٠	٣	١	٢١
الدفع لمعلمي المجتمع	تشاد	١,٦٨٠,٠٠٠	مدرسة	٥٢	١٤	١٢	١٨	٤
تدخل إنمائي منزلي للطفولة المبكرة	كولومبيا	٥٧٣,٠٠٠	١,٤٢٩ فردًا	٥٤	٣٦	٢	٢	٧
تجربة شبكة أمان اجتماعي للتغذية / العمل المتكامل	جيبوتي	٤٨٠,٠٠٠	١,١٥٠ فردًا	٧٥	٠	٠	٦	١٨
الإشراف والحوافز لزيادة التعلم: برنامج عالي الأداء TCAI	غانا	٤٩٨,٠٠٠	٤٨٠ مدرسة	٥١	٤٦	٣	٠	٠
الحوافز القائمة على الأداء للمعلمين	غينيا	٢,٠٥٥,٠٠٠	٤٢٠ مدرسة	٨٢	٩	٣	١	٤

(تابع)

تابع ٢-١٢ (تابع)

تقييم الأثر	الدولة	إجمالي التكلفة (دولار أمريكي)	حجم العينة	جمع البيانات (نسبة مئوية)	العاملون والاستشاريون (نسبة مئوية)	السفر (نسبة مئوية)	النشر وورش العمل (نسبة مئوية)	أخرى (نسبة مئوية)
دعم تقديم خدمات التعليم	هايتي	٤٣٦,٠٠٠	٢٠٠ مدرسة	٤٠	٣١	١٧	٣	٩
التحفيز غير المالي الخارجي والداخلي للمعلم	الهند	٤٤٨,٠٠٠	٣٦٠ مدرسة	٨٣	٥	١١	١	٠
تحفيز الطفولة المبكرة والتواصل الاجتماعي في إستراتيجية تنمية الطفل المتكاملة في الهند	الهند	٦٩٦,٠٠٠	٢,٢٥٠ فردًا	٤٩	٤٣	٥	٣	٠
مجموعات المساعدة الذاتية النسائية لتعزيز الصحة والتغذية والصرف الصحي والأمن الغذائي	الهند	٨٤٤,٠٠٠	٣,٠٠٠ أسرة	٥٢	٣٩	٥	١	٢
تنمية الطفولة المبكرة للفقراء	الهند	١,٧١٨,٠٠٠	٢,٥٨٨ أسرة	٤٦	٥٣	١	١	٠
التغذية في مرحلة الطفولة المبكرة، وتوافر مقدمي الخدمات الصحية، وحصيلة الحياة كشباب بالغين	إندونيسيا	٢,٤٩٠,٠٠٠	٦,٧٤٣ فردًا	٩٤	٠	٢	٤	٠
معالجة سوء التغذية المزمن	مدغشقر	٦٥١,٠٠٠	٥,٠٠٠ فردًا	٠	٠	٦٦	٢	٣٢
التربية المتكاملة والتغذية والوقاية من الملاريا	مالي	٩٤٩,٠٠٠	٣,٦٠٠ فردًا	٥٨	٢٢	٤	٥	١١

(تابع)

الجدول ١٢-٢ (تابع)

تقييم الأثر	الدولة	إجمالي التكلفة (دولار أمريكي)	حجم العينة	جمع البيانات (نسبة مئوية)	العاملون والاستشاريون (نسبة مئوية)	السفر (نسبة مئوية)	التشاور والعمل (نسبة مئوية)	أخرى (نسبة مئوية)
زيادة المساءلة عن التعليم من خلال المساعدين التربويين المجتمعيين	المكسيك	٢٦٨,٠٠٠	٢٣٠ مدرسة	٧٠	٢٦	٣	٢	٠
الوصول إلى نموذج التعليم الشامل الخاص	المكسيك	٤٢٠,٠٠٠	١٧٢ فردًا	٤٥	٤٨	٥	١	١
تقييم الأثر العشوائي لمختلف تدخلات مهارات القراءة والكتابة المبكرة	موزمبيق	١,٧٦٢,٠٠٠	١١٠ مدرسات	٧٨	٥	٤	٨	٦
تنمية الطفولة المبكرة المتكاملة والتغذية	موزمبيق	١,٩٠٨,٠٠٠	٦,٧٠٠ أسرة	٧٤	٨	٥	٧	٧
برنامج تجريبي للتأمين الصحي	نيبال	٤٨٥,٠٠٠	٦,٣٠٠ أسرة	٦١	٣٣	٣	٤	٠
المعلومات والتحويلات النقدية غير المشروطة حول حصاد التغذية	نيبال	٩٨٤,٠٠٠	٣,٠٠٠ فردًا	٥٧	٢٣	٩	١	١٠
التحويلات النقدية، تدريب الآباء، والتنمية الشاملة للطفولة المبكرة	النيجر	٩٨٤,٠٠٠	٤,٣٣٢ أسرة	٦٧	١٨	٧	١	٧
فهم ديناميكيات المعلومات من أجل المساءلة	نيجيريا	١,٠٥٢,٠٠٠	١٢٠ مدرسة	٥٩	٢٥	٨	٣	٦

(تابع)

الجدول ٢-١٢ (تابع)

أخرى (نسبة مئوية)	النشر وورش العمل (نسبة مئوية)	السفر (نسبة مئوية)	العاملون والاستشاريون (نسبة مئوية)	جمع البيانات (نسبة مئوية)	حجم العينة	إجمالي التكلفة (دولار أمريكي)	الدولة	تقييم الأثر
٢	٤	٦	١٣	٧٦	٥,٠٠٠ أسرة	٢,٧٧٥,٠٠٠	نيجيريا	برنامج إعادة الاستثمار والتمكين للدعم ومبادرة صحة الأم والطفل
١٨	٣	٦	١٥	٥٩	٢٨٧ مدرسة	٨٤٥,٠٠٠	باكستان	المشاركة المجتمعية للجنة المدرسة
٤٢	٢	٥	٢٥	٢٦	٢,٠٠٠ مدرسة	٢,١٢٤,٠٠٠	باكستان	تعزيز المدارس الخاصة لفقراء الريف
١١	١	٣	٧	٧٩	٣٠٠ مدرسة	٧٩٧,٠٠٠	رواندا	الاختيار وآثار عقود الأداء التحفيزية لمعلمي المدارس الابتدائية
٤	٣	٢٠	٢٤	٦٧	٢٠٠ مدرسة	٦٤٧,٠٠٠	جنوب إفريقيا	الحملة الإعلامية في المدارس الابتدائية
٠	٢	٧	٦	٨٦	٤٢٠ مدرسة	٧١٢,٠٠٠	تنزانيا	اختيار المعلومات من أجل المساواة والتدخلات التحفيزية للمعلمين
٠	٢	٢	١١	٨٥	٤٢٠ مدرسة	٨٨٩,٠٠٠	تنزانيا	تصميم برامج حوافز فعالة للمعلمين
٠	١	٢	٧	٩٠	٣,٦٠٠ مدرسة	١,٢٤٢,٠٠٠	تنزانيا	برنامج للنساء المعرضات لخطر الإصابة بفيروس نقص المناعة البشرية

(تابع)

الجدول ٢-١٢ (تابع)

تقييم الأثر	الدولة	إجمالي التكلفة (دولار أمريكي)	حجم العينة	جمع البيانات (نسبة مئوية)	العاملون والاستشاريون (نسبة مئوية)	السفر (نسبة مئوية)	النشر وورش العمل (نسبة مئوية)	أخرى (نسبة مئوية)
التدخلات في حجم الفصل وجودة المعلم	أوغندا	٦٣٩,٠٠٠	٢٠٠ مدرسة	٨٢	٩	٧	٢	٠
التجاعة المتباينة في تقديم خدمات التعليم في القطاعين العام والخاص	أوغندا	٧٣٧,٠٠٠	٢٨٠ مدرسة	٧٧	١٨	٣	٣	٠
المتوسط		١,٠٢٦,٠٠٠		٦٣	٢١	٧	٣	٧

المصدر: عينة من تقييمات الأثر الممولة من صندوق تقييم الأثر الإستراتيجي للبنك الدولي.
 أ. لا تعكس التكاليف المقدرة دائماً التكاليف الكاملة للتقييم، بما في ذلك وقت فريق السياسات.
 ب. النسبة المئوية للتكلفة الإجمالية للتقييم حسب الفئة. لا تشمل هذه التكلفة تكاليف موظفي المشروع المحليين، الذين غالباً ما شاركوا بدرجة كبيرة في تصميم التقييم والإشراف عليه، حيث لا يتم تسجيل بيانات دقيقة عن هذه التكاليف بشكل منتظم.

وضع ميزانية تقييم الأثر

هناك حاجة إلى العديد من الموارد لتنفيذ تقييم دقيق للأثر، لا سيما عند جمع البيانات الأولية. تتضمن بنود الميزانية أتعاب فريق العمل الذي يتضمن على الأقل محققًا/باحثًا رئيسيًا، ومنسقًا ميدانيًا، وخبير أخذ العينات، وفريق جمع البيانات. كما أن الوقت الذي يستغرقه طاقم المشروع في فريق السياسات ضروري أيضًا لتقديم التوجيه والدعم طوال عملية التقييم. قد تتكون هذه الموارد البشرية من باحثين وخبراء فنيين من منظمات دولية ومستشارين دوليين أو محليين وموظفي البرنامج المحليين. يجب أيضًا أن تُدرج تكاليف السفر والإقامة في الميزانية. ويجب اعتبار موارد النشر، غالبًا في شكل ورش عمل وتقارير وأوراق أكاديمية، عند التخطيط للتقييم.

كما لوحظ، عادةً ما تكون أكبر التكاليف في التقييم هي تكاليف جمع البيانات (بما في ذلك إنشاء المسح واختباره التجريبي)، ومواد ومعدات جمع البيانات، وتدريب مندوبي الإحصاء، والأجور اليومية لمندوبي الإحصاء، والمركبات والوقود، وعمليات إدخال البيانات. يتطلب حساب تكاليف جميع هذه المدخلات وضع بعض الافتراضات حول، على سبيل المثال، المدة التي سيستغرقها إكمال الاستبيان وأوقات السفر بين المواقع.

قد تتوزع تكاليف تقييم الأثر على مدار عدة سنوات. يوضح نموذج الميزانية في الجدول ١٢-٣ كيف يمكن تصنيف النفقات في كل مرحلة من مراحل التقييم حسب السنة لأغراض المحاسبة وإعداد التقارير. مرة أخرى، من المرجح أن تكون طلبات الميزانية أعلى خلال السنوات التي يتم فيها جمع البيانات.

خيارات تمويل التقييمات

يمكن أن يأتي تمويل التقييم من عدة مصادر، بما في ذلك موارد المشروع أو ميزانيات البرامج المباشرة أو المنح البحثية أو تمويل المانحين. تبحث فرق التقييم، في كثير من الأحيان، عن مجموعة من المصادر لتوليد الأموال اللازمة. على الرغم من أن تمويل التقييمات كان يأتي في المقام الأول من ميزانيات الأبحاث، إلا أن التركيز المتزايد على وضع السياسات القائمة على الأدلة أدى إلى زيادة التمويل من مصادر أخرى. في الحالات التي من المرجح أن يعالج فيها التقييم فجوة معرفية كبيرة تهم مجتمع التنمية على نطاق واسع، وحيث يمكن تطبيق تقييم موثوق وقوي، ينبغي تشجيع صانعي السياسات على البحث عن تمويل خارجي، بالنظر إلى الصالح العام الذي ستوفره نتائج التقييم. تتضمن مصادر التمويل الحكومة وبنوك التنمية والمنظمات متعددة الأطراف ووكالات الأمم المتحدة والمؤسسات وفاعلي الخير ومنظمات البحث والتقييم؛ مثل المبادرة الدولية لتقييم الأثر.

الجدول ٣-١٢ نموذج ميزانية تقييم الأثر

مرحلة التصميم		مرحلة بيانات خط الأساس				الوحدة	أ. رواتب طاقم العمل
التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)	التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)		
١٥,٠٠٠	٢	٧,٥٠٠	١٥,٠٠٠	٢	٧,٥٠٠	أسبوع	ب. أتعاب الاستشاري
٤١,٩٠٠			١٤,٢٥٠				المستشار الدولي (١)
٠	٠	٤٥٠	٦,٧٥٠	١٥	٤٥٠	يوم	المستشار الدولي (٢)
٣,٥٠٠	٠	٣٥٠	٣,٥٠٠	١٠	٣٥٠	يوم	مساعد باحث/منسق ميداني
٣٦,٤٠٠	٣٠	٢٨٠	٠	٠	٢٨٠	يوم	خبير إحصائي
٢,٠٠٠	٥	٤٠٠	٤,٠٠٠	١٠	٤٠٠	يوم	ج. السفر والإقامة
٣,٣٥٠	١	٣٣٥٠	٣,٣٥٠	١	٣٣٥٠	رحلة	طاقم العمل: السفر الجوي الدولي
٧٥٠	٥	١٥٠	٧٥٠	٥	١٥٠	يوم	طاقم العمل: الفندق والبدلات اليومية
٥٠	٥	١٠	٥٠	٥	١٠	يوم	طاقم العمل: النقل البري المحلي
٧,٠٠٠	٢	٣٥٠٠	٧,٠٠٠	٢	٣٥٠٠	رحلة	المستشارون الدوليون: السفر الجوي الدولي
٣,٠٠٠	٢٠	١٥٠	٣,٠٠٠	٥	١٥٠	يوم	المستشارون الدوليون: الفندق والبدلات اليومية
٥٠	٥	١٠	٥٠	٠	١٠	يوم	المستشارون الدوليون: النقل البري المحلي
١,٣٥٠	١	١٣٥٠	٠			رحلة	المنسق الميداني: السفر الجوي الدولي

(تابع)

(تابع)

التكاليف الإجمالية للمرحلة

مرحلة التصميم	مرحلة الأساسية	٤٣٤٠	مرحلة بيانات خط الأساس				٠٠١٧٦١
			الوحدة	التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)	
ورشة (ورش) العمل	طفل	٣٠٠٠	يوم	٠	٣	٠٠١	٠٠٠٧٨
	طفل	٣٠٠٠	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
نوع البيانات ٣: الحصائل الصحية	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
نوع البيانات ٢: الحصائل التعليمية	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
نوع البيانات ١: الموافقة	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
د. جمع البيانات	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
المنسق الميداني: النقل البري المحلي	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
اليومية	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	
المنسق الميداني: الفندق والبدلات	طفل	٣١	يوم	٠	٣	٠٠١	٠٠٠١٣
	طفل	٣١	يوم	٠	٣	٠٠١	
	مدرسة	٠٠١	يوم	٠	٣	٠٠١	

الجدول ٣-١٢ (تابع)



الجدول ٣-١٢ (تابع)

مرحلة التصميم			مرحلة بيانات خط الأساس		
التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)	التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)
١٥,٠٠٠	٢٢	٧,٠٠٠	١٥,٠٠٠	٢٢٢٢٢	٧,٠٠٠
٣٨,٠٠٠			٤٣,٧٥٠		
٤,٥٠٠	١٠	٤٥٠	٦,٧٥٠	١٥	٤٥٠
٣,٥٠٠	١٠	٣٥٠	٧,٠٠٠	٢٠	٣٥٠
٢٨,٠٠٠	١٠٠	٢٨٠	٢٨,٠٠٠	١٠٠	٢٨٠
٢,٠٠٠	٥	٤٠٠	٢,٠٠٠	٥	٤٠٠
٦,٧٠٠	٢	٣٣٥٠	٣,٣٥٠	١	٣٣٥٠
١,٥٠٠	١٠	١٥٠	١,٥٠٠	١٠	١٥٠
٥٠	٥	١٠	٥٠	٥	١٠
٧,٠٠٠	٢	٣٥٠٠	٧,٠٠٠	٢	٣٥٠٠
٣,٠٠٠	٢٠	١٥٠	٣,٠٠٠	٢٠	١٥٠
٥٠	٥	١٠	٥٠	٥	١٠
١,٣٥٠	١	١٣٥٠	١,٣٥٠	١	١٣٥٠

(تابع)

أ. رواتب طاقم العمل
ب. أتعاب الاستشاري
المستشار الدولي (١)
المستشار الدولي (٢)
مساعد باحث/منسق ميداني
خبير إحصائي
ج. السفر والإقامة
طاقم العمل: السفر الجوي الدولي
طاقم العمل: الفندق والبدلات اليومية
طاقم العمل: النقل البري المحلي
المستشارون الدوليون: السفر الجوي الدولي
المستشارون الدوليون: الفندق والبدلات اليومية
المستشارون الدوليون: النقل البري المحلي
المنسق الميداني: السفر الجوي الدولي



الجدول ٣-١٢ (تابع)

مرحلة التصميم		مرحلة بيانات خط الأساس				إجمالي تكاليف التقييم	
التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)	الوحدة	التكلفة الإجمالية (دولار أمريكي)	عدد الوحدات	التكلفة للوحدة (دولار أمريكي)	الوحدة
٤٥٠	٣	١٥١	يوم	٤٥٠	٣	١٥٠	يوم
٣٠	٣	١٠	يوم	٣٠	٣	١٠	يوم
٢٢٦٠٠٠	١٠٠	٢٠	مدرسة	١٢٦٠٠٠	١٠٠	١٢٠	مدرسة
٤٢٠٠٠	٣٠٠٠	٤١	طفل	٤٢٠٠٠	٣٠٠٠	٤١	طفل
٧٢٠٠٠	٣٠٠٠	٢٤	طفل	٧٢٠٠٠	٣٠٠٠	٢٤	طفل
٥٥٠٠							
٤٠٠٠٠							
٥٠١							
٢٤٥١٣٠			المتابعة، المرحلة II				المتابعة، المرحلة I
٦٩٧,٧٤٠			إجمالي تكاليف التقييم				

(تابع)

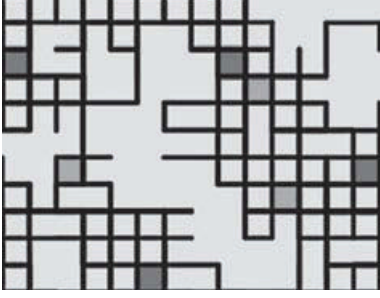
مصادر إضافية

- للحصول على المواد المرتبطة بهذا الفصل وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للوصول إلى عدة أدوات للمساعدة في تخطيط التقييم وتنفيذه، اطلع على بوابة تقييم بنك التنمية للبلدان الأمريكية (<http://www.iadb.org/evaluationhub>)، بما في ذلك ما يلي:
 - قسم التصميم: مخططات جانب للمساعدة في جدولة أنشطة تقييم الأثر، وأداة نموذج الميزانية لتقدير تكاليف تقييم الأثر، وقائمة مراجعة للأنشطة الأساسية التي يتعين تنفيذها.
 - قسم التنفيذ: عينة من الشروط المرجعية (TORs) للمحققين الرئيسيين، ومؤسسات جمع البيانات، والدعم الفني والإشراف.
- للحصول على إرشادات وأدوات للمساعدة في تخطيط التقييم وتنفيذه، راجع مجموعة أدوات تقييم الأثر لدى البنك الدولي (Vermeersch, Rothenbühler, and Sturdy 2012)، بما في ذلك ما يلي:
 - النموذج ٢: بناء الفريق: عينة من الشروط المرجعية للمحققين الرئيسيين ومنسقي التقييم ومحللي البيانات والباحثين المحليين وخبراء حساب القوة الإحصائية وخبراء جودة البيانات والعاملين الميدانيين وغيرهم.
 - كتيبات ميدانية وبرامج تدريبية للأسر والمنشآت الصحية.
 - النموذج ٣: التصميم: إرشادات حول كيفية مواءمة التوقيت، وتكوين الفريق، والميزانية الخاصة بتقييم الأثر؛ ونموذج الميزانية.
 - النموذج ٤: التحضير لجمع البيانات: معلومات عن جدولة أنشطة جمع البيانات والتوصل إلى اتفاقيات مع الأطراف المعنية بشأن ملكية البيانات؛ مخطط جانب؛ عينة ميزانية جمع البيانات.

ملاحظات

١. J-PAL هو مختبر عبد اللطيف جميل لمكافحة الفقر. SPD هو مكتب التخطيط الإستراتيجي وفعالية التنمية التابع لبنك التنمية للبلدان الأمريكية (IDB). RES هو قسم الأبحاث في بنك التنمية للبلدان الأمريكية.
٢. للحصول على مناقشة مفصلة لعوامل التوقيت المتعلقة بتقييم البرامج الاجتماعية، راجع (King and Behrman 2009).
٣. "هناك العديد من الأسباب التي تجعل التنفيذ ليس فوريًا ولا مثاليًا، ولماذا تختلف مدة التعرض للمعالجة ليس فقط بين مناطق البرنامج ولكن أيضًا بين المستفيدين النهائيين، ولماذا قد تؤدي فترات التعرض المتفاوتة إلى تقديرات مختلفة لأثر البرنامج (King and Behrman 2009, 56).
٤. في حين أن الجدولين ١٢-١ و ١٢-٢ يوفران أسس مقارنة مفيدة، إلا أنهما لا يمثلان جميع التقييمات التي أجراها برنامج SIEF أو البنك الدولي.
٥. في هذه الحالة، يتم احتساب التكلفة كنسبة مئوية من جزء تكلفة المشروع الذي يموله البنك الدولي.

- Bertrand, Marianne, Bruno Crépon, Alicia Marguerie, and Patrick Premand. 2016. "Impacts à Court et Moyen Terme sur les Jeunes des Travaux à Haute Intensité de Main d'oeuvre (THIMO) : Résultats de l'évaluation d'impact de la composante THIMO du Projet Emploi Jeunes et Développement des compétences (PEJEDEC) en Côte d'Ivoire." Washington, DC: Banque Mondiale et Abidjan, BCP-Emploi.
- Currie, Janet. 2001. "Early Childhood Education Programs." *Journal of Economic Perspectives* 15 (2): 213–38.
- Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85 (3): 341–64.
- . 2000. "School Quality and the Longer-Term Effects of Head Start." *Journal of Economic Resources* 35 (4): 755–74.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, and others. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187): 998–1001.
- Grantham-McGregor, Sally, Christine Powell, Susan Walker, and John Himes. 1994. "The Long-Term Follow-up of Severely Malnourished Children Who Participated in an Intervention Program." *Child Development* 65: 428–93.
- IPA (Innovations for Poverty Action). 2014. "Researcher Guidelines: Working with IPA." September 1. http://www.poverty-action.org/sites/default/files/researcher_guidelines_version_2.0.pdf.
- King, Elizabeth M., and Jere R. Behrman. 2009. "Timing and Duration of Exposure in Evaluations of Social Programs." *World Bank Research Observer* 24 (1): 55–82.
- King, Elizabeth M., Peter F. Orazem, and Elizabeth M. Paterno. 2008. "Promotion with and without Learning: Effects on Student Enrollment and Dropout Behavior." Policy Research Working Paper 4722, World Bank, Washington, DC.
- McEwan, Patrick J. 2014. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research*. doi:10.3102/0034654314553127.
- Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16 (2): 241–74.
- Sturdy, Jennifer, Sixto Aquino, and Jack Molyneaux. 2014. "Learning from Evaluation at the Millennium Challenge Corporation." *Journal of Development Effectiveness* 6 (4): 436–50.
- Vermeersch, Christel, Elisa Rothenbühler, and Jennifer Sturdy. 2012. *Impact Evaluation Toolkit: Measuring the Impact of Results-based Financing on Maternal and Child Health*. World Bank, Washington, DC. <http://www.worldbank.org/health/impactevaluationtoolkit>.



أخلاقيات وعلم تقييم الأثر

إدارة تقييمات أخلاقية ذات مصداقية

تركز أخلاقيات التقييم على حماية الأفراد أو الأشخاص الذين يشاركون في التقييم، بينما تساعد شفافية الأساليب في ضمان أن تكون نتائج التقييم غير متحيزة وموثوقة وذات مصداقية وتساهم في إنشاء معرفة أوسع نطاقًا. يتمتع صانعو السياسات والباحثون بمسؤولية ومصلحة مشتركة لضمان أن التقييم أخلاقي وأن نتائجه غير متحيزة وموثوقة وذات مصداقية. ويمكن أن يؤدي الإخفاق في تحقيق ذلك إلى إبطال التقييم ويتسبب في مشاكل خارج نطاق التقييم. تخيل تقييم الأثر الذي يعرض مجموعة من الأشخاص للخطر من خلال الإفصاح عن بيانات شخصية، أو تقييم يستخدم آلية تعيين برنامج غير عادلة حيث يستبعد العائلات الأكثر احتياجًا. تخيل تقييمًا يوضح أن أحد البرامج يحقق نجاحًا كبيرًا، لكنه لا يوفر أي بيانات لدعم تقييمه. يمكن أن تؤدي أي من هذه الحالات إلى احتجاج شعبي؛ أو شكاوى في وسائل الإعلام أو في المحاكم أو في أي مكان آخر، وإجراح صانعي السياسات والباحثين على حدٍ سواء. قد يمتد النقد الموجه للتقييم إلى البرنامج نفسه بل ويقوض تنفيذه. كما تمثل موثوقية نتائج التقييم واكتمالها أمرين مهمين للغاية: فعندما تنتج التقييمات تقديرات متحيزة أو جزئية لتأثير البرامج، فإن صانعي السياسات يكونون مقيدون في قدرتهم على اتخاذ قرار مستنير بالكامل.

بينما ترتبط تقييمات الأثر بالبرامج والمشاريع العامة، فهي تعد نشاطًا بحثيًا أيضًا، وبالتالي يتم إجراؤها في مجال العلوم الاجتماعية. وبناءً عليه، يجب على فريق التقييم الالتزام بعدد من مبادئ وقواعد العلوم الاجتماعية للتأكد من أن التقييم أخلاقي وتتسم طرقه ونتائج بالشفافية.

أخلاقيات إجراء تقييمات الأثر

عندما يخصص تقييم الأثر أشخاصًا لمجموعات المعالجة ومجموعات المقارنة ويجمع البيانات المتعلقة بهما ويحللها، يتحمل فريق التقييم مسؤولية تقليل أي مخاطر قد يتعرض لها هؤلاء الأفراد إلى أقصى حد ممكن والتأكد من أن هؤلاء الأفراد المشاركين في التقييم يشاركون بعد موافقتهم المبينة على معلومات مستنيرة.

أخلاقيات التعيين لمجموعات المعالجة والمقارنة

كما هو الحال مع قسم أبقراط في مهنة الطب، يجب أن يكون أول مبدأ لأخلاقيات التقييم هو عدم الإضرار. بحيث ينصب الاهتمام الأكبر في أنه قد يسبب تدخل البرنامج المراد تقييمه أذى للأفراد، سواء بشكل مباشر أو غير مباشر. على سبيل المثال، قد يؤدي مشروع إعادة تأهيل الطرق إلى تشريد الأسر التي تعيش على طول بعض أجزاء الطرق. أو أن مشروع محو الأمية الذي لا يأخذ في الاعتبار استخدام اللغات الأصلية قد يضر بمجتمعات السكان الأصليين. تستخدم العديد من الحكومات والجهات المانحة الدولية التي تمول مشاريع التنمية إطار عمل ضامن لمنع هذه الأنواع من المخاطر وتخفيفها. في حين تحمل منفذي البرنامج المسؤولية الأساسية لتطبيق تدابير حماية المشروع، إلا أنه يجب أن يكون فريق التقييم يقظًا للتحقق من أن المشروع يمثل لهذه الأطر المطلوبة.

هناك شاغل آخر متعلق بالضرر الذي قد ينشأ عن حجب التدخل عن المستفيدين المحتملين. المبدأ الأساسي هو أنه لا ينبغي استبعاد مجموعات من تدخل معروف أنه مفيد بغرض وحيد وهو إجراء التقييم. لا يجب إجراء التقييمات سوى في الحالات التي لا يعرف فيها فريق التقييم ما إذا كان التدخل مفيدًا أم لا ضمن الإطار المحدد الذي يتم تقييمه. بالإضافة إلى ذلك، إذا أظهر التقييم أن البرنامج فعال من حيث التكلفة، فيجب على ممولي البرنامج، سواء الحكومات أو المانحين أو المنظمات غير الحكومية، بذل جهود معقولة لتوسيع البرنامج ليشمل مجموعات المقارنة بمجرد الانتهاء من تقييم الأثر.

المفهوم الأساسي

لا ينبغي استبعاد مجموعات من تدخل معروف أنه مفيد بسبب أغراض إجراء التقييم فقط.

من المبادئ ذات الصلة التي ندافع عنها في هذا الكتاب هو أن التقييمات لا ينبغي أن تحدد كيفية تعيين البرامج؛ بل بدلاً من ذلك، يجب أن تتناسب التقييمات مع قواعد تعيين البرامج إلى الحد الذي تكون فيه تلك القواعد واضحة وعادلة. يمكن أن يساعد التقييم أيضًا على إعادة تعريف القواعد إن كانت غائبة أو غير عادلة. عند اتباع هذه العملية سيساعد ذلك على ضمان عدم نشوء مخاوف أخلاقية كبيرة من تقييم الأثر بحد ذاته بل من أخلاقيات القواعد المستخدمة لاختيار المستفيدين من البرنامج. ومع ذلك، من شأن تعيين الوحدات إلى مجموعتي المعالجة

والمقارنة أن يفاقم من المخاوف حول أخلاقيات حرمان المستفيدين المؤهلين من مزايا البرنامج. وينطبق ذلك بشكل خاص عند إجراء تعيين عشوائي لمزايا البرنامج. سبق أن أكدنا في الجزء الثاني وفي الفصل ١١ على أن التعيين العشوائي عبارة عن طريقة يمكن تطبيقها في سياقات تشغيلية محددة. بشكل خاص، حقيقة أن سير عمل معظم البرامج يتم وفقًا لموارد مالية وإدارية محدودة تجعل من المستحيل الوصول إلى كل المستفيدين المؤهلين دفعة واحدة. وذلك يعالج المخاوف الأخلاقية بما أن البرنامج نفسه يجب أن يضع قواعد التخصيص/التوزيع ويفرض شكلًا من أشكال توزيع الحصص في غياب تقييم الأثر. ومن وجهة نظر أخلاقية، هناك شرط حسن يتعين أن ينطبق على جميع الأشخاص المؤهلين للمشاركة في برنامج ما وهو أن يكونوا على سواء من حيث توافر الفرصة نفسها للاستفادة من البرنامج. جدير بالذكر أن التعيين العشوائي يلبي هذا المطلب. وفي السياقات التشغيلية الأخرى التي سيتم فيها تنفيذ البرنامج على مراحل بمرور الوقت، يمكن أن يستند طرح البرنامج إلى الاختيار العشوائي للترتيب الذي سيستفيد وفقًا له المستفيدون المستحقون أو مجموعات المستفيدين المستحقين من البرنامج بشكل متساوٍ. ونكرر ما ذكرناه آنفًا، سيتم ذلك كل مستفيد مؤهل الفرصة نفسها ليكون أول من يستفيد من البرنامج. في هذه الحالات، يمكن استخدام المستفيدين الذين التحقوا بالبرنامج لاحقًا كمجموعة مقارنة للمستفيدين السابقين؛ ما يؤدي إلى وضع تصميم قوي وفعال للتقييم، بالإضافة إلى طريقة شفافة وعادلة لتوزيع الموارد النادرة.

وأخيرًا، يمكن أن يوجد شاغل أخلاقي حول عدم إجراء تقييم عندما تستثمر البرامج موارد ضخمة في تدخلات ذات فعالية مجهولة. وفي هذا السياق، قد يُنظر إلى غياب التقييم نفسه على أنه أمر غير أخلاقي لأنه قد يطيل من أمد البرامج المبددة للموارد التي لا تنفع المجتمع، في حين يمكن استثمار هذه الأموال المهدورة في تدخلات ذات فعالية أكبر. ويمكن أن تؤدي المعلومات عن فعالية البرنامج المستمدة من تقييمات الأثر إلى استثمار أكثر فعالية وأخلاقية للموارد العامة.

حماية العناصر البشرية أثناء جمع البيانات ومعالجتها وتخزينها

هناك نقطة زمنية ثانية يمكن فيها أن تتضرر العناصر البشرية وهي في أثناء جمع البيانات ومعالجتها وتخزينها. إن الأسر المعيشية والمعلمين والأطباء والإداريين وغيرهم ممن يجيبون عن الاستبيانات أو يوفرون البيانات من خلال وسائل أخرى، يكونون معرضين للضرر إذا تمت مشاركة المعلومات التي يقدمونها مع الجمهور من دون تفعيل وسائل حماية كافية من أجل حماية سرية هويتهم. وقد يؤثر هذا الضرر في الأفراد أنفسهم أو المؤسسة التي ينتمون إليها. وإليك بعض الأمثلة:

- في أثناء تقديم المسح، تشارك امرأة معلومات بخصوص ممارسات تنظيم الأسرة وزوجها (الذي لا يحذ تنظيم الأسرة) يستمع إلى محادثتها مع مندوب الإحصاء.
 - تم انتهاك خصوصية الأسر المعيشية (وتعرضت سلامتها للخطر) عندما تمكن فرد ما من استخدام بيانات المسح التي تم نشرها على الإنترنت من أجل تحديد الدخل والأصول التي تتمتع بها عائلات معينة.
 - استخدمت دراسة ما مندوبي إحصاء غير مؤهلين بشكل لائق من أجل تقديم فحوصات بيومترية، مثل سحب عينة دم.
 - طلب أحد المجيبين عن المسح الانسحاب من الدراسة في منتصف المقابلة الشخصية، ولكن أمره مندوب الإحصاء بالانتهاء من الإجابة عن أسئلة المسح.
 - تم استخدام بيانات المسح لتحديد المؤسسات المجتمعية التي تعارض سياسات حكومية معينة في سبيل الانتقام منها.
- مع وضع هذه المخاطر بعين الاعتبار، تقع المسؤولية على المحققين الرئيسيين وغيرهم من العاملين في فريق الأبحاث أن يصونوا الحقوق والرفاهة للعناصر البشرية المتضمنة في تقييم الأثر بما يتوافق مع قواعد الأخلاقيات أو التشريعات الوطنية المناسبة والإرشادات الدولية.^١ توصي منظمة الصحة العالمية (WHO) بالمعايير الأساسية التالية من أجل تقييم المشاريع التي تتضمن عناصر بشرية:
- يجب حماية الحقوق والرفاهة للعناصر البشرية المتضمنة في تقييم الأثر كما ينبغي.
 - يجب أن يحصل الباحثون على موافقة مستنيرة ذات إرادة حرة من المشاركين.
 - يجب تقييم التوازن بين المخاطر والمنافع المحتملة واعتباره مقبولاً من قبل لجنة من الخبراء المستقلين.
 - يجب استيفاء أي متطلبات وطنية خاصة.
- يحدد تقرير بيلمونت: المبادئ والإرشادات الأخلاقية لحماية العناصر البشرية في الأبحاث (المفوضية الوطنية ١٩٧٨) ثلاثة مبادئ تشكل أساس السلوك الأخلاقي للأبحاث التي تتضمن عناصر بشرية:
- احترام الأشخاص. كيف سيحصل الباحثون على موافقة مستنيرة من العناصر البشرية المتضمنة في بحثهم؟

• *الإحسان.* كيف سيضمن الباحثون أن البحث (١) لا ينطوي على ضرر و(٢) يعظم من المنافع المحتملة ويحد من الضرر المحتمل؟

• *العدل.* كيف سيضمن الباحثون أن المنافع والأعباء الخاصة بالبحث متشاركة بشكل عادل على قدم المساواة؟

هناك عنصر أساسي يجب على المحققين الرئيسيين الالتزام به من أجل حماية العناصر البشرية، ألا وهو إرسال البحث وبروتوكولات جمع البيانات إلى لجنة مراجعة مؤسسية (IRB)، المعروفة أيضًا باللجنة الأخلاقية المستقلة (IEC) أو لجنة المراجعة الأخلاقية (ERB)، لتخضعه للمراجعة وتصرح بإجازته. لجنة المراجعة المؤسسية هي عبارة عن لجنة تم تعيينها بشكل رسمي لمراجعة واعتماد ومتابعة الأبحاث الطبية الأحيائية والسلوكية التي تتضمن عناصر بشرية. تتولى هذه اللجنة مراجعة بروتوكولات الأبحاث والمواد ذات الصلة قبل بدء الدراسة وأثناء تنفيذها وذلك من أجل تقييم أخلاقيات الأبحاث وطرقها. في سياق تقييمات الأثر، تكون مراجعة لجنة المراجعة المؤسسية لها أهمية خاصة عندما تتطلب الدراسة جمع بيانات عن الأسر المعيشية والأفراد. فعلى وجه الخصوص، تتحقق المراجعة التي تجريها اللجنة مما إذا كان المشاركون قادرين على الاختيار بأنفسهم المشاركة في أنشطة جمع البيانات وما إذا كان اختيارهم مستنيرًا وطوعيًا بالكامل. وأخيرًا، تراجع لجنة المراجعة المؤسسية ما إذا كان يوجد أي سبب للاعتقاد بأن سلامة المشاركين قد تكون في خطر.

يكون المحقق الرئيسي مسؤولاً عن تحديد جميع المؤسسات التي يجب عليها مراجعة الدراسة وإجازتها. تحظى الكثير من الدول بلجنة مراجعة أخلاقية وطنية وتحظى أغلب الجامعات بلجنة مراجعة مؤسسية. عادة يكون الفريق مطالبًا بالحصول على إجازة أخلاقية من لجنة المراجعة الأخلاقية الوطنية في البلد المعني ومن لجنة المراجعة المؤسسية لأي جامعة يتبعها المحققون. وهناك حالات تُنفذ فيها تقييمات الأثر في دول ليس بها لجنة مراجعة أخلاقية وطنية أو لا يكون للباحثين فيها لجنة مراجعة مؤسسية. وهنا، يجب على المحقق الرئيسي التعاقد مع لجنة مراجعة مؤسسية خارجية (ويصلح أن تكون تجارية). يمكن أن تستغرق عملية المراجعة والإجازة شهرين إلى ثلاثة أشهر، رغم أن الزمن يتفاوت حسب عدد المرات التي يجتمع فيها أعضاء لجنة المراجعة المؤسسية. يجب أن ينسق فريق السياسات والأبحاث إرسال البروتوكولات إلى لجنة المراجعة المؤسسية وأنشطة جمع البيانات بحيث يمكنهم الحصول على كل الإجازات المطلوبة قبل بدء عملية جمع البيانات التي تنطوي على عناصر بشرية.

تُعد المراجعة التي تجريها لجنة المراجعة المؤسسية ضرورية ولكنها شرط غير كافٍ لضمان حماية العناصر البشرية. وقد تتفاوت لجان المراجعة المؤسسية بشكل واسع من حيث قدراتها وخبراتها في التجارب العلمية الاجتماعية وكذلك في تركيز مراجعتها. إن لجان المراجعة المؤسسية، خاصة إذا كان موقعها بعيدًا عن حيث يُجرى التقييم، قد لا تكون واعية بدرجة كافية بالظروف المحلية لتكون قادرة على تحديد التهديدات السياقية للعناصر البشرية. وقد تضع تركيزًا كثيفًا على صياغة الاستبيانات ونماذج الموافقة. أو قد تكون لديها خبرة في منطقة موضوعية أكثر تركيزًا، مثل التجارب الطبية التي تختلف معاييرها كثيرًا عن التجارب الاجتماعية من حيث المخاطر المرتبطة بالعناصر البشرية. ولا يتعين التواني للحظة بشأن حماية العناصر البشرية بمجرد الحصول على موافقة لجنة المراجعة المؤسسية، بدلاً من ذلك يجب اعتبار هذه الموافقة نقطة بداية لضمان وصف التقييم بأنه أخلاقي.

تتطلب لجان المراجعة المؤسسية بشكل عام إرسال المعلومات التالية للمراجعة:

أدلة على التدريب. تتطلب الكثير من لجان المراجعة المؤسسية (وكذلك العديد من الإرشادات الأخلاقية الوطنية) أن يتم تدريب فريق الأبحاث على حماية العناصر البشرية، رغم أن الوحدات الموضوعية قد تختلف حسب الدولة. ونسرد بضعة خيارات للتدريب في قسم الموارد الإضافية في نهاية هذا الفصل.

المفهوم الأساسي

لجنة المراجعة المؤسسية (IRB) هي عبارة عن لجنة تم تعيينها لمراجعة واعتماد ومتابعة الأبحاث التي تنطوي على عناصر بشرية.

البروتوكول البحثي. يتضمن البروتوكول البحثي عناصر أساسية عادةً ما تحدد معالم خطة التقييم (على الأبرز تتمثل في الغرض من الدراسة وأهداف التقييم، وأسئلة السياسات الأساسية، ومنهجية التقييم المقترحة)، وتصف أيضًا كيف يمكن لفريق الأبحاث أن يضمن حماية العناصر البشرية. وبالتالي، يُعد هذا البروتوكول وثيقة غاية في الأهمية ضمن وثائق التقييم. عادةً ما يتضمن البروتوكول البحثي العناصر التالية من حيث التعامل مع العناصر البشرية: معايير اختيار مشاركي الدراسة (العناصر البحثية)، والمنهجية والبروتوكولات المطبقة

لحماية العناصر المعرضة للخطر، والإجراءات المستخدمة لضمان أن العناصر البشرية مدركة للمخاطر والمنافع المترتبة على المشاركة في الدراسة والإجراءات المستخدمة لضمان إخفاء هويتهم. إن البروتوكول البحثي يجب استخدامه من قبل الشركة المعنية بالمسح من أجل توجيه الإجراءات الميدانية. يمكن العثور على المزيد من المعلومات حول محتوى البروتوكول البحثي على الموقع الإلكتروني لمنظمة الصحة العالمية (WHO) وفي مجموعة أدوات تقييم الأثر (Impact Evaluation Toolkit).^١

الإجراءات المحددة لطلب الموافقة المستنيرة وتوثيقها. تشكل الموافقة المستنيرة ركناً أساسياً لحماية حقوق العناصر البشرية في أي دراسة. وهي تشترط أن يكون المجيبون على فهم واضح بالغرض والإجراءات والمخاطر والمنافع لعملية جمع البيانات التي يُطلب منهم المشاركة فيها. بصورة افتراضية، الموافقة المستنيرة التي يمنحها المجيب البالغ يجب أن تكون في صيغة وثيقة تتضمن قسمًا يسرد الأساليب المستخدمة لحماية سرية المجيب، وقسمًا آخر يوضح حق المجيب في الرفض أو التوقف عن المشاركة في أي وقت، وشرطًا للمخاطر والمنافع المحتملة، ومعلومات الاتصال في حالة كان يرغب المجيب في الاتصال بفريق جمع البيانات، ومساحة فارغة تسمح للمجيب بتسجيل موافقته الكتابية الرسمية على المشاركة في عملية جمع البيانات مع تذييل ذلك بتوقيعه. أحيانًا يتعذر على المشاركين في الدراسة اختيار المشاركة من عدمها. فمثلًا، عادةً ما يعتبر الأطفال غير قادرين على اتخاذ هذا الاختيار. ولذلك، على عكس البالغين القادرين، القصر الذين يتعذر عليهم الموافقة على المشاركة في مسح ما، قد يؤخذ لهم بالمشاركة بعد الحصول على إذن كتابي من ولي أمرهم أو الوصي عليهم.

في حين أن الخطوات المذكورة تشكل الإجراءات الافتراضية التي تعكس الموافقة المستنيرة، فإن العديد من تقييمات الأثر تطلب من لجنة المراجعة المؤسسية التابعة لها أن تُعفى من متطلب الحصول على موافقة كتابية رسمية من المجيبين. على سبيل المثال، عند التعامل مع مجتمع من الأميين، يتم إسقاط شرط الحصول على موافقة كتابية رسمية من المشاركين البالغين المحتملين المؤهلين غالبًا ويتم استبدالها بموافقة شفوية موثقة.^٣

مفهوم أساسي

الموافقة المستنيرة هي أحد الأركان الأساسية لحماية العناصر البشرية. وهي تشترط أن يكون المجيبون على فهم واضح بالغرض والإجراءات والمخاطر والفوائد لعملية جمع البيانات التي يُطلب منهم المشاركة فيها.

إجراءات لحماية سرية المجيبين. إن حماية سرية المجيبين تشكل ركنًا أساسيًا عند تخزين البيانات وإتاحة البيانات للجمهور. ويجب نزع الهوية من جميع المعلومات التي يقدمها المجيبون على مدار عملية جمع البيانات في سبيل حماية هويتهم. وعلى الرغم من احتمالية نشر نتائج الدراسة، فإن التقرير يجب كتابته بطريقة تجعل من المستحيل تحديد هوية فرد أو أسرة معيشية بعينها. فيما يخص ضمان السرية في البيانات، يجب ربط كل عنصر بشري في المسح برقم تعريف (ID) مشفر فريد، ويتعين حذف جميع الأسماء ومعرفات الهوية من قاعدة البيانات التي ستكون متاحة للجمهور. وتتضمن معرفات الهوية أي متغيرات تسمح بتحديد هوية الأفراد

أو الأسر المعيشية (مثل العنوان)، أو أي مجموعة من المتغيرات التي تؤدي الدور نفسه (مثل الجمع بين تاريخ الميلاد ومحل الميلاد والنوع الاجتماعي وسنوات التعليم). إذا كان فريق الأبحاث يرى أنه سيحتاج إلى معرفات الهوية من أجل متابعة المجيبين في مسح لاحق، فيمكنه الاحتفاظ بقاعدة بيانات منفصلة مؤمنة تربط الأرقام التعريفية المشفرة بالمعلومات المحددة لهوية المجيبين.٤ بالإضافة إلى تشفير الأرقام التعريفية للأفراد، لا بد أيضًا من تشفير المواقع والمؤسسات. فمثلًا، إذا كانت الأسر المعيشية والأفراد مرمزين بأرقام تعريفية مشفرة ولكن القرى محددة بشكل واضح، فقد يظل بالإمكان تحديد هوية الأسر المعيشية من خلال الخصائص المتضمنة في المسح. على سبيل المثال، قد تكون في قرية معينة أسرة واحدة فقط تمتلك دراجة نارية وسبع أبقار ومتجر طلاقة. أي شخص يمكنه الوصول إلى تلك البيانات ربما سيكون قادرًا على تحديد موقع الأسرة المعيشية، وهذا ينتهك سرية هذه الأسرة.

ضمان إجراء تقييمات موثوقة وذات مصداقية من خلال العلم المفتوح

هناك هدف أساسي لتقييم الأثر، وهو تقدير أثر برنامج ما على مجموعة من الحاصلات محل الاهتمام. يناقش الجزء ٢ سلسلة من الطرق التي يمكن بها ضمان دقة الآثار المقدّرة. يجب أن يضمن تقييم الأثر جيد التصميم والتنفيذ أن النتائج غير متحيزة وموثوقة وذات مصداقية وأنها تسهم في مجموعة معارف أوسع. عندما تتسم التقييمات بعدم التحيز والموثوقية والمصداقية ويمكن تفسيرها ضمن مجموعة ذات صلة من المعارف المرتبطة، فهي يمكن أن تسهم في صنع قرارات سياسة موفقة بما يحسن حياة الناس. ومع ذلك، في الواقع، يمكن لبضع مشكلات أن تعيق تحقيق هذا المبدأ المثالي.

في هذا القسم، سناقش كيف يمكن لمجموعة من المشكلات العلمية في تقييم الأثر أن تنعكس لتصير مشكلات صعبة بالنسبة لصانعي السياسات، وسناقش التدابير المحتملة لمنع حدوث هذه المشكلات أو تخفيفها. يتم وصف هذه التدابير بشكل عام تحت مصطلح *العلم المفتوح*، لأنها تهدف إلى جعل الطرق البحثية شفافة. ٥ يحتاج فريق الأبحاث إلى التعامل مع معظم هذه المشكلات، ولكن فريق السياسات الذي يشرف على تقييم الأثر يجب عليه أن يحيط بها علمًا في أثناء إدارة تقييمات الأثر. المشكلات وتداعيات السياسة والحلول المحتملة جميعها مذكورة بشكل مختصر في الجدول ١٣-١.

الجدول ١٣-١ التأكد من توفير معلومات موثوقة وذات مصداقية للسياسات من خلال العلم المفتوح

المشكلة البحثية	تداعيات السياسة	حلول المنع والتخفيف من خلال العلم المفتوح
<i>تحيز النشر.</i> يتم نشر النتائج الإيجابية فقط. ولا يتم نشر التقييمات التي تُظهر آثارًا محدودة أو لا تُظهر أي آثار، على نطاق واسع.	تقوم قرارات السياسات على مجموعة محرفة من المعارف. يفترض صانعو السياسات إلى معلومات حول ما لا يصلح ويتابعون تجربة/اعتماد سياسات ليس لها أثر.	سجلات التجارب
<i>تنقيب البيانات.</i> يتم تقسيم البيانات إلى أجزاء صغيرة حتى تظهر نتيجة انحدار إيجابية أو يتم تعديل الفرضية بناء على النتائج.	قرارات السياسات لاعتماد التدخلات قد تكون قائمة على تقديرات إيجابية غير مضمونة للآثار.	خطط التحليل المسبق
<i>اختبار الفرضيات المتعددة، تحليل المجموعة الفرعية.</i> يقوم الباحثون بتقسيم البيانات إلى أجزاء صغيرة حتى يجدوا نتيجة إيجابية لمجموعة ما. بشكل خاص، (١) يؤدي الاختبار المتعدد إلى خلاصة بوجود بعض الآثار في حين لا تكون هناك أي آثار في الواقع، أو (٢) لا يتم الإبلاغ عن أي آثار إلا إذا كانت بارزة وواضحة بشكل كبير.	قرارات السياسات لاعتماد التدخلات قد تكون قائمة على تقديرات إيجابية غير مضمونة للآثار.	خطط التحليل المسبق وأساليب التعديل الإحصائي الخاصة مثل اختبارات الفهارس ومعدل الخطأ الأسري وضابط معدل الاستكشاف الخاطئ
<i>الافتقار إلى التكرارية.</i> يتعذر تكرار النتائج لأن البروتوكول البحثي والبيانات وطرق التحليل ليست موثقة بشكل كافٍ.	قد تكون السياسة قائمة على نتائج متلاعب بها (سواء إيجابية أو سلبية) حيث إن النتائج قد تنبع عن أخطاء في الحسابات.	توثيق البيانات وتسجيلها، بما في ذلك بروتوكولات المشروع وتنظيم الرموز ونشر الرموز ونشر البيانات
قد يتعذر اكتشاف الأخطاء والتلاعب بالبيانات.	يتعذر مقارنة النتائج بين الدراسات المختلفة.	التغيرات في سياسات المجلة العلمية وسياسات التمويل لطلب توثيق البيانات وتشجيع التكرار
لا يكون الباحثون مهتمين بتكرار الدراسات، ولا تكون المجلات العلمية مهتمة بالنتائج "المتماثلة".	يتعذر اختبار صلاحية النتائج في سياق آخر.	
يتعذر تكرار التدخلات لأن بروتوكول التدخل غير موثوق بشكل كافٍ.	قد يتعذر على صانعي السياسات تكرار التدخل في سياق مختلف.	

أ. للإلمام بمقدمة أساسية حول مشكلة المقارنات المتعددة والتصحيحات الإحصائية المحتملة، يرجى مراجعة https://en.wikipedia.org/wiki/Multiple_comparisons_problem.

هناك نزعة فطرية لدى الباحثين الذين يعملون على تقييمات الأثر تجعلهم حريصين على نشر نتائج تقييماتهم في المجلات العلمية المحكمة؛ لأن ذلك يساعدهم في الارتقاء بحياتهم المهنية. ومع ذلك، معظم النتائج المنشورة في المجلات العلمية تُظهر آثارًا إيجابية. وهنا نطرح سؤالاً: "ماذا يحدث للتقييمات التي تُظهر نتائج سلبية أو التي تفشل في إظهار أي نتائج عديمة الدلالة؟". لا يكون هناك أي دافع لدى الباحثين لإعداد تقارير بنتائج عديمة الدلالة أو تقديمها للنشر في المجلات العلمية المحكمة؛ لأنها يتصورون أن هناك اهتمامًا ضئيلاً بالنتائج وأن المجلات العلمية سترفض أبحاثهم (Franco, Malhotra, and Simonovits 2014). ويُشار إلى تحيز النشر هذا بشكل عام بـ "مشكلة درج الملفات" لأن النتائج تظل في درج الملفات بدون توزيعها أو نشرها. قد تنشأ مشكلات تحيز نشر مماثلة لتقييمات أثر برامج محددة. على الأرجح أن تقوم فرق السياسات والممولون والحكومات بالنشر والدعاية للنتائج الإيجابية المستمدة من تقييم برنامج ما عن إيلاء الاهتمام نفسه للنتائج السلبية أو الأبحاث غير المؤدية إلى نتائج. وفي ظل هذه النزعات التحيزية، يصعب الحصول على صورة واضحة للتدخلات التي لا تصلح بما أن نتائج مثل هذه التدخلات عادة لا يكون لها وجود وتكون مجموعة الأدلة المتوفرة محدودة. إن صانعي السياسات الذين يحاولون تأسيس سياساتهم على الأدلة المتوفرة قد لا تكون لديهم إمكانية للوصول إلى الأبحاث غير المنشورة التي لم تتوصل إلى نتائج ذات دلالة، وكنتيجة لذلك، قد يستمرون في تجربة السياسات التي باءت بالفشل في أماكن أخرى.

يوجد حل جزئي لتحيز النشر وهو تسجيل التجارب. يجب أن تلقى فرق تقييم الأثر التشجيع لتسجيل تجاربها، ويكون فريق السياسات منوطاً بدور مهم لضمان تسجيل فريق الأبحاث لتقييم الأثر. إن تسجيل التجارب أمر شائع للغاية (وغالباً ما يكون شرطاً) في العلوم الطبية، ولكنه يزداد زحماً ويكتسب اهتماماً في مجال العلوم الاجتماعية، بما في ذلك تقييمات الأثر. تنطوي عملية التسجيل على أن يعلن الباحثين عن نيّتهم في تنفيذ تقييم قبل القيام بذلك فعلياً عن طريق تسجيل معلومات أساسية حول التقييم في سجل (انظر المربع ١٣-١٠). وبذلك، سيكون من الممكن الحصول على قائمة كاملة من تقييمات الأثر التي تم تنفيذها سواء كانت النتائج إيجابية أو خلاف ذلك.

تشكل السجلات خطوة كبيرة للأمام في ضمان اتصاف مجموعة المعارف المتاحة بأنها أقل تحيزاً. ولكن تظل هناك بعض التحديات القائمة. على سبيل المثال، حتى إذا كان يتضح من السجل أن التقييم تم تنفيذه، قد لا يكون بهذه السهولة الحصول على معلومات حول نتائج التقييم. قد تكون تقييمات الأثر توقفت أو لم يتم تنفيذها بشكل صحيح. وحتى إذا توافرت الأبحاث غير المؤدية إلى نتائج من تقييم ما، فهي غالباً ما تثير مجموعة إضافية من الأسئلة التي تصعب تفسير النتائج: هل لم يعثر الباحثون على نتائج لأن التقييم كان رديء التصميم والتنفيذ، أم لأن البرنامج لم يُنفذ بشكل جيد، أم لأن البرنامج ليس له أثر حقاً؟ كما يوضح الفصل ١٦، من شأن جمع بيانات تكميلية من خلال متابعة البرنامج أو أي مصادر بيانات بديلة أن يساعد على ضمان تفسير النتائج بشكل جيد.

المربع ١٣-١: سجلات التجارب للعلوم الاجتماعية

- "مركز العلم المفتوح" يدير "إطار عمل العلم المفتوح" (OSF) الذي يختلف قليلاً في التركيز ولكنه أيضاً يُعد سجلاً (<https://osf.io>). إن هذا الإطار عبارة عن نظام إدارة قائم على السحابة للمشاريع البحثية يسمح بإنشاء لقطات حية للأبحاث في أي نقطة زمنية مع ختمها بطابع زمني ورابط URL ثابت. يستطيع الباحثون تحميل البروتوكول والفرضيات البحثية والبيانات والرموز إلى "إطار عمل العلم المفتوح" ومشاركة رابط الإنترنت الخاص بالنتائج باعتباره دليلاً على التسجيل.
- يجب تسجيل تقييمات أثر السياسات العامة بشكل طبيعي مع السجلات العلمية الاجتماعية بدلاً من السجلات الطبية بسبب طبيعة البحث. وإليك بعض الأمثلة:
 - سجل "الجمعية الاقتصادية الأمريكية" للتجارب العشوائية المضبوطة: يمكن الوصول إليها من الموقع الإلكتروني <http://www.socialscienceregistry.org>. اعتباراً من يوليو ٢٠١٥، يسرد السجل ٤١٧ دراسة في ٧١ دولة.
 - "المبادرة الدولية لتقييم الأثر" (3ie) تدير "سجل تقييمات الأثر الإنمائية الدولية" (RIDIE) الذي يركز على تقييمات الأثر المرتبطة بالتنمية في البلدان المنخفضة والمتوسطة الدخل. وقد سجلت ٦٤ تقييمًا تقريباً اعتباراً من يوليو ٢٠١٥.

تنقيب البيانات والفرضيات المتعددة وتحليل المجموعة الفرعية

هناك مشكلة محتملة أخرى متعلقة بتقييم الأثر تُدعى *تنقيب البيانات*، وهي ممارسة تتم فيها معالجة البيانات بحثاً عن نتائج إيجابية. يمكن أن يتجلى تنقيب البيانات بأشكال مختلفة. على سبيل المثال، عندما تتوفر البيانات، ربما يستميل الباحثون إلى إجراء تحليلات انحدار على البيانات حتى تظهر أي نتائج إيجابية، ومن ثم يرتبون فرضية جذابة لهذه النتيجة. وهنا تنشأ مشكلة نوضح سببها كما يلي: عندما تجري اختبارات إحصائية لقياس دلالة الآثار، نحتاج إلى استخدام مستوى دلالة، لنقل ٥ بالمائة. من الناحية الإحصائية، في واحد من كل ٢٠ اختباراً للآثار تأتي الدلالة بمستوى ٥ بالمائة حتى إذا كان التوزيع المستخدم في الأساس لا يضمن وجود أثر بالفعل (انظر الفصل ١٥ للاطلاع على مناقشة حول أخطاء النوع الأول). مع تنقيب البيانات، لا أحد يستطيع التأكد من أن نتيجة الأثر نتيجة صحيحة أو إنها ناتجة عن الخصائص الإحصائية للاختبار بشكل بحت. ترتبط هذه المشكلة بمشكلة اختبار الفرضيات المتعددة: عندما يتضمن عملاً بحثياً عدة فرضيات مختلفة، توجد احتمالية مرتفعة أن واحدة منها سيتم تأكيدها باختبار إيجابي يعزو إلى الصدفة وحدها (بسبب الخصائص الإحصائية للاختبار) وليس بسبب وجود أثر حقيقي. وينشأ موقف مماثل بالنسبة لتحليل المجموعة الفرعية: عندما تكون العينة كبيرة بما يكفي، فقد يحاول الباحثون تقسيمها فرعياً حتى يعثروا على أثر /و/وحدة من المجموعات الفرعية. ونكرر قولنا، لا أحد يستطيع التيقن من أن نتيجة الأثر لهذه المجموعة الفرعية نتيجة صحيحة أو ما إذا كانت ناتجة عن الخصائص الإحصائية للاختبار بشكل بحت.

مثال آخر على تنقيب البيانات هو عندما يكون قرار متابعة أو إيقاف جمع البيانات متخذًا بشكل أساسي اعتمادًا على نتيجة وسطية: مثلاً، كان من المخطط إجراء مسح لأسر معيشية باستخدام عينة حجمها ٢,٠٠٠ أسرة معيشية وأتم العمل الميداني جمع البيانات من ١,٠٠٠ أسرة معيشية. إذا أخرجت هذه العينة المختزلة نتيجة تقدير أثر إيجابية وتم أخذ قرار بإيقاف جمع البيانات لتجنب مخاطرة تغير النتائج بفعل البيانات الإضافية، فهذا إذن يُسمى تنقيب البيانات. وهناك أمثلة أخرى يتم فيها استبعاد ملاحظات أو مجموعات معينة غير مناسبة، أو يخفي المسؤولون النتائج التي يرونها غير متناسبة بصورة انتقائية. على الرغم من أنه لا يوجد سبب للاعتقاد بأن هذه الممارسات واسعة الانتشار، فهناك بضعة حالات فاضحة ذائعة الصيت من الممكن أن تقوض الصلاحية العلمية لتقييم الأثر. بالإضافة إلى ذلك، حتى الحالات الأخف من تنقيب البيانات من المحتمل أن تحرف مجموعة الأدلة التي يستخدمها صانعو السياسات لتحديد متى تبدأ التدخلات أو تستمر أو تنقطع.

هناك توصية عامة لتجنب تنقيب البيانات وهي استخدام خطة تحليل مسبق. تحدد مثل هذه الخطط طرق التحليل قبل تنفيذ تحليل تقييم الأثر، وبالتالي تحدد تركيز التقييم وتحدد من احتمالية تغيير الطرق بمجرد بدء التحليل. يجب أن تحدد خطة التحليل المسبق الحقائق المطلوب قياسها، والمتغيرات المقرر تشكيلها واستخدامها، والمجموعات الفرعية التي سيشملها التحليل، والمقاربات التحليلية الأساسية المقرر استخدامها لتحديد الآثار. يتعين أن تتضمن خطط التحليل المسبق أيضًا التصحيحات المقترحة من الباحثين لاختبار الفرضيات المتعددة واختبار المجموعة الفرعية، إذا لزم الأمر. فمثلاً، اختبار أثر تدخل تعليمي على ست نتائج اختبارات مختلفة (الرياضيات، الإنجليزية، الجغرافيا، التاريخ، العلوم، الفرنسية) لخمس مجموعات مدرسية مختلفة (من الصف الأول حتى الخامس) ونوعين اجتماعيين (الذكور والإناث) سنتج عنه ٦٠ فرضية مختلفة ولزاماً ستكون نتيجة واحدة أو بعض منها في الاختبار ذات دلالة بمحض الصدفة. بدلاً من ذلك، يستطيع الباحثون اقتراح احتساب مؤشر واحد أو أكثر يجمع المؤشرات معاً، بحيث يتم الحد من عدد الفرضيات والمجموعات الفرعية.^٦

ربما تحدد خطة التحليل المسبق من المخاوف المحيطة بتنقيب البيانات، إلا أن هناك شاغلاً من أن هذه الخطة ستقلل من درجة المرونة المطلوبة في نوع التحليل الذي ينفذه الباحثون. فمثلاً، قد تحدد خطة التحليل المسبق القنوات المتوقعة لأثر التدخل عبر سلسلة النتائج بالكامل. ولكن، بمجرد تنفيذ التدخل فعلياً، قد تظهر من العدم مجموعة إضافية من العوامل غير المتوقعة. على سبيل المثال، إذا كانت الحكومة تفكر في تنفيذ طريقة جديدة لسداد مستحقات مقدمي الرعاية الصحية، قد يستطيع المرء التوصل إلى قنوات محتملة للأثر. ولكن، سيكون من الصعب جداً توقع كل تأثير محتمل يمكن أن ينشأ. في بعض الحالات، تكون المقابلات الشخصية النوعية مع مقدمي الخدمات لازمة للوصول إلى فهم دقيق لكيفية اعتماد التغييرات وكيف يمكن أن يؤثر ذلك في الأداء. سيكون من الصعب للغاية تضمين كل هذه الاحتمالات في خطة التحليل المسبق بصورة سابقة. وفي هذه الحالة، سيضطر الباحثون إلى العمل خارج إطار خطة التحليل المسبق، ويتعين عدم مجازاتهم على فعلهم ذلك. بمعنى آخر، يمكن أن تفرض خطة التحليل المسبق قيوداً إضافية على التقييمات عن طريق تحويلها إلى مجرد تأكيدات على فرضية معينة، بدلاً من أن تُصاغ في سياق بحث استكشافي. ولكن يجب تمكين الباحثين لمواصلة استكشاف خيارات جديدة يمكن أن تتحول إلى بحث تأكيد في تقييمات لاحقة.

يوجد نوعان من التكرار يحملان أهمية خاصة بالنسبة لتقييمات الأثر. أولاً، بالنسبة لدراسة معينة، يجب على الباحثين بخلاف فريق الأبحاث الأصلي أن يكونون قادرين على تكرار النتائج نفسها (أو الخروج بنتائج مماثلة لها على الأقل) التي توصل لها الباحثون الأصليون، وذلك عند استخدام هذه البيانات والتحليلات. تفيدنا عمليات تكرار نتائج معينة لتقييم الأثر في التحقق من صلاحيتها الداخلية وعدم تحيزها. عندما يتعذر تكرار الدراسات أو النتائج بسبب عدم إتاحة معلومات بشأن الترميز أو البيانات، توجد مخاطرة تتمثل في عدم القدرة على اكتشاف الأخطاء والتلاعبات في التحليل واستمرار تأثر السياسة بالنتائج غير الدقيقة. لحسن الحظ، تحدث تطورات كبيرة من حيث إتاحة البيانات والترميز والبروتوكولات. وبدأ عدد متزايد من المجلات العلمية الاجتماعية في فرض شرط إتاحة البيانات والترميز إلى جانب نشر النتائج. تتغير الممارسات والحوافز وإن كانت بوتيرة بطيئة بفضل توجيهات وإرشادات مثل تلك التي وضعها "مركز العلم المفتوح"، ونقصد بذلك "توجيهات تعزيز الشفافية والمصارحة". لضمان إمكانية التكرار، تحتاج فرق تقييم الأثر إلى إتاحة البيانات للجمهور وضمان أن جميع البروتوكولات (بما يتضمن ذلك بروتوكول التوزيع العشوائي)، ومجموعات البيانات، ورموز التحليل لتقييم الأثر موثقة ومخزنة بشكل آمن ومفصلة بدرجة كافية.

ثانياً، بمجرد إكمال تقييم الأثر، يجب إتاحة المجال لصانعي السياسات والباحثين للأخذ بالبروتوكولات الأصلية الخاصة بالتدخل والتقييم وتطبيقها في سياق مختلف أو في فترة زمنية مختلفة للتثبت من مدى تحقق النتائج نفسها في ظل ظروف مختلفة. مع انعدام تكرارية نتائج التقييم، تنشأ مشكلة جدية بالنسبة لصانعي السياسات. لنقل فرضاً أن هناك تقييماً يُظهر أن إدخال أجهزة الكمبيوتر في المدارس له نتائج وفوائد عظيمة، ولكن هذه هي الدراسة الوحيدة التي خرجت بمثل هذه النتائج، ولم يتمكن غيرهم من الباحثين من الحصول على النتائج الإيجابية نفسها في تقييمات لاحقة لبرامج مماثلة. ماذا ينبغي لصانع السياسات أن يفعل في مثل هذه الحالات؟ إن انعدام إمكانية تكرار النتائج له أسباب عدة. أولاً، ربما يكون من الصعب تنفيذ تقييمات على سبيل محاولة تكرار نتائج تم الحصول عليها في دراسة مماثلة: قد لا يبدي الباحثون أو الممولون اهتماماً نحو الدراسات "المماثلة". ثانياً، بغرض أن هناك رغبة وموارد مالية لتكرار الدراسات، فإن التكرار لا يكون ممكناً على الدوام لأن البروتوكولات (بما تشمل بروتوكول التوزيع العشوائي)، والبيانات، ورموز التحليل للدراسة الأصلية قد لا تكون متوفرة أو موضحة بالتفصيل الكافي. هناك جهود متزايدة في أوساط المؤسسات التي تدعم تقييمات الأثر هدفها تشجيع التكرارات عبر الأماكن المختلفة، مثلاً، عن طريق تطوير مجموعات فئوية من الدراسات حول موضوعات مماثلة أو تعزيز تقييمات الأثر متعددة المواقع.

قائمة المراجعة: تقييم أثر أخلاقي وذو مصداقية

يؤدي صانعو السياسات دورًا مهمًا لضمان أن كل شيء مجهز لإجراء تقييم أثر أخلاقي ذي مصداقية. وعلى وجه الخصوص، يتحمل صانعو السياسات المسؤولية الأساسية لضمان أن قواعد تعيين البرنامج عادلة، وعليهم أن يحملوا فريق الأبحاث مسؤولية شفافية طرق البحث. ونقترح اتباع قائمة المراجعة التالية التي تتضمن أسئلة يجب طرحها:

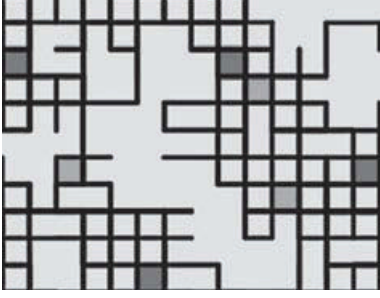
- ✓ هل هناك عدالة في التعيين في مجموعتي المعالجة والمقارنة؟ هل توجد أي فئات أشد احتياجًا بشكل خاص يجب أن تتلقى البرنامج على أي حال؟ من سيتم استثناءه من تقييم الأثر؟
- ✓ هل حدد فريق الأبحاث لجنة المراجعة المؤسسية أو لجنة المراجعة الأخلاقية الوطنية ذات الصلة؟
- ✓ هل يسمح الجدول الزمني لتقييم الأثر بوقت كافٍ لتحضير بروتوكول البحث وإرساله إلى لجنة المراجعة المؤسسية والحصول على الموافقة قبل بدء جمع البيانات من العناصر البشرية؟
- ✓ هل أرسل فريق الأبحاث بروتوكول البحث وخطة التحليل المسبق إلى سجل تجارب علمية اجتماعية؟
- ✓ هل يتم العمل بإجراء يضمن توثيق العناصر الأساسية للتدخل مع حدوثها وليس فقط كما هو مخطط لها؟
- ✓ هل يفهم صانعو السياسات أن نتائج التقييم قد تُظهر أن التدخل لم يكن ذا فعالية، وهل يوافقون على نشر مثل هذه النتائج وعدم حظرها؟
- ✓ هل حدد فريق التقييم الطريقة التي ستتم بها إتاحة بيانات التقييم والنتائج حتى إذا لم يتمكن فريق الأبحاث من نشر النتائج في مجلة علمية محكمة؟

يمكن للمبادئ والمشكلات وقائمة المراجعة المذكورة في هذا الفصل أن تساعد على ضمان أن يتصف تقييم الأثر الخاص بك بالمصداقية والأخلاقية.

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- تدريب "العناصر البشرية" من المعهد الوطني للصحة (NIH) الأمريكي
 - يوفر المعهد الوطني للصحة تدريبًا عبر الإنترنت وعلى الرغم من أنه يركز على العلوم الطبية والولايات المتحدة، فهو لا يزال زائرًا بالمعلومات ويستغرق حوالي ساعة واحدة لإكماله. انظر <http://www.ohsr.od.nih.gov> و <http://phrp.nihtraining.com/users/login.php>
- تدريب "العناصر البشرية" من خلال مبادرة التدريب المؤسسية التعاونية في جامعة ميامي (CITI)
 - توفر هذه المبادرة دورات تدريبية دولية بعدة لغات لكل من المؤسسات والأفراد على حد سواء، رغم أن البرنامج يفرض رسومًا (بدلاً من ١٠٠ دولار أمريكي للشخص). انظر <http://www.citiprogram.com>
- المؤلف الدولي لمعايير الأبحاث البشرية
 - في كل عام، تنشر إدارة الصحة والخدمات البشرية الأمريكية مجموعة من القوانين واللوائح والإرشادات التي تحكم الأبحاث المتضمنة عناصر بشرية. يشتمل الإصدار ٢٠١٥ على ١١٣ دولية وكذلك معايير من عدد من المؤسسات الدولية والإقليمية. تذكر الوثيقة لجان المراجعة المؤسسية الوطنية والدولية (<http://www.hhs.gov/ohrp/international>)
- إجراءات حماية العناصر البشرية في الأبحاث المدعومة من USAID (الوكالة الأمريكية للتنمية الدولية) (<http://www.usaid.gov/policy/ads/200/humansub.pdf>)
- دليل أفضل الممارسات في الأبحاث العلمية الاجتماعية الشفافة بواسطة جاريث كريستنسن بمساعدة من Courtney Soderberg (مركز العلم المفتوح) (<https://github.com/garretchristensen/BestPracticesManual>)
- هذا عبارة عن دليل تشغيلي حول أحدث وأفضل الممارسات للأبحاث العلوم الاجتماعية الكمية الشفافة. ويخضع هذا الدليل للتحديث بصفة دورية.
- إرشادات تعزيز الشفافية والمصارحة (<http://centerforopenscience.org/top/>) (TOP).
- يمكن الاطلاع على هذه الإرشادات على الموقع الإلكتروني لمركز العلم المفتوح. للحصول على روابط إلى أبرز لجان المراجعة المستقلة وخدمات لجان المراجعة المؤسسية المستقلة، تفضل بزيارة "بوابة تقييم بنك التنمية للبلدان الأمريكية" (<http://www.iadb.org/evaluationhub>)
- لمعرفة المزيد حول جمع البيانات، ارجع إلى "بوابة تقييم بنك التنمية للبلدان الأمريكية" (<http://www.iadb.org/evaluationhub>)
- راجع قسم جمع البيانات تحت عنوان "حماية العناصر البشرية". لاحظ الرابط المؤدي إلى جمعية اعتماد برامج حماية البحوث البشرية (AAHRPP). توفر AAHRPP تدريبات وشهادات
- للجان المراجعة المؤسسية. يمكن العثور على قائمة بالمؤسسات المعتمدة على موقعها.
- للاطلاع على إرشادات لحماية المشاركين في الأبحاث البشرية، انظر الوحدة ٤ في "مجموعة أدوات تقييم الأثر من البنك الدولي" (<http://www.worldbank.org/health>) World Bank Impact Evaluation Toolkit, Module 4

- ١ في غياب الإرشادات الأخلاقية الوطنية، يجب على المحقق والفريق الاسترشاد بإعلان هلسنكي المعتمد من قبل الجمعية الطبية العالمية التاسعة والعشرين في طوكيو (أكتوبر ١٩٧٥) والمادة ٧ من العهد الدولي للحقوق المدنية والسياسية المعتمد من قبل الجمعية العامة للأمم المتحدة في ١٦ ديسمبر عام ١٩٦٦. تتوفر إرشادات إضافية من قبل منظمة الصحة العالمية وتقرير بلمونت حول المبادئ والإرشادات الأخلاقية لحماية العناصر البشرية (١٩٧٤) (<http://www.hhs.gov/>) (<http://www.hhs.gov/ohrp/policy/belmont.html>). هي عبارة عن مؤلف دولي لمعايير الأبحاث البشرية التي يمكن العثور عليها في الموقع الإلكتروني <http://www.hhs.gov/ohrp/international>.
- ٢ يمكن العثور على إرشادات منظمة الصحة العالمية حول كيفية كتابة بروتوكول للأبحاث المتضمنة مشاركين في الموقع الإلكتروني http://www.who.int/rpc/research_ethics/guide_rp/en/index.html.
- ٣ تتوفر المزيد من المعلومات حول إجراءات الموافقة أثناء جمع البيانات في مجموعة أدوات تقييم الأثر من البنك الدولي.
- ٤ للاطلاع على المزيد من المعلومات حول تعيين المعرفات في مجموعة أدوات تقييم الأثر من البنك الدولي.
- ٥ للمزيد من المعلومات حول توصيات العلم المفتوح في سياق تقييم الأثر، يرجى مراجعة Miguel and others (٢٠١٤).
- ٦ تتوفر أساليب أخرى. انظر على سبيل المثال Anderson (٢٠٠٨).

- Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.
- Christensen, Garret, with Courtney Soderberg. 2015. *The Research Transparency Manual*. Berkeley Initiative for Transparency in the Social Sciences. <https://github.com/garretchristensen/BestPracticesManual>.
- Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–5.
- Miguel, Edward, C. Camerer, Katherine Casey, Joshua Cohen, Kevin M. Esterling, and others. 2014. "Promoting Transparency in Social Science Research." *Science* 343: 30–31.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1978. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. U. S. Department of Health, Education, and Welfare Publication No. (OS) 78-0012. Washington, DC: Government Printing Office.
- Vermeersch, Christel, Elisa Rothenbühler, and Jennifer Sturdy. 2012. *Impact Evaluation Toolkit: Measuring the Impact of Results-Based Financing on Maternal and Child Health*. World Bank, Washington, DC. <http://www.worldbank.org/health/impactevaluationtoolkit>.



نشر النتائج وتحقيق أثر السياسة

قاعدة أدلة قوية للسياسات

لقد أتممت أخيرًا المهمة الشاقة لتقييم برنامجك من البداية إلى النهاية، وهو جهد يستغرق عدة سنوات يُبذل فيه الوفير من الموارد المالية والبشرية. وقمت بتسليم مخرجات التقييم النهائية بما يشمل ذلك تقريرًا من مئتي صفحة متممًا بملحقات متعددة. هل أنجزت المهمة هكذا؟

في الحقيقة، تبدأ مرحلة جديدة الآن من أجل ضمان أن تؤتي كل هذه الجهود ثمارها في صورة أثر على السياسة. تهدف تقييمات الأثر في الأساس إلى توفير حس المحاسبة للاستثمارات السابقة وتوجيه قرارات السياسات في المستقبل نحو تطوير ذي فعالية أكبر من حيث التكلفة بحيث تحقق الموارد الشحيحة أعلى عوائد اجتماعية ممكنة. وستتأثر قرارات السياسات هذه بمجموعة من العوامل، بدءًا من الاقتصاد السياسي إلى الأوضاع الأيديولوجية. ولكن يمكن (بل يجب) أن تؤثر تقييمات الأثر في السياسات عن طريق توفير قاعدة أدلة قوية توجه الموارد صوب تدخلات فعالة ذات جدوى مثبتة. في المراحل المبكرة من البرنامج، من المفترض أن تؤدي الأدلة المستخلصة من تقييمات الأثر المتوفرة وذات الصلة دورا مركزيا في إرشاد تصميم البرنامج وتوجيه المجموعة التالية من أسئلة التقييم.

المفهوم الأساسي

يجب أن تجيب تقييمات
الأثر عن أسئلة السياسات
ذات الصلة بطريقة
دقيقة وجادة، وتأتي
بالأدلة القابلة للتنفيذ إلى
الأطراف المعنية الأساسية
في التوقيت المناسب،
ونشر الأدلة في صيغة
يستطيع صانعو السياسات
الوصول إليها واستخدامها
بسهولة.

ومع ذلك، من العادي أن لا تحدث العملية المؤثرة في السياسات بشكل تلقائي من خلال مجموعة من الأدلة وحدها. أولاً وقبل كل شيء، يتعين أن تجيب تقييمات الأثر عن أسئلة السياسات ذات الصلة بطريقة دقيقة وجادة وتأتي بالأدلة القابلة للتنفيذ إلى الأطراف المعنية الأساسية في التوقيت المناسب. ولكن صانعي السياسات ومديري البرنامج قد لا يكون لديهم الوقت والطاقة للتعلم في تفاصيل تقرير يتألف من مائتي صفحة في محاولة لاستخلاص النتائج والتوصيات الأساسية. يجب تجميع المعلومات المستخرجة من خلال تقييمات الأثر ونشرها بطريقة تسمح لصانعي السياسات بالوصول إليها واستخدامها بسهولة.

سنتناول في هذا الفصل الطرق التي يمكن أن يؤثر بها تقييم الأثر الخاص بك في السياسات، ومؤسسات الدولة التي يحاذ الوصول إليها، والإستراتيجيات الخاصة بتوصيل ونشر المعلومات إلى الجماهير المستهدفة بحيث يؤثر التقييم على السياسة العامة.

تبدأ المرحلة الأولى للتأثير على السياسة من اختيار أسئلة التقييم ذات الصلة التي ستكون مفيدة لاتخاذ قرارات السياسات، كما ناقشنا في الجزء الأول من الكتاب. سيبدأ صانعو السياسات والمقيمون، على الأرجح، بإعداد قائمة من الأسئلة المرجوة في أثناء المراحل المبكرة لتصميم تقييم الأثر. يجب تدقيق هذه الأسئلة مع المجموعة الأساسية من الأطراف المعنية وصانعي السياسات الذين سيستخدمون في نهاية المطاف تقييم الأثر لصياغة الأسئلة. عادةً يتم تعديل هذه القائمة المرجوة وتحسينها بمرور الوقت لتضمن عدد محدود أكثر من الأسئلة حسنة الصياغة التي تكون ذات صلة بالسياسة وقابلة للتعديل ليتم الإجابة عنها من خلال تقييم الأثر باستخدام الطرق التي خضعت للمناقشة في الجزء الثاني من هذا الكتاب. وفي الوقت ذاته، هناك خطوة أولى حاسمة للتأثير في السياسة وهي إشراك صانعي السياسات لتحديد أهم الأسئلة وتشجيع فريق التقييم على استجماع الجدوى الفنية للإجابة عن هذه الأسئلة. بمجرد تفعيل البرنامج على أرض الواقع، من المحتمل أن يبدأ تقييم الأثر في إنتاج مدخلات تحليلية مهمة يمكنها أن تدعم السياسات بالمعلومات قبل وقت طويل من اتيان البرنامج وتقييم الأثر لثمارهما. ومن بين الأمثلة الشائعة على ذلك نتائج مسح خط الأساس أو تحليل للنتائج قصيرة الأجل. غالبًا ما تنتج مسوحات خط الأساس أولى البيانات الشاملة في إطار المجتمع الإحصائي لبرنامج معين، ما يوفر الإحصائيات الوصفية التي يمكنها تغذية تصميم البرنامج وحوار السياسات. في حين أن البرنامج قد يكون له وصف عام لمجتمعه الإحصائي المستهدف من خلال المسوحات الوطنية أو الدراسات التشخيصية، فإن مسح خط الأساس قد يوفر أولى المعلومات التفصيلية لمجتمعات إحصائية أو مناطق جغرافية محددة التي سيتم فيها تنفيذ البرنامج. على سبيل المثال، قد يتضمن برنامجاً مصمماً لتحسين تغذية الأطفال من خال المكملات الغذائية إحصائيات حول معدلات التقزم والهزال من مسوحات حالية على المستوى الوطني، ولكن قد يوفر مسح خط الأساس أولى القياسات لحالة التغذية وعادات الأكل لفئة الأطفال التي سيخدمها البرنامج. ويمكن أن يكون هذا النوع من البيانات عالي القيمة لتكييف تصميم التدخل ويجب إتاحتها لفريق السياسات بالتوقيت المناسب (بشكل مثالي قبل طرح التدخل) من أجل التأثير في تصميم البرنامج. يقدم المربع ١٤-١ مثالاً من موزمبيق.

يمكن لبعض تقييمات الأثر - خاصة تلك التي تعتمد على مصادر البيانات الإدارية أو المسوحات الدورية - أن تخرج بنتائج مرحلية تغذي البرنامج بالمعلومات بينما يُجرى تنفيذ البرنامج. وتوفر هذه النتائج معلومات وتوصيات عالية القيمة بشأن كيف تتغير المؤشرات على طول المسار السببي بمرور الوقت، ما يسمح بتعديل كل من تنفيذ البرنامج وتوقيت أنشطة التقييم وفقًا لذلك. فمثلاً، إذا كان البرنامج في منتصف الطريق، فمن الطبيعي ألا تكون هناك آثار على الحصائل قصيرة الأجل، وقد ينصح القائمون على البرنامج بتنفيذ تقييم تشغيلي للكشف عن المعوقات واتخاذ ما يلزم من الإجراءات التصحيحية. يمكن تعديل الخط الزمني للتقييم في سبيل تجنب إجراء مسح خط نهاية عالي التكلفة قبل أن تكون لنتائج التدخل فرصة للظهور. في مثال تغذية الأطفال، إذا كان تحليل البيانات الإدارية بشأن توزيع المكملات الغذائية يُظهر أن المكملات لا تصل إلى المستفيدين المقصودين، فمن ثم يمكن تنبيه فريق السياسات بمرور مراجعة سلسلة الإمداد لديه. إن مسح المتابعة عالي التكلفة المستخدم لقياس الوزن والطول للأطفال يمكن تأجيله لبعد مرور بضعة أشهر من تشغيل البرنامج بشكل فعال، لأنه لا يوجد سبب وجيه للاعتقاد بأن برنامج التغذية سيولد آثاراً عما قريب إذا لم يصل إلى المشاركين فيه.

المربع ١٤-١: أثر سياسة نموذج مبتكر لمرحلة رياض الأطفال في موزمبيق (تابع من الفصل ١)

العديد من الأطفال في عدة اتجاهات، بدءاً من اللغة والتواصل للتنمية المعرفية والاجتماعية العاطفية. لقد تم تقديم مسح خط الأساس من قبل فريق التقييم في الندوات وورش العمل حيث تمت مناقشة النتائج مع شخصيات رفيعة المستوى من صانعي السياسات والمانحين الدوليين والأطراف المعنية الأساسية من أوساط تنمية الطفولة المبكرة. وكان لهذه البيانات المكتسبة من تقييم الأثر دور تأكيد على ضرورة الاستثمار في هذه المنطقة، ودور تحفيزي لحشد الدعم من أجل تحقيق أجندة الطفولة المبكرة في البلاد. وتم توزيع التقييم المكتمل في النهاية عبر منافذ مختلفة، بما في ذلك مذكرات السياسات ومقاطع الفيديو والمدونات، وتم جمع عدد منها على الموقع الإلكتروني لمؤسسة المبادرة الدولية لتقييم الأثر (3ie).

تذكر المثال المذكور في الفصل ١ (المربع ١-٢): كان برنامج مرحلة ما قبل المدرسة المجتمعي المنفذ من منظمة إنقاذ الطفل في موزمبيق مدخلاً مهماً لسياسة تنمية الطفولة المبكرة الوطنية في البلاد. ومع ذلك، حتى قبل انتهاء البرنامج، توصل التقييم إلى معلومات جديدة وكاشفة تغذي حوار السياسات في ذلك السياق بالدولة. من مسح خط الأساس المنفذ ضمن التقييم، توصلوا إلى أولى قياسات حصائل تنمية الأطفال على مستوى المجتمع الإحصائي باستخدام اختبارات متخصصة لتنمية الأطفال مكيفة حسب السياق الموزمبيقي وتم جمعها من قبل القائمين على المسح المتخصصين. على الرغم من أن البيانات كانت مستمدة من فئة مختارة من المجتمعات في إحدى ولايات موزمبيق، فإن إحصائيات خط الأساس وفرت لمحة عامة أولى عن حصائل تنمية الأطفال في الدولة وأظهرت تأخر

تميل تقييمات الأثر إلى إنتاج أحجام كبيرة من المعلومات بدءًا من الدعائم الفنية لتصميم التقييم إلى الإحصائيات الوصفية وتحاليل الأثر المتممة بمجموعات البيانات والرموز الإحصائية والتقارير. لا بد أن يبذل فريق التقييم جهدًا لتوثيق جميع المعلومات خلال دورة التقييم، وإلى أقصى حد ممكن، ووضع الوثائق الفنية (غير السرية) في النطاق العام: على سبيل المثال: من خلال موقع إلكتروني مخصص. وفي النهاية، ستعتمد مصداقية نتائج التقييم على المنهجية والدقة المتبعة عند تنفيذ التقييم. تعزز الشفافية الكاملة مصداقية التقييم وإمكاناته على التأثير في السياسات.

في حين أن الاكتمال والشفافية أمران مهمان، فإن معظم مستخدمي المعلومات لن يتعمقوا في التفاصيل. الأمر مرهون باستخراج فريق التقييم لمجموعة يسهل التعامل معها من الرسائل الأساسية التي تلخص معظم النتائج والتوصيات ذات الصلة بالسياسات وتوصل هذه الرسائل بشكل متسق عبر الجماهير. ويشكل تسلسل أنشطة النشر جزءا بالغ الأهمية في أثر السياسات. ما لم يتفق فريق السياسات على خلاف ذلك، يجب إجراء الجولة الأولى من العروض التقديمية والمشاورات داخليا في وجود فريق ومديري البرنامج وصانعي السياسات. وقد يضر تسرب نتيجة مبكرة إلى النطاق العام بمسموعة البرنامج مع إحداث ضرر دائم لأثر السياسات.

تصميم إستراتيجية إعلام لأنواع الجماهير المختلفة

توجد على الأقل ثلاثة جماهير رئيسية لاستنتاجات تقييم الأثر: العاملون والمديرون في البرنامج المشاركون في البرنامج الخاضع للتقييم؛ والشخصيات رفيعة المستوى من صانعي السياسات الذين سيستخدمون التقييم لدعم قرارات التمويل وتصميم السياسات بالمعلومات؛ وأوساط الممارسين، وتشمل على نطاق واسع المجتمع الأكاديمي والممارسين في المجال الإنمائي والمجتمع المدني (الإعلام من ضمنه) والمشاركين في البرنامج. كل من هذه الجماهير ستكون لديه اهتمامات مختلفة في نتائج التقييم، وسيطلب إستراتيجيات إعلامية مصممة خصيصًا له من أجل إكمال هدف إعلام السياسات والتأثير عليها (الجدول ١٤-١).

الفنيون والمديرون. يتضمن الجمهور الأساسي الأول الفريق الفني والتشغيلي، والمديرين الذين صمموا البرنامج ونفذوه وكذلك الأفراد من المؤسسات (مثل الوزارة أو مؤسسة التمويل) المرتبطين بالمشروع من كُتب. ستكون هذه المجموعة من الأفراد، بشكل أساسي، أول من يرى نتائج تقييم الأثر ويوفر تعليقات على تفسيرات التقييم وتوصياته.

الجدول ١٤-١ إشراك الفئات المستهدفة الرئيسية لمعرفة أثر السياسات: لماذا، ومتى، وكيف

المجموعات المجتمعية المدنية والأكاديمية الإنمائية	صانعو السياسات رفيعو المستوى	العاملون والمديرون في البرنامج	
يحتاجون إلى أدلة بشأن أثر البرامج الإنمائية من أجل اتخاذ قرارات، وتصميم برامج جديدة، وتكرار البرامج الناجحة في بلدان أخرى، وإجراء الأبحاث التي يمكن أن تساعد على تحسين نوعية حياة البشر.	يمكن أن يصبحوا مؤيدين لتقييم الأثر والاستفادة الفعلية من الأدلة.	يمكن أن يصبحوا مؤيدين لتقييم الأثر والاستفادة الفعلية من الأدلة.	لماذا؟
اعتمادًا على البرنامج الجاري تقييمه، يمكن أن تشكل فئات المجتمع المدني والخبراء في مجال التنمية دور المؤيدين المحليين، وهو دور بالغ الأهمية. يجب توزيع المعلومات بمجرد اختتام النتائج وتدقيقها من قبل فريق البرنامج وصانعي السياسات.	في وقت مبكر عند تحديد أسئلة التقييم وقبل بدء التقييم ومرة ثانية عند اختتام النتائج. من الضروري أن يفهم قادة صانعي السياسات ما السبب وراء إجراء تقييم الأثر وكيف يمكن أن تساعد النتائج.	في وقت مبكر، حتى قبل طرح البرنامج ومع التدخلات المستمرة والمتكررة في جميع الأوقات. يمكن استخدام بيانات خط الأساس لعمل تصميم مخصص للتدخل. هم أول من يحق لهم التعليق على نتائج التقييم.	متى؟
الأحداث والمنتديات العامة، بما في ذلك الندوات والمؤتمرات ووثائق العمل ومقالات المجلات العلمية والتغطية الإعلامية والمواد القائمة على الإنترنت، وكل ذلك منافذ للوصول إلى هذه الجماهير.	يجب تقديمهم من خلال ورش العمل الوطنية وإعداد اجتماعات مباشرة مع صانعي السياسات من المستوى الإداري لشرح آلية العمل. شجّع مديري البرنامج والفريق الفني وصانعي السياسات من الفئة المتوسطة على إحاطة الوزارات بكل جديد فيما يخص تقييم الأثر. عندما تتم صياغة الأدلة بشكل نهائي، قدمها إلى قادة صانعي السياسات. وحيثما أمكن، قم بتحليل جدوى التكلفة أو فعالية التكلفة واقتراحات للخطوات التالية.	قم بتعريف دور الأدلة في صنع السياسات من خلال ورشة عمل لإشراك مديري البرنامج في تصميم التقييم. يجب إجراء اجتماعات متابعة في المراحل الأساسية: فورًا بعد جمع بيانات خط الأساس، وبعد جمع النتائج الوسيطة، وفي خط النهاية.	كيف؟

بما أن هذه هي المرة الأولى التي تظهر فيها النتائج عادة إلى النور، فإن توقيت إصدار المعلومات لهذه المؤسسات الأساسية له دور حاسم. من ناحية أخرى، من الضروري مشاركة النتائج بصورة مبكرة، بحيث يستطيع صانعو القرار المعنيون بالبرنامج إدخال أي تغييرات واتخاذ قرارات بخصوص السياسة، مثل تعزيز التدخل (أو تقليل مداه) أو تعديل مكونات البرنامج بهدف تحسين استخدام الموارد وتحقيق أثر أعظم. وفي المقابل، يجب أن نحذر من مشاركة النتائج الأولية استنادًا إلى تحليل جزئي أو غير مكتمل. فقد تكون هذه النتائج خاضعة للتغيير. ومن شأن هذا الإصدار المبكر أن يشكل حدود التوقعات لفريق البرنامج ويسرع باتخاذ قرارات سابقة لأوانها في مجال السياسات والتي يمكن أن يكون التراجع عنها مكلفًا في المستقبل. ولذلك، يجب مراعاة التوازن ما بين التوقيت والاكتمال في النشر المبدئي للنتائج مع فريق المشروع. ويحدث ذلك عادةً عندما يكون فريق التقييم قد أجرى تحليلًا شاملاً وعمليات فحص دقيقة، ولكن قبل صياغة الصورة النهائية للنتائج والتفسيرات والتوصيات.

في الغالب يكون فريق ومديرو البرنامج مهتمين بكل من التفاصيل الفنية لمنهجية التقييم وتحليله، والتفصيلات الدقيقة للاستنتاجات والتوصيات الأولية. قد تكون المناقشات الأولية للناتج مع هذه المجموعة مناسبة جدًا للاجتماعات التي تكون على طراز ورش العمل، مع العروض التقديمية من قبل فريق التقييم، والوقت الوفير للاستفهام وإبداء التعليقات من جميع الأطراف. ومن شأن هذه المناقشات الأولية أن تثير، عادةً، التحليل النهائي، وتدعم تفسير النتائج بالمعلومات، وتساعد على تخصيص التوصيات النهائية بحيث تكون ملائمة على نحو أفضل لتوجيه أهداف سياسة البرنامج. وستكون المناقشات الأولية مع فريق ومديري البرنامج فرصة جيدة لمناقشة النتائج المثيرة للجدل غير المتوقعة أو المحتملة ولاقتراح توصيات السياسة والاستجابة تحسبًا للكشف العام لتقييم الأثر.

ربما تكون النتائج السلبية (بما في ذلك عدم اكتشاف أي أثر) أو النتائج غير المتوقعة مثيرة للإحباط بالنسبة لفريق ومديري البرنامج الذين استثمروا وقتًا وجهدًا ثمينين في هذا البرنامج، ولكنهما قد تشكلان وظيفة أساسية لتعزيز السياسة المقرر إعادة تشكيلها. على سبيل المثال، إذا تبين أن البرنامج قد فشل في تحقيق هدفه الأساسي بسبب تحديات التنفيذ، فإن من الممكن أخذ تدابير للتعامل مع هذه النقاط ومن ثم يمكن إعادة تقييم البرنامج في وقت لاحق. إذا لم يُظهر البرنامج أي آثار على المدى القصير أو أظهر آثارًا في مجموعة فرعية فحسب من سلسلة النتائج، وهناك سبب للاعتقاد بأنه يلزم وقت إضافي لبلوغ الحاصلات النهائية، فمن ثم يمكن تقديم التقييم والدفاع عن النتائج الأولية، ويمكن تخطيط قياسات إضافية في تاريخ مستقبلي. وأخيرًا، إذا كان من الواضح أن التدخل يفشل في الوصول إلى المزايا المقصودة أو كان يسبب ضررًا على غير المتوقع، فهنا إذن يستطيع مديرو البرنامج اتخاذ خطوات فورية لإيقاف التدخل أو إعادة صياغة تصميمه. بهذه الطريقة، عندما تُتاح نتائج التقييم للجمهور، يستطيع صانعو السياسات المسؤولون عن البرنامج الإعلان عن التدابير التصحيحية وتشكيل الاستجابة بشكل مسبق تحسبًا لأي أسئلة صعبة قد يتم طرحها في المجالس الحوارية للسياسات أو في الإعلام.

كبار صانعي السياسات. الجهة المستهدفة الأساسية الثانية هي كبار صانعي السياسات الذين سيتخذون قرارات متعلقة بالسياسات استنادًا إلى نتائج تقييم الأثر، مثل ما إذا كان سيتم توسيع نطاق التمويل أو الحفاظ عليه كما هو أو تقليله بالنسبة للتدخل المعني. وقد تشمل هذه الجهة المشرعين الوطنيين، أو الرؤساء ورؤساء الوزارة، أو الوزراء والأمناء الأساسيين، أو مجلس الإدارة، أو المانحين. وتحصل هذه الفئة من الأطراف المعنية على نتائج التقييم بشكل أساسي بمجرد إعداد الصياغة النهائية لها ومراجعتها من قبل فريق ومديري البرنامج وتدقيقها من جانب خبراء فنيين خارجيين. وفي هذه المرحلة، سيحتاج فريق التقييم إلى التركيز في توصيل النتائج والتوصيات الأساسية بطريقة ميسورة التناول، حيث ربما تكون التفاصيل الفنية للتقييم ذات أهمية ثانوية. سيهتم كبار صانعي السياسات بترجمة الآثار إلى قيم ذات دلالة اقتصادية من خلال تحليل جدوى التكلفة أو المقارنة بالتدخلات البديلة من خلال تحليل الفعالية من حيث التكلفة.

وستساعد هذه المعلومات على تزويد صانعي السياسات بالمعلومات لبيان ما إذا كان البرنامج ذا جدوى وجدريًا باستحقاق الموارد المحدودة المستثمرة فيه من أجل الاستمرار في تحقيق هدف إنمائي مهم. قد يكون كبار صانعي السياسات مهتمين أيضًا باستخدام النتائج للمضي قدمًا بأجندتهم متعلقة بالسياسات، مثل ممارسة الضغط تأييدًا (أو دحضًا) لسياسة عامة معينة يدعمها (أو لا يدعمها) التقييم. بإمكان فريق التقييم التعاون مع خبراء الإعلام لضمان أن النتائج والتوصيات ذات الصلة مفسرة بشكل صحيح وأن الرسائل في إستراتيجية الإعلام لا تزال متسقة مع استنتاجات التقييم. *أوساط الممارسين*. الجهة المستهدفة الأساسية الثالثة لتحقيق أثر السياسة على نطاق واسع تتضمن مستخدمي التقييم خارج النطاق المباشر للبرنامج أو السياق القطري. وهذه الفئة غير المتجانسة تضم أوساط الممارسين في قطاعات وثيقة الصلة بالتقييم، بما في ذلك الممارسون في مجال التنمية، والأكاديميون، والمجتمع المدني، وصانعو السياسات في دول أخرى. ربما يكون الممارسون في مجال التنمية خارج نطاق البرنامج المعني، مهتمين باستخدام نتائج التقييم لوضع تصميم مستدير لبرامج جديدة أو حالية. وسيكون هؤلاء الممارسون مهتمين بتفاصيل التقييم (الطرق والنتائج والتوصيات) وبالدروس والتوصيات التشغيلية التي يمكن أن تساعد في تنفيذ مشاريعهم الخاصة بنجاعة أعلى. من ناحية أخرى، قد يبدي الوسط الأكاديمي اهتمامًا أكبر بالتقييم من حيث المنهجية والبيانات والاستنتاجات التجريبية.

وضمن المجتمع المدني على نطاق واسع، تنفرد جهتان مستهدفتان أساسيتان: الإعلام والمشاركون في البرنامج. إن إحاطة الجمهور العام علمًا بنتائج التقييم من خلال الإعلام يمكن أن يلعب دورًا أساسيًا في تحقيق المساءلة بخصوص الإنفاق العام، وتوطيد دعم العامة لتوصيات التقييم، والمحافظة على السياسات الفعالة. وينطبق ذلك بشكل خاص على السياسات الجديدة والمبتكرة حيث تكون الحصلة في البداية غير يقينية أو خاضعة للجدل في مجالس حوار السياسات. إذا ألقى التقييم ضوءًا تجريبيًا على ما كان يُعد نقطة جدال نظرية أو أيديولوجية بشكل عام حتى تاريخه، فقد يُنظر إلى هذا التقييم على أنه أداة قوية لتحريك السياسات.

وأخيرًا، يجب تضمين المشاركين في البرنامج في جهود نشر نتائج التقييم. لقد استثمر هؤلاء المشاركون وقتهم وطاقاتهم في البرنامج، وربما قضا وقتًا كبيرًا في توفير المعلومات لغرض هذا التقييم. تأكد من أن المشاركين في البرنامج قادرين على الوصول إلى نتائج التقييم وعلى اطلاع بكل جديد فيما يخص هذه النتائج وفي ذلك لفئة بسيطة ولكنها كبيرة الأثر ويمكن أن تسهم في الحفاظ على اهتمامهم بالبرنامج ورغبتهم في المشاركة في أي تقييمات مستقبلية.

الأمر التالي المطروح للنقاش هو الإستراتيجيات المختلفة التي يمكن وضعها في الاعتبار لإعلام هذه الجهات المستهدفة الأساسية ولتحقيق أثر السياسة. في الأوضاع المثالية، ستتضمن المراحل المبكرة من تخطيط التقييم إستراتيجية للنشر أو أثر السياسة. ويتعين الاتفاق على هذه الإستراتيجية من البداية وفيها يتم تحديد بشكل واضح هدف سياسة التقييم (مثلاً، توسعة نموذج تدخل ذي فعالية أكبر من حيث التكلفة)، والجمهور الأساسية التي يعتزم التقييم بلوغها، وإستراتيجيات الإعلام المقرر استخدامها، وميزانية لإجراء أنشطة التوزيع. تختلف الصيغة والمحتوى لأنشطة ومنتجات النشر على أساس كل حالة على حدة، ولكننا نوفر بعض النصائح والإرشادات العامة فيما تبقى من هذا الفصل. يسرد المربع ٢-١٤ بعض أدوات التوعية والنشر.

التقارير هي في الأساس المنفذ الأول لتقديم المجموعة الكاملة لنتائج التقييم. ونوصي بالحفاظ على طول متوسط لهذه التقارير، في نطاق من ٣٠ إلى ٥٠ صفحة، بما في ذلك الخلاصة في صفحة واحدة أو أقل، وملخص تنفيذي إلى جانب النتائج والتوصيات الأساسية المكتوبة على صفتين إلى أربع صفحات. أما التفاصيل الفنية والتوثيقات ذات الصلة والتحليلات الداعمة، مثل اختبارات الدقة والدحض، فيمكن تقديمها في ملاحق أو مرفقات.

المربع ٢-١٤: أدوات التوعية والنشر

فيما يلي بعض الأمثلة لنماذج توزيع تقييمات الأثر:

- شريحة عرض تقديمي تُظهر نتائج البرنامج والتقييم.
- مقاطع فيديو تستعرض المستفيدين وهم يعبرون عن رأيهم حول البرنامج ويوضحون كيف أثر على حياتهم.
- مذكرات سياسات موجزة تشرح التقييم وتلخص التوصيات المتعلقة بالسياسات. مدونات ينشئها الباحثون وصانعو السياسات يوضحون من خلالها أهمية التقييم
- تقارير كاملة بعد التوصل إلى النتائج النهائية مع ملخصات تنفيذية لضمان سرعة إلمام القارئ بالاستنتاجات الرئيسية
- دعوات إعلامية تسمح للصحافة برؤية البرنامج على أرض الواقع وإعداد تقارير بالنتائج.

يتطلب الأمر جهودًا كبيرة لنشر تقييم الأثر في صورة بحث أكاديمي أو مقالة في مجلة علمية محكمة، ولكنها خطوة نهائية جديرة بالتعب لاختتام نتائج التقييم. إن مراجعة النظراء الصارمة مطلوبة لعملية النشر، وستوفر ملاحظات عالية القيمة من أجل تحسين التحليلات وتفسير النتائج، وقد يمثل النشر إشارة قوية لصانعي السياسات إلى مدى جودة نتائج التقييم ومصادقيتها.

استنادًا إلى إستراتيجية النشر المتفق عليها، يمكن نشر التقارير والأوراق البحثية من خلال منافذ عدة، وهي تشمل على الموقع الإلكتروني الخاص بالبرنامج، ومن خلال الموقع الإلكتروني الخاص بمؤسسة التقييم، وكجزء من سلسلة أبحاث ومجلات علمية أكاديمية محكمة والكتب.

رغم أن التقارير والأبحاث الأكاديمية للتقييم تشكل أساسًا لإستراتيجية النشر والتوزيع، فإن مدى وصولها إلى الجمهور الأوسع خارج أوساط الممارسين والأكاديميين قد يكون محدودًا بسبب طولها وصيغتها الفنية. هناك خطوة مهمة ربما يستطيع فريق التقييم القيام بها بالتعاون مع خبراء الإعلام وهي إنتاج مقالات قصيرة مكتوبة في سرد قصصي أو صحفي بلغة بسيطة وواضحة من أجل نشرها على قاعدة جماهيرية أوسع. ويمكن نشر المقالات القصيرة في صورة موجزات سياسات ونشرات إخبارية ونشرات ورسوم بيانية. بالنسبة لهذه المنشورات، سيكون من المفيد بشكل خاص إزالة أي مصطلحات فنية معقدة وترجمة النتائج في أشكال تمثيلية جذابة مرئيًا، أي على شاكلة الصور والمخططات والرسوم البيانية (المربع ٣-١٤).

المربع ٣-١٤: نشر تقييمات الأثر بفعالية

تُظهر المنشورات المختلفة نتائج تقييمات الأثر بصيغة سهلة التداول والفهم. وهي تتضمن تحديثين مع تركيز إقليمي:

- يتم تقديم نتائج تقييم الأثر من البرامج المنفذة في جميع أنحاء أمريكا اللاتينية ومنطقة الكاريبي في Development Effectiveness Overview، وهي تقارير سنوية ينشرها مكتب التخطيط الإستراتيجي والفعالية الإنمائية التابع لبنك التنمية للبلدان الأمريكية (IDB). وفي هذه التقارير، يتم تلخيص النتائج في صورة مقالات موجزة سهلة القراءة تتضمن ملخصات رسومية من صفحة واحدة وتوضح سؤال تقييم الأثر الأساسي والطرق والنتائج وتوصيات السياسة باستخدام الأشكال ورسومات الجرافيك والأيقونات التي تسمح للقراء باستيعاب الرسائل الأساسية بصورة سهلة وسريعة للغاية. يتضمن Development Effectiveness Overview لعام ٢٠١٤ نتائج من تقييمات أثر برامج متنوعة في المجال: السياحة في الأرجنتين، والتدريب المهني في جمهورية الدومينيكان، والإنتاجية الزراعية في بوليفيا، وشباب الأوركسترا في بيرو.
- يجمع Africa Impact Evaluation Update الذي يصدره البنك الدولي أحدث الأدلة من المنطقة. وكان يركز على النوع الاجتماعي في ٢٠١٣ والزراعة والأراضي في ٢٠١٤.

المصدر: <http://deo.iadb.org> and <http://www.worldbank.org>

يستطيع فرق التقييم إنشاء مجموعة من العروض التقديمية التي ترافق التقارير والمقالات القصيرة المكتوبة. ويجب تخصيص هذه العروض التقديمية حسب الجمهور المستهدف. كخطوة أولية جيدة، من الممكن إنتاج عرض تقديمي فني لفريق المشروع والجمهور الأكاديمية، وعرض تقديمي آخر أقصر طولاً وذي نزعة فنية أقل لصانعي السياسات والمجتمع المدني. رغم أن الاستنتاجات وتوصيات السياسة الأساسية ستكون واحدة، فإن التنظيم والمحتوى لهذين النوعين من العروض التقديمية سيختلفان في الأهمية. ينبغي للعرض التقديمي الفني أن يركز على بناء المصداقية للنتائج من خلال تقديم طرق التقييم والبيانات والتحليلات قبل التطرق إلى النتائج والتوصيات. أما العرض التقديمي الذي يستهدف صانعي السياسات ينبغي له التأكيد على مشكلة التنمية التي يعتزم التدخل معالجتها والتداعيات العملية للاستنتاجات ولكن مع التخفيف من التفاصيل الفنية.

للاستفادة من الوصول الموسع للإنترنت في الدول النامية والبدائل منخفضة التكلفة لإنتاج وسائط متعددة، تستطيع فرق التقييم أيضاً النظر في مجموعة متنوعة من الوسائط لنشر وتوزيع استنتاجات التقييم، بدءاً من المواقع الإلكترونية إلى الأعمال الصوتية والمرئية. قد تكون مقاطع الفيديو القصيرة طريقة قوية لتحويل الأفكار المعقدة إلى عناصر صوتية ومرئية بسيطة، ما يسمح بالكشف عن قصة التقييم بطريقة أسرع تضمن فهمًا أشمل بالمقارنة بوسائل الإعلام المطبوعة الاعتيادية (المربع ١٤-٤).

المربع ١٤-٤: نشر تقييمات الأثر على الإنترنت

نقدم هنا بعض الأمثلة الجيدة بالذكر لتوزيع نتائج تقييم الأثر عبر الإنترنت:

- تنظم مؤسسة المبادرة الدولية لتقييم الأثر (Zie) الأدلة المستمدة من تقييمات الأثر حسب القطاع، بما يتضمن ذلك مذكرات السياسات والمراجعات المنهجية وخرائط فجوات الأدلة.
- يقوم مختبر عبد اللطيف جميل لمكافحة الفقر (J-PAL) بنشر الأدلة المستمدة من تقييمات الأثر التي يجريها الباحثون التابعون، بما في ذلك مذكرات السياسات وتحليل جدوى التكلفة وروابط إلى الأبحاث الأكاديمية.
- تقدم مجموعة Development Impact Evaluation (DIME) التابعة للبنك الدولي مذكرات ونشرات إخبارية وتقارير تسلط الضوء على النتائج المكتسبة من تقييمات الأثر لمشاريع البنك الدولي.
- ويتضمن Strategic Impact Evaluation Fund (SIEF) التابع للبنك الدولي مقاطع فيديو ومذكرات ومقابلات شخصية.

المربع ١٤-٥: مدونات تقييم الأثر

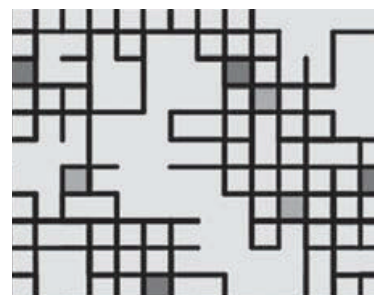
فيما يلي بعض الأمثلة على مدونات تقدم بصفة منتظمة نتائج تقييمات الأثر:

- World Bank Development Impact Blog
- Inter-American Bank Development Effectiveness Blog
- Innovations for Poverty Action Blog

وأخيرًا، عند التسلح بمختلف منتجات النشر والتوزيع، يجب على فريق التقييم أن يتخذ خطوات سابقة بشأن نشر هذه المنتجات إلى المستهلكين ضمن البرنامج والحكومة وأوساط الممارسين بشكل أوسع بحيث تصل إلى المستخدمين المستهدفين ويتم إدماجها في عملية صنع القرار وحوار السياسات. تحدث عملية النشر والتوزيع من خلال الاجتماعات الشخصية بين فريق التقييم ومدير البرنامج، وممارسة الضغط مع كبار صانعي السياسات، والعروض التقديمية في الندوات والمؤتمرات حيث يتعلم الأكاديميون والأشخاص المنتمون إلى أوساط الممارسين عن أحدث التطورات في البحوث والتقييمات الإنمائية، والمقابلات الشخصية، والبرامج الإخبارية على الراديو والتلفزيون، وعبر الإنترنت بصورة متزايدة. يمكن أن تكون المدونات والوسائط الاجتماعية بشكل خاص من الطرق الفعالة من حيث التكلفة التي تمكّن الوصول إلى أعداد أكبر من المستخدمين المحتملين وكسب الجمهور وتوجيه القراء صوب مجموعة متنوعة من المنتجات المتوفرة من تقييم معين (المربع ١٤-٥). في حين أن الإستراتيجيات الخاصة ستتفاوت على أساس كل حالة على حدة، فإننا نكرر توصيتنا بالتخطيط ووضع الميزانية لمنافذ وأنشطة النشر والتوزيع في مرحلة مبكرة، بحيث يمكن أن تصل نتائج التقييم إلى جماهيرها المقصودة بطريقة سريعة وفعالة وبالتالي يتم تعظيم أثر السياسة.

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة الموقع الإلكتروني "تقييم الأثر في الواقع العملي" (<http://www.worldbank.org/ieinpractice>).
- طورت مؤسسة المبادرة الدولية لتقييم الأثر (3ie) ومعهد التنمية لما وراء البحار (ODI) مجموعة أدوات أثر السياسة المتوفرة عبر الإنترنت بهدف مساعدة الباحثين على نشر واستخدام الأدلة المكتسبة من تقييمات الأثر من أجل صنع القرار.



الجزء ٤

كيفية الحصول على البيانات لإجراء تقييم الأثر

يوفر الجزء الرابع من الكتاب توجيهات حول كيفية الحصول على البيانات لإجراء تقييم أثر، بما يتضمن ذلك اختيار العينة والعثور على مصادر ملائمة من البيانات.

كذلك يناقش الفصل ١٥ كيفية سحب العينة من المجتمع الإحصائي محل الاهتمام، وكيفية إجراء حسابات للقوة الإحصائية من أجل تحديد الحجم المناسب من عينة تقييم الأثر. ويركز هذا الفصل على وصف المنطق البديهي الرئيسي وراء أخذ العينات وحسابات القوة الإحصائية. ويسلط الضوء أيضًا على العناصر التي يحتاج فريق السياسات إلى توفيرها إلى فريق الأبحاث أو الخبير الفني المسؤول عن تنفيذ أخذ العينات وحسابات القوة الإحصائية.

أما الفصل ١٦ فهو يتناول مختلف مصادر البيانات التي يمكن استخدامها مع تقييمات الأثر. بالإضافة إلى أنه يلقي بالضوء على الحالات التي يمكن فيها استخدام مصادر البيانات الحالية، بما يشمل ذلك البيانات الإدارية. وحيث إن العديد من التقييمات تتطلب جمع بيانات جديدة، فإن هذا الفصل يناقش خطوات جمع بيانات المسح الجديدة: تحديد الشخص المسؤول عن جمع البيانات، ووضع أدوات جمع البيانات الأولية، وإجراء العمل الميداني وضمان الجودة، ومعالجة البيانات وتخزينها.

وأخيرًا، يوفر الفصل ١٧ خلاصة للكتاب برمته. ويراجع باختصار العناصر الأساسية لتقييم الأثر جيد التصميم وكذلك بعض النصائح المتعلقة بتخفيف المخاطر الشائعة المصاحبة لتقييم الأثر. ويطرح أيضًا منظورات معينة حول النمو الذي تحقق حديثًا في استخدام تقييم الأثر والجهود المؤسسية ذات الصلة.

اختيار العينة

أخذ العينات وحسابات القوة الإحصائية

بمجرد اختيارك للطريقة التي ستحدد بها مجموعة مقارنة وتقدر بها الواقع المضاد، هناك خطوات تالية في سبيل إجراء تقييم أثر، وتلك تشمل تحديد المطلوب من البيانات والعينة اللازمين لتقدير الفرق في الحاصل بشكل دقيق بين مجموعتي المعالجة والمقارنة. وناقش في هذا الفصل كيف يمكنك سحب عينة من المجتمع الإحصائي محل الاهتمام (أخذ العينات)، وكيف يمكنك تحديد حجم العينة المطلوب لتوفير تقديرات دقيقة لأثر البرنامج (حسابات القوة الإحصائية). يتطلب كل من أخذ العينات وحسابات القوة الإحصائية مهارات فنية محددة وغالبًا ما يتولى تنفيذها خبير متخصص. وكذلك نسرد في هذا الفصل أساسيات إجراء أخذ العينات وحسابات القوة الإحصائية، ونسلط الضوء على العناصر التي يحتاج فريق السياسات إليها من أجل تزويد الخبراء الفنيين بها.

سحب العينة

أخذ العينات هو عملية سحب وحدات من المجتمع الإحصائي محل الاهتمام من أجل تقدير خصائص هذا المجتمع الإحصائي. وغالبًا ما يكون أخذ العينات ضروريًا حيث إنه يتعذر في الأساس ملاحظة الحاصل وقياسها بشكل مباشر للمجتمع الإحصائي محل الاهتمام بالكامل. فمثلًا، إذا كنت مهتمًا بمعرفة متوسط طول الأطفال دون سن العامين في دولة ما، فسيكون الأمر غاية في الصعوبة وباهظ التكلفة ومضيعة للوقت إذا كنا سنزور جميع الأطفال ونقيس أطوالهم في هذا المجتمع الإحصائي. بدلاً من ذلك، يتم سحب عينة من الأطفال الذين ينتمون إلى هذا المجتمع الإحصائي كي يتم استخدامها للاستدلال على متوسط خصائص المجتمع الإحصائي (الشكل ١٥-١).

يجب إيلاء اهتمام خاص لهذه العملية التي تنطوي على سحب عينة من المجتمع الإحصائي المعني. توفر مبادئ أخذ العينات توجيهات حول سحب العينات التمثيلية. عملياً، توجد ثلاث خطوات رئيسية لسحب العينة:

١. تحديد المجتمع الإحصائي محل الاهتمام.

٢. تعريف إطار تحديد العينة.

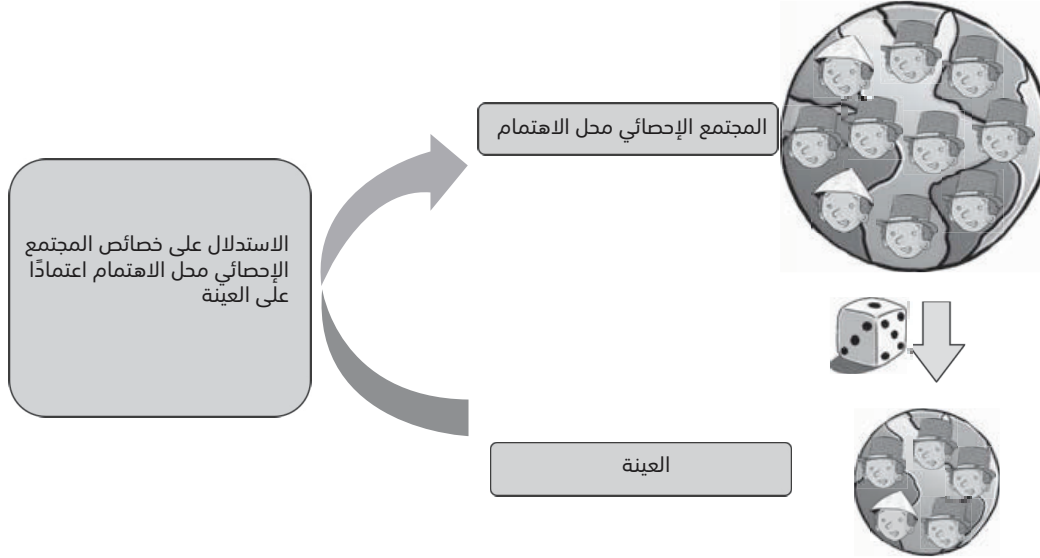
٣. سحب أكبر قدر من الوحدات من إطار تحديد العينة حسبما يكون مطلوباً عن طريق حسابات القوة الإحصائية. أولاً، يتعين تعريف المجتمع الإحصائي محل الاهتمام بوضوح. ويتطلب ذلك تحري الدقة عند تحديد الوحدة ضمن المجتمع الإحصائي محل الاهتمام التي سنعتمد الحاصل من أجلها، وتحديد التغطية الجغرافية بوضوح أو أي سمات أخرى ذات صلة تميز هذا المجتمع الإحصائي. فمثلاً، إذا كنت تدير برنامجاً لتنمية الطفولة المبكرة، قد تكون مهتماً بقياس أثر البرنامج على الحاصل المعرفية للأطفال الصغار بين العمرين: ثلاث وست سنوات في الدولة بأكملها، أو للأطفال في المناطق الريفية فقط أو أولئك الملتحقين بمرحلة رياض الأطفال.

ثانياً، بمجرد تعريف المجتمع الإحصائي محل الاهتمام، يجب وضع إطار تحديد العينة. إن إطار

المفهوم الأساسي

إطار تحديد العينة هو
أشمل حصر للوحدات
المأخوذة من المجتمع
الإحصائي محل الاهتمام.
ويحدث تحيز في التغطية
إذا كان إطار تحديد العينة
لا يتداخل بشكل مثالي مع
المجتمع الإحصائي محل
الاهتمام.

الشكل ١٥-١ استخدام عينة للاستدلال على متوسط خصائص المجتمع الإحصائي محل الاهتمام

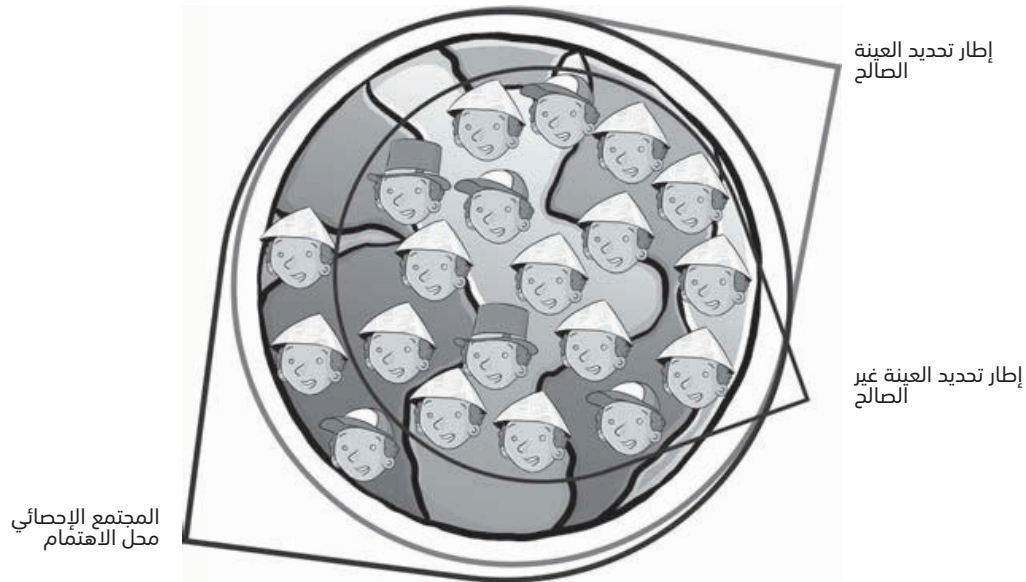


تحديد العينة هو أشمل حصر للوحدات المأخوذة من المجتمع الإحصائي محل الاهتمام. في العموم، يجب أن يتوافق إطار تحديد العينة مع المجتمع الإحصائي محل الاهتمام على أتم وجه. ومن شأن إجراء تعداد كامل ومحدّث تمامًا للمجتمع الإحصائي محل الاهتمام أن يشكل إطارًا مثاليًا لتحديد العينة. ففي الواقع العملي، غالبًا ما تُستخدم القوائم الفعلية (مثل التعدادات السكانية أو تعدادات المنشآت أو قوائم التسجيل) كإطارات تحديد عينة.

ويلزم وضع إطار تحديد عينة مناسب لضمان أن الخلاصات الناتجة عن تحليل عينة ما يمكن تعميمها على المجتمع الإحصائي بالكامل. بلا شك، عندما لا يتوافق إطار تحديد العينة بشكل تام مع المجتمع الإحصائي محل الاهتمام، يحدث تحيز في التغطية، كما هو موضح في الشكل ١٥-٢. إن وقع تحيز في التغطية، فإن النتائج التي نحصل عليها من العينة لا تكون لها صلاحية خارجية للمجتمع الإحصائي محل الاهتمام بالكامل، ولكن للمجتمع الإحصائي المتضمن في إطار تحديد العينة فقط. إذا أردنا تعميم الإحصائيات المحسوبة استنادًا إلى العينة، فإن حجم التحيز في التغطية (أي فقدان التداخل بين إطار تحديد العينة والمجتمع الإحصائي محل الاهتمام) هو ما يحدد إلى أي مدى يمكن تطبيق هذا التعميم على المجتمع الإحصائي محل الاهتمام ككل.

ترتبط التحيزات في التغطية بمخاطر، ولهذا تشكيل أطر العينة يتطلب بذل عناية فائقة. فعلى سبيل المثال، قد تحتوي بيانات التعداد على قائمة بجميع الوحدات الموجودة في مجتمع إحصائي معين. ومع ذلك، إذا كان قد مضى الكثير من الوقت بين التعداد وجمع بيانات العينة، قد لا يكون إطار أخذ العينات محدثًا بشكل كامل. علاوة على ذلك، قد لا تحتوي بيانات التعداد على معلومات كافية حول سمات محددة لصياغة إطار تحديد عينة. إذا كان المجتمع الإحصائي محل الاهتمام يتألف من أطفال ملتحقين بمرحلة رياض الأطفال، ولا يتضمن التعداد بيانات حول الالتحاق بمثل هذه المؤسسات، فقد يتطلب الأمر بيانات تكميلية حول الالتحاق أو قوائم بالمنشآت.

الشكل ١٥-٢ إطار تحديد عينة صالح يغطي المجتمع الإحصائي محل الاهتمام بالكامل



المفهوم الأساسي

أخذ العينات هي عملية يتم فيها سحب وحدات من إطار تحديد العينة. عند استخدام طريقة أخذ العينات الاحتمالية، فإنه يتم تخصيص احتمال معرف جيداً لكل وحدة مطلوب سحبها.

وبمجرد تعريف المجتمع الإحصائي محل الاهتمام وإطار تحديد العينة، يجب عليك اختيار طريقة لسحب العينة. يمكن استخدام العديد من الإجراءات التبادلية.

وتُعد طرق أخذ العينات الاحتمالية هي الأكثر صرامة ودقة، حيث إنها تخصص احتمالاً معرفاً جيداً لكل وحدة مطلوب سحبها. ونذكر الطرق الرئيسية الثلاث لأخذ العينات الاحتمالية فيما يلي:

• **أخذ العينات العشوائية.** كل وحدة في المجتمع الإحصائي تتشارك في احتمال تضمينها في العينة المسحوبة.^١

• **أخذ العينات العشوائية الطبقية.** يتم تقسيم المجتمع الإحصائي إلى فئات (مثل تقسيمه

إلى ذكور وإناث)، وتؤخذ العينات العشوائية ضمن كل فئة. نتيجة لذلك، كل وحدة في كل فئة (أو طبقة) تتشارك في احتمال تضمينها في العينة المسحوبة. بشرط أن تكون كل فئة كبيرة بما يكفي، تتيح طريقة أخذ العينات الطبقية إمكانية اكتساب استدلالات حول الحاصل، ليس على مستوى المجتمع الإحصائي فحسب بل أيضاً ضمن كل فئة. وتُعد هذه الطريقة مفيدة عندما تود زيادة حجم العينة للفئات الفرعية في المجتمع الإحصائي، التي تكون صغيرة العدد (مثل الأقليات) من أجل دراستها بمزيد من العناية. الطبقية شرط أساسي للتقييمات التي تستهدف مقارنة آثار البرامج بين كل من الفئات الفرعية.

• **أخذ العينات العنقودية.** تنقسم الوحدات إلى مجموعات عنقودية، ويتم سحب عينة عشوائية من المجموعات العنقودية. وبالتالي، إما كل الوحدات في هذه المجموعات العنقودية تشكل العينة أو يتم سحب عينة عشوائية من عدد من الوحدات ضمن المجموعة العنقودية. وهذا يعني أن كل مجموعة عنقودية تتمتع باحتمال محدد بدقة أن يتم اختيارها، ويكون لدى الوحدات ضمن المجموعة العنقودية المختارة أيضاً احتمال محدد بدقة أن يتم سحبها. في سياق تقييم الأثر، يتحدد إجراء سحب العينة، غالباً، بشروط أهلية البرنامج الخاضع للتقييم. كما سنوضح في مناقشتنا لمسألة حجم العينة، إذا كانت أصغر وحدة صالحة قابلة للتطبيق أكبر من وحدة الملاحظة، فإن التعيين العشوائي للفوائد ستنشأ عنه مجموعات عنقودية. ولهذا السبب، غالباً ما تُستخدم طريقة أخذ العينات العنقودية في دراسات تقييم الأثر.

أخذ العينات غير الاحتمالية يمكن أن تتسبب في إحداث أخطاء فادحة في أخذ العينات. على سبيل المثال، افترض أن مسجلاً وطنياً يُجرى عن طريق طلب من مجموعة من المداورين جمع بيانات عن الأسر المعيشية في المساكن الأقرب إلى المدرسة في كل قرية. في حال استخدام طريقة أخذ العينات غير الاحتمالية، فعلى الأرجح ألا تكون العينة ممثلة للمجتمع الإحصائي محل الاهتمام ككل. وبالأخص، سينشأ تحيز في التغطية، حيث إن المساكن البعيدة لن يشملها المسح.

ومن الضروري أن تولي اهتماماً متأنياً إلى إطار تحديد العينة وطريقة أخذ العينات لتحديد ما إذا كانت النتائج التي حصلنا عليها من عينة محددة يمكن تعميمها على المجتمع الإحصائي محل الاهتمام ككل.

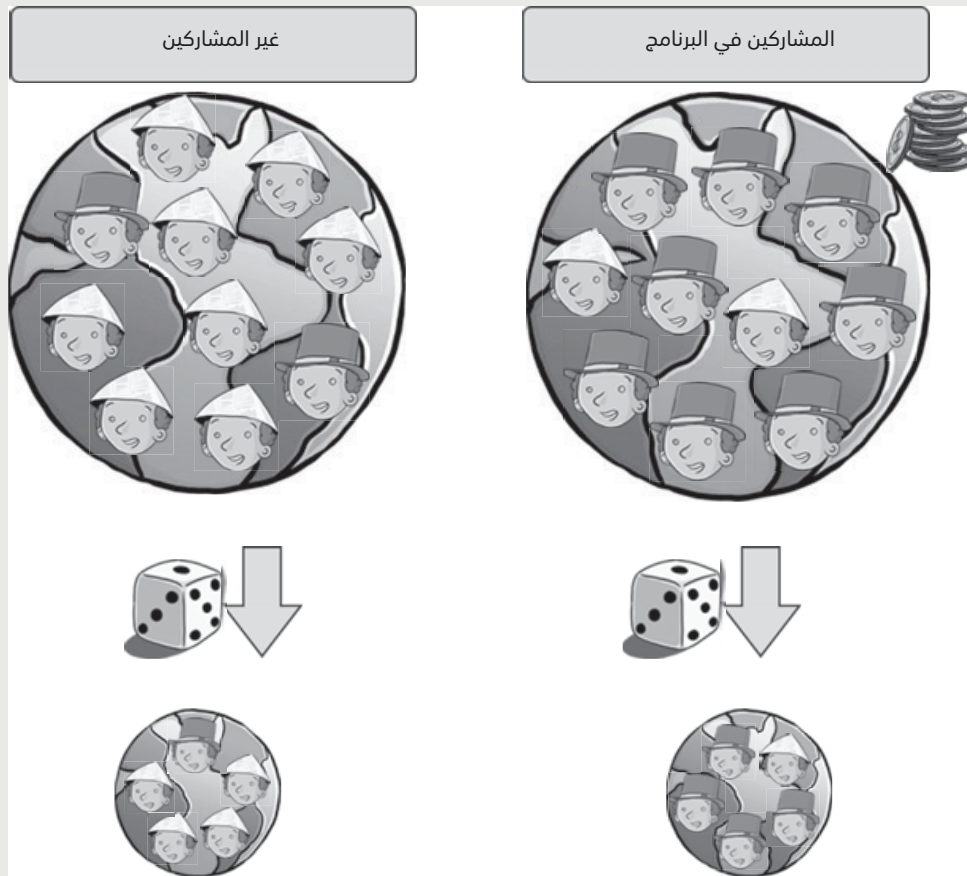
حتى إذا كان إطار تحديد العينة يتمتع بتغطية مثالية وتم استخدام طريقة أخذ العينات الاحتمالية، فإن الأخطاء غير المرتبطة بأخذ العينات يمكن أن تؤثر في الصلاحية الداخلية والخارجية لتقييم الأثر. ويتم تناول الأخطاء غير المرتبطة بأخذ العينات في الفصل ١٦. وأخيرًا، يوجد خلط في بعض الأحيان بين أخذ العينات العشوائية والتعيين العشوائي. ويوضح المربع ١٥-١ أن أخذ العينات العشوائية يختلف تمامًا عن التعيين العشوائي.

المربع ١٥-١: أخذ العينات العشوائية ليس كافيًا لإجراء تقييم الأثر

من الأفراد غير مشاركة في هذا البرنامج. ماذا إن كان من المقرر أن تأخذ عينة عشوائية من كل من هاتين المجموعتين؟ يوضح الشكل الأول أنك ستحصل على عينة عشوائية من المشاركين وعينة عشوائية من غير المشاركين.

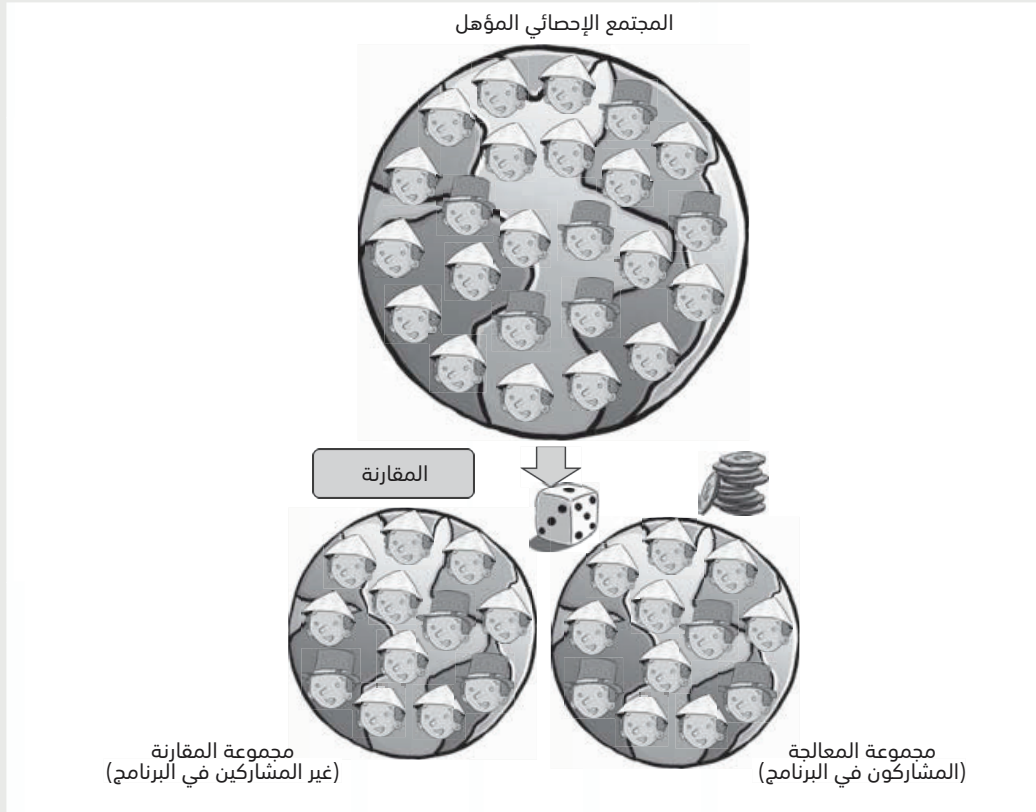
الخلط يحدث أحيانًا بين أخذ العينات العشوائية والتعيين العشوائي. ماذا إذا أخبرك شخص ما بكل فخر أنه ينفذ تصميم تقييم الأثر عن طريق محاورة عينة عشوائية من المشاركين وغير المشاركين؟ افترض أنك تلاحظ مجموعة من الأفراد المشاركين في برنامج توظيف، ومجموعة

الشكل ب ١٥-١-١ أخذ العينات العشوائية فيما بين المجموعات غير القابلة للمقارنة للمشاركين وغير المشاركين



(تابع)

المربع ١٥-١: أخذ العينات العشوائية ليس كافياً لإجراء تقييم الأثر (تابع)
الشكل ب ١٥-١-٢ التعيين العشوائي لفوائد البرنامج بين مجموعة معالجة ومجموعة مقارنة



الوحدات (عادةً تتألف من أشخاص أو مجموعات من الأشخاص، مثل الأطفال في مدرسة ما) من المجتمع الإحصائي المؤهل إلى مجموعة معالجة ستتلقى التدخل ومجموعة مقارنة لن يشملها التدخل. عملية التوزيع العشوائي للبرنامج في الشكل الثاني يختلف عن عملية أخذ العينات العشوائية المذكورة في الشكل الأول. كما ناقشنا في الجزء ٢، عند تنفيذ تعيين عشوائي بشكل جيد، فإنه يساهم في الصلاحية الداخلية لتقييم الأثر. ويمكن الانتفاع من أخذ العينات العشوائية لضمان الصلاحية الخارجية إلى حد أن تكون العينة مسحوبة عشوائياً من المجتمع الإحصائي محل الاهتمام.

إن اختلفت خصائص كل من المشاركين وغير المشاركين، فستكون مختلفة كذلك العينة المأخوذة من المشاركين وغير المشاركين. إن أخذ العينات العشوائية لا يجعل المجموعتين غير القابلتين للمقارنة مجموعتين قابلتين للمقارنة، وبالتالي لا يوفر صلاحية داخلية لتقييم الأثر. ولهذا السبب، لا يُعد أخذ العينات العشوائية كافياً لإجراء تقييم الأثر.

كما يجب أن يتضح من المناقشة في الجزء ٢ أن التعيين العشوائي لفوائد البرامج يختلف عن أخذ العينات العشوائية. تبدأ عملية التعيين العشوائي من مجتمع إحصائي محل اهتمام مؤهل، وتستخدم طريقة عشوائية لتعيين

فيما تبقى من هذا الفصل، نناقش كيف يهتم حجم العينة بالنسبة لدقة تقييم الأثر. كما سيتضح لاحقاً فإن العينات الأكبر حجمًا نسبيًا تكون مطلوبة للحصول على تقديرات دقيقة لخصائص المجتمع الإحصائي. وتكون العينات الأكبر مطلوبة أيضًا لتمكين الحصول على تقديرات أكثر دقة للفروق بين مجموعات المعالجة ومجموعات المقارنة، أي لتقدير أثر البرنامج.

التحديد القائم على حجم العينة لتقييم الأثر: حسابات القوة الإحصائية

كما ناقشنا، يصف أخذ العينات عملية سحب عينة من وحدات المجتمع الإحصائي محل الاهتمام من أجل تقدير خصائص هذا المجتمع الإحصائي. وتمنح العينات الكبيرة دقة أكبر لتقديرات خصائص المجتمع الإحصائي. ما الحجم بالضبط المطلوب لإجراء تقييم أثر؟ إن الحسابات التي تحدد الحجم المطلوب للعينة تُسمى حسابات القوة الإحصائية. ونناقش المنطق البديهي الأساسي وراء حسابات القوة الإحصائية عن طريق التركيز على أبسط حالة: تقييم يتم إجراؤه باستخدام طريقة التعيين العشوائي واختبار فعالية البرنامج مقابل مجموعة مقارنة لا يشملها التدخل، وافترض أن عدم الامتثال ليس مشكلة. ٢ ونناقش باختصار الاعتبارات الإضافية وراء هذه الحالة البسيطة في نهاية الفصل.

الأساس المنطقي لحسابات القوة الإحصائية

حسابات القوة الإحصائية تشير إلى الحد الأدنى لحجم العينة المطلوب لإجراء تقييم أثر والإجابة عن سؤال السياسة محل الاهتمام بشكل مقنع. وعلى وجه الخصوص، يمكن استخدام حسابات القوة الإحصائية من أجل

- تقييم ما إذا كانت مجموعات البيانات كبيرة بما يكفي لإجراء تقييم أثر.
- تجنب جمع كم ضئيل من البيانات. إذا كانت العينة صغيرة للغاية، فربما لا تستطيع الكشف عن الأثر الإيجابي - حتى إن وُجد - وبالتالي قد تخلص إلى أن البرنامج ليس له مفعول. ومن شأن ذلك أن يؤدي إلى قرار على مستوى السياسة بإنهاء البرنامج، وهذا أمر ليس بجيد.
- المساعدة في صنع القرارات بشأن حجم العينة المناسب. توفر الأحجام الأكبر من العينات تقديرات تتسم بدقة أكبر لآثار البرنامج، ولكن جمع المعلومات قد يكون مكلفًا للغاية. وتوفر حسابات القوة الإحصائية مدخلات أساسية من أجل تقييم المفاضلات بين التكاليف المطلوبة لجمع بيانات والمكاسب الإضافية التي سنحصلها من زيادة دقة تقييم الأثر.

المفهوم الأساسي

توفر حسابات القوة الإحصائية مؤشرًا لأصغر عينة من الممكن استخدامها لتقدير أثر برنامج ما بدقة، أي أصغر عينة تسمح لنا بالكشف عن الفروق ذات الدلالة في الحاصل بين مجموعتي المعالجة والمقارنة.

تمنحنا حسابات القوة الإحصائية مؤشرًا لأصغر عينة (وأدنى ميزانية) من الممكن استخدامها لقياس أثر برنامج ما، أي أصغر عينة تسمح بالكشف عن الفروق ذات الدلالة في الحاصل بين مجموعتي المعالجة والمقارنة. وبالتالي تُعد حسابات القوة الإحصائية ضرورية لتحديد أي البرامج ناجحة وأيها غير ذلك.

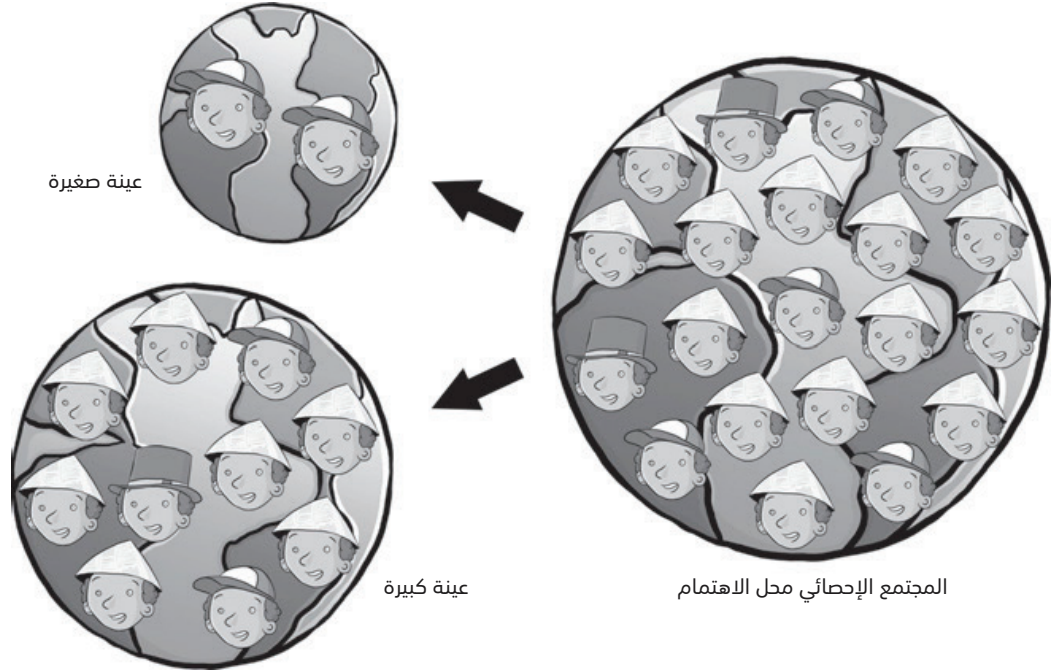
كما ناقشنا في الفصل ١، سؤال التقييم الأساسي الذي يتناوله تقييم الأثر هو "ما الأثر أو المفعول السببي للبرنامج على الحصلة محل اهتمام؟". يمكننا إعادة صياغة الفرضية البسيطة الكامنة في هذا السؤال كما يلي: هل تغير أثر البرنامج عن قيمته الصفرية؟ في حالة التعيين العشوائي، للإجابة عن هذا السؤال يلزم اتباع خطوتين، وهما:

١. تقدير متوسط الحاصل لمجموعتي المعالجة والمقارنة.
٢. تقييم ما إذا كان هناك فرق موجود بين متوسط الحصلة لمجموعة المعالجة ومتوسط الحصلة لمجموعة المقارنة. وسنناقش هنا كيفية تقدير متوسط الحاصل لكلا المجموعتين، ثم كيفية اختبار البيانات للكشف عن وجود فرق بين المجموعتين.

تقدير متوسط الحاصل لمجموعتي المعالجة والمقارنة

لنفترض أنك مهتم بتقدير أثر برنامج تغذية على وزن الأطفال البالغين من العمر عامين، وكان هناك ٢٠٠ ألف طفل مؤهلين لهذا البرنامج. ومن بين جميع الأطفال المؤهلين، تم تعيين ١٠٠ ألف عشوائيًا للمشاركة في البرنامج. أما بالنسبة للمائة ألف طفل المؤهلين الذين لم يتم تعيينهم عشوائيًا إلى البرنامج، فسيمثلون مجموعة المقارنة. كخطوة أولى، ستحتاج إلى تقدير متوسط وزن الأطفال المشاركين ومتوسط وزن الأطفال غير المشاركين. ومن أجل تحديد متوسط وزن الأطفال المشاركين، يمكن وزن كل طفل من المائة ألف طفل المشاركين ثم احتساب متوسط الأوزان. وبالطبع سينطوي ذلك على تكاليف باهظة للغاية. لحسن الحظ، ليس من الضروري قياس وزن كل طفل. وذلك لأنه يمكن تقدير المتوسط باستخدام متوسط وزن عينة مسحوبة من المجتمع الإحصائي المتمثل في الأطفال المشاركين. ٣ كلما كثر عدد الأطفال المتضمنين في العينة، كان متوسط العينة أقرب إلى المتوسط الفعلي. وعندما تكون العينة صغيرة، يشكل متوسط الوزن تقديرًا أبعد ما يكون عن الدقة لمتوسط المجتمع الإحصائي. على سبيل المثال، العينة المكونة من طفلين لن تعطيك تقديرًا دقيقًا. وفي المقابل، العينة التي تشمل ١٠ آلاف طفل ستعطي تقديرًا أكثر دقة هو أقرب بكثير من متوسط الوزن الفعلي. وبصفة عامة، كلما كثرت الملاحظات في العينة، زادت دقة الإحصائيات التي نحصل عليها من العينة (الشكل ١٥-٣).^٤

الشكل ١٥-٣ (أ) العينة الأكبر على الأرجح أن تمثل المجتمع الإحصائي محل الاهتمام



إذن، نعلم الآن أن العينة الأكبر حجمًا تعطينا صورة أكثر دقة وتحديدًا للمجتمع الإحصائي المتمثل في الأطفال المشاركين. وينطبق الأمر نفسه على الأطفال غير المشاركين: كلما كبرت عينة الأطفال غير المشاركين، كانت معرفتنا أكثر تحديدًا بما يبدو عليه المجتمع الإحصائي. ما الداعي للاهتمام هنا؟ إذا كنت تستطيع تقدير متوسط الحصيلة (الوزن) للأطفال المشاركين وغير المشاركين بشكل أكثر دقة، فإنه سيكون باستطاعتك أيضًا أن تكتشف بالتحديد الفرق في الوزن بين المجموعتين، أي ما يمثل تقدير أثر البرنامج. بعبارة أخرى، إذا كانت لديك معرفة ضبابية حول متوسط وزن الأطفال في مجموعة المشاركين (المعالجة) ومجموعة غير المشاركين (المقارنة)، فإذًا كيف يمكنك الحصول على فكرة واضحة حول الفرق في الوزن بين المجموعتين؟ هذا صحيح، لا يمكنك. وسنستعرض هذه الفكرة في القسم التالي بطريقة أكثر شمولية.

مقارنة متوسط الحاصل لمجموعتي المعالجة والمقارنة

بمجرد أن تقوم بتقدير متوسط الحصيلة (الوزن) لمجموعة المعالجة (الأطفال المشاركين المختارين بالتعيين العشوائي) والمجموعة المقارنة (الأطفال غير المشاركين المختارين بالتعيين العشوائي)، يمكنك المتابعة لتحديد ما إذا كانت الحصيلتان مختلفتين. هذا الجزء واضح: ستطرح المتوسطين، وتتحقق من الفرق. ومن الناحية الإحصائية، يختبر تقييم الأثر الفرضية الصفرية (أو الافتراضية) مقابل الفرضية البديلة.

تُعرّف الفرضية الصفرية بأنها الفرضية التي تنص على أن البرنامج ليس له أثر. ويتم التعبير عنها بالصيغة:

H_0 : الأثر أو الفرق بين الحاصل في مجموعتي المعالجة والمقارنة = ٠ .

H_a : الأثر أو الفرق بين الحاصل في مجموعتي المعالجة والمقارنة \neq ٠ .

تخيل أنك، في مثال برنامج التغذية، تبدأ بعينة لطفلين خاضعين للمعالجة وطفلين خاضعين للمقارنة. مع مثل هذه العينة الصغيرة، لن تكون هناك موثوقية كبيرة بخصوص تقديرك لمتوسط وزن الأطفال ضمن المعالجة والمقارنة، وبالتالي تقديرك للفرق بين المجموعتين. ويمكنك التحقق من ذلك عن طريق سحب عينات من الطفلين ضمن المعالجة والطفلين ضمن المقارنة. ما ستجده أولاً هو أن الأثر المقدر للبرنامج متقلب جداً.

على النقيض، لنقل إنك ستبدأ بعينة من ١,٠٠٠ طفل ضمن المعالجة و١,٠٠٠ طفل ضمن المقارنة. فكما ناقشنا، تقديراتك لمتوسط الوزن لكلتا المجموعتين ستكون أكثر دقة بكثير. لذلك، تقديرك للفرق بين المجموعتين سيكون أيضاً أكثر دقة. على سبيل المثال، لنقل أنك وجدت أن متوسط الوزن في عينة الأطفال ضمن المعالجة (المشاركين) هو ١٢,٢ كيلوجراماً (كجم)، وأن المتوسط في عينة الأطفال ضمن المقارنة (غير المشاركين) هو ١٢,٠ كيلوجراماً. الفرق بين المجموعتين هو ٠,٢ كجم. إذا أتت هذه الأرقام من عينات من ملاحظتين لكل من المجموعتين، فإنك لن تكون شديد الثقة من أن أثر البرنامج إيجابي تماماً لأن التقدير الكامل البالغ ٠,٢ كجم ناتج عن انخفاض دقة تقديراتك. ومع ذلك، إذا أتت هذه الأرقام من عينات من ١,٠٠٠ ملاحظة لكل من المجموعتين، فستكون واثقاً أكثر أنك قريب جداً من أثر البرنامج الفعلي، وهو في هذه الحالة إيجابي.

السؤال الأساسي هنا هو "بالضبط ما الحجم المطلوب للعينة لكي تسمح لك بمعرفة أن الأثر المقدر الإيجابي ناتج عن أثر البرنامج الفعلي وليس لقلة الدقة في تقديراتك؟".

خطأ محتمل في تقييمات الأثر

عندما تختبر ما إذا كان برنامج ما له أثر، يمكن أن يحدث نوعان من الأخطاء. النوع الأول من الأخطاء يحدث عندما يخلص التقييم إلى أن البرنامج له أثر، في حين لا يكون له أي أثر في الواقع. وفي حالة مثال التدخل الغذائي، سيحدث ذلك إذا خلصت، بصفتك أحد أعضاء فريق التقييم، أن متوسط وزن الأطفال في عينة المعالجة أعلى من متوسط وزن الأطفال في عينة المقارنة، على الرغم من أن متوسط وزن الأطفال في كلا المجتمعين الإحصائيين متساويان في الواقع والفروق التي لوحظت كانت بمحض الصدفة. وفي هذه الحالة، الأثر الإيجابي الذي رأيته كان نابغاً بشكل بحت من قلة دقة تقديراتك.

المفهوم الأساسي

النوع الأول من الأخطاء
يحدث عندما يخلص التقييم
إلى أن البرنامج له أثر، في
حين لا يكون له أي أثر في
الواقع.

النوع الثاني من الأخطاء
يحدث عندما يخلص التقييم
إلى أن البرنامج ليس له أثر،
في حين يكون له أثر في
الواقع.

النوع الثاني من الأخطاء هو النوع المقابل لذلك الخطأ. يحدث النوع الثاني من الأخطاء عندما يخلص التقييم إلى أن البرنامج ليس له أثر، في حين يكون له أثر في الواقع. وفي حالة التدخل الغذائي، سيحدث ذلك إذا خلصت إلى أن متوسط وزن الأطفال في العينتين متساويان، على الرغم من أن متوسط وزن الأطفال في مجموعة المعالجة أكبر في الحقيقة من متوسط وزن الأطفال في مجموعة المقارنة. مرة أخرى، سيكون الأثر إيجابيًا ولكن بسبب قلة دقة تقديراتك، خلصت إلى أن البرنامج ليس له أي أثر على الإطلاق (الفرضية الصفرية). عند اختبار فرضية وجود أثر للبرنامج، يستطيع الإحصائيون تقييد حجم النوع الأول من الأخطاء. فمن الممكن تعيين احتمال حدوث خطأ النوع الأول باستخدام معلمة تسمى مستوى الدلالة الإحصائية. غالبًا ما يكون مستوى الدلالة الإحصائية ثابتًا عند 0.05 بالمائة، ما يعني أنك تستطيع الوثوق بنسبة 95 بالمائة من أن البرنامج كان له أثر. أما إذا كنت تخشى من ارتكاب النوع الأول من الأخطاء، فيمكنك تحديد مستوى دلالة إحصائية أقل بشكل متحفظ، مثل 0.01 بالمائة بحيث تكون واثقًا بنسبة 99% من خلوصك إلى أن البرنامج كان له أثر.

ومع ذلك، تشكل أخطاء النوع الثاني مصدرًا للقلق أيضًا لصانعي السياسات. وهناك عوامل كثيرة تؤثر في احتمالية ارتكاب خطأ النوع الثاني، ويشكل حجم العينة أهمية حاسمة. فإذا كان متوسط وزن 50,000 طفل ضمن مجموعة المعالجة هو نفسه متوسط وزن 50,000 طفل ضمن مجموعة المقارنة، فإذا يمكنك على الأرجح أن تخلص - وأنت واثق - إلى أن البرنامج ليس له أثر. وعلى النقيض، إذا كانت عينة لوزن طفلين ضمن مجموعة المعالجة في المتوسط مساوية لوزن عينة لطفلين ضمن مجموعة المقارنة، فمن الصعب التوصل إلى خلاصة ذات موثوقية. هل متوسط الوزن مماثل لأن التدخل لم يكن له أثر أم لأن البيانات لم تكن كافية لاختبار الفرضية في مثل هذه العينة الصغيرة؟ إن سحب عينات كبيرة يقلل احتمالية ملاحظة الأطفال الذين لهم وزن مماثل كنتاج عن (سوء) الحظ في الاختيار فحسب. في العينات الكبيرة، يوفر الفرق في المتوسط بين عينة المعالجة وعينة المقارنة تقديرًا أفضل للفرق الفعلي في المتوسط بين جميع وحدات المعالجة وجميع وحدات المقارنة.

المفهوم الأساسي

القوة الإحصائية هي
احتمال الكشف عن أثر،
عندما يوجد أثر في الواقع.
ويتصف تقييم الأثر بأن له
قدرة عالية إذا كانت هناك
مخاطر قليلة تنطوي على
عدم اكتشاف أي آثار فعلية
للبرنامج، بمعنى أن يتم
ارتكاب خطأ النوع الثاني.

إن القوة (أو القوة الإحصائية) لتقييم الأثر هي احتمال اكتشاف فرق بين مجموعتي المعالجة والمقارنة، عندما يوجد فرق في الواقع. ويتصف تقييم الأثر بأن له قدرة عالية إذا كانت هناك مخاطر قليلة تنطوي على عدم اكتشاف أي آثار فعلية للبرنامج، بمعنى أن يتم ارتكاب خطأ النوع الثاني. وتوضح الأمثلة السابقة أن حجم العينة عامل محدد أساسي لقدرة تقييم الأثر. وسنوضح في الأقسام التالية هذه النقطة بشكل أكثر تفصيلاً.

أسباب أهمية حسابات القوة الإحصائية بالنسبة للسياسات

الغرض من حسابات القوة الإحصائية هو تحديد حجم العينة المطلوب لتجنب استنتاج أن البرنامج ليس له أثر، في حين يكون له أثر في الواقع (أي لتجنب حدوث خطأ النوع الثاني). ونحصل على قيمة القوة الإحصائية لاختبار ما عند طرح ١ من احتمال خطأ النوع الثاني.

ويتصف تقييم الأثر بأن له قدرة عالية إذا كان خطأ النوع الثاني من المستبعد حدوثه، أي أنه من المستبعد تعرضك للإبطاء على إثر النتائج التي تُظهر أن البرنامج الخاضع للتقييم ليس له أثر، عندما يكون في الواقع له أثر. ومن منظور السياسات، فإن تقييمات الأثر منخفضة القوة الإحصائية التي من المحتمل بشدة تعرضها لأخطاء النوع الثاني، لا تكون عديمة الجدوى فحسب بل قد تكون مكلفة للغاية. لا شك في أن الاحتمال المرتفع لحدوث خطأ النوع الثاني يهدد احتمال أن يحدد تقييم الأثر نتائج ذات دلالة إحصائية. ولذلك، يكون حشد الموارد في تقييمات الأثر منخفضة القوة الإحصائية استثمارًا محفوفًا بالمخاطر.

إلى جانب ما سبق، قد تؤدي تقييمات الأثر منخفضة القوة الإحصائية إلى عواقب خطيرة أيضًا على أرض الواقع. فمثلًا، في مثال التدخل الغذائي الافتراضي الذي سبق أن ذكرناه، إذا خلصت إلى أن البرنامج لم يكن ذا فعالية، على الرغم من أنه كان ذا فعالية، فربما ينهي صانعو السياسات البرنامج الذي يفيد الأطفال بصورة حقيقية. ومن هنا، نخلص إلى الأهمية القصوى لبذل كل ما هو ممكن للحد من احتمال حدوث أخطاء النوع الثاني عن طريق استخدام عينات كبيرة بما يكفي في تقييمات الأثر. وهذا ما يجعل تنفيذ حسابات القوة الإحصائية أمرًا حاسمًا للغاية ومهمًا إلى أبعد درجة.

حسابات القوة الإحصائية خطوة بخطوة

سننتقل الآن إلى المبادئ الأساسية لحسابات القوة الإحصائية مع التركيز على الحالة البسيطة التي تنطوي على البرامج المعينة عشوائيًا. إن تنفيذ حسابات القوة الإحصائية يتطلب معالجة الأسئلة الرئيسية الخمسة التالية بشكل متأن:

١. هل يؤثر البرنامج على مجموعات عنقودية؟
٢. ما مؤشر الحصيلة (أو مؤشرات الحصيلة)؟
٣. ما المستوى الأدنى للأثر الذي من شأنه أن يبرر الاستثمار في التدخل؟
٤. ما متوسط الحصيلة للمجتمع الإحصائي محل الاهتمام؟ ما التباين الكامن لمؤشر الحصيلة؟
٥. ما المستويات المعقولة للقدرة الإحصائية والدلالة الإحصائية بالنسبة للتقييم الخاضع للإجراء؟

الفائدة	المستوى الذي عنده يتم تعيين الفوائد (المجموعة العنقودية)	الوحدة التي يتم عندها قياس الحصيلة
التحويلات النقدية	القرية	الأسر المعيشية
علاج الملاريا	المدرسة	الأفراد
البرنامج التدريبي	الحي	الأفراد

ينطبق كل سؤال من الأسئلة السابقة على سياق سياسة محدد تقرر فيه إجراء تقييم الأثر. عند حساب القوة الإحصائية، تتمثل الخطوة الأولى في تحديد ما إذا كان البرنامج الذي تريد تقييمه تنتج عنه مجموعات عنقودية عند تنفيذه. التدخل الذي يكون فيه مستوى التدخل (غالبًا ما يكون أماكن) مختلفًا عن المستوى الذي عنده سترغب في قياس الحصائل (غالبًا ما يكون أشخاصًا)، تنتج عنه مجموعات عنقودية حول موقع التدخل. فعلى سبيل المثال، قد يكون من الضروري تنفيذ برنامج على مستوى المستشفى أو المدرسة أو القرية (بمعنى آخر، عبر مجموعات عنقودية)، ولكنك تقيس أثره على المرضى أو الطلاب أو القرويين (انظر الجدول ١٠-١٥).^٥ عندما ينطوي تقييم الأثر على مجموعات عنقودية، حينئذ يكون عدد المجموعات العنقودية هو ما يحدد بشكل كبير الحجم الملائم للعينة. وعلى النقيض، لعدد الأفراد ضمن المجموعات العنقودية أهمية أقل. وسنناقش هذه النقطة بالتحديد فيما يلي.

إن كانت بيانات العينة مستمدة من برامج قائمة على مجموعات عنقودية، فإن طبيعتها تكون مختلفة قليلًا عن تلك العينات التي نحصل عليها من البرامج التي لا تكون كذلك. ونتيجة لذلك، ستنطوي حسابات القوة الإحصائية على خطوات مختلفة قليلًا، اعتمادًا على ما إذا كان البرنامج يعين بشكل عشوائي الفوائد فيما بين المجموعات العنقودية أو ببساطة يعين الفوائد بشكل عشوائي فيما بين جميع الوحدات في المجتمع الإحصائي. وسنناقش كلتا الحالتين واحدة تلو الأخرى. ونبدأ بمبادئ حسابات القوة الإحصائية عندما لا تكون هناك أي مجموعات عنقودية، أي عندما تكون المعالجة معينة على مستوى تكون فيه الحصائل قيد الملاحظة. ثم سنمضي لنخوض في حسابات القدرة عندما تكون هناك مجموعات عنقودية.

حسابات القوة الإحصائية في غياب المجموعات العنقودية

افترض أنك أجبت عن السؤال الأول بإثبات أن فوائد البرنامج غير معينة حسب مجموعات عنقودية. بمعنى آخر، البرنامج المقرر تقييمه يعين بشكل عشوائي الفوائد فيما بين جميع الوحدات في المجتمع الإحصائي المؤهل. وفي الخطوة الثانية، يجب عليك تحديد أهم مؤشرات الحصائل التي تم تصميم البرنامج لتحسينها. ويتم اشتقاق هذه المؤشرات من هدف البرنامج، ونظرية التغيير، وسؤال بحث التقييم الأساسي، كما ناقشنا في الجزء الأول. وستعطينا حسابات القوة الإحصائية أيضًا رؤى حول نوع المؤشرات التي من أجلها تستطيع تقييمات الأثر تحديد الآثار. وبالطبع كما سنناقش بمزيد من التفصيل، العينات التي تختلف في الأحجام قد تكون مطلوبة لقياس الآثار على مؤشرات مختلفة.

وثالثاً، يجب عليك تحديد المستوى الأدنى للأثر الذي من شأنه أن يبرر الاستثمار في التدخل. وهذا في الأساس عبارة عن سؤال سياسي لا فني. هل برنامج التحويل النقدي استثمار مجزٍ إذا كان سيحد من الفقر بمعدل ٥ بالمائة أم ١٠ بالمائة أو ١٥ بالمائة؟ هل برنامج سوق العمل النشطة يستحق التنفيذ إذا كان سيزيد الدخل بمعدل ٥ بالمائة أم ١٠ بالمائة أم ١٥ بالمائة؟ تتسم الإجابة بأنها ذات طبيعة خاصة للغاية وفقاً للسياق، ولكن في جميع السياقات يكون من الضروري تحديد التغيير في مؤشرات الحاصل التي ستبرر جدوى الاستثمار في هذا البرنامج. بصيغة أخرى، ما مستوى الأثر الذي دونه يجب اعتبار التدخل غير ناجح؟ عندما تجيب عن هذا السؤال ستتعرف على الحد الأدنى للتأثير القابل للاكتشاف الذي يجب أن يكون تقييم الأثر قادراً على تحديده. ولا تعتمد هذه الإجابة على تكلفة البرنامج ونوع الفوائد التي يوفرها فحسب، بل أيضاً على تكلفة الفرصة لعدم استثمار الأموال في تدخل بديل.

في حين أن الحد الأدنى للتأثيرات القابلة للاكتشاف يمكن أن يستند إلى أهداف السياسة، فإنه يمكن استخدام مقاربات أخرى لإثباتها. لربما يكون من المفيد وضع أساس مقارنة للحد الأدنى للتأثير القابل للاكتشاف مقابل النتائج الواردة عن دراسات حول برامج مماثلة لإلقاء الضوء على حجم الآثار التي يمكن توقعها. فعلى سبيل المثال، التدخلات التعليمية غالباً ما

تقيس المكاسب من حيث درجات الاختبار المعيارية. وتُظهر الدراسات الحالية أن الزيادة بانحراف معياري ٠,١ صغيرة نسبياً، بينما الزيادة بانحراف معياري ٠,٥ كبيرة نسبياً. وبدلاً من ذلك، يمكن إجراء عمليات المحاكاة السابقة بهدف تقييم مجموعة من الآثار التي تكون واقعية بموجب فرضيات مختلفة. ترد أمثلة على عمليات المحاكاة السابقة في الفصل ١ لبرامج التحويل النقدي الشرطي. وأخيراً، يمكن أن يلقي التحليل الاقتصادي السابق الضوء على حجم الآثار التي ستكون مطلوبة لمعدل المنفعة على استثمار معين ليكون مرتفعاً بشكل كافٍ. على سبيل المثال، الأرباح السنوية التي يتم اكتسابها عن طريق برنامج التدريب الوظيفي يجب أن تكون أعلى من سعر فائدة السوق السائد.

بصورة بديهية، من الأسهل تحديد فرق كبير بين المجموعتين مقارنة بتمييز وجود فرق صغير بين المجموعتين. ولكي يكون تقييم الأثر مؤهلاً لاكتشاف فرق صغير بين مجموعتي المعالجة والمقارنة، سيكون مطلوباً الإتيان بتقدير بالغ الدقة للفرق في متوسط الحاصل بين المجموعتين. ويتطلب ذلك عينة كبيرة. وبعوضاً عن ذلك، بالنسبة للتدخلات التي يُنظر أنها عالية القيمة فقط إذا كانت تؤدي إلى تغييرات كبيرة في مؤشرات الحاصل، ستكون العينات المطلوبة لإجراء تقييم أثر أصغر. وبغض النظر عن ذلك، يجب تعيين الحد الأدنى للتأثير القابل للاكتشاف بشكل متحفظ، بما أن أي أثر أصغر من الحد الأدنى للأثر المطلوب سيكون أقل ترجيحاً لاكتشافه.

رابعاً، لإجراء حسابات القوة الإحصائية، يجب عليك أن تطلب من أحد الخبراء تقدير بعض المعلومات الأساسية، مثل متوسط خط الأساس وتباين مؤشرات الحاصل.

المفهوم الأساسي

الحد الأدنى للتأثير القابل للاكتشاف

(MDE) هو حجم المفعول الذي تم تصميم تقييم الأثر من أجل تقديره لمستوى معين من الدلالة والقوة الإحصائية. ومع فرض استواء جميع القيم، تكون العينات الكبيرة مطلوبة لتقييم الأثر من أجل اكتشاف الفروق الصغيرة بين مجموعتي المعالجة والمقارنة أو لاكتشاف الفروق في حصة أكثر تبايناً.

من المستحسن الحصول على قيم أساس المقارنة هذه من البيانات الحالية التي تم جمعها في مكان مماثل لذلك المكان حيث من المقرر تنفيذ البرنامج الخاضع للدراسة أو من مسح تجريبي في المجتمع الإحصائي محل الاهتمام.^٦ جدير بالذكر أنه كلما زاد تباين الحاصل محل الاهتمام بشكل مثبت، كانت العينة بحاجة إلى أن تكون أكبر من أجل تقدير مفعول المعالجة بشكل دقيق. وفي مثال التدخل الغذائي الافتراضي، وزن الأطفال هو الحصلة محل الاهتمام. فإذا كان وزن جميع الأفراد هو نفسه عند خط الأساس، فسيكون من السهل تقدير أثر التدخل الغذائي في عينة صغيرة. وعلى النقيض، إذا كانت الأوزان عند خط الأساس شديدة التباين، وجب حينها الاعتماد على عينة أكبر من أجل تقدير أثر البرنامج.

خامسًا، يحتاج فريق التقييم إلى تحديد مستوى قوة إحصائية ومستوى دلالة إحصائية معقولين لتقييم الأثر المخطط له. وكما ذكرنا آنفًا، نحصل على قيمة القوة الإحصائية لاختبار ما عند طرح ١ من احتمال خطأ النوع الثاني. لذلك، تتراوح القدرة بين ٠ و ١ حيث يشير ارتفاع القيمة إلى انخفاض مخاطر الفشل في تحديد أثر قائم. تُستخدم القدرة التي قيمتها ٠,٨ بشكل واسع كأساس للمقارنة في حسابات القوة الإحصائية. وهذا يعني أنك ستجد أثرًا في ٨٠٪ من الحالات التي يكون فيها أثر. وغالبًا ما يوفر ارتفاع مستوى القوة الإحصائية إلى ٠,٩ (أو ٩٠ بالمائة) أساس مقارنة مفيدًا، ولكنه يكون أكثر تحفظًا، الأمر الذي يزيد من حجم العينة المطلوب.

إن مستوى الدلالة الإحصائية يعني احتمال ارتكاب خطأ النوع الأول. وعادة ما يتم تعيين هذا المستوى عند ٥ بالمائة، بحيث يمكنك أن تكون واثقًا بنسبة ٩٥ بالمائة في الخلو إلى أن البرنامج له أثر إذا وجدت بالفعل أثرًا ذا دلالة. ويوجد مستويان شائعان للدلالة الإحصائية وهما: ١ و ١٠ بالمائة. كلما كان مستوى الدلالة أصغر، زادت الثقة في مدى واقعية الأثر المقدّر.

بمجرد معالجة هذه الأسئلة الخمسة، يستطيع خبير حسابات القوة الإحصائية حساب حجم العينة المطلوب باستخدام برنامج إحصائي معياري.^٧ وستشير حسابات القوة إلى حجم العينة المطلوب، اعتمادًا على المعلومات المذكورة في الخطوات من ١ إلى ٥. تتسم هذه الحسابات بأنها مباشرة بمجرد تحديد المعلومات ذات الصلة بالسياسة (بالأخص في الخطوتين الثانية والثالثة).^٨ وإذا كنت مهتمًا بإجراء حسابات القوة الإحصائية، فستجد على الموقع الإلكتروني للكتاب خبيرًا فنيًا سيزودك بأمثلة على حسابات القوة باستخدام برنامجي Stata و Optimal Design.

فعندما تطلب نصيحة من خبراء إحصائيين، ينبغي لفريق التقييم طلب تحليل لمدى تأثير حساب القوة الإحصائية بأي تغيير يطرأ على الافتراضات. أي أنه من الضروري فهم حجم العينة المطلوب الذي قد يستدعي الأمر زيادته بموجب الافتراضات الأكثر تحفظًا (مثل الأثر الأقل توقعًا أو ارتفاع التباين في مؤشر الحصلة أو ارتفاع مستوى القوة الإحصائية). ومن الممارسات الجيدة أيضًا إجراء حسابات قوة إحصائية مختلفة لمؤشرات حائل متنوعة، حيث إن أحجام العينة المطلوبة يمكن أن تتفاوت بشكل أساسي إذا كانت بعض مؤشرات الحاصل أكثر تباينًا من نظيراتها.

وأخيرًا، يمكن أن تشير حسابات القوة الإحصائية أيضًا إلى حجم العينة المطلوب لإجراء مقارنة لآثار البرنامج عبر مجموعات فرعية محددة (على سبيل المثال، الرجال أو النساء أو مجموعات فرعية أخرى للمجتمع الإحصائي محل الاهتمام). وسيلزم وقتها أن تكون كل مجموعة فرعية مطابقة لحجم العينة المطلوب.

تقييم أثر برنامج HISP: تحديد حجم العينة المطلوب لتقييم برنامج HISP الموسع



لنعد إلى مثالنا الذي ذكرناه في الجزء ٢، لنفترض أن وزارة الصحة سرتّها الجودة والنتائج لتقييم برنامج معونة التأمين الصحي (HISP). ومع ذلك، قبل التوسع في تنفيذ البرنامج، قررت الوزارة تجريب إصدار موسع للبرنامج أطلقوا عليه HISP+. بالنسبة لبرنامج HISP الأصلي، فهو يدفع جزءًا من تكلفة التأمين الصحي للأسر المعيشية الريفية الفقيرة لتغطية تكاليف الرعاية الأولية والعقاقير، ولكنها لا تغطي الإقامة في المستشفى. ويتساءل وزير الصحة ما إذا كان HISP+ الموسع الذي يغطي أيضًا الإقامة في المستشفى سيقول أيضًا النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية الفقيرة. ولهذا السبب، طلب منك الوزير تصميم تقييم أثر الغرض منه تقييم ما إذا كان HISP+ سيحد من النفقات الصحية للأسر المعيشية الريفية الفقيرة. وفي هذه الحالة، لن تواجه صعوبة عند اختيار تصميم تقييم الأثر: الموارد شحيحة بالنسبة لبرنامج HISP+، ويتعذر تطبيقه على العموم بصورة فورية. كنتيجة لذلك، لقد خلصت إلى أن التعيين العشوائي سيكون طريقة تقييم الأثر الأكثر صلاحية وفعالية. ويفهم وزير الصحة إلى أي مدى طريقة التعيين العشوائي يمكن أن تفلح وتكون داعمة.

ومن أجل اختتام تصميم تقييم الأثر، لقد استأجرت إحصائيًا لمساعدتك على تحديد الحجم المطلوب للعينة. وقبل أن يشرع في العمل، سألك الإحصائي عن بعض المدخلات الأساسية. واستخدم قائمة تدقيق مكونة من خمسة أسئلة.

١. هل ستنتج عن برنامج HISP+ مجموعات عنقودية؟ أنت لا تكون على تمام التيقن في هذه المرحلة. وترى أن باستطاعتك - على نحو محتمل - توزيع حزمة الفوائد الموسعة على مستوى الأسر المعيشية فيما بين الأسر الريفية الفقيرة التي تستفيد بالفعل من برنامج HISP. ومع ذلك، أنت مدرك أن وزير الصحة قد يفضل تعيين البرنامج الموسع على مستوى القرى، ومن ثم ستنتج مجموعات عنقودية. واقترح الإحصائي إجراء حسابات قوة إحصائية لحالة أساس مقارنة لا تنطوي على مجموعات عنقودية، وبعد ذلك الأخذ بالاعتبار كيف ستتغير النتائج في ظل وجود مجموعات عنقودية.

٢. ما المقصود بمؤشر الحصيلة؟ لقد أوضحت أن الحكومة مهتمة بمؤشر محدد جيدًا: النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية الفقيرة. ومن ثم بحث الإحصائي عن أحدث مصدر للحصول على قيم أساس مقارنة لمؤشره، واقترح استخدام مسح متابعة بالاستناد إلى تقييم HISP. وذكر أنه فيما بين الأسر المعيشية التي تلقت برنامج HISP، كان متوسط النفقات الصحية المدفوعة من المال الخاص لكل فرد سنويًا ٧,٨٤ دولار.

٣. ما المستوى الأدنى للأثر الذي من شأنه أن يدعم الاستثمار في هذا التدخل؟ بمعنى آخر، ما الانخفاض في النفقات الصحية المدفوعة من المال الخاص دون المتوسط البالغ ٧,٨٤ دولار الذي سيجعل التدخل جيد الأثر؟ أكد الإحصائي على أن ذلك ليس فقط اعتبارًا فنيًا، إنما سؤال سياسة حقيقي، وهو لماذا يجب على صانع سياسات مثلك تحديد الحد الأدنى للتأثير الذي يجب أن يكون التقييم قادرًا على اكتشافه. وهنا تذكرت أن استنادًا إلى التحليل الاقتصادي السابق، يعتبر برنامج HISP+ ذا فعالية إذا قلل النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية بمعدل دولارين. رغم ذلك، أنت تعلم أن لغرض التقييم، قد يكون من الأفضل أن تكون متحفظًا في تحديد الحد الأدنى للتأثير القابل للاكتشاف بما أن أي أثر صغير من المستبعد أن يُلاحظ. ورغبةً منك في فهم كيف يختلف الحجم المطلوب للعينة استنادًا إلى الحد الأدنى للتأثير القابل للاكتشاف، اقترحت أن يجري الإحصائي حسابات للحد الأدنى للانخفاض في النفقات الصحية المدفوعة من المال الخاص عندما تبلغ قيمته دولارًا ودولارين وثلاثة دولارات.

٤. ما المقصود بتباين مؤشر الحصيلة في المجتمع الإحصائي محل الاهتمام؟ عاد الإحصائي إلى مجموعة البيانات الخاصة بالأسر المعيشية المعالجة ببرنامج HISP مشيرًا إلى أن الانحراف المعياري للنفقات الصحية المدفوعة من المال الخاص هو ٨ دولارات.

٥. ماذا سيكون مستوى القوة الإحصائية المعقول ليتم إجراء التقييم؟ أضاف الإحصائي أن حسابات القوة الإحصائية عادةً ما تُجرى للقوة التي تتراوح بين ٠,٨ و ٠,٩. وأوصى بالوصول إلى القيمة ٠,٩، ولكنه عرض إجراء عمليات تحقق صارمة في وقت لاحق لمستوى أقل تحفظًا يبلغ ٠,٨.

وبعد اقتناء كل هذه المعلومات، أجرى الإحصائي حسابات القوة الإحصائية. وكما وافقنا، بدأ الإحصائي مع حالة أكثر تحفظًا قوتها الإحصائية تبلغ ٠,٩. واستخرج النتائج الموضحة في الجدول ١٥-٢.

وقد خلص الإحصائي إلى أنه لكشف انخفاض بمعدل دولارين في النفقات الصحية المدفوعة من المال الخاص بقوة إحصائية ٠,٩، يجب أن تحتوي العينة على ٦٧٢ وحدة على الأقل (٣٣٦ من وحدات المعالجة و٣٣٦ من وحدات المقارنة في غياب أي تجمعات عنقودية). وذكر أنك إذا كنت راضيًا عن اكتشاف انخفاض بمعدل ٣ دولارات في النفقات الصحية المدفوعة من المال الخاص، فمن الممكن حينئذ استخدام عينة أصغر بشرط ألا تقل عن ٣٠٠ وحدة.

الجدول ١٥-٢ تقييم برنامج HISP+: حجم العينة المطلوب للكشف عن الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف، مع افتراض القوة الإحصائية = ٠,٩

الحد الأدنى للتأثير القابل للاكتشاف	مجموعة المعالجة	مجموعة المقارنة	إجمالي العينة
دولار	١,٣٤٤	١,٣٤٤	٢,٦٨٨
دولاران	٣٣٦	٣٣٦	٦٧٢
٣ دولارات	١٥٠	١٥٠	٣٠٠

ملاحظة: الحد الأدنى للتأثير القابل للاكتشاف يصف الحد الأدنى للانخفاض في النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية ويمكن اكتشافها بواسطة تقييم الأثر. القوة الإحصائية = ٠,٩، بدون تجمعات عنقودية..

الجدول ١٥-٣ تقييم برنامج HISP+: حجم العينة المطلوب للكشف عن الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف، مع افتراض القوة الإحصائية = ٠,٨

الحد الأدنى للتأثير القابل للاكتشاف	مجموعة المعالجة	مجموعة المقارنة	إجمالي العينة
دولار	١,٠٠٤	١,٠٠٤	٢,٠٠٨
دولاران	٢٥١	٢٥١	٥٠٢
٣ دولارات	١١٢	١١٢	٢٢٤

ملاحظة: الحد الأدنى للتأثير القابل للاكتشاف يصف الحد الأدنى للانخفاض في النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية ويمكن اكتشافها بواسطة تقييم الأثر. القوة الإحصائية = ٠,٨، بدون تجمعات عنقودية..

(١٥٠ وحدة في كل مجموعة) سيكون كافياً. على النقيض، يجب توفير عينة أكبر بمعدل ٢,٦٨٨ وحدة على الأقل (١,٣٤٤ في كل مجموعة) للكشف عن انخفاض دولار واحد في النفقات الصحية المدفوعة من المال الخاص. ومن ثم استخرج الإحصائي جدولاً آخر لمستوى قوة إحصائية يبلغ ٠,٨. ويظهر الجدول ١٥,٣ أن أحجام العينة المطلوبة أصغر لقدرة تبلغ ٠,٨ عن القدرة البالغة ٠,٩. وللكشف عن انخفاض قدره دولاران في النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية، يكفي أن يكون إجمالي العينة ٥٠٢ وحدة كحد أدنى. أما للكشف عن انخفاض قدره ٣ دولارات، فيجب توفير ٢٢٤ وحدة على الأقل. إلا أنه للكشف عن انخفاض قدره دولار واحد، سيكون مطلوباً ٢,٠٠٨ وحدات على الأقل في العينة. وأكد الإحصائي أن النتائج التالية نموذجية بالنسبة لحسابات القوة الإحصائية:

- كلما ارتفع مستوى القوة الإحصائية (أي زاد تحفظاً)، زاد حجم العينة المطلوب.
- كلما صغر الأثر المراد اكتشافه، كبرت العينة المطلوبة.

الجدول ١٥،٤ تقييم برنامج HISP+: حجم العينة المطلوب لاكتشاف الحد الأدنى للتأثيرات المختلفة القابلة للاكتشاف (الزيادة في معدل الاستشفاء)

القوة الإحصائية = ٨٠٪، بدون تجمعات عنقودية

الحد الأدنى للتأثير القابل للاكتشاف (نقطة مئوية)	مجموعة المعالجة	مجموعة المقارنة	إجمالي العينة
١	٧,٢٥٧	٧,٢٥٧	١٤,٥١٤
٢	١,٨١٥	١,٨١٥	٣,٦٣٠
٣	٨٠٧	٨٠٧	١,٦١٤

ملاحظة: يصف الحد الأدنى للتأثير المطلوب الحد الأدنى للتغيير في معدل استخدام الاستشفاء (المعبر عنه بالنقاط المئوية) الذي يمكن اكتشافه بواسطة تقييم الأثر.

سألك الإحصائي ما إذا كنت تود إجراء حسابات قوة إحصائية للحصول محل الاهتمام الأخرى. واقترحت أيضًا اعتبار حجم العينة المطلوب للكشف عما إذا كان برنامج HISP+ له تأثير على معدل الاستشفاء. وفي عينة القرى المعالجة ببرنامج HISP، يزور فرد الأسرة المعيشية المستشفى في سنة معينة بنسبة ٥ بالمائة من الأسر، وهذا يوفر معدل أساس مقارنة. ومن ثم أنشئ الإحصائي جدولاً جديداً يُظهر أن العينات الأكبر نسبياً ستكون مطلوبة للكشف عن التغييرات في معدل الاستشفاء (الجدول ١٥-٤) للنقاط المئوية ١ أو ٢ أو ٣ من معدل خط الأساس البالغ ٥ بالمائة.

يُظهر الجدول ١٥-٤ أن متطلبات حجم العينة أكبر لتلك الحصة (معدل الاستشفاء) مقارنة بالنفقات الصحية المدفوعة من المال الخاص. وبالتالي، خلص الإحصائي إلى أنك إذا كنت مهتماً بالكشف عن الآثار على كلتا الحصيلتين، فيجب عليك استخدام أحجام أكبر من العينات المتناسبة مع حسابات القوة الإحصائية التي تم إجراؤها على معدلات الاستشفاء. بما أنه يتم استخدام أحجام عينات مستنبطة من حسابات القوة الإحصائية التي أجريت للنفقات الصحية المدفوعة من المال الخاص، فقد اقترح الإحصائي إعلام وزير الصحة أن التقييم لن تكون له قوة إحصائية كافية لاكتشاف المفاعيل ذات الصلة بالسياسة حول معدلات الاستشفاء.

السؤال ٨ الخاص ببرنامج HISP



- أ. أي حجم عينة توصي به لتقدير أثر برنامج HISP+ على النفقات الصحية المدفوعة من المال الخاص؟
ب. هل سيكون حجم العينة كافياً لاكتشاف التغييرات في معدل الاستشفاء؟

حسابات القوة الإحصائية في وجود المجموعات العنقودية

قدمنا في المناقشة السابقة مبادئ إجراء حسابات قوة إحصائية للبرامج التي لا تنتج عنها مجموعات عنقودية. ومع ذلك، كما ناقشنا في الجزء ٢، بعض البرامج تقوم بتعيين الفوائد على مستوى المجموعة العنقودية. وها نحن نصف بإيجاز كيف أن المبادئ الأساسية لحسابات القوة الإحصائية يجب تكييفها للعينات المنطوية على مجموعات عنقودية.

وفي وجود التجمعات العنقودية، هناك مبدأ توجيهي مهم ينص على أن عدد المجموعات العنقودية يهتم عادةً أكثر من عدد الأفراد داخل هذه المجموعات. ويلزم توافر عدد كافٍ من المجموعات العنقودية لإجراء اختبار مقنع يُظهر ما إذا كان برنامج ما له أثر عن طريق مقارنة الحاصلات في العينات المتمثلة في وحدات مجموعتي المعالجة والمقارنة. إن عدد المجموعات العنقودية هو ما يحدد بشكل كبير الحجم الفعّال أو المجدي للينة. فإذا كنت تقوم بتعيين المعالجة عشوائيًا فيما بين عدد صغير من المجموعات العنقودية، فمن المستبعد تطابق هذه المجموعات فيما بين مجموعتي المعالجة والمقارنة. ولن يضمن التعيين العشوائي بين منطقتين إداريتين أو بين مدرستين أو بين مستشفين تماثل المجموعتين العنقوديتين. وعلى النقيض، التعيين العشوائي لتدخل ما فيما بين ١٠٠ منطقة إدارية أو ١٠٠ مدرسة أو ١٠٠ مستشفى على الأرجح أن يضمن تماثل مجموعتي المعالجة والمقارنة. بإيجاز، العدد الكافي من المجموعات العنقودية ضروري لضمان تحقيق التوازن. وعلاوة على ذلك، عدد المجموعات العنقودية تنعكس أهميته على دقة تقدير مفاعيل المعالجة. ويصير شرطًا توافر العدد الكافي من المجموعات العنقودية من أجل اختبار الفرضية بأن برنامج ما له أثر بقوة إحصائية كافية. ولذلك، عند تنفيذ تقييم أثر اعتمادًا على التعيين العشوائي، يكون من المهم جدًا ضمان أن عدد المجموعات العنقودية كبيرًا ما يكفي.

يمكنك تحديد عدد المجموعات العنقودية اللازم لإجراء اختبار دقيق للفرضية باستخدام حسابات القوة الإحصائية. ومن أجل إجراء حسابات القوة الإحصائية للعينات المتضمنة مجموعات عنقودية، تستدعي الضرورة طرح الأسئلة الخمسة ذاتها المذكورة أعلاه إلى جانب سؤال إضافي: ما مدى تباين مؤشر الحصيلة داخل المجموعات العنقودية؟ بشكل كبير للغاية، تتسم كل الحاصلات ضمن المجموعة العنقودية بأنها مترابطة بشكل مثالي. فمثلاً، قد يكون دخل الأسرة المعيشية ليس متباينًا بشكل خاص داخل القرى نفسها، إنما يحدث التفاوت الكبير في الدخل بين القرى بعضها البعض. وفي هذه الحالة، إذا كنت تضع في اعتبارك إضافة فرد إلى عينة التقييم الخاص بك، فإن إضافة فرد من قرية جديدة سيوفر قوة إحصائية إضافية أكثر بكثير من إضافة فرد من قرية ممثلة بالفعل. وبما أن الحاصلات مترابطة بشكل تام ضمن المجموعة العنقودية، فإن إضافة فرد جديد من المجموعة الحالية لن تضيف أي معلومات جديدة. وبالطبع، في هذه الحالة، القروي الثاني من المرجح أن يكون مماثلًا بشدة للقروي الأصلي المدرج بالفعل في العينة. كقاعدة عامة، كلما ارتفع الترابط البيئي العنقودي في الحاصلات (أي ارتفاع الترابط في الحاصلات أو الخصائص بين الوحدات التي تنتمي إلى المجموعة العنقودية نفسها)، زاد عدد المجموعات العنقودية المطلوبة لتحقيق مستوى قوة إحصائية معين.

المفهوم الأساسي

عدد المجموعات يهتم أكثر بالنسبة لحسابات القوة الإحصائية عن عدد الأفراد داخل المجموعات العنقودية نفسها. وفي أغلب الأحيان، يتطلب الأمر مجموعات عنقودية يتراوح عددها بين ٣٠ و ٥٠ مجموعة في كل من مجموعتي المعالجة والمقارنة، رغم أن متطلبات حجم العينة ستختلف على أساس كل حالة على حدة، وتكون حسابات القدرة مطلوبة لضمان توفير حجم العينة المناسب.

في العينات المتضمنة مجموعات عنقودية، تسلط حسابات القوة الإحصائية الضوء على المفاضلات بين إضافة مجموعات عنقودية وإضافة ملاحظات ضمن المجموعات العنقودية. إن الزيادة النسبية في القوة الإحصائية الناتجة عن إضافة وحدة إلى مجموعة عنقودية جديدة تكاد تكون دائماً أكبر من تلك القدرة الناتجة عن إضافة وحدة إلى مجموعة عنقودية حالية. وعلى الرغم من أن الزيادة في القوة الإحصائية الناتجة عن إضافة مجموعة عنقودية جديدة يمكن أن تكون كبيرة، فإن إضافة مجموعات عنقودية قد تنطوي أيضاً على تداعيات تشغيلية وتزيد من تكلفة تنفيذ البرنامج أو جمع البيانات. ونوضح لاحقاً في هذا الفصل كيفية إجراء حسابات القوة الإحصائية في ظل وجود مجموعات عنقودية في حالة برنامج HISP+، ونناقش بعضاً من المفاضلات التي تؤول إليها.

في العديد من الحالات، يجب أن يتراوح عدد المجموعات العنقودية بين ٤٠ و ٥٠ مجموعة في كل من المجموعتين: المعالجة والمقارنة للإتيان بقوة إحصائية كافية ولضمان التوازن بين خصائص خط الأساس عند استخدام طرق تعيين عشوائية. ومع ذلك، قد يتفاوت العدد اعتماداً على المعلومات المختلفة التي ناقشناها بالفعل، وكذلك الترابط البيئي العنقودي. وبالإضافة إلى ذلك، كما سنبين فيما يلي، العدد سيزداد على الأرجح عند استخدام طرق بخلاف التعيين العشوائي (بافتراض ثبات كل المتغيرات الأخرى).

تقييم أثر برنامج HISP: تحديد حجم العينة المطلوب لتقييم برنامج HISP الموسع في وجود المجموعات العنقودية



بعد مناقشتك الأولى مع الإحصائي حول حسابات القوة الإحصائية لبرنامج HISP+، قررت التحدث بإيجاز إلى وزير الصحة بشأن تنفيذ التعيين العشوائي لفوائد برنامج HISP+ الموسع فيما بين كل الأفراد في المجتمع الإحصائي الذي سيتلقى خطة HISP الأساسية. وكشفت الاستشارة أن مثل هذا الإجراء لن يكون في المستطاع من الناحية السياسية: في ذلك السياق، سيكون من الصعب تفسير لماذا شخص ما سيتلقى الفوائد الموسعة في حين لن يتلقى جاره الفوائد نفسها.

ولذلك، بدلاً من التوزيع العشوائي على المستوى الفردي، اقترحت الاختيار العشوائي لعدد القرى المستفيدة من برنامج HISP لتجريب HISP+. وقتئذ، سيكون جميع القرويين في القرية المختارة مؤهلين. وستنتج عن هذا الإجراء مجموعات عنقودية، الأمر الذي يحتم إجراء حسابات قوة إحصائية جديدة. أنت الآن ترغب في تحديد حجم العينة المطلوب لتقييم أثر برنامج HISP+ عند تعيينه بشكل عشوائي حسب المجموعة العنقودية. ولذلك، استشرت الإحصائي الخاص بك مرة ثانية. وطمأنك أن عليك بذل القليل من الجهد الإضافي. وفي قائمة التدقيق الخاصة به، تبقى سؤال واحد فقط غير مجاب عنه.

الجدول 10-0 تقييم برنامج HISP+ حجم العينة المطلوب لاكتشاف الحد الأدنى للتأثيرات القابلة للاكتشاف (الانخفاض في النفقات الصحية لدى الأسر المعيشية)

القوة الإحصائية = ٠,٨، الحد الأقصى هو ١٠٠ مجموعة عنقودية

الحد الأدنى للتأثير القابل للاكتشاف	عدد المجموعات العنقودية	الوحدات في كل مجموعة عنقودية	إجمالي العينة مع المجموعات العنقودية	Total sample without clusters
دولار	١٠٠	١٠٢	١٠,٢٠٠	٢,٠٠٨
دولاران	٩٠	٧	٦٣٠	٥٠٢
٣ دولارات	٨٢	٣	٢٤٦	٢٢٤

ملاحظة: الحد الأدنى للتأثير القابل للاكتشاف يصف الحد الأدنى للانخفاض في النفقات الصحية المدفوعة من المال الخاص التي تتحملها الأسر المعيشية ويمكن اكتشافها بواسطة تقييم الأثر. عدد المجموعات العنقودية هو إجمالي عدد المجموعات العنقودية، نصفها ينتمي إلى مجموعة المقارنة والنصف الآخر إلى مجموعة المعالجة.

وأعرب الإحصائي عن حاجته إلى معرفة مدى تباين مؤشر الحصيلة ضمن المجموعات العنقودية. ولحسن الحظ، هذا أيضًا سؤال يستطيع الإجابة عنه باستخدام بيانات برنامج HISP. ووجد أن الارتباط الداخلي للقرية فيما يتعلق بالنفقات الصحية المدفوعة من المال الخاص يساوي ٠,٠٤.

وتساءل أيضًا عما إذا تم وضع حد أعلى لعدد القرويين الذين من الممكن أن يشملهم البرنامج التجريبي الجديد حال تنفيذه. وبما أن البرنامج يتضمن الآن ١٠٠ قرى مستفيدة من برنامج HISP، فقد أوضحت أنه يمكنك تضمين ٥٠ قرية معالجة و ٥٠ قرية مقارنة كحد أقصى لبرنامج HISP+. ومع تلك المعلومات المعطاة، أجرى الإحصائي حسابات القوة الإحصائية كما هي موضحة في الجدول ١٥,٥ لقدرة تساوي ٠,٨.

وقد خلص الإحصائي إلى أنه لكشف انخفاض مقداره دولاران في النفقات الصحية المدفوعة من المال الخاص، يجب أن تتضمن العينة ما لا يقل عن ٦٣٠ وحدة؛ أي ٧ وحدات لكل مجموعة عنقودية في ٩٠ مجموعة عنقودية (٤٥ في مجموعة المعالجة و ٤٥ أخرى في مجموعة المقارنة). وذكر أن هذا العدد يزيد عن ذلك في العينة الخاضعة للتعين العشوائي على مستوى الأسر المعيشية، التي تتطلب فقط إجمالي ٥٠٢ وحدة (٢٥١ في مجموعة المعالجة و ٢٥١ في مجموعة المقارنة، انظر الجدول ١٥-٣). لكشف عن انخفاض مقداره ٣ دولارات في النفقات الصحية المدفوعة من المال الخاص، يجب أن تتضمن العينة ما لا يقل عن ٢٤٦ وحدة؛ أي ٣ وحدات في كل من المجموعات العنقودية البالغ عددها ٨٢ (٤١ في مجموعة المعالجة و ٤١ أخرى في مجموعة المقارنة).

بعد ذلك، أظهر الإحصائي كيف أن إجمالي عدد الملاحظات المطلوبة في العينة يختلف مع اختلاف إجمالي عدد المجموعات العنقودية. وقرر تكرار الحسابات للحد الأدنى للتأثير القابل للاكتشاف البالغ دولارين والقوة الإحصائية البالغة ٠,٨. إن حجم إجمالي العينة المطلوب لتقدير مثل هذا المفعول يزداد بقوة عندما يقلص عدد المجموعات العنقودية (الجدول ١٥-٦).

بالنسبة لـ ١٢٠ مجموعة عنقودية، سيكون من المطلوب الحصول على عينة مكونة من ٦٠٠ ملاحظة. وإذا توافرت ٣٠ مجموعة عنقودية فحسب، فإن إجمالي العينة يجب أن يتضمن ١,٥٠٠ ملاحظة. وعلى النقيض، إذا توافرت ٩٠ مجموعة عنقودية، فالمطلوب حينئذ هو ٦٣٠ ملاحظة فقط.

الجدول ٦-١٥ تقييم برنامج HISP+: حجم العينة المطلوب لاكتشاف الحد الأدنى البالغ دولارين لأعداد مختلفة من المجموعات العنقودية

القوة الإحصائية = ٨٠,٠

الحجم الأدنى للتأثير	عدد المجموعات العنقودية	الوحدات في كل مجموعة عنقودية	إجمالي العينة مع المجموعات العنقودية
دولاران	٣٠	٥٠	١,٥٠٠
دولاران	٥٨	١٣	٧٥٤
دولاران	٨١	٨	٦٤٨
دولاران	٩٠	٧	٦٣٠
دولاران	١٢٠	٥	٦٠٠

ملاحظة: عدد المجموعات العنقودية هو إجمالي عدد المجموعات العنقودية، نصفها ينتمي إلى مجموعة المقارنة والنصف الآخر إلى مجموعة المعالجة. وإذا كان التصميم لا يتضمن أي مجموعات عنقودية، فإن ٢٥١ وحدة في كل مجموعة ستكون مطلوبة لتحديد الحد الأدنى للتأثير البالغ دولارين (انظر الجدول ٣-١٥).

السؤال ٩ الخاص ببرنامج HISP



- أ.** أي إجمالي حجم عينة توصي به لتقدير أثر برنامج HISP+ على النفقات الصحية المدفوعة من المال الخاص؟
ب. في كم من القرى ستصحح وزير الصحة بتنفيذ برنامج HISP+؟

تجاوز حالة أساس المقارنة

في هذا الفصل، ركزنا على حالة أساس المقارنة لتقييم الأثر الذي تم إجراؤه باستخدام طريقة التعيين العشوائي مع الامتثال الكامل. وهذا هو أبسط سيناريو، وبالتالي هو الأكثر ملاءمة لتوصيل المنطق البديهي وراء حسابات القوة الإحصائية. وعلى الرغم من ذلك، فإننا لم نناقش بعد جوانب عملية كثيرة مرتبطة بحسابات القوة الإحصائية، كما أن الانحرافات عن الحالات الأساسية التي ناقشناها هنا نحتاج إلى اعتبارها بعناية. ونناقش فيما يلي بعضاً من هذه الانحرافات.

باستخدام الأساليب شبه التجريبية. مع افتراض ثبات المتغيرات، عادةً ما تستدعي أساليب تقييم الأثر شبه التجريبي (مثل انقطاع الانحدار أو المطابقة أو الاختلاف في الاختلافات) توافر عينات أكبر من تلك العينات التي يتطلبها التعيين العشوائي في إطار أساس المقارنة. فعلى سبيل المثال، عند استخدام تصميم انقطاع الانحدار، سلط الفصل ٦ الضوء على أنه لا يمكن استخدام إلا الملاحظات حول حد الأهلية. كما أنه يلزم توافر عينة كبيرة بما يكفي حول الحد.

وتكون حسابات القوة الإحصائية مطلوبة لتقدير العينة المطلوبة لعمل مقارنات ذات دلالة حول الحد. ومن ناحية أخرى، توافر الجولات المتعددة للبيانات يمكن أن يساعد على زيادة قدرة تقييم الأثر لحجم عينة محدد. فمثلاً، بيانات خط الأساس حول الحاصل والخصائص الأخرى يمكن أن تساعد على إضفاء المزيد من الدقة على تأثيرات المعالجة. ويمكن أيضاً الانتفاع من توافر التدابير التكرارية للحاصل بعد بدء المعالجة.

فحص مختلف أشكال البرنامج أو ابتكارات التصميم. وفي الأمثلة المقدمة في هذا الفصل، تم تقسيم إجمالي حجم العينة بين مجموعتي المعالجة والمقارنة. وفي بعض الحالات، قد يستلزم سؤال السياسة الرئيسي للتقييم مقارنة آثار البرنامج بين أشكال البرنامج أو ابتكارات التصميم. وإذا كانت هذه هي الحالة، فقد يكون الأثر المتوقع أصغر نسبياً إن كانت مجموعة المعالجة المستفيدة من البرنامج من المقرر مقارنتها مع مجموعة المقارنة التي لا تستفيد من البرنامج على الإطلاق. وعلى هذا النحو، قد يكون الحد الأدنى للتأثير بين مجموعتي المعالجة أصغر من الحد الأدنى للتأثير المطلوب بين مجموعتي المعالجة والمقارنة. وقد يؤدي التوزيع الأمثل للعينة إلى حصولنا على مجموعات معالجة أكبر نسبياً من مجموعة المقارنة.^٩ في تقييمات الأثر مع أذرع المعالجة المتعددة، قد يتطلب الأمر إجراء حسابات القوة الإحصائية لتقدير حجم كل من مجموعتي المعالجة والمقارنة بشكل منفصل، اعتماداً على أسئلة السياسة الرئيسية محل الاهتمام.

المقارنة بين المجموعات الفرعية. وفي حالات أخرى، قد تركز بعض أسئلة تقييم الأثر على تقييم ما إذا كانت آثار البرنامج تتباين بين مجموعتي فرعيتين مختلفتين على أساس، مثلاً، النوع الاجتماعي أو السن أو الدخل. فإذا كانت هذه الحالة، ستكون متطلبات حجم العينة أكبر، ويجب تعديل حسابات القوة الإحصائية بما يوافق ذلك. كمثال، قد يتمحور سؤال السياسة الأساسي حول ما إذا كان برنامج تعليمي معين يؤثر على الطالبات بصورة أكبر من تأثيره على الطلاب. وبديهيًا، ستحتاج إلى عدد كافٍ من الطلاب لكل نوع (ذكور وإناث) في مجموعتي المعالجة والمقارنة لكي يتسنى لك اكتشاف أثر لكل مجموعة فرعية. وإذا آل القرار إلى مقارنة آثار البرنامج بين مجموعتين فرعيتين، فمن شأن ذلك أن يضاعف من حجم العينة المطلوب. فإن اعتبار عدم التجانس بين عدد أكبر من المجموعات (المصنفة حسب العمر، مثلاً) يمكن أن يرفع من حجم العينة المطلوب بشكل كبير. وإذا كانت مثل هذه المقارنات عبر المجموعات مصممة في سياق تقييم أثر معتمد على التعيين العشوائي، فمن المستحسن أيضاً الأخذ بذلك الأمر في الاعتبار عند تنفيذ التوزيع العشوائي، ولا سيما لإجراء التوزيع العشوائي ضمن الكتل أو الطبقات (أي ضمن كلتا المجموعتين الفرعيتين المطلوب مقارنتهما). وفي الواقع العملي، حتى إن كان من غير المقرر عقد مقارنة عبر مجموعات فرعية، فإن التوزيع العشوائي حسب الكتل أو الطبقات يمكن أن يساعد في تعظيم القوة الإحصائية لحجم عينة معين.

تحليل الحاصلات المتعددة. يجب اتخاذ العناية الواجبة عند إجراء حسابات القوة الإحصائية في الحالات التي يسعى تقييم الأثر فيها لاختبار ما إذا كان برنامج معين يثير تغيرات في حواصل متعددة. فإذا شرعت في تضمين عدة حواصل مختلفة في اعتبارك، فسيكون هناك احتمال أعلى نسبيًا بأن تقييم الأثر سيكتشف آثارًا في واحدة من الحاصلات بصورة عرضية. ولمعالجة هذا الأمر، سيحتاج فريق تقييم الأثر إلى النظر في إجراء اختبار دلالة إحصائية مشتركة للتغيرات الواقعة في حواصل مختلفة. عوضًا عن ذلك، يمكن إعداد بعض المؤشرات لمجموعات الحاصلات. وهذه المقاربات الغرض منها معالجة اختبار فرضيات متعددة لحسابات القوة الإحصائية وحجم العينة، وعلى هذا النحو يتعين أخذها في الاعتبار عند تحديد العينة المطلوبة لتقييم الأثر.^{١٠}

التعامل مع الامتثال غير المثالي أو التناقص. غالبًا ما توفر حسابات القوة الإحصائية الحجم الأدنى المطلوب للعينة. وفي الواقع العملي، كثيرًا ما تشير مشكلات التنفيذ ضمنيًا إلى أن حجم العينة الفعلي أصغر من حجم العينة المخطط له. فمثلًا، قد يشير الامتثال غير المثالي ضمنيًا إلى أنه تم قبول البرنامج من قبل نسبة فقط من المستفيدين المستهدفين. وتزداد متطلبات حجم العينة عندما ينشأ الامتثال غير المثالي. بالإضافة إلى ذلك، حتى إذا قبل جميع الأفراد البرنامج، فقد يحدث بعض التناقص في مسح المتابعة إذا تعذر تتبع كل هؤلاء الأفراد. وحتى إذا كان مثل هذا عدم الامتثال أو التناقص عشوائيًا ولا يؤثر على اتساق تقديرات الأثر، فإن هذين الأمرين يؤثران بالفعل في قدرة تقييم الأثر الإحصائية. ولذلك، يُنصح عامةً بإضافة هامش إلى حجم العينة المتوقع عن طريق حسابات القوة الإحصائية لاحتساب هذين العاملين. وعلى نحو مماثل، تكون البيانات منخفضة الجودة مرهونة بوقوع خطأ أكبر في القياس وإثارة التباين بصورة أكبر في الحاصلات محل الاهتمام، ومن ثم فهي تستدعي أيضًا أحجام عينة أكبر.

إن الاعتبارات الأكثر تقدمًا المذكورة في هذا القسم تتجاوز نطاق هذا الكتاب، ولكن الموارد الإضافية المبينة في نهاية هذا الفصل قد تعينك على الإلمام بها. وعلى أرض الواقع، تحتاج فرق التقييم إلى ضم (أو التعاقد مع) خبير باستطاعته إجراء حسابات قوة إحصائية وتقديم النصائح حول المشكلات الأكثر تقدمًا.

مصادر إضافية

- للحصول على المواد المرتبطة بهذا الفصل وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للاطلاع على أمثلة حول كيفية إجراء حسابات القوة الإحصائية في البرنامجين Stata و Optimal Design لحالة برنامج HISP بعينها التي يتناولها هذا الفصل، يرجى مراجعة الموارد الفنية المتوفرة في الموقع الإلكتروني للكتاب (<http://www.worldbank.org/ieinpractice>). تتضمن الموارد الفنية المواد الفنية الإضافية للقراء ذوي الخبرة عن الإحصائيات والمقاييس الاقتصادية.

- للحصول على مناقشات تفصيلية حول أخذ العينات (بما في ذلك الطرق الأخرى مثل أخذ العينات المنهجية أو أخذ العينات متعددة المراحل) فيما وراء المفاهيم الأساسية التي ناقشناها هنا، انظر الموارد التالية:
Cochran, William G. ١٩٧٧. Sampling Techniques, third edition. New York: John Wiley. -
Kish, Leslie. ١٩٩٠. Survey Sampling. New York: John Wiley. -
Lohr, Sharon. ١٩٩٩. Sampling: Design and Analysis. Pacific Grove, CA: Brooks Cole -
Thompson, Steven K. ٢٠٠٢. Sampling, second edition. New York: John Wiley -
أو على مستوى أساسي أكثر،
Kalton, Graham. ١٩٨٣. Introduction to Survey Sampling. Beverly Hills, CA: Sage
يمكن العثور على توجيهات عملية لأخذ العينات في المصادر التالية:
Grosh, Margaret, and Juan Muñoz. ١٩٩٦. "A Manual for Planning and -
Implementing the Living Standards Measurement Study Survey." LSMS Working Paper ١٢٦, World -
Bank, Washington, DC
UN (United Nations). ٢٠٠٠. Household Sample Surveys in Developing and Transition Countries. New -
York: United Nations
Iarossi, Giuseppe. ٢٠٠٦. The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting -
Results, and Influencing Respondents. Washington, DC: World Bank
Fink, Arlene G. ٢٠٠٨. How to Conduct Surveys: A Step by Step Guide, fourth edition. Beverly Hills, CA: -
Sage
- للحصول على جدول بيانات لحسابات القوة الإحصائية بحسب القوة لحجم عينة معين
بعد إدخال خصائص معينة، انظر مركز Inter-American Development Bank Evaluation في قسم Design المندرج تحت
Tools (www.iadb.org/evaluationhub)
- لمعرفة المزيد من المعلومات حول حسابات القوة الإحصائية وحجم العينة، راجع الوحدة ٣ "التصميم" (Design) في World
Bank Impact Evaluation Toolkit (Vermeersch, Rothenbühler, and Sturdy 2012). وتتضمن هذه الوحدة أيضًا دليلًا
لإجراء حسابات قوة إحصائية مسبقة، وأبحاثًا حول حسابات القوة الإحصائية مع متغيرات ثنائية، فضلًا عن مجموعة من المراجع
المفيدة للاستزادة في هذا الموضوع. (<http://www.worldbank.org/health/impacetevaluationtoolkit>)
- للاطلاع على منشورات مدونة أكثر تتحدث عن حسابات القوة الإحصائية، راجع مدونة World Bank Development Impact
(<http://blogs.worldbank.org/impacetevaluations/>). لمناقشة بعض الاعتبارات لحسابات القوة الإحصائية في
التصاميم
- الأكثر تعقيدًا من حالة أساس المقارنة للتعيين العشوائي في وجود امتثال مثالي، انظر ما يلي:
Spybrook, Jessaca, Stephen Raudenbush, Xiaofeng Liu, Richard Congdon, and Andrés Martinez. ٢٠٠٨. -
Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design"
Software. New York: William T. Grant Foundation.
Rosenbaum, Paul. ٢٠٠٩. "The Power of Sensitivity Analysis and Its Limit." Chapter ١٤ in Design of -
Observational Studies, by Paul Rosenbaum. New York: Springer Series in Statistics.

- للتعلم في موضوع اختبار الفرضيات المتعددة، انظر ما يلي:

Duflo, E., R. Glennerster, M. Kremer, T. P. Schultz, and A. S. John. ٢٠٠٧. "Using Randomization in Development Economics Research: A Toolkit." Chapter ٦١ in Handbook of Development Economics, Vol. ٤, ٣٨٩٥-٩٦٢. Amsterdam: Elsevier.

Schochet, P. Z. ٢٠٠٨. Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions. - Prepared by Mathematica Policy Research Inc., for the Institute of Education Sciences, U.S. Department of Education, Washington, DC.

- يتوفر عدد من الأدوات لأولئك المهتمين باستكشاف تصاميم العينة بشكل أعمق. على سبيل المثال طورت W. T. Grant Foundation برنامج Optimal Design المتوفر مجانًا للأبحاث متعددة المستويات والأبحاث الطولية، وهو مفيد في تحليل القوة الإحصائية في ظل وجود المجموعات العنقودية. ويمكن تنزيل برنامج Optimal Design والدليل الخاص به من الموقع الإلكتروني <http://hlmssoft.net/od>.

ملاحظات

١. على وجه التحديد، تُسحب العينات من أطر تحديد العينة. وفي مناقشتنا، نفترض أن إطار تحديد العينة يتداخل بشكل مثالي مع المجتمع الإحصائي.
٢. كما ناقشنا في الجزء ٢، يفترض الامتثال أن كل الوحدات المعيّنة إلى مجموعة المعالجة تُعالج على عكس جميع الوحدات المعيّنة إلى مجموعة المقارنة التي لا تُعالج.
٣. في هذا السياق، لا يشير المجتمع الإحصائي إلى تعداد السكان في الدولة، بل إلى المجموعة الكاملة من الأطفال الذين يقعون في محل اهتمامنا (ما يُطلق عليه "المجتمع الإحصائي محل الاهتمام").
٤. هذا المنطق البديهي يُصاغ بموجب نظرية تُسمى نظرية الحد المركزي. رسميًا، للحصول على y ، تنص نظرية الحد المركزي على أن متوسط العينة \bar{y} في المتوسط يشكل تقديرًا صالحًا لمتوسط المجتمع الإحصائي. وبالإضافة إلى ذلك، بالنسبة لعينة حجمها n ولتباين مجتمع إحصائي σ^2 ، يكون تباين متوسط العينة متناسبًا مع حجم العينة بصورة عكسية:
$$var(\bar{y}) = \frac{\sigma^2}{n}$$
٥. فكلما ازداد حجم العينة n ، يميل تباين تقديرات العينة إلى القيمة ٠. بمعنى آخر، نحصل على تقدير أدق للمتوسط في العينات الكبيرة عنه بالنسبة للعينات الصغيرة.
٥. في أغلب الأحيان، تستدعي الضرورة تخصيص الفوائد حسب المجموعات العنقودية وفقًا للاعتبارات الاجتماعية أو السياسية التي تجعل التعيين العشوائي ضمن المجموعات العنقودية مستحيلًا. وفي سياق تقييم الأثر، غالبًا ما يصبح التجميع (العنقودي) أمرًا حتميًا بسبب التأثيرات غير المباشرة المحتملة أو انتشار فوائد البرنامج بين الأفراد ضمن المجموعات العنقودية. راجع مناقشتنا الواردة في الفصل ١١.
٦. عند حوسبة القوة الإحصائية اعتمادًا على بيانات خط الأساس، فإن ترابط الحاصل الذي يحدث بمرور الوقت يجب أيضًا أن يؤخذ في الاعتبار في حسابات القوة الإحصائية.




٧. على سبيل المثال، قدّم Spybrook and others (٢٠٠٨) برنامج Optimal Design، وهو برنامج ذو واجهة بسيطة لإجراء حسابات القوة الإحصائية.
٨. بصفة عامة، من المحبذ توفير مجموعتي معالجة ومقارنة متساويتين في الحجم. بالنظر إلى عدد الملاحظات في عينة ما، تكون القوة الإحصائية في أقصى حدودها بالطبع عند تعيين نصف الملاحظات إلى مجموعة المعالجة ونصفها إلى مجموعة المقارنة. ومع ذلك، ليس شرطًا أن تكون مجموعتا المعالجة والمقارنة متساويتين في الحجم دائمًا. انظر المناقشة في نهاية الفصل.
٩. يمكن أيضًا أخذ تكاليف المعالجة في الاعتبار، ويؤدي ذلك إلى عدم تساوي مجموعتي المعالجة والمقارنة. راجع، مثلاً، Duflo and others (٢٠٠٧).
١٠. راجع، على سبيل المثال، Duflo and others (٢٠٠٧) أو Schochet (٢٠٠٨).

المراجع

- Cochran, William G. 1977. *Sampling Techniques*, third edition. New York: John Wiley & Sons.
- Duflo, E., R. Glennerster, and M. Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, Vol. 4, edited by T. Paul Schultz and John Strauss, 3895–962. Amsterdam: Elsevier.
- Fink, Arlene G. 2008. *How to Conduct Surveys: A Step by Step Guide*, fourth edition. Beverly Hills, CA: Sage.
- Grosh, Margaret, and Paul Glewwe, eds. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington, DC: World Bank.
- Grosh, Margaret, and Juan Muñoz. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey." LSMS Working Paper 126, World Bank, Washington, DC.
- Iarossi, Giuseppe. 2006. *The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents*. Washington, DC: World Bank.
- Kalton, Graham. 1983. *Introduction to Survey Sampling*. Beverly Hills, CA: Sage.
- Kish, Leslie. 1995. *Survey Sampling*. New York: John Wiley.
- Lohr, Sharon. 1999. *Sampling: Design and Analysis*. Pacific Grove, CA: Brooks Cole.
- Rosenbaum, Paul. 2009. *Design of Observational Studies*. New York: Springer Series in Statistics.
- Schochet, P. Z. 2008. *Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions*. NCEE 2008-4018. National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences.
- Washington, DC: U.S. Department of Education.
- Spybrook, Jessaca, Stephen Raudenbush, Xiaofeng Liu, Richard Congdon, and Andrés Martinez. 2008. *Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software*. New York: William T. Grant Foundation.





Thompson, Steven K. 2002. Sampling, second edition. New York: John Wiley. Vermeersch, Christel, Elisa Rothenbühler, and Jennifer Sturdy. 2012. *Impact Evaluation Toolkit: Measuring the Impact of Results-Based Financing on Maternal and Child Health*. World Bank, Washington, DC. <http://www.worldbank.org/health/impactevaluationtoolkit>.

العثور على مصادر بيانات مناسبة

أنواع البيانات المطلوبة

سننتقل في هذا الفصل إلى مصادر البيانات المتنوعة التي يمكن أن تستخدمها تقييمات الأثر. وناقش أولاً مصادر البيانات الحالية، لاسيما البيانات الإدارية، ونذكر بعض الأمثلة لتقييمات الأثر التي تستخدم البيانات الحالية. بما أن العديد من التقييمات تتطلب جمع بيانات جديدة، فإننا سناقش بعد ذلك الخطوات الضرورية لجمع بيانات مسح جديدة. ويتعين استيعاب هذه الخطوات بوضوح لما في ذلك من ضمان أن تقييم الأثر الخاص بك يعتمد على بيانات عالية الجودة لا تشكل أي مخاطر على تصميم التقييم. وكخطوة أولى، ستحتاج إلى الشروع في التخطيط لاستبيان ملائم. وفي الوقت نفسه، عليك طلب المساعدة من شركة أو وكالة حكومية متخصصة في جمع البيانات. ومن ثم سيتولى الكيان المسؤول عن جمع البيانات توظيف عدد من العاملين الميدانيين وتدريبهم، ويطرحون الاستبيان على سبيل اختبار عملي. وبعد إجراء التعديلات الضرورية، ستكون الشركة أو الوكالة قادرة على متابعة العمل الميداني وجمع البيانات وتحويلها رقمياً ومعالجتها، قبل أن يتولى فريق التقييم تقديمها وتخزينها وتحليلها. إن البيانات عالية الجودة مطلوبة لتقييم أثر التدخل على الحاصل محل الاهتمام. وتوفر سلسلة النتائج التي خضعت للمناقشة في الفصل ٢ أساساً لتحديد أي المؤشرات يجب قياسها ومتى. ومن ثم تكون المؤشرات المطلوبة عبر سلسلة النتائج.

المفهوم الأساسي

تكون المؤشرات مطلوبة
عبر سلسلة الحاصلات
لقياس الحاصلات
النهائية، والحاصلات
المتوسطة وكذلك فوائد
البرنامج وجودة التنفيذ.

البيانات حول الحاصلات. أهم شيء يتعين الحصول عليه هو البيانات حول مؤشرات الحاصلات المتأثرة بشكل مباشر بالبرنامج. وترتبط مؤشرات الحاصلات هذه بالأهداف التي يسعى البرنامج لتحقيقها. فكما ناقشنا في الفصل ٢، من الأفضل اختيار مؤشرات حاصلات توفر السمات الخمس الرئيسية: محددة وقابلة للقياس وقابلة للإسناد وواقعية ومستهدفة. ومع ذلك، يجب عدم استخدام تقييم الأثر بشكل حصري لقياس تلك الحاصلات التي يؤدي البرنامج إليها بشكل مباشر. إن البيانات حول مؤشرات الحاصلات التي يؤثر فيها البرنامج بشكل غير مباشر أو المؤشرات التي تحصد تأثيرات البرنامج غير المقصودة، ستعظم من قيمة المعلومات التي يولدها تقييم الأثر وكذلك فهم الفعالية الكلية للبرنامج.

بيانات حول الحاصلات المتوسطة: بالإضافة إلى أنواع البيانات السالف ذكرها، فإن للبيانات حول الحاصلات المتوسطة فائدة كبيرة في دعم فهمنا للقنوات التي من خلالها كان للبرنامج الخاضع للتقييم أثر - أو لم يكن له أثر - فيما يتعلق بالحاصلات النهائية محل الاهتمام. لقد جرت العادة على إجراء تقييمات الأثر على فترات زمنية متعددة، وعليك تحديد الموعد المناسب لقياس مؤشرات الحاصلات. ومع وضع سلسلة النتائج في الاعتبار، يمكنك تنظيم تسلسل هرمي لمؤشرات الحاصلات التي تتراوح بين المؤشرات قصيرة المدى التي يمكن قياسها بينما لا يزال المشاركون في البرنامج (مثل معدلات الحضور المدرسي التي تُجمع في مسح متابعة قصير المدى في سياق برنامج تعليمي) إلى المسوحات طويلة المدى (مثل الإنجاز الطلابي أو حاصلات سوق العمل التي يمكن قياسها في مسح متابعة طويل المدى بعد انتهاء مشاركة المستفيدين من البرنامج). ومن أجل قياس الأثر بطريقة مقنعة على فترات زمنية متباعدة، يجب توافر بيانات تبدأ من خط الأساس قبل تنفيذ البرنامج أو الابتكار المراد تقييمه. يسلط القسم الذي يدور حول توقيت التقييمات في الفصل ١٢ الضوء على موعد جمع البيانات.

وكما ناقشنا في الفصل ١٥ في سياق حسابات القوة الإحصائية، قد لا تستجيب بعض المؤشرات لتقييم الأثر في حال كانت العينات صغيرة. ومن ثم، قد يتطلب الأمر عينات كبيرة تنطوي على تكاليف باهظة من أجل اكتشاف آثار مؤشرات الحاصلات التي تكون شديدة التغير أو أحداثًا نادرة أو تكون على الأرجح متأثرة هامشيًا فحسب بتدخل ما . على سبيل المثال، تحديد أثر تدخل ما على معدلات الوفيات النفاسية (معدلات وفيات الأمهات) لن يكون ميسورًا إلا في عينة تحتوي على عشرات الآلاف من النساء الحوامل، بما أن الوفاة (حمدًا لله) حدث نادر. وفي مثل هذه الحالة، قد يكون من الضروري إعادة تركيز تقييم الأثر على مؤشرات أكثر وسطية في الحدث، تلك المرتبطة بالحاصلات النهائية وتتوفر لها قوة إحصائية كافية لاكتشاف آثارها. وفي حالة كان التدخل يهدف إلى الحد من الوفيات النفاسية، فقد يكون أحد المؤشرات المتوسطة مرتبطًا بمعدل استخدام الخدمات الاستشفائية خلال الحمل والولادة في المؤسسات، التي ترتبط بالوفيات. إن حسابات القوة الإحصائية التي ناقشناها في الفصل ١٥ يمكن أن تلقي بالضوء على المؤشرات التي يمكن اكتشاف الآثار عليها، والمؤشرات التي قد يكون من الصعب اكتشاف الآثار عليها من دون استخدام عينات كبيرة للغاية.

البيانات حول أنشطة البرنامج ومخرجاته: إن المؤشرات مطلوبة أيضًا لجزء سلسلة النتائج الذي يسرد الأنشطة والمخرجات للبرنامج. وعلى وجه الخصوص، يمكن أن توفر بيانات متابعة البرنامج معلومات أساسية حول طريقة تقديم التدخل، بما يشمل ذلك تحديد مَن المستفيدين وأي فوائد أو مخرجات للبرنامج قد يتلقونها. وكحد أدنى، يجب توفير بيانات المتابعة لمعرفة متى يبدأ البرنامج ومَن يحصل على الفوائد وكذلك لتوفير قياس لقدرة التدخل أو جودته. وتزداد أهمية ذلك الأمر في الحالات التي قد لا يتم فيها تقديم البرنامج إلى جميع المستفيدين على النسق نفسه من حيث المحتوى أو الجودة أو المدة. ويتعين الوصول إلى فهم جيد لمدى تقديم التدخل وطرحه على المستفيدين وهذا أمر أساسي لتفسير نتائج تقييم الأثر التي قد تسلط الضوء على فعالية البرنامج المنفذ طبقًا للتصميم أو المنفذ في ظل بعض العيوب في التصميم.

بيانات إضافية: قد يتطلب تقييم الأثر بيانات أخرى، ويعتمد ذلك على المنهجية المستخدمة. وكذلك قد تستدعي الحاجة توافر بيانات حول عوامل أخرى ربما تؤثر في النتيجة محل الاهتمام، وذلك بهدف ضبط المؤثرات الخارجية. وهذا الجانب مهم بشكل خاص عند استخدام طرق تقييم تعتمد على الافتراضات بشكل يفوق الطرق المعتمدة على التوزيع العشوائي. وفي بعض الأحيان، يكون من الضروري أيضًا الحصول على بيانات حول الحاصل والعوامل الأخرى بمرور الوقت لحساب الاتجاهات كما هي الحال عند اتباع طريقة الاختلاف في الاختلافات. يتعين أيضًا مراعاة العوامل الأخرى والاتجاهات السابقة لأن ذلك من شأنه أن يساعد على زيادة القوة الإحصائية. وحتى في ظل التعيين العشوائي، فإن لتوافر بيانات حول خصائص أخرى يمكن أن يسمح بتقدير تأثيرات المعالجة بشكل أكثر دقة. ويمكن استخدام هذه البيانات لتضمين وحدات ضابطة إضافية أو تحليل عدم التجانس لتأثيرات البرنامج على الخصائص ذات الصلة. سيؤثر التصميم المختار لتقييم الأثر أيضًا على متطلبات البيانات. فعلى سبيل المثال، عند اختيار طريقة المطابقة أو الاختلاف في الاختلافات، سيكون من الضروري جمع بيانات حول مجموعة عريضة من الخصائص لكل من مجموعتي المعالجة والمقارنة ما يجعل من الممكن تنفيذ مجموعة متنوعة من اختبارات الدقة، كما هو موصوف في الجزء ٢ أو الفصل ١١ (انظر الجدول ١١-٢).

بالنسبة لكل تقييم، يكون من المفيد تصميم مصفوفة تضم السؤال محل الاهتمام، ومؤشرات الحاصل لكل سؤال، والأنواع الأخرى من المؤشرات المطلوبة، ومصدر البيانات، كما هو محدد في الشكل ١-٢ الذي يوضح سلسلة النتائج في الفصل ٢. ويشكل إعداد خطة تقييم الأثر وخطة التحليل القبلية فرصتين مهمتين لتخصيص قائمة دقيقة للمؤشرات الأساسية المطلوبة لتقييم الأثر.

استخدام البيانات الكمية الحالية

من أول الأسئلة الواجب وضعها في الاعتبار عند تصميم تقييم الأثر هو "ما مصادر البيانات المستخدمة في هذا التقييم؟". وهناك اعتبار أساسي هنا وهو تحديد ما إذا كان تقييم الأثر سيعتمد على البيانات الحالية أم إنه يتطلب جمع بيانات جديدة.

فبعض البيانات الحالية تكاد تكون دائمًا مطلوبة في مستهل تقييم الأثر من أجل تقدير قيم أساس المقارنة للمؤشرات أو إجراء حسابات القوة الإحصائية، كما ناقشنا في الفصل ١٥. وبما يتجاوز مراحل التخطيط، يمكن لتوافر البيانات الحالية أن يحد بشكل كبير من تكلفة إجراء تقييم الأثر. في حين أن البيانات الحالية، لا سيما البيانات الإدارية، تكون منقوصة الاستخدام بشكل محتمل في تقييم الأثر بشكل عام، فإن جدوى استخدام البيانات الحالية لاحتياجات تقييم الأثر يجب تقديرها بعناية.

وبالطبع، كما ناقشنا في الفصل ١٢، جمع البيانات غالبًا ما يكون عنصر التكلفة الأكبر عند تنفيذ أي تقييم أثر. ومع ذلك، من أجل تحديد ما إذا كانت هناك إمكانية لاستخدام بيانات حالية في تقييم أثر معين، يجب التعامل مع مجموعة من الأسئلة:

- **أخذ العينات.** هل البيانات الحالية متوفرة لكل من مجموعتي المعالجة والمقارنة؟ هل العينات الحالية مسحوبة من إطار تحديد عينة يتوافق مع المجتمع الإحصائي محل الاهتمام؟ هل الوحدات المسحوبة من إطار تحديد العينة قائمة على إجراء أخذ العينات الاحتمالية؟
- **حجم العينة.** هل مجموعات البيانات الحالية كافية للكشف عن التغيرات في مؤشرات الحاصلات بقوة إحصائية كافية؟ وتعتمد الإجابة عن هذا السؤال على مؤشرات الحاصلات المختارة وكذلك على نتائج حسابات القوة الإحصائية التي ناقشناها في الفصل ١٥.
- **توافر بيانات خط الأساس.** هل البيانات الحالية متوفرة لكل من مجموعتي المعالجة والمقارنة قبل طرح البرنامج أو الابتكار المقرر تقييمه؟ إن توافر بيانات خط الأساس مهم لتوثيق التوازن في الخصائص القبلية للبرنامج بين مجموعتي المعالجة والمقارنة عند استخدام طرق قائمة على التوزيع العشوائي، ولها دور حيوي في تنفيذ التصاميم شبه التجريبية.
- **الوتيرة.** هل يتم جمع البيانات الحالية بشكل متواتر بما فيه الكفاية؟ هل هي متوفرة لجميع الوحدات في العينة بمرور الوقت، بما في ذلك في الأوقات التي يجب فيها قياس مؤشرات الحاصلات وفقًا لسلسلة النتائج ومنطق التدخل؟
- **النطاق.** هل تحتوي البيانات على جميع المؤشرات المطلوبة للإجابة عن أسئلة السياسة محل الاهتمام، بما يشمل ذلك مؤشرات الحاصلات الأساسية والحاصلات المتوسطة محل الاهتمام؟

• *الارتباطات بمعلومات متابعة البرنامج.* هل يمكن ربط البيانات الحالية ببيانات متابعة تنفيذ البرنامج، بما يشمل ذلك من أجل ملاحظة أي وحدات تكون في مجموعتي المعالجة والمقارنة، وما إذا كانت كل الوحدات معينة إلى مجموعة المعالجة التي تلقت الفوائد نفسها؟

• *المعرفات الفريدة.* هل توجد معرفات فريدة لربط مصادر البيانات المختلفة؟

كما تبرز الأسئلة الموضحة أعلاه، متطلبات البيانات الحالية مهمة للغاية وليس من الشائع أن تكون البيانات الحالية كافية لتقييمات الأثر. وعلى الرغم من ذلك، في ظل التوسع سريع الوتيرة لنطاق نظم المعلومات وامتداد تأثيرها وكذلك التحول الشامل صوب عالم يتم فيه تخزين البيانات الواردة عن مجموعة عريضة ومتنوعة من المصادر بشكل روتيني، يتزايد عدد تقييمات الأثر التي يمكنها توظيف البيانات الحالية. ويمكن استخدام مجموعة متنوعة من المصادر المحتملة للبيانات الحالية لتقييم الأثر، على سبيل المثال لا الحصر بيانات التعداد السكاني أو المسوحات الوطنية أو البيانات الإدارية.

يمكن أن توفر بيانات التعداد السكاني بيانات شاملة عن المجتمع الإحصائي بالكامل. وبالإستطاعة استخدامها في تقييمات الأثر عندما تتوفر على مستوى مجزأ بشكل كافٍ وتتضمن تفاصيل توضح أي الوحدات تنتمي إلى مجموعة المعالجة وأنها تنتمي إلى مجموعة المقارنة، مثل المعرفات الجغرافية أو الشخصية. وعلى الرغم من أن بيانات التعداد السكاني تُجمع ميدانيًا بشكل أقل تواترًا وعادة ما تنطوي على مجموعة صغيرة من المؤشرات الأساسية، فهي أحيانًا تُجمع لتغذية نظم معلومات أو سجلات توفر الأساس للبرامج العامة المستهدفة، بما يتضمن ذلك المعرفات الفريدة التي يمكنها دعم تحقيق الارتباط مع مجموعات البيانات الحالية الأخرى.

يمكننا أيضًا أن نضع في اعتبارنا استخدام المسوحات ذات التمثيل الوطني مثل مسوحات الأسر المعيشية ومسوحات قياس مستويات المعيشة ومسوحات القوى العاملة والمسوحات الديموغرافية والصحية والمسوحات المؤسسية ومسوحات المرافق. وقد تحتوي هذه الموارد على مجموعة شاملة من متغيرات الحاصل، ولكنها نادرًا ما تحتوي ما يكفي من الملاحظات بالنسبة لمجموعتي المعالجة والمقارنة لإجراء تقييم أثر. افترض، مثلاً، أنك مهتم بتقييم برنامج وطني كبير يصل إلى ١٠ بالمائة من الأسر المعيشية في دولة معينة. فإذا كان يتم تنفيذ مسح ذي تمثيل وطني يضم ٥ آلاف أسرة معيشية كل عام، فقد يشمل ٥٠٠ أسرة معيشية تقريبًا تتلقى البرنامج محل الاهتمام. هل هذه العينة كبيرة بما يكفي لإجراء تقييم أثر؟ تستطيع حسابات القوة الإحصائية الإجابة عن هذا السؤال ولكن في العديد من الحالات تكون الإجابة لا.

بالإضافة إلى استكشاف ما إذا كان يمكنك استخدام مسوحات حالية أم لا، يجب عليك أيضًا بيان ما إذا كانت هناك أي جهود مزمنة لجمع بيانات وطنية حديثة. إذا كان من المخطط إجراء مسح سيغطي المجتمع الإحصائي محل الاهتمام، فقد تكون قادرًا على إدخال سؤال أو سلسلة من الأسئلة كجزء من هذا المسح. فإذا كان من المخطط بالفعل إجراء مسح يقيس المؤشرات المطلوبة، فقد تكون هناك فرص للحصول على عينة تفوق الحاجة لمجتمع إحصائي معين لضمان التغطية المناسبة لمجموعتي المعالجة والمقارنة وتكييفها مع تقييم الأثر. على سبيل المثال، تقييم الصندوق الاجتماعي النيكاراغوي نفذ "دراسة قياس مستويات المعيشة الوطنية" مع عينة إضافية من المستفيدين (Pradhan and Rawlings 2002).

تُجمع البيانات الإدارية عادةً من قبل وكالات عامة أو خاصة كجزء من عملياتها الدورية وفي الغالب تكون بشكل أكثر دورية لمتابعة الخدمات المقدمة أو تسجيل التفاعلات مع المستخدمين. وفي بعض الحالات، تحتوي البيانات الإدارية على مؤشرات الحاصلات المطلوبة لإجراء تقييم أثر. فمثلاً، تجمع الأنظمة التعليمية السجلات المدرسية بشأن التحاق الطلاب أو حضورهم أو درجات امتحانهم، ويمكنها أيضًا جمع معلومات حول المدخلات المدرسية والمعلمين. وعلى نحو مماثل، تستطيع الأنظمة الصحية جمع بيانات حول الخصائص والأماكن للمراكز الصحية وإمداد الخدمات الصحية وتخصيص الموارد. ويمكنها أيضًا دمج البيانات التي يتم جمعها في المراكز الصحية حول السجلات الطبية للمرضى والقياسات الأنثروبومترية وسجلات التلقيح، وبشكل أكثر توسعًا، بيانات حول حدوث الأمراض والإحصاءات الحيوية. وتجمع وكالات المرافق العامة بيانات حول استخدام المياه أو الكهرباء. وقد تجمع الوكالات المعنية بالضرائب بيانات حول الدخل والضرائب. وتجمع أنظمة النقل بيانات حول الركاب وأوقات السفر. أما وكالات الأنظمة المالية، فتجمع بيانات حول معاملات العملاء أو تاريخهم الائتماني. كل هذه مصادر لبيانات حالية يمكن الاستفادة منها بشكل محتمل لإجراء تقييمات الأثر. وفي بعض الأحيان تتضمن سلسلة زمنية طويلة يمكنها المساعدة في تتبع الوحدات بمرور الزمن.

تقييم توافر البيانات والجودة يُعد ركنًا أساسيًا عند اعتبار صلاحية الاعتماد على البيانات الإدارية من عدمها. في بعض الحالات، تكون مصادر البيانات الإدارية ذات موثوقية أكبر بالمقارنة مع بيانات المسوحات. على سبيل المثال، اكتشفت دراسة في مالايو أن المجيبين بالغوا في تقدير الحضور المدرسي والالتحاق بالمدارس في البيانات المبلغ عنها ذاتيًا الواردة عن مسح أسر معيشية بالمقارنة مع السجلات الإدارية التي تم الحصول عليها من المدارس، وبالتالي ستكون نتائج تقييم الأثر أكثر موثوقية إذا كانت معتمدة على البيانات الإدارية (Baird and Özler 2012). إلا أنه في الوقت نفسه، في سياقات كثيرة، تُجمع البيانات الإدارية عن طريق عدد كبير من مقدمي الخدمات ما قد يجعلها متفاوتة في الجودة. وبالتالي ينبغي لنا تقييم موثوقيتها بشكل كامل قبل اتخاذ قرار الاعتماد على البيانات الإدارية لتقييم الأثر أم لا. وهناك جانب شديد الأهمية علينا مراعاته وهو ضمان أن المعارف الفريدة موجودة ومعمول بها لربط البيانات الإدارية بمصادر البيانات الأخرى، بما في ذلك بيانات متابعة البيانات التي توثق أي وحدات تلقت فوائد البرنامج. عندما توجد مثل هذه المعارف، مثل أرقام التعريف الوطنية المستخدمة بشكل متسق، قد نتجنب بآغا طويلاً من العمل على تحضير البيانات وتنقيتها. وفي جميع الحالات، تشكل حماية سرية المعلومات ركنًا أساسيًا من بروتوكول تحضير البيانات وإدارتها. وتنطبق أيضًا المبادئ الأخلاقية التي تنص على حماية العناصر البشرية (انظر المناقشة في الفصل ١٣) على استخدام البيانات الحالية.

هناك بعض التقييمات الرجعية المهمة التي اعتمدت على السجلات الإدارية (Galiani, Gertler, and Schargrotsky 2005) في سياسة المياه في الأرجنتين، و [2008] Ferraz and Finan في عمليات التدقيق وأداء السياسيين، و Chetty, Friedman, and Saez [2013] في الائتمانات الضريبية في الولايات المتحدة). يوفر المربع ١٦،١ مثالاً على تقييم أثر صحي في الأرجنتين. يوضح المربع ١٦-٢ استخدام البيانات الإدارية في تقييم الأثر لبرنامج تحويل نقدي في هندوراس. في بعض الحالات، يمكن جمع البيانات المطلوبة لإجراء تقييم الأثر عن طريق طرح معلومات أو أنظمة بيانات إدارية جديدة. ومثل هذا الطرح من الممكن تنسيقه مع تنفيذ تصميم التقييم بحيث يتم جمع مؤشرات الحاصلات لمجموعتي المعالجة والمقارنة في مرات متفرقة. وحينها قد يلزم الانتهاء من إعداد نظم المعلومات قبل إطلاق التدخلات الجديدة بحيث تستخدم المراكز الإدارية في مجموعة المقارنة نظام المعلومات الجديد قبل الاستفادة من التدخل المقرر تقييمه. ومع احتمالية تباين جودة البيانات الإدارية، تنشأ أهمية التدقيق والتحقق من الصلاحية الخارجيين من أجل ضمان موثوقية التقييم. ومن شأن جمع بيانات تقييم الأثر من خلال المصادر الإدارية بدلاً من إجراء المسوحات أن يحد بشكل كبير من تكلفة التقييم، ولكن لا يتيسر ذلك في كل الحالات.

المربع ١٦-١: إنشاء مجموعة بيانات في تقييم Plan Nacer في الأرجنتين

مهمة غاية في الصعوبة تمثلت في الحصول على السجلات الطبية حول الولادات الفردية في مستشفيات الولادة العامة. صحيح أن كل مستشفى ولادة قد جمعت بيانات حول حوائط رعاية ما قبل الولادة والولادات، إلا أن نصف السجلات فقط كان رقمياً. أما البقية، فكانت ورقية، ما تطلب من فريق التقييم دمج السجلات الورقية في النظام المحوسب.

وبوجه عام، تمكن الفريق من جمع قاعدة بيانات شاملة لحوالي ٧٨٪ من الولادات التي حدثت في أثناء فترة التقييم. وأثمر ذلك مجموعات بيانات كبيرة سمحت للفريق بفحص أثر برنامج Plan Nacer على أحداث نادرة نسبياً، مثل وفيات الأطفال حديثي الولادة. وكان تحقيق ذلك مستحيلاً بشكل أساسي في التقييمات التي تنطوي على عينات أصغر حجماً يتم جمعها من خلال المسوحات. وقد توصل التقييم إلى أن المستفيدين من برنامج Plan Nacer كانت لهم فرصة أقل بنحو ٧٤٪ للتعرض لوفاة الأطفال حديثي الولادة في المستشفى مقارنة بغير المستفيدين.

عند تقييم برنامج التمويل الصحي القائم على النتائج في الأرجنتين، جمع Plan Nacer, Gertler, Giovagnoli, and Martinez (2014) بيانات إدارية من عدة مصادر لتشكيل قاعدة بيانات كبيرة وشاملة لغرض التحليل. من بعد فشل عدة إستراتيجيات سابقة للتقييم، تحول الباحثون إلى مقارنة المتغيرات المساعدة (instrumental variables). وتطلب ذلك قدرًا كبيرًا من البيانات من عموم سجلات الميلاد في الولايات السبع الخاضعة للدراسة.

فقد كان الباحثون يحتاجون إلى بيانات بخصوص حوائط رعاية ما قبل الولادة والولادات، ومثل تلك البيانات يمكن العثور عليها في سجلات المواليد والمستشفيات العامة. ثم سيحتاجون إلى تحديد ما إذا كانت الأم مستفيدة من برنامج Plan Nacer وما إذا كانت العيادة التي زارتها كانت متضمنة في هذا البرنامج في وقت الزيارة أم لا. ولإنشاء قاعدة بيانات تستوعب كل هذه المعلومات، عكف فريق التقييم على ربط خمسة مصادر بيانات مختلفة، بما يتضمن ذلك قواعد بيانات مستشفيات الولادة العامة، وبيانات تنفيذ برنامج Plan Nacer،

والسجلات الصيدلانية، وتعداد السكان لعام ٢٠٠١، والمعلومات الجغرافية للمرافق الصحية. وواجه الفريق

المصدر: Gertler, Giovagnoli, and Martinez 2014.

المربع ٢-١٦: استخدام بيانات التعداد لإعادة تقييم برنامج PRAF في هندوراس

ومن ثم حصل الباحثون على عينة كبيرة الحجم سمحت لهم باختبار دقة الاستنتاجات بالإضافة إلى التأثيرات غير المباشرة. وعلاوة على ذلك، بما أن الباحثين كانت بحوزتهم بيانات التعداد من كل البلديات، فقد تمكنوا من تطبيق تصميمين مختلفين من تصاميم انقطاع الانحدار (RDD) باستخدام مجموعات مقارنة بديلة. بالنسبة لتصميم RDD الأول، استخدم الباحثون نقطة انقطاع الأهلية، أما للثاني فقد استخدموا الحدود البلدية.

وكما هي الحال مع تقييم الأثر السابق، اكتشف Galiani and McEwan آثارًا إيجابية ذات دلالة إحصائية ناتجة عن البرنامج. إلا أن تقديرتهما أشارت ضمناً إلى أن برنامج PRAF كان له أثر أكبر بكثير من الأثر الذي اكتشفه التقييم الأصلي. فقد وجد أن برنامج PRAF قد زاد من معدل الالتحاق بالمدارس بالنسبة للأطفال المؤهلين بنسبة ١٢٪ أكثر من أولئك الذين ينتمون إلى مجموعة المقارنة. وبصفة عامة، تم التأكد من دقة الاستنتاجات من خلال النتائج التي تم التوصل إليها من تصميمي انقطاع الانحدار المتبادلتين.

كان يهدف برنامج Programa de Asignación Familiar (PRAF) في هندوراس إلى تحسين الحاصلات التعليمية والصحية للأطفال الصغار الذين يعيشون في سياقات فقيرة. ومن ثم كان يقدم التحويلات النقدية إلى الأسر المعيشية المؤهلة بشرط الحضور المدرسي المنتظم والزيارات الدورية إلى المركز الصحي. وبدأ البرنامج في ١٩٩٠. وتم تضمين مكون تقييم في المرحلة الثانية من برنامج PRAF في ١٩٩٨. وأفاد Glewwe and Olinto (2004) and Morris and others (2004) أن هناك آثار محتملة على الحاصلات التعليمية والصحية.

وبعد بضع سنوات، أعاد Galiani and McEwan (2013) تقييم أثر البرنامج باستخدام مصدر بيانات مختلف. في حين أن تقييم الأثر الأصلي تم جمع له بيانات مسح من ٧٠ بلدية من أصل ٢٩٨ بلدية، استخدم Galiani and McEwan بيانات من تعداد هندوراس لعام ٢٠٠١. وقاما بدمج البيانات على المستوى الفردي والأسر المعيشية من التعداد مع البيانات على مستوى البلديات بشأن المجتمعات الإحصائية المعالجة.

المصدر: Galiani and McEwan 2013.

على الرغم من أن البيانات الحالية ليست كافية لإتمام كل خطوات تقييم الأثر، فإنه في بعض الأحيان يمكن استخدامها لإتمام أجزاء من تقييم الأثر. فمثلاً، في بعض الحالات، تجمع البرامج بيانات مستهدفة تفصيلية عن المستفيدين المحتملين من أجل تحديد الأشخاص المؤهلين. أو قد تتوفر بيانات التعداد لفترة وجيزة بعد طرح البرنامج. وفي مثل هاتين الحالتين، يمكن أحياناً استخدام هذه البيانات الحالية من أجل توثيق توازن خط الأساس في الخصائص السابقة لتطبيق البرنامج في مجموعتي المعالجة والمقارنة، على الرغم من أنه لا تزال هناك حاجة لجمع بيانات متابعة إضافية بهدف قياس المجموعة الأوسع لمؤشرات الحاصلات.

من النادر نسبيًا أن تكون البيانات الحالية كافية لإجراء تقييم الأثر بالكامل. فإذا كانت البيانات الإدارية غير كافية لإجراء تقييمك، فعلى الأرجح ستكون مضطرًا إلى الاعتماد على بيانات المسح. وكنتيجه لذلك، يجب عليك بنسبة كبيرة وضع ميزانية لجمع بيانات جديدة. وعلى الرغم من أن جمع البيانات غالبًا ما يمثل التكلفة الأكبر في تقييم الأثر، فإنه أيضًا يمكن أن يكون استثمارًا مرتفع المنفعة تعتمد عليه جودة التقييم في الغالب. فمع جمع البيانات الجديدة، تتوفر المرونة التي تضمن قياس كل المؤشرات الضرورية لإجراء تقدير شامل لأداء برنامج ما.

وتتطلب معظم تقييمات الأثر جمع بيانات مسح، بما في ذلك مسح خط أساس قبل تنفيذ التدخل أو الابتكار المراد تقييمه، ومسح متابعة بعد تنفيذ هذا التدخل أو الابتكار. وقد تأتي بيانات المسح بأشكال مختلفة اعتمادًا على البرنامج المراد تقييمه ووحدة التحليل. فمثلًا، تستخدم المسوحات المؤسسية الشركات كوحدة رئيسية للملاحظة، وتستخدم مسوحات المرافق المراكز الصحية أو المدارس كوحدة رئيسية للملاحظة، وعلى هذا المنوال، مسوحات الأسر المعيشية تستخدم الأسر كوحدة رئيسية للملاحظة. وتعتمد معظم التقييمات على مسوحات الأفراد أو الأسر المعيشية كمصدر بيانات رئيسي. وفي هذا القسم، سنستعرض بعض المبادئ العامة لجمع بيانات المسح. على الرغم من أن هذه المبادئ ترتبط في المقام الأول بمسوحات الأسر المعيشية، فإنها تنطبق أيضًا على معظم الأنواع الأخرى لبيانات المسح.

تتمثل الخطوة الأولى في تحديد مقاربة أخذ العينة وذلك بناءً على ما إذا كان من الممكن استخدام البيانات الحالية أو كان الأمر يتطلب جمع بيانات مسح جديدة (كما ناقشنا في الفصل ١٥). بمجرد أن تقرر مسألة جمع بيانات المسح من أجل التقييم، ستحتاج إلى

- تحديد مَنْ المسؤول عن جمع البيانات
- تطوير أداة جمع البيانات واختبارها
- إجراء العمل الميداني وتطبيق تدابير ضبط الجودة
- معالجة البيانات وتخزينها

ويتم تنفيذ هذه الخطوات المختلفة عادةً بإسناد المهام إلى كيان معين، ولكن فهم نطاق كل نقطة والمكونات الأساسية شرط أساسي لتقديم تقييم أثر عالي الجودة بطريقة تتسم بالفعالية.

تحديد مَن المسؤول عن جمع البيانات

ستحتاج إلى تكليف الوكالة المسؤولة عن جمع البيانات بصورة مبكرة. ويجب اعتبار بعض المفاضلات المهمة عند تحديد مَن يجب عليه جمع بيانات تقييم الأثر. من بين المرشحين المحتملين لهذه المهمة ما يلي:

- المؤسسة المسؤولة عن تنفيذ البرنامج
 - مؤسسة حكومية أخرى تتمتع بخبرة في جمع البيانات (مثل الهيئة الوطنية للإحصاء)
 - شركة أو مركز أبحاث مستقل متخصص في جمع البيانات
- ويحتاج الكيان المسؤول عن جمع البيانات، دائمًا، إلى التنسيق الوثيق مع الوكالة المنفذة للبرنامج. ويلزم هذا التنسيق الوثيق لضمان عدم تنفيذ عمليات البرنامج قبل جمع بيانات خط الأساس أولاً. وعندما تكون بيانات خط الأساس مطلوبة لعملية البرنامج (على سبيل المثال، البيانات الخاصة بمؤشر أهلية في سياق تقييم قائم على تصميم انقطاع الانحدار)، فيجب على الكيان المسؤول عن جمع البيانات أن يكون قادرًا على معالجة البيانات بسرعة ونقل هذه البيانات إلى المؤسسة المسؤولة عن تنفيذ عمليات البرنامج. كذلك التنسيق الوثيق مطلوب أيضًا في تحديد الوقت المناسب لجمع بيانات مسح المتابعة. فمثلاً، إذا اخترت الطرح العشوائي، حينها يجب تنفيذ مسح المتابعة قبل طرح البرنامج على مجموعة المقارنة لتجنب تداخل العينات. وهناك عامل في غاية الأهمية في سياق تحديد مَن يجب عليه جمع البيانات وهو أن إجراءات جمع البيانات نفسها يجب استخدامها لمجموعتي المقارنة والمعالجة. غالبًا لا تتواصل وكالة التنفيذ إلا مع مجموعة المعالجة فحسب، ولذلك لا تكون في موضع جيد لجمع بيانات عن مجموعات المقارنة. ولكن استخدام وكالات جمع البيانات المختلفة لمجموعتي المعالجة والمقارنة ينطوي على مخاطرة كبيرة حيث إن ذلك يخلق فروقًا في الحاصلات المقيسة في المجموعتين والسبب البديهي وراء ذلك هو اختلاف الإجراءات المتبعة في جمع البيانات. إذا تعذر على وكالة التنفيذ جمع البيانات بشكل فعال لمجموعتي المعالجة والمقارنة، فيجب إدًا اعتبار تكليف مؤسسة أو وكالة خارجية بتلك المهمة.

المفهوم الأساسي

يجب استخدام إجراءات جمع البيانات نفسها مع مجموعتي المقارنة والمعالجة.

ونوصي أيضًا في بعض السياقات بتفويض وكالة مستقلة لجمع البيانات لضمان أن البيانات تتسم بالموضوعية. ليس هناك ما يضمن انتفاء أي دواعي للقلق من عدم قيام وكالة تنفيذ البرنامج بجمع بيانات موضوعية، ولكن الكيان المستقل المسؤول عن جمع البيانات الذي لا مصلحة له في نتائج التقييم يمكن أن يعزز مصداقية جهود تقييم الأثر ككل. كذلك يمكن لهذا الأمر أن يضمن عدم تصور المجيبين أن المسح جزء من البرنامج وبالتالي يحد من مخاطر تقديم المجيبين لإجابات إستراتيجية في محاولة لزيادة فرصهم في المشاركة في البرنامج.

نظرًا إلى أن جمع البيانات ينطوي على تسلسل معقد من العمليات، فمن الموصى به أن يكون هناك كيان متخصص وخبير مسؤول عن جمع البيانات. تتمتع قلة من وكالات تنفيذ البرامج بخبرة في جمع بيانات عالية الجودة على نطاق واسع، تلك التي تكون مطلوبة لإجراء تقييم الأثر. وفي معظم الحالات، ستضطر إلى التفكير في تفويض مؤسسة محلية، مثل وكالة إحصائية وطنية أو شركة أو مركز أبحاث متخصص.

فعند تفويض مؤسسة محلية، مثل وكالة إحصائية وطنية، يتسنى لهذه المؤسسة التعرض لدراسات تقييم الأثر ومن ثم بناء إمكانياتها بشكل أكبر، الأمر الذي قد يكون في حد ذاته فائدة جانبية لتقييم الأثر. ومع ذلك، لا تتمتع الوكالات الإحصائية الوطنية على الدوام بالإمكانيات اللوجستية التي تسمح لها بقبول أعباء عمل إضافية تتجاوز أنشطتها الاعتيادية. وكذلك قد تفتقر أيضًا إلى الخبرة الضرورية في المسوحات الميدانية التي تتطلبها تقييمات الأثر، مثل خبرة في تتبع الأفراد بنجاح بمرور الوقت أو في تنفيذ أدوات المسح غير التقليدية. فإن برزت مثل هذه القيود، كان الحل الأكثر عمليًا هو التعاقد مع شركة أو مركز أبحاث مستقل متخصص في جمع البيانات.

وليس عليك بالضرورة أن توظف الكيان نفسه لجمع المعلومات في استقصائي خط الأساس والمتابعة اللذين قد يختلفان في النطاق. على سبيل المثال، بالنسبة لتقييم أثر متعلق ببرنامج تدريبي، الذي فيه المجتمع الإحصائي محل الاهتمام يتألف من الأفراد الذين سجلوا في الدورة التدريبية، فإن المؤسسة المسؤولة عن الدورة التدريبية تستطيع جمع بيانات خط الأساس لدى تسجيل الأفراد. ومع ذلك، من المستبعد أن تكون الوكالة نفسها الخيار الأفضل لجمع معلومات المتابعة لكل من مجموعتي المعالجة والمقارنة. وفي هذا السياق، التعاقد مع كيان خارجي لإجراء جولات جمع بيانات منفصلة له مزاياه، ولكن يجب أن تُبذل الجهود لمنع فقدان أي معلومات فيما بين الجولات، تلك التي ستكون مفيدة في تتبع الأسر المعيشية أو الأفراد، وكذلك في ضمان أن بيانات خط الأساس والمتابعة خاضعة لقياس متسق.

ومن أجل تحديد أفضل مؤسسة لجمع بيانات تقييم الأثر، يصير من اللازم تقييم كل هذه العوامل (الخبرة في جمع البيانات، والقدرة على التنسيق مع وكالة تنفيذ البرنامج، والاستقلالية، وفرص لبناء الإمكانيات، والتكيف مع سياق تقييم الأثر) بشكل مناسب جنبًا إلى جنب مع التكلفة المتوقعة والجودة المحتملة للبيانات التي يتم جمعها في كل حالة. وهناك طريقة فعالة لتحديد أفضل مؤسسة باستطاعتها جمع بيانات عالية الجودة، ألا وهي صياغة الشروط المرجعية بوضوح ومطالبة المؤسسات بإرسال المقترحات الفنية والمالية.

نظرًا إلى أن التسليم العاجل للبيانات وضمن جودتها غالبًا ما يشكلان جانبين حاسمين لموثوقية تقييم الأثر، فإن التعاقد مع الوكالة المسؤولة عن جمع البيانات يجب أن يُصاغ بعناية. فإنه يجب توضيح نطاق الأعمال والمنجزات المتوقعة بغاية الحرص. بالإضافة إلى ذلك، يُنصح غالبًا بتضمين بند للحوافز في العقود وربط هذه الحوافز بمؤشرات واضحة تنم عن جودة البيانات.

فمثلاً معدل عدم الاستجابة يُعد مؤشراً أساسياً لجودة البيانات. من أجل إنشاء دوافز لوكالات جمع البيانات هدفها الحد من معدل عدم الاستجابة، يمكن أن ينص العقد على تحديد تكلفة معينة لكل وحدة ضمن أول ٨٠٪ من العينة وتكلفة أعلى لكل وحدة في نطاق ٨٠ و ٩٠٪ من العينة أخيراً تكلفة أعلى لكل وحدة في نطاق ٩٠ و ١٠٠٪ من العينة. بدلاً من ذلك، يمكن كتابة عقد منفصل ينص على أن شركة المسح ملتزمة بتتبع غير المجيبين. بالإضافة إلى ذلك، قد يتضمن عقد جمع البيانات دوافز أو شروطاً ذات صلة بالتحقق من جودة البيانات، مثل من خلال عمليات الفحص الاستراتيجية أو عمليات تدقيق الجودة ضمن عينة فرعية من مسح تقييم الأثر.

تطوير أداة جمع البيانات وتجربتها

عند الشروع في عملية جمع البيانات، على فريق التقييم أداء دور مهم في توفير توجيهات محددة بشأن محتوى أدوات جمع البيانات أو الاستبيانات. يجب أن تستخرج أدوات جمع البيانات كل المعلومات المطلوبة للإجابة عن سؤال السياسة الذي حدده تقييم الأثر. كما ناقشنا، يجب قياس المؤشرات عبر سلسلة النتائج بالكامل، بما يتضمن ذلك المؤشرات الخاصة بالحاصلات النهائية والحاصلات المتوسطة وقياسات فوائد البرنامج وجودة التنفيذ. وكذلك من الضروري أن تكون انتقائياً بشأن أي مؤشرات يتعين قياسها. فمن شأن هذا السلوك الانتقائي أن يحد من تكاليف جمع البيانات، ويبسط مهمة الوكالة المعنية بجمع البيانات، ويحسن من جودة البيانات التي يتم جمعها عن طريق تقليل الطلبات المكلف بها مندوبو الإحصاء وتقليل الوقت المطلوب من المجيبين. فإن جمع المعلومات التي تتصف بأنها ليست ذات صلة أو من المستبعد استخدامها، يكون مكلفاً للغاية. تتطلب البيانات الإضافية المزيد من الوقت للتحرير والتدريب والجمع والمعالجة. ومع محدودية الإتاحة وفترات الاحتفاظ بالانتباه، قد يوفر المجيبون معلومات متناقضة في الجودة مع طول المسح، وسيحاول المحاورون التقصير في أداء مهامهم من أجل الإيفاء بأهداف المسح. وهذه الأسئلة الخارجة عن نطاق المسح ليست "مجانية". حدد أولوية المعلومات الضرورية بالنسبة لك من خلال توضيح أهداف تقييم الأثر بما يتماشى مع أهداف البرنامج المحددة بشكل جيد. وعليك أيضاً صياغة خطة تحليل مكتوبة بشكل مسبق (انظر المناقشة في الفصولين ١٢ و ١٣)، وبهذا ستضمن أن المسح يجمع البيانات المطلوبة لتحليل الأثر، وستتجنب تضمين معلومات إضافية خارجية (ذات تكلفة كبيرة).

من الأفضل جمع بيانات بشأن مؤشرات الحاصلات وخصائص الوحدات الضابطة بشكل متسق عند مرحلتي خط الأساس والمتابعة. ويحبذ بشدة توافر بيانات خط الأساس. حتى إذا كنت تستخدم التعيين العشوائي أو تصميم انقطاع الانحدار، ففي الحالات التي يمكن فيها استخدام فروق ما بعد التدخل البسيطة - مبدئياً - من أجل تقدير أثر البرنامج، تكون بيانات خط الأساس ضرورية لاختبار ما إذا كان تصميم تقييم الأثر مناسباً أم لا (راجع المناقشة في الجزء ٢). بعبارة أخرى، إن امتلاك بيانات خط الأساس يعطيك ضماناً عندما لا يفلح التوزيع العشوائي،

وفي هذه الحالة يمكن استخدام طريقة الاختلاف في الاختلافات كحل بديل. وكذلك بيانات خط الأساس مهمة في أثناء مرحلة تحليل الأثر، بما أن المتغيرات الضابطة لخط الأساس يمكن أن تساعد على زيادة القوة الإحصائية وتسمح لك بتحليل الآثار على مجتمعات إحصائية فرعية مختلفة. وأخيرًا، يمكن استخدام بيانات خط الأساس من أجل تعزيز تصميم البرنامج. على سبيل المثال، تتيح بيانات خط الأساس في بعض الأحيان إمكانية تحليل نجاعة الاستهداف أو توفير معلومات إضافية حول المستفيدين إلى الوكالة المنفذة للبرنامج. وفي بعض الحالات، قد يتضمن مسح المتابعة مجموعة أوسع من المؤشرات عن مسح خط الأساس.

المفهوم الأساسي

قياس المؤشرات فن في حد ذاته وأفضل من يؤديه هم المتخصصون، بمن في ذلك فريق أبحاث تقييم الأثر والوكالة المستأجرة لجمع البيانات وخبراء المسح والخبراء في قياس المؤشرات المعقدة المخصصة.

فمجرد أن تحدد البيانات الأساسية التي يتعين جمعها، تتمثل الخطوة التالية في أن تحدد بالضبط طريقة قياس هذه المؤشرات. إن القياس فن في حد ذاته وأفضل من يؤديه هم المتخصصون، بمن في ذلك فريق أبحاث تقييم الأثر والوكالة المستأجرة لجمع البيانات وخبراء المسح والخبراء في قياس المؤشرات المعقدة المخصصة. ويجب أن تكون مؤشرات الحاصل متسقة قدر الإمكان مع أفضل الممارسات المحلية والدولية. ويكون من المفيد دائمًا اعتبار كيف يتم قياس المؤشرات محل الاهتمام في مسوحات مماثلة على الصعيدين المحلي والدولي. وباستخدام مؤشرات مماثلة (بما في ذلك وحدات أو أسئلة المسح نفسها)، نضمن قابلية المقارنة بين البيانات الموجودة مسبقًا والبيانات التي تم جمعها لتقييم الأثر. إن اختيار

مؤشر لا يكون قابلاً للمقارنة بالكامل أو غير مقيس بشكل جيد قد يحد من مدى فائدة نتائج التقييم. وفي بعض الحالات، يكون من المنطقي أن تستثمر في الموارد لجمع مؤشر الحصيلة الابتكاري الجديد، وكذلك بديل أكثر اعتمادية. يجب إيلاء اهتمام خاص إلى ضمان أن المؤشرات كافة يمكن قياسها بالطريقة نفسها بالنسبة لكل الوحدات في كل من مجموعتي المعالجة والمقارنة. وتنشأ مخاطر التحيز عند استخدام طرق جمع بيانات مختلفة (مثل استخدام مسح هاتفي لمجموعة ومسح شخصي للمجموعة الأخرى). وهكذا ينطبق الأمر أيضًا عند جمع بيانات في أوقات مختلفة للمجموعتين (مثل جمع بيانات لمجموعة المعالجة خلال موسم الأمطار وجمع بيانات لمجموعة المقارنة في موسم الجفاف). ولهذا السبب بالتحديد يجب صياغة الإجراءات المستخدمة في قياس أي مؤشر حصيلة بشكل دقيق للغاية. فعملية جمع البيانات يجب أن تكون متماثلة بالضبط بالنسبة لجميع الوحدات. وضمن كل استبيان، يجب تقديم كل وحدة مرتبطة بالبرنامج من دون التأثير على سير أو تأطير الإجابات في الأجزاء الأخرى من الاستبيان. وفي الواقع، من الأفضل تجنب تحديد أي تمييز بين مجموعتي المعالجة والمقارنة في عملية جمع البيانات، حيثما أمكن. ففي معظم الحالات، لا يوجد سبب يستدعي معرفة الوكالة التي تجري جمع البيانات (أو على الأقل الأفراد القائمين بالمسح) ما إذا كان الفرد الخاضع للمسح ينتمي إلى مجموعة المعالجة أم مجموعة المقارنة.

هناك قرار مهم يجب اتخاذه وهو يتعلق بالكيفية التي نقيس بها مؤشرات الحاصل، فهل سنقوم بذلك من خلال المسوحات التقليدية القائمة على الاستبيان والأسئلة المبلغ عنها ذاتيًا أم من خلال طرق أخرى؟ في السنوات الأخيرة، طرأت عدة تطورات في كيفية قياس الحاصل أو السلوكيات الأساسية ذات الصلة بتقييم الأثر. وتتضمن تلك التطورات تنقيح طرق جمع البيانات المبلغ عنها ذاتيًا من خلال الاستبيانات، وكذلك الأساليب المستخدمة في قياس الحاصل الأساسية بشكل مباشر.

تصميم الاستبيان كان ولا يزال موضوع الكثير من الأبحاث. فقد تم تأليف كتاب كامل حول أفضل الطرق لقياس مؤشرات محددة في سياقات خاصة، بما في ذلك طريقة صياغة الأسئلة المطروحة في مسوحات الأسر المعيشية.^١ وهناك أيضًا قاعدة أدلة متنامية حول أفضل الطرق لقياس الاستبيانات الرامية إلى جمع بيانات زراعية أو بيانات استهلاكية أو بيانات توظيف من أجل تعظيم دقتها.^٢ وقد وردت أدلة حديثة من تجارب عشوائية تختبر طرقًا مختلفة لصياغة الاستبيانات ومقارنة موثوقيتها.^٣ ووفقًا لذلك، يتطلب تصميم الاستبيان الانتباه إلى أفضل الممارسات الدولية وكذلك الخبرات المحلية في قياس المؤشرات. يمكن للتغيرات الطفيفة في صياغة الأسئلة أو تسلسلها أن تكون لها تأثيرات كبيرة في البيانات التي تم جمعها، ولذلك الانتباه الأكبر للتفاصيل يُعد أساسيًا في تطوير الاستبيان. وتزداد أهمية ذلك الأمر بشكل خاص عند محاولة ضمان قابلية المقارنة عبر المسوحات، بما في ذلك، على سبيل المثال، من أجل قياس الحاصل بشكل متكرر بمرور الزمن. يناقش المربع ١٦-٣ الإرشادات المرتبطة بتصميم الاستبيان ويوفر مراجع إضافية.

لقد تم تطوير مجموعة من التقنيات للحصول على *قياس مباشر للحاصل*. فمثلًا، في القطاع الصحي، تُستخدم التوجيهات الطبية أحيانًا لتعريف العاملين في المجال الصحي بأعراض معينة، ولتقييم ما إذا كان مقدم الخدمات الصحية يوصي بالعلاج المناسب استنادًا إلى الإرشادات والبروتوكولات المحددة. ومثل هذه التوجيهات الطبية توفر قياسًا مباشرًا لمعرفة مقدمي الخدمات الصحية. وتعتمد التقييمات الحديثة على المرضى المعياريين (المعروفين أيضًا باسم المرضى المقنعين أو الصوريين) من أجل زيارة المراكز الصحية والوصول المباشر إلى جودة الخدمات المقدمة.^٤ في القطاع التعليمي، تسعى العديد من التقييمات إلى تقييم آثار البرنامج على تعلم الطلاب. وللقيام بذلك، يتم استخدام مجموعة من تقديرات التعلم ومقاييس مباشرة لمهارات الطلاب. وعلى نحو مماثل، تم تطوير مجموعات اختبارات متنوعة من أجل إجراء قياس مباشر للتطور المعرفي أو اللغوي أو الحركي فيما بين الأطفال الصغار في سياق تقييمات الأثر لتدخلات التنمية في الطفولة المبكرة. وقد تم إحراز تقدم للحصول على قياسات مباشرة للمهارات فيما بين البالغين، بما يشمل ذلك المهارات الاجتماعية العاطفية أو السمات الشخصية. وإلى جانب قياس المهارات، هناك عدد متنامي من تقييمات الأثر التي تسعى إلى الحصول على قياسات جودة التدريس من خلال الملاحظات المباشرة لسلوكيات المعلمين في قاعة التدريس.

المربع ١٦-٣: تصميم وصياغة الاستبيانات

هذه التوصيات تمامًا على معظم أدوات جمع البيانات الأخرى:

١. يجب كتابة كل سؤال حرفيًا بشكل كامل بحيث يستطيع المحاور إجراء المقابلة عن طريق قراءة كل سؤال كلمة بكلمة.

٢. يتعين أن يتضمن الاستبيان تعريفات دقيقة لجميع المفاهيم الأساسية المستخدمة في المسح، بحيث يستطيع المحاور الرجوع إلى التعريف في أثناء المقابلة عند الضرورة.

٣. يجب أن تكون صياغة كل سؤال قصيرة وبسيطة قدر الإمكان وتحتوي مصطلحات شائعة مألوفة.

٤. يجب تصميم الاستبيان بحيث تكون الإجابات لجميع الأسئلة تقريبًا مرمزة مسبقًا.

٥. يجب أن يكون مخطط ترميز الأسئلة متسقًا عبر جميع الأسئلة.

٦. يجب أن يتضمن المسح أنماط التخطي أو التجاوز التي تحدد أي أسئلة من المفترض عدم طرحها بناءً على الإجابات التي تلقاها المحاور عن أسئلة سابقة.

بمجرد صياغة الاستبيان من قبل الشخص المفوض للعمل على هذه الأداة، يجب تقديمه إلى فريق الخبراء ليخضع للمناقشة. بعد ذلك، يجب استشارة جميع الأشخاص المشاركين في فريق التقييم (صانعو السياسات والباحثون ومحللو البيانات وجامعو البيانات) حول ما إذا كان الاستبيان يجمع كل المعلومات المطلوبة بطريقة ملائمة. صحيح أن مراجعة الاستبيان من قبل فريق من الخبراء أمر ضروري، إلا أنها ليست كافية؛ لأنه يلزم دائمًا إجراء اختبار ميداني مكثف.

على الرغم من أن تصميم الاستبيان في تقييمات الأثر جزء لا يتجزأ من جودة البيانات، فإنه غالبًا ما يتم تجاهله. وقد يعزو سبب ذلك إلى أن تصميم الاستبيان ينطوي على عملية معقدة وطويلة وتكرارية ومليئة بالقرارات العديدة التي تُتخذ طوال المسار حول ما يمكن قياسه وبأي كيفية. وتوفر الدورة التدريبية حول الطرق التطبيقية لتقييم الأثر، المقدمة في جامعة كاليفورنيا في بيركلي (<http://aie.cega.org>) دليلًا توجيهيًا يعرفك على تصميم الاستبيان والمراحل الثلاث للتصميم: المحتوى والصياغة والاختبار. وخلال هذه المراحل، تسلط الوحدة الضوء على أهمية تضمين الأطراف المعنية ذوي الصلة، وإتاحة الوقت الكافي للعمليات التكرارية، والاختبار المتأني:

١. المحتوى. حدد محتوى المسح عن طريق تعريف أولاً التأثيرات التي بحاجة إلى القياس، ووحدات الملاحظة، والارتباطات مع العوامل الأخرى. ومن ثم يجب ترجمة هذه التعريفات المفاهيمية إلى مؤشرات ملموسة.

٢. الصياغة. صوغ الأسئلة من أجل قياس المؤشرات المحددة. وهذه خطوة غاية في الأهمية حيث إن جودة البيانات تعتمد عليها. وتوفر الوحدة المزيد من التوصيات التفصيلية حول صياغة الأسئلة، وتنظيم المسح، والتنسيق، فضلاً عن اعتبارات أساسية أخرى.

٣. الاختبار. اختبر الاستبيان على ثلاثة مستويات: السؤال والوحدة والمسح بالكامل.

جدير بالذكر أن تنسيق الاستبيان له أهمية أيضًا لضمان جودة البيانات. نظرًا إلى أن استخدام الطرق المختلفة في طرح سؤال المسح نفسه يمكن أن يثمر إجابات مختلفة، تبرز الحاجة إلى اتساق تأطير وتنسيق الأسئلة عبر جميع الوحدات منعًا لأي تحيز من جانب المجيب أو مندوب الإحصاء. وضعت الأمم المتحدة (٢٠٠٥) ست توصيات خاصة فيما يخص تنسيق الاستبيانات لمسوحات الأسر المعيشية. وتنطبق

إن الملاحظة المباشرة للحصائل الأساسية مهمة بشكل خاص عندما يصعب - على نحو محتمل - استخلاص الحصائل محل الاهتمام من المجيبين بشكل صريح. فعلى سبيل المثال، لتجنب الاعتماد على البيانات المبلغ عنها ذاتيًا لقياس الحصائل المرتبطة بالجريمة أو العنف، تتضمن بعض تقييمات الأثر باحثين مدربين في المجتمعات الإحصائية المشمولة في العينة من أجل ملاحظة سلوك موضوعات الدراسة (أي الأفراد) بشكل مباشر باستخدام الطرق الإثنوغرافية. مثل هذه الملاحظة المباشرة يمكن أن تطف من كم المشكلات من خلال السلوكيات المبلغ عنها ذاتيًا، وقد توفر المزيد من المعلومات الدقيقة عند إجرائها بشكل صحيح. تسمح أحدث التطورات التكنولوجية بإجراء قياسات مباشرة لنطاق من السلوك البشري، وبالتالي يمكن أن تساعد في الحد من استخدام البيانات المبلغ عنها ذاتيًا. وتتضمن الأمثلة على ذلك الملاحظة المباشرة لتوقيت ومدى استخدام مواقع الطهي المحسّنة، والقياسات المباشرة لجودة الماء، واستخدام المراحيض العامة، ودرجات الحرارة الداخلية باستخدام المستشعرات الإلكترونية.

تعتمد تقييمات الأثر بشكل أساسي على مزيج من المسوحات التقليدية القائمة على الاستبيانات والطرق الأخرى الهادفة إلى تحقيق ملاحظة مباشرة للحصائل محل الاهتمام. فمثلًا، في سياق تقييم الأثر للتمويل القائم على النتائج في القطاع الصحي، يتم قياس مجموعة واسعة من المؤشرات من خلال مصادر تكميلية (Vermeersch, 2012, Rothenbühler, and Sturdy). ويتضمن مسح المرفق الصحي تقييم المرفق من أجل قياس الخصائص الرئيسية له، ومحاورة عامل صحي من أجل قياس خصائص العامل الصحي، ومحاورة مريض بعد الخروج من المستشفى من أجل قياس الخدمات المقدمة، فضلًا عن مؤشرات جودة الرعاية من خلال مزيج من التوجيهات الطبية والملاحظة المباشرة. ويتضمن مسح الأسر المعيشية بيانات على مستوى الأسر المعيشية بشأن السلوك الأسري والفرد (مثل عدد مرات زيارة المرافق، والرعاية التي يحصلون عليها، والنفقات الصحية) وكذلك وحدات على مستوى الأفراد حول صحة الإناث والأطفال. بالإضافة إلى القياس الأنثروبومتري، يتم جمع اختبارات بيومترية من أجل قياس انتشار الأنيميا أو الملاريا أو فيروس العوز المناعي البشري (HIV) بشكل مباشر. وأخيرًا، تجمع الاستبيانات المجتمعية الخصائص المجتمعية والخدمات والبنية التحتية وإمكانية الوصول إلى الأسواق والأسعار والخدمات على مستوى المجتمع. وإضافة إلى تطوير المؤشرات والعتور على أنسب طريقة لقياسها، هناك قرار حاسم يجب اتخاذه عند جمع بيانات جديدة وهو تحديد التقنية المستخدمة في جمع البيانات. وفقًا للطرق التقليدية لجمع البيانات، يتم جمع قواعد البيانات الموجودة على الورق، وتحويل هذه البيانات رقميًا في مرحلة تالية، وغالبًا ما يتم ذلك باتباع مقارنة إدخال بيانات مزدوج التعمية، الذي ينطوي على تفويض وكيلين منفصلين لرقمنة البيانات نفسها للتحقق من حالات عدم الدقة. وباتباع هذه التطورات التكنولوجية الحديثة، أصبح من الشائع استخدام أدوات جمع البيانات المعززة بالحاسوب. وفي سبيل تسريع معالجة البيانات، يتم تثبيت تطبيقات لجمع البيانات على الهواتف الذكية أو الأجهزة اللوحية (التابلت)، وهناك ميزة إضافية أيضًا تتمثل في إتاحة الفرصة لإجراء فحص لجودة البيانات والتحقق من صحة البيانات في الوقت الفعلي. يناقش المربع ١٦-٤ بعض المزايا والعيوب لجمع البيانات الإلكترونية.

المربع ١٦-٤: بعض المزايا والعيوب لجمع البيانات الإلكترونية

- توفر المقابلات الشخصية المعززة بالحاسوب (CAPI) بديلاً للمقابلات الشخصية بالورقة والقلم (PAPI) التقليدية. عند استخدام المقابلات الشخصية المعززة بالحاسوب، يتم تحميل المسح على جهاز إلكتروني، مثل جهاز لوحي (تابلت) أو هاتف ذكي. ويقرأ المحاور الأسئلة من الشاشة، ويُدخل (يكتب) الإجابات بصورة فورية في الجهاز. وقد تم تطوير مجموعة متنوعة من البرمجيات والتطبيقات لجمع البيانات عن طريق المقابلات الشخصية المعززة بالحاسوب. وينبغي لفريق التقييم أن يدرس بعناية المزايا والعيوب المرتبطة بإجراء مقابلات شخصية معززة بالحاسوب (أو CAPI اختصاراً).

بعض المزايا:

- يستطيع جمع البيانات الإلكترونية أن يحسّن جودة البيانات. في تجربة عشوائية مصممة لمقارنة المقابلات الشخصية المعززة بالحاسوب (CAPI) والمقابلات الشخصية بالورقة والقلم (PAPI) على أساس مسح بخصيص الاستهلاك في تنزانيا، وجد Caeyers, Chalmers, and De Weerd (2012) أن البيانات الواردة عن المسوحات الورقية ظهرت بها أخطاء تجنبها المسوحات الإلكترونية. وكذلك اكتشف الباحثون أن الأخطاء في بيانات المقابلات الشخصية بالورقة والقلم كانت مترابطة بخصائص أسرية معينة، الأمر الذي قد يثير تحيزاً في بعض تحليلات البيانات.

بعض العيوب:

- من الممكن تضمين عمليات مؤتمتة (أو آلية) لفحص الاتساق في برامج جمع البيانات الإلكترونية. ويمكن لإجابات معينة إحداث رسائل تحذيرية بهدف تقليل أي أخطاء في إدخال البيانات إلى أقصى حد وتوضيح أي مشكلة مع المجيب في أثناء المقابلة الشخصية. على سبيل المثال، قام Fafchamps and others (2012) بدراسة فوائد عمليات فحص الاتساق في مسوحات بخصوص المؤسسات الصغرى في غانا. وتبين لهم أنه عند تضمين عمليات فحص الاتساق، كان الانحراف المعياري لبيانات الربح والمبيعات أقل. ووجدوا أيضاً أن في معظم الأوقات، لم يكن يلزم إجراء أي تصحيح: في ٨٥ إلى ٩٧ بالمائة من الوقت، أكد المجيبون الإجابة الأصلية.

- يمكن أن تكون المقابلات الشخصية أقصر وأسهل في إجرائها. عند استخدام المقابلات الشخصية المعززة بالحاسوب، يمكن تخصيص سير عمل الاستبيان لتوجيه المحاورين بشكل أفضل فيما يخص أنماط التخطي ولحد من معدل الخطأ والنسيان في الاستبيان إلى أقصى حد. في مسح أسر معيشية في تنزانيا، كان محاورو المقابلات الشخصية المعززة بالحاسوب، في المتوسط، أقل بمعدل ١٠٪ من القائمين على إجراء استبيانات مماثلة يتم جمعها على الورق، حسبما وجد Caeyers و Chalmers (2012) و De Weerd.

- إن جمع البيانات الإلكترونية يلغي الحاجة إلى قوى عاملة لإعادة إدخال البيانات. ويمكن لذلك الأمر أن يحد من التكاليف ويسرع من معالجة البيانات.
- من خلال استخدام التكنولوجيا الحديثة، ننهل من مجموعة متنوعة من الفوائد غير المباشرة. فمثلاً، باستخدام الأجهزة اللوحية أو الهواتف الذكية، من السهل جمع إحداثيات نظام تحديد المواقع العالمي (GPS) أو النقاط الصور الفوتوغرافية. ويمكن أيضاً إدخال التنوع التجريبي في محتوى المسح. وباستخدام بعض البرامج، يمكن تسجيل أجزاء من المقابلة الشخصية من أجل تسهيل عمليات فحص الجودة والمتابعة.

- عادة ما تكون التكاليف الثابتة أعلى بالنسبة للمقابلات الشخصية المعززة بالحاسوب مقارنة بالمقابلات الشخصية بالورقة والقلم، على الرغم من أن التكاليف المتغيرة يمكن أن تكون أقل. علاوة على ذلك، التكلفة المدفوعة مقدماً لشراء الأجهزة الإلكترونية وبرمجتها قد لا تكون خياراً متاحاً لتقييمات الأثر ذات الميزانيات المحدودة. ويلزم أيضاً توافر وقت كافٍ في البداية لضمان إتمام برمجة واختبار الاستبيانات الإلكترونية على أنسب وجه، ويتأتى ذلك في الغالب بعد وضع استبيانات ورقية.

ويمكن أن تنشأ مشكلات أثناء جمع البيانات عندما تكون للجهاز الإلكتروني شاشة أو واجهة صغيرة على نحو غير مألوف بالنسبة للمحاورين. وتكون مخاطر السرقة أعلى بالنسبة للأجهزة الإلكترونية بالمقارنة بالمسوحات الورقية. وأخيرًا، الاحتفاظ بالبيانات ومزامنتها في مكان آمن يتطلبان بروتوكولات واضحة للحد من مخاطر فقدان البيانات. صحيح أن التحويلات الإلكترونية للبيانات تسهل الكثير، إلا أنها تتطلب مستوى أدنى من الاتصال.

- تكون بعض الخبرات الفنية مطلوبة لبرمجة الاستبيانات الإلكترونية وإعداد المعالجات من أجل إدارة سير عمل البيانات التي يتم جمعها إلكترونيًا. في البلدان النامية التي تعوز إلى إمكانيات تكنولوجيا المعلومات، قد يصعب تحقيق ذلك. ويصير الأمر أكثر صعوبة فيما يتعلق بتطوير برمجيات أو استبيانات ليست باللغة الإنجليزية أو اللغات الرومانسية.
- من الممكن أن تمثل المشكلات التكنولوجية عائقًا أمام جمع البيانات أو الاحتفاظ بها في مكان آمن.

المصادر: Caeyers, Chalmers, and De Weerd 2012; Fafchamps and others 2012.

من الأهمية بمكان أن يتم تجريب أداة جمع البيانات واختبارها ميدانيًا على نطاق واسع قبل إطلاقها في صورتها النهائية. عند إجراء تجريب موسع للأداة، سيتم فحص مدى ملاءمتها للسياق المحلي ومحتواها وأي خيارات بديلة للتنسيق والصياغة وكذلك بروتوكولات جمع البيانات بما يتضمن ذلك التكنولوجيا. كذلك، لا مفر من إجراء اختبار ميداني لأداة جمع البيانات بشكل شامل في ظروف واقعية من أجل التحقق من مدتها والتثبت من صحة تنسيقها ما إذا كان متسقًا وشاملاً بشكل كافٍ لإنتاج قياسات دقيقة لجميع المعلومات ذات الصلة. ويُعد الاختبار الميداني جزءًا لا يتجزأ من إعداد أدوات جمع البيانات.

إجراء العمل الميداني وضمان الجودة

حتى عندما تشرع في جمع البيانات، يجب قبل كل شيء الوصول إلى فهم صحيح وواضح لجميع الخطوات التي تنطوي عليها عملية جمع البيانات وسيساعدك ذلك على ضمان أن آليات ضمان الجودة والحوافز الصحيحة في نصابها الصحيح. سيحتاج الكيان المسؤول عن جمع البيانات إلى تنسيق عمل مجموعة أكبر من الفاعلين المختلفين بمن في ذلك مندوبو الإحصاء والمشرّفون والمنسقون الميدانيون وفريق الدعم اللوجستي بالإضافة إلى فريق إدخال البيانات المتألف من المبرمجين والمشرّفين ومدخلي البيانات. يجب وضع خطة عمل واضحة لتنسيق عمل جميع هذه الفرق، وستكون خطة العمل هذه منجزًا أساسيًا.

وقبل بدء جمع البيانات، يجب أن تتضمن خطة العمل تدريبًا لفريق جمع البيانات. ويجب إعداد *دليل مرجعي* كامل للتدريب ولاستخدامه في خلال العمل الميداني. ويمثل التدريب ركناً أساسياً لضمان جمع البيانات بشكل متسق من قبل جميع المشاركين في المسح. وتشكل عملية التدريب أيضًا فرصة جيدة لتحديد مندوبي الإحصاء أصحاب الأداء الأفضل ولإجراء آخر تجربة للأدوات والإجراءات تحت الظروف الطبيعية. لا تبدأ عملية جمع البيانات إلا فور سحب (أو أخذ) العينة، وتصميم الأدوات وتجربتها، وتدريب الفرق. وتتمثل إحدى الممارسات الجيدة في ضمان التزام كل فريق من فرق المسح العاملة ميدانيًا بجمع البيانات على العدد نفسه من وحدات المعالجة والمقارنة.

وكما ناقشنا في الفصل 10، لا بد من أخذ العينات بطريقة مناسبة لضمان جودة العينة. ومع ذلك، يمكن أن تحدث الكثير من الأخطاء غير المرتبطة بأخذ العينات في أثناء جمع البيانات. وفي سياق تقييم الأثر، يكون هناك شاغل خاص في أن هذه الأخطاء قد لا تكون متماثلة في مجموعتي المعالجة والمقارنة.

المفهوم الأساسي

تنشأ حالات عدم الاستجابة عندما تكون البيانات مفقودة أو غير كاملة بالنسبة لبعض الوحدات المشمولة في العينة. وتستطيع حالات عدم الاستجابة أن تثير التحيز في نتائج التقييم.

تنشأ عدم الاستجابة عندما يكون من المستحيل جمع بيانات كاملة لبعض الوحدات المشمولة في العينة. ولأن العينات الفعلية المستخدمة في التحليل محصورة على تلك الوحدات التي يمكن جمع البيانات من أجلها، فإن الوحدات التي تقرر عدم الرد على مسح ما قد تجعل العينة ذات قدرة تمثيلية أقل ويمكنها أن تثير تحيزًا في نتائج التقييم. ويُعد *التناقص* شكلًا شائعًا لعدم الاستجابة الذي يحدث عندما ينخفض عدد الوحدات من العينة بين جولات جمع البيانات: فمثلاً المهاجرون لا يمكن تتبعهم كما ينبغي.

إن تناقص العينة الناتج عن عدم الاستجابة يمثل إشكالية خاصة في سياق تقييمات الأثر لأنها يمكن أن تثير فروقًا بين مجموعة المعالجة ومجموعة المقارنة. على سبيل المثال، قد يكون *التناقص* مختلفًا في المجموعتين: إذا وقعت عملية جمع البيانات بعد بدء تنفيذ البرنامج، يمكن أن يكون معدل الاستجابة بين وحدات المعالجة أعلى من المعدل الخاص بوحدات المقارنة. وربما يحدث ذلك لأن وحدات المقارنة غير راضية عن عدم اختيارها أو كان من المرجح أكثر أن تهجر. ويمكن أن تحدث حالات عدم الاستجابة أيضًا ضمن الاستبيان نفسه، ويقع ذلك عادةً لأن بعض المؤشرات تكون مفقودة أو أن البيانات لا تكون كاملة لوحدة معينة.

سنتطرق الآن إلى خطأ القياس، وهو نوع آخر من المشكلات التي قد تثير التحيز إذا كانت تتم بصورة منهجية. نعرّف *خطأ القياس* على أنه الفرق بين قيمة خاصة ما كما يوفرها المجيب والقيمة الحقيقية (المجهولة) (Kasprzyk 2005). هناك إمكانية لتتبع مثل هذا الفرق وفقًا لطريقة صياغة الاستبيان أو طريقة جمع البيانات المختارة. ليس هذا فقط، بل قد يحدث خطأ القياس أيضًا بسبب المحاور الذي يجري المسح ميدانيًا أو الشخص المجيب عن الأسئلة.

المفهوم الأساسي

تهدف أفضل ممارسات تقييمات الأثر إلى الاحتفاظ بمعدلات عدم الاستجابة والتناقص عند أدنى حد.

وتعتمد جودة تقييم الأثر بشكل مباشر على جودة البيانات التي يتم جمعها. يجب توضيح معايير الجودة لكل الأطراف المعنية في عملية جمع البيانات، والتأكيد على المعايير بشكل خاص في أثناء تدريب مندوبي الإحصاء وفي الأدلة الإرشادية المرجعية. فمثلاً، لا بد من وضع إجراءات تفصيلية للحد من عدم الاستجابة أو (إن كان مقبولاً) لاستبدال الوحدات في العينة.

وكذلك يجب على وكالة جمع البيانات أن تفهم دون أي مجال للشك المعدلات المقبولة لعدم الاستجابة والتناقص. ومن أجل توفير أساس مقارنة، تهدف العديد من تقييمات الأسرة إلى الاحتفاظ بمستوى عدم الاستجابة والتناقص دون 5 بالمائة. وسيعتمد الهدف على توقيت تقييم الأثر ووحدة التحليل: من المتوقع أن يكون التناقص أدنى نسبيًا بالنسبة للمسح الذي يتم تنفيذه بعد مسح خط الأساس بوقت قصير، وأعلى نسبيًا بالنسبة لتقييم الأثر طويل المدى الذي يتتبع الأفراد بعد مرور عدة سنوات. ومن المتوقع أيضًا أن تكون معدلات التناقص أعلى في المجتمعات الإحصائية دائمة الهجرة. وأحيانًا، يتم دفع تعويض (أو مقابل) لمجبي المسح من أجل الحد من معدلات عدم الاستجابة، على الرغم من أن تقديم مثل هذا التعويض يحتاج إلى الكثير من التفكير المتأن. وفي أحيان أخرى، بمجرد تحديد الوحدات المطلوب تتبعها، يتم اختيار عينة فرعية من هذه الوحدات بشكل عشوائي لتتبعها بشكل مكثف للغاية وهذا الأمر لا يتأتى إلا من خلال بذل جهود إضافية أو دفع مقابل بشكل آخر. في أي حالة، يجب أن يتضمن عقد وكالة جمع البيانات حوافز واضحة، مثل ارتفاع قيمة التعويض إذا ظل معدل عدم الاستجابة دون حد مقبول.

يتعين تأسيس إجراءات ضمان جودة محددة بشكل جيد لجميع مراحل عملية جمع البيانات، بما في ذلك تصميم إجراء أخذ العينات والاستبيان، ومراحل الإعداد، وجمع البيانات، وإدخال البيانات، وتنقية البيانات وتخزينها.

يجب منح عمليات فحص الجودة في أثناء العمل الميداني أولوية كبيرة للغاية من أجل تقليل الأخطاء لكل وحدة إلى أقصى حد. ولا بد من وضع إجراءات واضحة للرجوع إلى الوحدات التي لم توفر معلومات أو وفرت معلومات غير كاملة. ويجب تقديم عوامل تصفية متعددة في عملية ضمان الجودة: مثل مطالبة مندوبي الإحصاء والمشرفين والمنسقين الميدانيين (إن استدعى الأمر) بإعادة زيارة الوحدات غير المستجيبة للتحقق من حالتها. وكذلك لا يزال يجب ترميز وتسجيل الاستبيانات المرتبطة بالمقابلات الشخصية مع غير المستجيبين بكل وضوح. بمجرد الانتهاء من رقمنة البيانات، يمكن تلخيص معدلات عدم الاستجابة واحتساب جميع الوحدات ضمن العينة.

ويجب أيضًا إتمام أي عمليات فحص للجودة على أي بيانات غير مكتملة لوحدات معينة ضمن المسح. ومرة ثانية، يتعين تضمين عوامل تصفية متعددة في عملية ضمان الجودة. ويكون مندوب الإحصاء مسؤولاً عن التحقق من البيانات فور جمعها. ويؤدي المشرف والمنسق الميداني عمليات فحص عشوائية في مرحلة لاحقة.

تتسم عمليات فحص الجودة لأخطاء القياس بصعوبة أكبر إلا أن لها أهمية ماسة لتقييم مدى دقة البيانات التي تم جمعها. ويمكن إدماج عمليات فحص الاتساق في الاستبيان. بالإضافة إلى ذلك، يحتاج المشرفون ومراقبو الجودة إلى إجراء عمليات فحص مفاجئة عن طريق المشاركة في المقابلات الشخصية لضمان جمع مندوبي الإحصاء للبيانات وفقًا لمعايير الجودة المتفق عليها. وكذلك يمكن إجراء عمليات فحص استرجاعية أو عمليات تدقيق جودة ضمن عينة فرعية لمسح تقييم الأثر لضمان دقة البيانات التي تم جمعها.

ويتم هذا الأمر أحياناً عن طريق مطالبة مراقب الجودة بجمع مجموعة فرعية من الاستبيان مع مجيب ما، ومقارنة الإجابة مع تلك التي تم الحصول عليها سابقاً بواسطة مندوب الإحصاء مع المجيب نفسه.

ولا بد من مساهمة المنسقين الميدانيين أو أعضاء فريق التقييم في عمليات فحص الجودة من أجل الحد من تضارب المصالح ضمن شركة المسح إلى أقصى حد. وربما ينبغي لك التفكير في التعاقد مع وكالة خارجية لتدقيق جودة أنشطة جمع البيانات. ومن شأن ذلك أن يجنبك بشكل كبير مجموعة واسعة من المشكلات التي يمكن أن تنشأ كنتيجة لغياب عملية الإشراف على فريق جمع البيانات أو عدم كفاية إجراءات مراقبة الجودة.

وفي النهاية، لا بد من أن تنص الشروط المرجعية صراحة على كل الخطوات التي تنطوي عليها عملية التحقق من الجودة، وذلك عند الشروع في جمع البيانات.

معالجة البيانات وتخزينها

تُعد معالجة البيانات والتحقق من صحتها جزءاً لا يتجزأ من عملية جمع بيانات مسح جديد. وهي تتضمن خطوات تتم فيها رقمنة المعلومات الموجودة في المسوحات المنفذة بالورقة والقلم، وكذلك خطوات التحقق من صحة البيانات لكل من مسوحات الورقة والقلم وجمع البيانات الإلكتروني باستخدام أجهزة الحاسوب المحمولة أو الهواتف الذكية أو الأجهزة اللوحية (التابلت) أو غير ذلك من الأجهزة. عند العمل على مسوحات منفذة بالورقة والقلم، يجب تطوير برنامج إدخال بيانات وتفعيل نظام لإدارة سير البيانات المطلوب رقمتها. وفي تلك الأثناء، يجب تأسيس المعايير والإجراءات، وتدريب مدخلي البيانات بشكل كامل من أجل ضمان اتساق إدخال البيانات. وإلى أقصى حد ممكن، يجب إدماج عملية إدخال البيانات مع عمليات جمع البيانات (بما في ذلك في أثناء مرحلة الاختبار التجريبي)، بحيث يمكن تحديد أي مشكلات بشأن البيانات التي تم جمعها والتحقق من صحة هذه البيانات بشكل فوري. بصفة عامة، يجب أن يكون أساس مقارنة الجودة لعملية إدخال البيانات متمثلاً في تكرار البيانات غير المنقحة الموجودة في صورة مادية (ورق) في صورة رقمية دون إجراء أي تعديلات عليها عند إدخالها. ولتقليل أخطاء إدخال البيانات، يمكن اتباع مقارنة إدخال البيانات مزدوجة التعمية من أجل تحديد أي أخطاء متبقية وتصحيحها. ويمكن استخدام مقارنة الإدخال الميداني المعزز بالحاسوب (CAFE) وفيه يتم جمع البيانات في مسح بالورقة والقلم ثم تتم رقمنة هذه البيانات في الميدان ويتم التحقق من صحتها فوراً من أجل اكتشاف الأخطاء وحالات عدم الاتساق.

بالنسبة لمسوحات الورقة والقلم والمسوحات المعتمدة على جمع البيانات الإلكترونية، يمكن تطوير برامج لأداء عمليات فحص تلقائية للكشف عن الأخطاء غير المرتبطة بعملية أخذ العينات (كل من معدلات عدم الاستجابة وحالات عدم الاتساق) التي قد تحدث في الميدان وللتحقق من صحة البيانات. إذا تم إدماج عملية التحقق من صحة البيانات في إجراءات العمل الميداني، فيمكن إرجاع أي بيانات غير مكتملة أو غير متسقة إلى العاملين الميدانيين للتحقق من صحتها في الميدان.

ولا يخلو هذا النوع من الإدماج من التحديات بالنسبة لسير العمل المؤسسي للعمليات الميدانية، ولكنها تدر الكثير من المنافع من حيث الجودة والتقليل من أخطاء القياس وزيادة القوة الإحصائية لتقييم الأثر. ولذلك، يجب التفكير بتأنٍ في احتمالية استخدام مقاربة الإدماج ذلك في مرحلة التخطيط لجمع البيانات. ومن الممكن إدخال تقنيات جديدة لتسهيل عمليات التحقق من الجودة هذه.

وكما ناقشنا من قبل، تتألف عملية جمع البيانات من مجموعة من العمليات التي يجب أخذ تعقيداتها بعين الاعتبار. يناقش المربع ١٦-٥ كيف أدت عملية جمع البيانات لتقييم برامج Atención a Crisis التدريبية

المربع ١٦-٥: جمع البيانات لتقييم برامج Atención a Crisis التجريبية في نيكاراغوا

لقد تم جمع بيانات لمسح خط أساس في عام ٢٠٠٥، ومن ثم تلاه أول مسح متابعة في ٢٠٠٦، ومسح متابعة ثانٍ في ٢٠٠٨، أي بعد انتهاء التدخل بعامين. وتم تطبيق عمليات فحص جودة صارمة في جميع مراحل عملية جمع البيانات. أولاً، تم اختبار الاستبيانات ميدانياً على نطاق واسع، وتلقى مندوبو الإحصاء تدريباً في قاعات التدريس وفي الميدان على حد سواء. وثانياً، تم إعداد إشراف ميداني لمراجعة جميع الاستبيانات لعدة مرات من قبل منفذي المسح والمشرفين والمنسقين الميدانيين والمراجعين الآخرين. وثالثاً، تم استخدام نطاق إدخال بيانات مزدوج التعمية جنباً إلى جنب مع برنامج فحص جودة شامل بإمكانه التعرف على الاستبيانات غير الكاملة أو غير المتسقة. وفي الحالات التي كانت فيها الاستبيانات تنقصها بعض المعلومات في أسئلة معينة أو في حالة عدم الاتساق، كان يتم إرسال هذه الإدخالات مرة أخرى إلى الميدان للتحقق من صحتها. وكانت هذه الإجراءات والمتطلبات محددة صراحة في الشروط المرجعية للشركة المسؤولة عن جمع البيانات.

في عام ٢٠٠٥، أطلقت الحكومة النيكاراغوية برنامج Atención a Crisis التجريبي. وتم إعداد دراسة لتقييم أثر الجمع بين برنامج التحويل النقدي (CCT) والتحويلات الإنتاجية، مثل المنح المخصصة للاستثمار في الأنشطة غير الزراعية أو التدريب المهني. وتم تنفيذ برنامج Atención a Crisis التجريبي من قبل وزارة الأسرة بدعم من البنك الدولي.

وتم استخدام تعيين عشوائي ثنائي المرحلة للتقييم. أولاً، تم تعيين ١٠٦ مجتمعات إحصائية مستهدفة عشوائياً إلى مجموعة مقارنة أو مجموعة معالجة. ثانياً، ضمن المجتمعات الإحصائية المعالجة، تم تعيين الأسر المعيشية المؤهلة عشوائياً إلى واحدة من ثلاث حزم من الفوائد: التحويل النقدي الشرطي، والتحويل النقدي الشرطي بالإضافة إلى منحة دراسية تسمح لأحد أفراد الأسرة باختيار واحدة من بين عدد من الدورات التدريبية المهنية، والتحويل النقدي الشرطي بالإضافة إلى منحة استثمار إنتاجية لتشجيع المتلقين على بدء نشاط غير زراعي الغاية منه إنشاء أصول ودخل متنوع (Macours, Premand, and Vakis 2012).

(تابع)

المربع ١٦,٥: جمع البيانات لتقييم برامج Atención a Crisis التجريبية في نيكاراغوا (تابع)

علاوة على ذلك، تم تأسيس إجراءات تتبع تفصيلية للحد من التناقص قدر الإمكان. ففي البداية، تم إجراء تعداد كامل للأسر المعيشية التي تقيم في المجتمعات الإحصائية المعالجة والضابطة في ٢٠٠٨ بالتعاون الوثيق مع قادة المجتمع. ولأن الهجرة داخل البلاد كانت شائعة، تم منح الشركة المعنية بتنفيذ المسح حوافز لتتبع المهاجرين الأفراد في جميع أنحاء البلاد. ونتيجة لذلك، لم يتعذر إجراء المقابلة إلا مع ٢ بالمائة فقط من إجمالي ٤,٣٥٩ أسرة معيشية في ٢٠٠٩. وشرعت الشركة المعنية بالمسح أيضًا في تتبع جميع الأفراد الذين ينتمون إلى الأسر المعيشية التي خضعت للمسح في ٢٠٠٥. وللمرة الثانية، تعذر تتبع ٢ بالمائة فقط من الأفراد الذين استهدفهم برنامج التحويلات (بالإضافة إلى ٢ بالمائة من الأفراد قد وافتهم المنية). وكانت نسبة التناقص ٣ بالمائة لجميع أطفال الأسر المعيشية الخاضعة للمسح في ٢٠٠٥ و ٥ بالمائة

لجميع الأفراد في الأسر المعيشية الخاضعة للمسح في ٢٠٠٥. وتوفر نسب التناقص وعدم الاستجابة مؤشرًا جيدًا لجودة المسح. ولإبقاء التناقص عند مثل هذه المعدلات شديدة الانخفاض، تطلب الأمر جهودًا مكثفة من قبل الشركة المعنية بجمع البيانات فضلًا عن الحوافز الصريحة بالتأكد. وكانت التكلفة لكل وحدة للأسر المعيشية أو الأفراد المتتبعين أعلى بكثير أيضًا. بالإضافة إلى ذلك، أدت عمليات فحص الجودة الشاملة إلى زيادة التكلفة والوقت المستغرق في جمع البيانات. وعلى الرغم من ذلك، في سياق برنامج Atención a Crisis التجريبي، ظلت العينة ممثلة على مستوى الأسر المعيشية والأفراد على حد سواء بعد مرور فترة من ثلاثة إلى أربعة أعوام من بعد مسح خط الأساس، وتم الوصول بخطأ القياس إلى أدنى حد مع ضمان موثوقية بيانات التقييم. وكنتيجة لذلك، أمكن تحليل الآثار طويلة الأجل لبرنامج Atención a Crisis التجريبي على نحو مرضٍ.

المصدر: Macours, Premand, and Vakis 2012.

في نيكاراغوا إلى بيانات عالية الجودة مع انخفاض كبير في التناقص ومعدلات عدم الاستجابة وقلّة من أخطاء القياس والمعالجة. ويمكن الحصول على مثل هذه البيانات ذات الجودة العالية عند اتباع إجراءات جودة بيانات وحوافز سليمة في وقت بدء عملية جمع البيانات.

وفي نهاية عملية جمع البيانات، يجب تقديم مجموعة البيانات بوثائق تفصيلية (بما في ذلك دفتر رموز كامل وقاموس بيانات) مخزنة في مكان آمن (انظر المربع ١٦-٦). إذا كان يتم جمع البيانات لتقييم أثر ما، فيجب إذًا أن تشمل مجموعة البيانات معلومات تكميلية حول حالة المعالجة ومشاركة البرنامج. ومن شأن المجموعة الكاملة من الوثائق أن تسرع وتيرة تحليل بيانات تقييم الأثر، وتساعد في التوصل إلى نتائج يمكن استخدامها في صنع القرار على نحو عاجل، وتيسير مشاركة المعلومات والتكرار المحتمل.

المربع ١٦,٦: المبادئ التوجيهية لتوثيق البيانات وحفظها

لا تكون مرتبطة إلا بمعلومات المجيب في ملفات ضبط معلومات الهوية. ملفات ضبط معلومات الهوية تضم المعلومات الكاملة قبل إزالة الهوية الشخصية منها. ويحظر حفظها إلا في خادم آمن مع الحرص على عدم تضمينها أبدًا في أي من كتالوجات البيانات.

- **البيانات الكلية** تتضمن جميع المستندات الداعمة ذات الصلة بتفسير البيانات الجزئية، مثل قاموس بيانات، ودفتر رموز، ووصف لتصميم الدراسة، والاستبيانات.

ويساعد تنظيم كتالوجات البيانات الكلية والجزئية على حماية أمن البيانات ويتبع المعايير الدولية لتخزين البيانات. وتكون كتالوجات البيانات المركزية أقل عرضة للعبث أو الاختراق بالمقارنة مع أقراص التخزين الصلبة للحاسوب أو أجهزة التخزين المحمولة. وضمن كتالوجات بيانات معينة، يمكن حماية البيانات بكلمة مرور لفترة من الزمن قبل إتاحتها للجمهور.

عندما نتحدث عن توثيق البيانات، علينا ذكر توجيه أساسي، ألا وهو الاحتفاظ بسجل لبيانات تقييم الأثر كافة. ويتضمن ذلك بروتوكولات جمع البيانات، والاستبيانات، والأدلة التدريبية، وما شابه. يمتلك البنك الدولي وبنك التنمية للبلدان الأمريكية ومؤسسة تحدي الألفية (Millennium Challenge Corporation) إلى جانب مؤسسات أخرى مبادرات بيانات مفتوحة حيث تكون هذه البيانات متاحة للجمهور عن طريق كتالوج بيانات.

ويمكن تقسيم تخزين البيانات إلى ثلاث فئات: البيانات الجزئية والبيانات الكلية وملفات ضبط معلومات الهوية (ID).

- **البيانات الجزئية** هي بيانات على مستوى وحدة الملاحظة التي يتم طمس الهوية الشخصية منها ولا تتضمن أي معلومات تفيد بهوية الأفراد. ويتم طمس هوية المتغيرات المعرفة للهوية ذات الصلة باستخدام معرفات

مصادر إضافية

- للحصول على المواد المرفقة بهذا الكتاب وروابط المصادر الإضافية، يرجى زيارة موقع "تقييم الأثر في الواقع العملي" الإلكتروني (<http://www.worldbank.org/ieinpractice>).
- للوصول إلى دليل حول تصميم الاستبيان، راجع الوحدة المعدة بشأن "Applied Fieldwork Techniques" (التقنيات التطبيقية للعمل الميداني) في الدورة التدريبية الخاصة بطرق تقييم الأثر التطبيقية التي تقدمها جامعة كاليفورنيا. (<http://aie.cega.org>).
- بالنسبة لمنشورات المدونة حول جمع البيانات، راجع القائمة المختارة بعناية بشأن مدونة الأثر الإنمائي للبنك الدولي (World Bank Impact Evaluation Blog) (<http://blogs.worldbank.org/impactevaluations>) (Bank Development Impact). للمزيد من المعلومات حول جمع البيانات، يرجى الاطلاع على ما يلي:
- Fink, Arlene G., and Jacqueline Kosecoff. ٢٠٠٨. How to Conduct Surveys: A Step by Step Guide, fourth edition. London: Sage.
- Iarossi, Giuseppe. ٢٠٠٦. The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents. Washington, DC: World Bank.
- Leeuw, Edith, Joop Hox, and Don Dillman. ٢٠٠٨. International Handbook of Survey Methodology. New York: Taylor & Francis Group.
- للمزيد من المعرفة حول أنشطة جمع البيانات والإشراف على جودة البيانات، راجع World Bank Impact Evaluation Toolkit الوحدة ٥ Data Collection (جمع البيانات) (<http://www.worldbank.org/health/impactevaluationtoolkit>). تسرد هذه الوحدة بضعة أمثلة على تقارير صيرورة المسوحات، وأدلة إرشادية للعمل ميداني، وبرامج تدريبية للأسر المعيشية والمرافق الصحية.

- لاستعراض مختلف المواد التي ستؤهلك لتحضير مسح، راجع مركز Inter- American Development Bank Evaluation (تقييم بنك التنمية للبلدان الأمريكية) (<http://www.iadb.org/evaluationhub>). في قسم Data Collection (جمع البيانات)، يمكنك تنزيل
 - دليل مصممي الاستبيان
 - دليل إدخال البيانات
 - نماذج الموافقة، واستبيانات العينة، وبرامج إدخال البيانات، والأدلة الميدانية لأنواع مختلفة من المسوحات (بما يتضمن ذلك مسوحات للأسر المعيشية والمجتمعات والمرافق الصحية والمدارس والمزارعين)
 - روابط إلى المزيد من الأمثلة على أسئلة المسح والاستبيانات
 - روابط إلى أدلة إرشادية لجمع بيانات عالية الجودة
- روابط إلى الأدوات المتوفرة على الموقع الإلكتروني للشبكة الدولية لمسوحات الأسر المعيشية International Household Survey Network, IHSN (تخزين وإدارة البيانات. لمعرفة المزيد حول مدى أهمية توثيق البيانات، وكيفية القيام بذلك ومَن ضمن فريق التقييم مسؤول عنه، راجع World Bank Impact Evaluation Toolkit، الوحدة ٦ حول Data Storage (تخزين البيانات) (<http://www.worldbank.org/health/impactevaluationtoolkit>).

ملاحظات

١. اطلع على المراجع في (2000) Grosh and Glewwe والأمم المتحدة (٢٠٠٥). انظر أيضًا أبحاث (2005) Muñoz، و (2006) Fink and Kosecoff، و (2008) Leeuw, Hox, and Dillman، التي تضع بين يديك الوفير من التوجيهات العملية لجمع البيانات.
٢. راجع أبحاث (2012) McKenzie and Rosenzweig لاكتساب نظرة عامة حول أحدث التطورات. للأمثلة على مثل هذه التجارب، انظر (2012) McKenzie and Rosenzweig بشأن المشكلات العامة، و (2012) Beegle, Carletto, and Himeleing بشأن البيانات الزراعية، و (2012) Beegle and others بشأن استهلاك الأسر المعيشية، و (2011) Bardasi and others بشأن بيانات العمل.
٣. وللإطلاع على أمثلة للابتكارات في قياس الحاصل، ارجع إلى (2013) Holla، و (2007) Das and Hammer، و (2015) Planas and others.

المراجع

- Baird, S., and B. Özler. 2012. "Examining the Reliability of Self-reported Data on School Participation." *Journal of Development Economics* 98 (1): 89–93.
- Bardasi, E., K. Beegle, A. Dillon, A., and P. Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *The World Bank Economic Review* 25 (3): 418–47.
- Beegle, K., C. Carletto, and K. Himelein. 2012. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98 (1): 34–41.

- Beegle, K., J. De Weerdt, J. Friedman, and J. Gibson. 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98 (1): 3–18.
- Caeyers, Bet, Neil Chalmers, and Joachim De Weerdt. 2012. "Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment." *Journal of Development Economics* 98 (1): 19–33.
- Chetty, R., J. N. Friedman, and E. Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–721.
- Das, J., and J. Hammer. 2007. "Money for Nothing: The Dire Straits of Medical Practice in Delhi, India." *Journal of Development Economics* 83 (1): 1–36.
- Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff. 2012. "Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels." *Journal of Development Economics* 98 (1): 51–57.
- Ferraz, C., and F. Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *The Quarterly Journal of Economics* 123 (2): 703–45.
- Fink, A. G., and J. Kosecoff. 2008. *How to Conduct Surveys: A Step by Step Guide*, fourth edition. London: Sage.
- Galiani, S., P. Gertler, and E. Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113 (1): 83–120.
- Galiani, Sebastian, and Patrick McEwan. 2013. "The Heterogeneous Impact of Conditional Cash Transfers." *Journal of Public Economics* 103: 85–96.
- Gertler, Paul, Paula Giovagnoli, and Sebastian Martinez. 2014. "Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer." Policy Research Working Paper 6884, World Bank, Washington, DC.
- Glewwe, Paul. 2005. "An Overview of Questionnaire Design for Household Surveys in Developing Countries." In *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations.
- Glewwe, Paul, and Pedro Olinto. 2004. "Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras' PRAF Program." Final report. University of Minnesota and IFPRI-FCND.
- Grosh, Margaret, and Paul Glewwe, eds. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington, DC: World Bank.
- Holla, Alaka. 2013. "Measuring the Quality of Health Care in Clinics." World Bank, Washington, DC. http://www.globalhealthlearning.org/sites/default/files/page-files/Measuring%20Quality%20of%20Health%20Care_020313.pdf.
- Iarossi, G. 2006. *The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents*. Washington, DC: World Bank.
- Kasprzyk, Daniel. 2005. "Measurement Error in Household Surveys: Sources and Measurement." In *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations.

- Leeuw, E., J. Hox, and D. Dillman. 2008. *International Handbook of Survey Methodology*. New York: Taylor & Francis Group.
- Macours, Karen, Patrick Premand, and Renos Vakis. 2012. "Transfers, Diversification and Household Risk Strategies: Experimental Evidence with Implications for Climate Change Adaptation." Policy Research Working Paper 6053, World Bank, Washington, DC.
- McKenzie, David, and Mark Rosenzweig. 2012. "Symposium on Measurement and Survey Design." *Journal of Development Economics* 98 (May 148–1 :1).
- Morris, Saul S., Rafael Flores, Pedro Olinto, and Juan Manuel Medina. 2004. "Monetary Incentives in Primary Health Care and Effects on Use and Coverage of Preventive Health Care Interventions in Rural Honduras: Cluster Randomized Trial." *Lancet* 37–2030 :364.
- Muñoz, Juan. 2005. "A Guide for Data Management of Household Surveys." In *Household Sample Surveys in Developing and Transition Countries*, chapter 15. New York: United Nations.
- Planas, M-E, P. J. García, M. Bustelo, C. P. Carcamo, S. Martinez, H. Nopo, J. Rodriguez, M-F Merino, and A. Morrison. 2015. "Effects of Ethnic Attributes on the Quality of Family Planning Services in Lima, Peru: A Randomized Crossover Trial." *PLoS ONE* 10 (2): e0115274.
- Pradhan, M., and L. B. Rawlings. 2002. "The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund." *World Bank Economic Review* 16 (2): 275–95.
- UN (United Nations). 2005. *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations.
- Vermeersch, Christel, Elisa Rothenbühler, and Jennifer Sturdy. 2012. *Impact Evaluation Toolkit: Measuring the Impact of Results-Based Financing on Maternal and Child Health*. World Bank, Washington, DC. <http://www.worldbank.org/health/impactevaluationtoolkit>.

الخلاصة

تقييمات الأثر: تدريبات معقدة جدية بالاهتمام

يتعلق تقييم الأثر بالتوصل إلى أدلة تفيد بمدى صلاحية البرامج أو عدم صلاحيتها وكيفية تحسينها لتحقيق حصائل إنمائية أفضل. ويمكن إنجاز ذلك في إطار عمل تقييم أثر تقليدي، أي عن طريق مقارنة الحصائل بين مجموعتي المعالجة والمقارنة. ويمكن أيضًا إجراء تقييمات الأثر لاستكشاف بدائل التنفيذ ضمن برنامج محدد لغرض اختبار الابتكارات أو معاينة أكثر من برنامج لتقييم الأداء المقارن.

ونرى أن تقييمات الأثر استثمار مُجدٍ للغاية بالنسبة للعديد من البرامج. عند إقران تقييمات الأثر بالمتابعة وأشكال التقييم الأخرى، تصير في وضع أفضل لتعزيز فهمنا لمدى نجاعة سياسات معينة، والإسهام في تحسين المساءلة لمديري البرنامج والحكومات والممولين والجمهور العام، وتزويد القرارات بالمعلومات الضرورية حول كيفية تخصيص الموارد الإنمائية المحدودة بشكل أكثر فعالية، وهي تضيف أيضًا إلى مخزون المعرفة العالمي بشأن ما يصلح وما لا يصلح في مجال التنمية.

قائمة المراجعة: العناصر الأساسية لتقييم أثر جيد التصميم

- ✓ إن تقييمات الأثر عبارة عن مجموعة من التعهدات المتشابكة التي تساهم في دفعها أطراف عديدة. وتسلط قائمة المراجعة التالية الضوء على العناصر الأساسية لتقييم الأثر جيد التصميم:
- ✓ سؤال ملموس وذو صلة متعلق بالسياسة - يتأسس في نطاق نظرية التغيير - يمكن الإجابة عنه باستخدام تقييم الأثر.
- ✓ منهجية دقيقة مشتقة من القواعد التشغيلية للبرنامج من أجل تقدير واقع مضاد يُظهر العلاقة السببية بين البرنامج والحصائل محل الاهتمام.
- ✓ فريق تقييم تم تشكيله على أساس جيد، يؤدي وظائفه بموجب علاقة شراكة بين فريق السياسات وفريق الأبحاث.
- ✓ احترام المعايير الأخلاقية واعتبار العناصر البشرية في تصميم وتنفيذ التقييم وجمع البيانات ذات الصلة وكذلك الانتباه إلى المبادئ العالمية المفتوحة لضمان الشفافية
- ✓ عينة ذات قوة إحصائية كافية لتسمح باكتشاف الآثار ذات الصلة بالسياسة
- ✓ منهجية وعينة توفران نتائج قابلة للتعميم للمجتمع الإحصائي محل الاهتمام
- ✓ بيانات عالية الجودة توفر المعلومات المناسبة المطلوبة لتقييم الأثر، بما في ذلك بيانات لمجموعتي المعالجة والمقارنة، وبيانات عند مرحلتي خط الأساس والمتابعة، ومعلومات حول تنفيذ البرنامج، والتكاليف.
- ✓ إستراتيجية إشراك تدعم الحوار حول السياسات من خلال تنفيذ تقييم الأثر، وكذلك تقرير تقييم الأثر وموجزات السياسة ذات الصلة المنشورة على الجماهير الأساسية بصورة حسنة التوقيت.

قائمة المراجعة: نصائح لتخفيف المخاطر الشائعة المصاحبة لإجراء تقييم الأثر

- نحن نسلط الضوء أيضًا على بعض النصائح التي يمكنها التخفيف من المخاطر الشائعة التي تنشأ في عملية إجراء تقييم الأثر:
- ✓ من الأفضل تصميم تقييمات الأثر في مرحلة مبكرة من دورة المشروع، وهي تشكل جزءًا من تصميم البرنامج في الظروف المثالية، ولكن يجب تصميمها على الأقل قبل تنفيذ البرنامج المراد تقييمه. ويسمح التخطيط المبكر بتصميم التقييم المرتقب استنادًا إلى أفضل منهجية متاحة ويوفر الوقت الضروري لتخطيط وتنفيذ عملية جمع بيانات خط الأساس في مجالات التقييم قبل بدء البرنامج.

- ✓ يجب تغذية نتائج تقييم الأثر بتقييم صيرورة تكميلي وبيانات المتابعة التي تعطي صورة واضحة لتنفيذ البرنامج. عند نجاح البرامج، من الضروري فهم سر نجاحها. وعندما تفشل البرامج، من الضروري التمييز بين التنفيذ الركيك للبرنامج والتصميم المنقوص للبرنامج.
- ✓ يجب جمع بيانات خط الأساس ووضع منهجية احتياطية ودمجها في تصميم تقييم الأثر. إذا كان تصميم التقييم الأصلي ليس له صلاحية - مثلاً بسبب أن مجموعة المقارنة الأصلية حصلت على فوائد البرنامج -، فإن توافر خطة احتياطية يمكن أن يساعدك على تجنبك الاضطرار إلى الاستغناء عن التقييم برمته.
- ✓ يجب الاحتفاظ بالمعرفات المشتركة فيما بين مصادر البيانات المختلفة لوحداث الملاحظة لديك بحيث يمكن ربطها بسهولة في أثناء مرحلة التحليل. فمثلاً، يجب أن تكون لأسرة معيشية معينة المعرف نفسه في أنظمة المتابعة وفي استقصائي تقييم الأثر عند خط الأساس والمتابعة على حد سواء.
- ✓ تُعد تقييمات الأثر مفيدة في الإلمام بكيفية نجاح برنامج ما، وفي اختبار بدائل البرنامج، حتى بالنسبة للبرامج المستمرة المنفذة على نطاق واسع. تستطيع تقييمات الأثر جيدة التصميم المساعدة في اختبار الابتكارات أو توفير رؤى حول مدى فعالية مختلف السلع والخدمات المقدمة في حزمة بموجب البرامج الحالية. وعند تضمين ابتكار برنامج إضافي كإطلاق تجريبي صغير في سياق تقييم أكبر، فيمكن لذلك أن يستفيد من التقييم لإنتاج معلومات ستكون لها قيمة عالية عند صنع القرار في المستقبل.
- ✓ يجب النظر إلى تقييمات الأثر على أنها مكون آخر لعملية البرنامج، ويتعين تزويدها بفرق عمل وميزانية مناسبة بدعم من الموارد الفنية والمالية المطلوبة. كن واقعياً بشأن التكاليف ومدى تعقد تنفيذ تقييم الأثر. إن عملية تصميم تقييم ما وجمع بيانات خط الأساس من الصفر يمكن أن يستغرق عاماً أو أكثر عادةً. وبمجرد بدء البرنامج، تحتاج مجموعة المعالجة إلى فترة كافية من التعرض للتدخل من أجل أن تؤثر في الحاصل. واعتماداً على البرنامج، قد يستغرق ذلك زمناً يتراوح بين عام وخمسة أعوام أو أكثر بالنسبة للحصائل طويلة المدى. فإن جمع البيانات لمسح متابعة واحد أو أكثر وإجراء التحليل ونشر النتائج سيتطلب أيضاً جهوداً مضنية طيلة أشهر عديدة، بل سنوات. إذا جمعنا كل ذلك معاً، قد تستغرق عادةً دورة تقييم الأثر الكاملة من البداية إلى النهاية ما لا يقل عن ثلاثة إلى أربعة أعوام من العمل المكثف والمشاركة. وتشكل الموارد المالية والفنية المناسبة ركناً أساسياً في كل خطوة على الطريق.

في النهاية، توفر تقييمات الأثر الفردية إجابات ملموسة عن أسئلة سياسة محددة. وعلى الرغم من أن هذه الإجابات توفر معلومات تتسم بأنها مخصصة للكيان المحدد القائم على التقييم ويموله، فهي توفر أيضًا معلومات ذات قيمة للغير من حول العالم ممن يستطيعون التعلم وصنع القرار استنادًا إلى تلك الأدلة. فعلى سبيل المثال، برامج التحويلات النقدية في إفريقيا وآسيا وأوروبا لها دروس مستفادة من تقييمات أصلية مثل Familias en Acción في كولومبيا و Progresa في المكسيك وبرنامج تحويلات نقدية شرطية في بلدان أخرى في أمريكا اللاتينية. وبهذه الطريقة، تكون تقييمات الأثر جزئيًا ذات منفعة عامة عالمية. فضلًا عن ذلك، فإن الأدلة التي يتم تأسيسها من خلال تقييم الأثر تضيف إلى المعرفة العالمية حول الموضوع المعني. ويمكن لقاعدة المعرفة تلك أن تغذي قرارات السياسات بالمعلومات بعد ذلك في دول أخرى وسياقات مختلفة مع المراعاة المدروسة لمدى الصلاحية الخارجية. ولهذا الغرض، يقطع المجتمع الدولي أشواطًا نحو توسيع رقعة الدعم لإجراء تقييمات ذات دقة عالية.

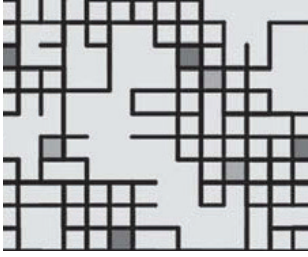
على مستوى الدولة، يتزايد عدد الحكومات الأكثر تطورًا وطلبًا التي تتطلع إلى إظهار النتائج وأن تكون في وضع مساهمة أفضل تحت راية تشريعاتها الأساسية. وكذلك يتزايد عدد التقييمات الجاري إجرائها من قبل الوزارات التنفيذية على الصعيدين الوطني ودون الوطني، وتتجهز الهيئات الحكومية لقيادة أجندة تقييم وطنية، مثل المجلس الوطني لتقييم سياسة التنمية الاجتماعية في المكسيك، ودائرة متابعة الأداء والتقييم في جنوب إفريقيا. وكذلك يتم استخدام الأدلة المكتسبة من تقييمات الأثر كمعلومات داعمة لتخصيصات الميزانية التي يعدها الكونجرس والبرلمان على المستوى الوطني. وفي الأنظمة حيث يتم تحديد مدى صلاحية البرامج استنادًا إلى الأدلة الدامغة والحصائل النهائية، تكون البرنامج ذات قاعدة الأدلة القوية المؤهلة للدفاع عن النتائج الإيجابية قادرة على الازدهار في حين أن البرامج التي تعوز إلى مثل هذه الأدلة تجد صعوبة أكبر في الاحتفاظ بالتمويل.

هناك مؤسسات متعددة الأطراف، مثل البنك الدولي وبنك التنمية للبلدان الأمريكية إلى جانب الوكالات الإنمائية الوطنية، والحكومات المانحة، والمؤسسات الخيرية، تتطلب أدلة أكثر وأفضل حول الاستخدام الفعال للموارد الإنمائية. ومثل هذه الأدلة تكون مطلوبة لتحديد حجم المساهمة بالنسبة لأولئك الذي يقرضون المال أو أولئك الذي يتبرعون بالمال وكذلك في صنع القرار حيث تكون الإستراتيجية الأفضل هي تخصيص موارد إنمائية شحيحة.

يوجد عدد متنامي من المؤسسات المتخصصة بشكل رئيسي في إنتاج تقييمات أثر عالية الجودة وهي آخذة في التوسع وتشمل تلك التي تنتمي إلى الأوساط الأكاديمية، مثل Poverty Action Lab و Innovations for Poverty Action، و Center for Effective Global Action، والوكالات المستقلة التي تدعم تقييمات الأثر، مثل International Initiative for Impact Evaluation (3ie). وهناك عدد من الجمعيات التي تجلب معًا مجموعات من ممارسي التقييم والباحثين وصانعي القرار المهتمين بالموضوع المعني، بما في ذلك "شبكة شبكات تقييم الأثر" والجمعيات الإقليمية، مثل "الجمعية الإفريقية للتقييم"، وشبكة تقييم الأثر للجمعية الاقتصادية في أمريكا اللاتينية والكاريبي. وتعكس كل تلك الجهود الأهمية المتزايدة لتقييم الأثر في السياسة الإنمائية الدولية.

وبأخذ هذا النمو الذي يشهده تقييم الأثر في الاعتبار، يُعد التمكن من لغة تقييم الأثر مهارة لا غنى عنها يتزايد الطلب عليها ويحتاج إليها أي ممارس في مجال التنمية سواء كنت تجري تقييمات ضمن مهام وظيفتك أو تتعاقد على تقييمات الأثر أو تستخدم نتائج تقييمات الأثر لصنع القرارات. عندما تتوفر أدلة قوية ناشئة عن تقييمات أثر بشأن موضوع ما، فيمكن لها أن تكون عوامل محرّكة لحوار سياسة التنمية؛ حيث إنها ستوفر الأساس الذي يدعم أو يعارض الاستثمارات في برامج وسياسات إنمائية. وتسمح الأدلة التي نستمدّها من تقييمات الأثر لصانعي السياسات ومديري المشاريع باتخاذ قرارات مستنيرة حول كيفية تحقيق الحاصل بالطريقة الأكثر فعالية من حيث التكلفة. وبقّما يتجهز فريق السياسات بالأدلة القائمة على تقييم الأثر، يتسنى لهم رأب الفجوات عن طريق إدخال هذه النتائج في عملية صنع القرار. وهذا النوع من الأدلة يمكن أن يشكل قاعدة مستنيرة للمناقشات والآراء وأخيرًا قرارات تخصيص الموارد البشرية والمالية للحكومات والمؤسسات متعددة الأطراف والجهات المانحة.

إن صنع القرار القائم على الأدلة يكون في الأساس متعلّقًا بالإلمام بتصميم البرنامج وتحسين تخصيص الميزانيات من أجل توسيع البرامج الفعالة من حيث التكلفة وتقليص البرامج غير الفعالة، وتقديم تحسينات إلى تصاميم البرنامج استنادًا إلى أفضل الأدلة المتاحة. لا يُعد تقييم الأثر مأخذًا أكاديميًا بحثًا. فإن تقييمات الأثر مدفوعة بالحاجة إلى أجوبة عن أسئلة سياسات تؤثر في الحياة اليومية للبشر. هناك إمكانيات غير مسبقة يمكنها تحسين رفاهية الأشخاص حول العالم من خلال القرارات التي تحدد الطريقة الأفضل لتخصيص الموارد الشحيحة للبرامج المكافحة للفقر، والنقل، والطاقة، والصحة، والتعليم، وشبكات الأمان، والقروض الصغيرة، والكثير من المبادرات الإنمائية الأخرى. لا بد من اتخاذ مثل هذه القرارات بالاستناد إلى أكثر الأدلة صحة ودقة.



قائمة المصطلحات

المصطلحات المسردة بخط مائل ضمن التعريفات تنفرد بتعريفها الخاص في قائمة المصطلحات.

SMART: محددة وقابلة للقياس وقابلة للإسناد وواقعية ومستهدفة. *المؤشرات الجيدة* تتسم بهذه الخصائص.

الأثر. يُعرف أيضًا باسم *التأثير السببي*. في سياق *تقييمات الأثر*، الأثر هو التغير في الحاصل الذي يمكن إسناده بشكل مباشر إلى برنامج أو طريقة برنامج أو ابتكار تصميمي.

اختبار الدلالة. اختبار يحدد ما إذا كانت *الفرضية البديلة* تحقق مستوى الدلالة المحدد بشكل مسبق من أجل قبولها كأفضلية على *الفرضية الصفرية*. إذا حصلت من اختبار الدالة على القيمة p الأدنى من مستوى الدلالة الإحصائية (α)، يتم رفض الفرضية الصفرية.

الاختبار الوهمي. اختبار قابلية الدحض ويُستخدم لتقييم ما إذا كانت الافتراضات التي تقترحها إحدى الطرق صحيحة أم لا. فمثلاً، عند تطبيق طريقة *الاختلاف في الاختلافات*، يمكن تنفيذ اختبار وهمي باستخدام مجموعة معالجة وهمية أو حصيلة وهمية: أي مجموعة أو حصيلة تعلم أنها غير متأثرة بالبرنامج. لا تستطيع الاختبارات الوهمية التأكيد على صحة الافتراضات، إلا أنها تستطيع تمييز الحالات التي تكون فيها الافتراضات غير صحيحة.

الاختيار. يحدث عندما تكون المشاركة في البرنامج قائمة على أساس التفضيلات أو القرارات أو الخصائص غير الملحوظة للمشاركين أو مسؤولي البرنامج.

الارتباط. قياس إحصائي يشير إلى مدى تأثير اثنين أو أكثر من المتغيرات ببعضها البعض.

الأساليب المختلطة. مقارنة تحليلية يجمع بين البيانات الكمية والنوعية.

إطار تحديد العينة. هو عبارة عن قائمة شاملة بالوحدات في المجتمع الإحصائي محل الاهتمام. ويلزم وضع إطار تحديد عينة مناسب لضمان أن الخلاصات الناتجة عن تحليل عينة ما يمكن تعميمها على المجتمع الإحصائي بالكامل. بفعل الفروق بين إطار تحديد العينة والمجتمع الإحصائي محل الاهتمام، ينشأ تحيز في التغطية. في وجود تحيز التغطية، تفقد النتائج الواردة عن العينة صلاحيتها الخارجية للمجتمع الإحصائي محل الاهتمام بالكامل.

افتراض قيمة معالجة الوحدة المستقرة (SUTVA). الشرط الأساسي الذي ينص على أن الحصيلة الخاصة بوحدة واحدة يجب عدم تأثرها بالتعيين المخصص للعلاجات إلى وحدات أخرى. وذلك الأمر ضروري لضمان أن *التعيين العشوائي* يحمل تقديرات غير متحيزة للأثر.

الامتثال غير المثالي. التعارض بين حالة المعالجة المعيّنة وحالة المعالجة الفعلية. ويحدث الامتثال غير المثالي عندما تشارك بعض الوحدات المعيّنة إلى مجموعة *الامتثال* في البرنامج أو لا تشارك بعض الوحدات المعيّنة إلى مجموعة *المعالجة* في البرنامج.

الامتثال. يحدث الامتثال عندما *تلتزم الوحدات* بتعيينها إلى مجموعة المعالجة أو مجموعة المقارنة.

البيانات الإدارية. هي بيانات يتم جمعها بصفة دورية من جانب وكالات عامة أو خاصة كجزء من إدارة برنامج محدد، وعادة ما تتكرر تلك العملية على أساس منتظم وغالبًا تكون عند تقديم الخدمة، بما في ذلك الخدمات المقدمة والتكاليف والمشاركة في البرنامج. وتُعد بيانات *المتابعة* شكلًا من أشكال *البيانات الإدارية*.

بيانات المسح. البيانات التي تغطي عينة من المجتمع الإحصائي محل الاهتمام. قارن ذلك بمصطلح بيانات التعداد السكاني.

التأثير السببي. راجع *الأثر*.

تأثير المعالجة على المعالجين (Treatment-On-the-Treated). تقيس تقديرات المعالجة على المعالجين (TOT) الفرق في الحاصل بين الوحدات التي تنتفع فعليًا من المعالجة ومجموعة المقارنة.

تأثير جون هنري. يحدث تأثير جون هنري عندما تسعى وحدات المقارنة جاهدة للتعويض عن كونها غير مشمولة في المعالجة. عندما نقارن وحدات معالجة مع وحدات المقارنة الكاذبة هذه، سيكون تقدير أثر *البرنامج متحيزًا*؛ أي أننا سنقوم بتقدير أثر أصغر للبرنامج عن الأثر الحقيقي مقارنة بالتقدير الذي كنا سنحصل عليه إذا لم تبذل تلك الوحدات جهدًا إضافيًا.

تأثير هوثورن. يحدث هذا التأثير عندما تتصرف الوحدات بشكل مختلف لمجرد أنها خاضعة للملاحظة.

تأثيرات التوازن العام. *التأثيرات غير المباشرة* التي تحدث عندما *تؤثر التدخلات* في العرض والطلب للسلع والخدمات وبالتالي تغير سعر السوق لتلك السلع أو الخدمات.

تأثيرات توازن السياق. *التأثيرات غير المباشرة* التي تحدث عندما *يؤثر تدخل* ما على المعايير السلوكية أو الاجتماعية ضمن سياق معين، مثل المكان الذي تُجرى فيه المعالجة.

التأثيرات غير المباشرة. تحدث عندما تؤثر مجموعة المعالجة بصورة مباشرة أو غير مباشرة على الحاصل في مجموعة المقارنة (أو العكس).

تجربة الآلية. هي تقييم أثر يختبر آلية سببية معينة ضمن نظرية التغيير الخاصة ببرنامج ما، بدلاً من اختبار التأثير السببي (الأثر) الخاص للبرنامج ككل.

تحليل الانحدار. هو طريقة تحليلية يتم استخدامها لتحليل العلاقات بين متغير تابع (المتغير المراد تفسيره) والمتغيرات التفسيرية. ولا يُعد تحليل الانحدار كافياً بشكل عام لاكتشاف التأثيرات السببية. في تقييم الأثر، يُعد تحليل الانحدار طريقة لتمثيل العلاقة بين قيمة مؤشر الحصيلة Y (المتغير التابع) ومتغير مستقل يسترعى التعيين إلى مجموعة المعالجة أو المقارنة مع إبقاء الخصائص الأخرى ثابتة في الوقت نفسه. كل من التعيين إلى مجموعتي المعالجة والمقارنة والخصائص الأخرى عبارة عن متغيرين تفسيرين. يمكن أن يكون تحليل الانحدار أحادي التغير (إذا كان لا يوجد إلا متغير تفسيري واحد، في حالة تقييم الأثر، المتغير التفسيري الوحيد يكون التعيين إلى مجموعة المعالجة أو المقارنة) أو متعدد المتغيرات (إذا كانت توجد بضعة متغيرات تفسيرية).

تحليل التكلفة والمنفعة. يقدّر هذا التحليل إجمالي المنافع المتوقعة لبرنامج ما، مقارنة بإجمالي التكاليف المتوقعة. ويسعى إلى تحديد جميع تكاليف وعائدات البرنامج من الناحية المالية، ويقيم ما إذا كانت المنافع تفوق التكاليف أم لا.

تحليل الحساسية. إلى أي مدى يكون التحليل حساساً للتغيرات التي تطرأ على الافتراضات. في سياق حسابات القوة الإحصائية، يساعد ذلك الإحصائيين على فهم الزيادة المطلوبة في حجم العينة بموجب افتراضات أكثر تحفظاً (مثل الأثر الأقل توقعاً أو التباين الأعلى في مؤشر الحصيلة أو القوة الأعلى في المستوى).

تحليل فعالية التكلفة. يقارن هذا التحليل التكلفة النسبية لاثنتين أو أكثر من البرامج أو بدائل البرنامج للوصول إلى حصيلة مشتركة، مثل المحاصيل الزراعية أو درجات اختبارات الطلاب.

تحيز الاختيار. يتعرض الأثر المقدّر للتحيز في الاختيار عندما يكون منحرفاً عن الأثر الحقيقي في وجود الاختيار. عادةً ما يحدث تحيز الاختيار عندما تكون هناك أسباب غير ملحوظة لمشاركة البرنامج مترابطة مع الحاصل. يحدث هذا التحيز بشكل شائع عندما تكون مجموعة المقارنة غير مؤهلة أو مستبعدة بشكل ذاتي من المعالجة.

تحيز الاستبدال. هو تأثير سلوكي غير مقصود يؤثر في مجموعة المقارنة. الوحدات التي لم يتم اختيارها للانتفاع بالبرنامج قد تكون قادرة على إيجاد بدائل جيدة للمعالجة، اعتماداً على جهودها الخاصة.

تحيز التغطية. يحدث عندما لا يتوافق إطار تحديد العينة مع المجتمع الإحصائي محل الاهتمام.

التحيز. في سياق تقييم الأثر، التحيز هو الفرق بين الأثر الذي يتم حسابه (أو تقديره) والأثر الحقيقي للبرنامج.

التدخل. في سياق تقييم الأثر، التدخل هو المشروع أو البرنامج أو الابتكار التصميمي أو السياسة المقرر تقييمها. ويُعرف أيضاً باسم المعالجة.

الترابط البيئي المجمع. يُعرف أيضًا باسم *الترابط البيئي الفئوي*. هذه هي درجة التماثل في *الحصائل* أو الخصائص فيما بين الوحدات ضمن المجموعات أو *المجموعات العنقودية* الموجودة مسبقًا بالنسبة إلى الوحدات الموجودة في مجموعات عنقودية أخرى. على سبيل المثال، الأطفال الملتحقون بالمدرسة نفسها سيكونون بالأساس أكثر تماثلًا أو ترابطًا من حيث منطقة سكنهم وخلفيتهم الاجتماعية والاقتصادية في مقابل الأطفال غير الملتحقين بالمدرسة نفسها.

الترويج العشوائي. طريقة تستخدم *المتغير المساعد* من أجل تقدير آثار البرنامج. تعيّن هذه الطريقة - بصورة عشوائية - شكلًا من *أشكال الترويج* أو التشجيع لمجموعة فرعية من الوحدات لحثها على المشاركة في البرنامج. يسعى الترويج العشوائي إلى زيادة الاشتراك في برنامج طوعي في عينة فرعية مختارة عشوائيًا من المجتمع الإحصائي. ويمكن أن يأخذ الترويج شكل حافز إضافي أو محفز أو معلومات تحث الوحدات على المشاركة في البرنامج دون التأثير بشكل مباشر على الحصيلة ذات الاهتمام. وبهذه الطريقة، يمكن ترك البرنامج متاحًا لجميع الوحدات المؤهلة.

التصميم المتقاطع. يُسمى أيضًا التصميم التقاطعي. ويتجسد هذا التصميم عندما يكون هناك تعيين عشوائي مع تدخلين أو أكثر، ما يسمح بتقدير أثر التدخل مفردًا وجمعًا.

تصميم انقطاع الانحدار (RDD). هي واحدة من طرق *تقييم الأثر شبه التجريبية* التي يمكن استخدامها للبرامج التي تعتمد على مؤشر متصل لتصنيف المشاركين المحتملين ولها نقطة انقطاع على طول المؤشر الذي يحدد ما إذا كان المشاركون المحتملون مؤهلين للاستفادة من البرنامج أم لا. يوفر حد الانقطاع لأهلية البرنامج نقطة فاصلة بين *مجموعة المعالجة ومجموعة المقارنة*. حصائل المشاركين على جانب واحد من هذا الحد مقارنة بحصائل غير المشاركين على الجانب الآخر من هذا الحد. عندما تمثل الوحدات كافة للتعيين الذي يتوافق معها على أساس مؤشر أهليتها، يُقال إن تحليل انقطاع الانحدار "شديد الدقة". أما إذا كان هناك عدم امتثال على أي من جانبي نقطة الانقطاع، يُقال إن تحليل انقطاع الانحدار "مضطرب".

التعداد السكاني. تعداد وإحصاء كامل للسكان. وتغطي بيانات التعداد السكاني جميع الوحدات المتضمنة في المجتمع الإحصائي. قارن ذلك بمصطلح العينة.

التعيين العشوائي أو التجارب العشوائية المضبوطة. طريقة *تقييم الأثر* حيث ترتبط كل وحدة (على سبيل المثال، الفرد أو الأسرة المعيشية أو الشركة أو المدرسة أو المستشفى أو المجتمع) باحتمال اختيارها للمعالجة عن طريق البرنامج. عند توازن عدد كبير بما يكفي *للوحدات*، تضمن عملية التعيين العشوائي التكافؤ في كل من الخصائص الملحوظة وغير الملحوظة بين *مجموعة المعالجة ومجموعة المقارنة*، وبهذا يتم استبعاد *التحيز في الاختبار*. يُعد التعيين العشوائي الطريقة الأكثر دقة لتقدير *الواقع المضاد*، وهو غالبًا ما يُشار إليه بالمعيار الذهبي *لتقييم الأثر*.

تقدير النية للعلاج (ITT). تقيس تقديرات النية للعلاج (ITT) الفرق في الحصائل بين الوحدات المعينة إلى *مجموعة المعالجة* والوحدات المعينة إلى *مجموعة المقارنة*، بغض النظر عما إذا كانت الوحدات المعينة إلى أي من المجموعتين تتلقى المعالجة بالفعل أم لا.

تقييم الأثر. هو تقييم ينشئ رابطة سببية بين البرنامج أو *التدخل* ومجموعة من *الحصائل*. ويجب تقييم الأثر عن السؤال: *ما الأثر* (أو التأثير السببي) للبرنامج على الحصيلة محل الاهتمام.

التقييم السابق. التقييمات المصممة والجاري تطبيقها من قبل تنفيذ البرنامج. يتم دمج التقييمات السابقة في خطط تنفيذ البرنامج. قارن ذلك بمصطلح *التقييم القائم على أحداث ماضية*.

التقييم اللاحق. هو تقييم مصمم بعد تنفيذ برنامج ما (لاحق). قارن ذلك بمصطلح *التقييم السابق*.

تقييم الصيرورة. هو تقييم يركز على كيفية تنفيذ البرنامج وتشغيله، وتقييم ما إذا كان يتوافق مع تصميمه الأصلي، وتوثيق تطويره وتشغيله. قارن ذلك بمصطلح *تقييم الأثر*.

التقييم. تقدير موضوعي ودوري لمشروع أو برنامج أو سياسة مخططة أو جارية أو مكتملة. وتستخدم التقييمات للإجابة عن أسئلة محددة، وغالبًا ما تكون ذات صلة بالتصميم، أو التنفيذ، أو النتائج.

التناقص. يحدث التناقص عندما ينخفض عدد وحدات العينة بين جولة واحدة من جمع البيانات وأخرى، مثلًا عندما ينتقل الأفراد المتضمنون في العملية الأولى إلى مكان آخر أو يتعذر تحديد موقعهم. ويُعد التناقص حالة من حالات عدم استجابة الوحدات. ويمكن للتناقص أن يثير التحيز في تقدير الأثر.

تنقيب البيانات. ممارسة تنطوي على معالجة البيانات بحثًا عن نتائج معينة.

جمع العينات الاحتمالية. واحدة من عمليات جمع عينات يتم فيها تعيين احتمال جيد التعريف لكل وحدة من المقرر سحبها من إطار تحديد العينة. وهي تشمل جمع العينات العشوائية، وجمع العينات العشوائية الطباقية، وجمع العينات العنقودية.

جمع العينات. عملية يتم من خلالها سحب عينات من إطار تحديد العينة التي تم جمعها من المجتمع الإحصائي محل الاهتمام. يمكن استخدام العديد من الإجراءات التبادلية لجمع العينات. وتُعد طرق جمع العينات الاحتمالية هي الأكثر صرامة ودقة لأنها تخصص احتمالًا معرفيًا جيدًا لكل وحدة مطلوب سحبها. يُعد كل من جمع العينات العشوائية، وجمع العينات العشوائية الطباقية، وجمع العينات العنقودية من طرق جمع العينات الاحتمالية. قد تنشأ أخطاء مرتبطة بجمع العينات عند استخدام طريقة جمع العينات غير الاحتمالية (مثل جمع العينات الهادفة أو الملائمة).

حجم التأثير. قوة التغيير في الحصيلة بفعل التدخل.

الحد الأدنى للتأثير القابل للاكتشاف. إن الحد الأدنى للتأثير القابل للاكتشاف عبارة عن مدخل لحسابات القوة الإحصائية؛ أي إنه يوفر حجم التأثير الذي من أجله تم تصميم تقييم الأثر بالنسبة لمستوى معين من الدلالة والقوة الإحصائيتين. يجب أن تكون عينات التقييم كبيرة بما يكفي لاكتشاف الحد الأدنى للتأثير القابل للاكتشاف ذي الصلة بالسياسة في ظل وجود قوة إحصائية كافية. يتم تحديد هذا الحد باعتبار التغيير في الحاصل الذي سيبرر الاستثمار في تدخل ما.

حسابات القوة الإحصائية. هي حسابات نحدد بها الحجم المطلوب للعينة لكي يتمكن تقييم الأثر من تقدير أثر البرنامج بدقة؛ أي أصغر عينة ستسمح لنا باكتشاف الحد الأدنى للتأثير القابل للاكتشاف. تعتمد حسابات القوة الإحصائية أيضًا على معلمات مثل القوة (أو احتمال وقوع خطأ النوع الثاني)، ومستوى الدلالة، والمتوسط، والتباين، والترابط البيئي المجمّع للحصيلة محل الاهتمام.

الحصيلة. نتيجة محل اهتمام تُقاس على مستوى المستفيدين من البرنامج. الحاصل هي النتائج المقرر تحقيقها بمجرد أن يستخدم المجتمع الإحصائي المستفيد مخزجات المشروع. لا تخضع الحاصل للتحكم المباشر من قبل الوكالة المنفذة للبرنامج؛ حيث إنها تتأثر بكل من تنفيذ البرنامج (*الأنشطة والمخرجات* التي يقدمها) والاستجابات السلوكية من المستفيدين المعرضين لذلك البرنامج (الاستفادة التي يحققها المستفيدون من الفوائد أو المزايا التي يتعرضون لها من خلال البرنامج). يمكن أن تكون الحصيلة متوسطة أو نهائية (طويلة الأجل). الحاصل النهائية هي الحاصل الأبعد. يمكن تفسير المسافة من حيث الوقت (تستغرق فترة زمنية أطول للوصول إلى الحصيلة) أو من حيث السببية (الروابط السببية العديدة المطلوبة للوصول إلى الحصيلة والعوامل المتعددة التي تؤثر فيها).

خط الأساس. الحالة قبل *التدخل* مقابل أي تقدم يمكن تقييمه أو مقارنة يصلح إجراؤها. تُجمع بيانات خط الأساس قبل تنفيذ البرنامج أو السياسة من أجل تقييم الحالة *القبليّة*. إن توفر بيانات خط الأساس مهم لتوثيق التوازن في الخصائص السابقة لتنفيذ البرنامج فيما بين مجموعتي المعالجة والمقارنة. ويلزم توفير بيانات خط الأساس لبعض التصاميم شبه التجريبية.

خطأ النوع الأول. يُعرف أيضًا باسم الخطأ *الإيجابي الكاذب*. هو خطأ يتم ارتكابه عند رفض *الفرضية الصفرية* رغم كونها صحيحة في واقع الأمر. في سياق *تقييم الأثر*، يحدث خطأ النوع الأول عندما يخلص التقييم إلى أن البرنامج كان له أثر (أي أن الفرضية الصفرية بعدم وجود أثر مرفوضة)، في حين لا يكون للبرنامج أثر في واقع الأمر (أي أن الفرضية الصفرية صحيحة). إن مستوى *الدلالة الإحصائية* يعني احتمال ارتكاب خطأ النوع الأول.

خطأ النوع الثاني. يُعرف أيضًا باسم الخطأ *السلبي الكاذب*. هو خطأ يتم ارتكابه عند قبول (وليس رفض) *الفرضية الصفرية*، في حين تكون غير صحيحة في واقع الأمر. في سياق *تقييم الأثر*، يحدث خطأ النوع الثاني عند الخلو إلى أن البرنامج لم يكن له أثر (أي أن الفرضية الصفرية بعدم وجود أثر غير مرفوضة)، في حين يكون للبرنامج أثر في واقع الأمر (أي أن الفرضية الصفرية غير صحيحة). احتمال ارتكاب خطأ النوع الثاني يساوي ١ مطروح منه *مستوى القوة*.

دراسة الفعالية. تقيّم مدى أداء البرنامج تحت العمل في ظروف طبيعية على نطاق واسع. عند تصميم هذه الدراسات وتنفيذها بشكل سليم، يمكن أن تكون النتائج التي نستخلصها منها ذات قابلية أكبر للتعميم بالمقارنة مع *دراسات النجاعة*.

دراسة النجاعة. تقيّم مدى أداء برنامج ما تحت العمل في ظروف مثالية. وتنفذ هذه الدراسات تحت ظروف خاصة للغاية، على سبيل المثال، مع المشاركة الفنية المكثفة من الباحثين في أثناء تنفيذ البرنامج. وغالبًا ما تُجرى هذه الدراسات لاختبار مدى قدرة برنامج ما على الاستمرار. وقد لا تكون نتائجها صالحة للتعميم فيما يتخطى نطاق التقييم.

درجة الميل. في سياق *تقييمات الأثر* باستخدام طرق المطابقة، درجة الميل هي احتمال أن وحدة ما ستشارك في البرنامج استنادًا إلى الخصائص الملحوظة. وهذه الدرجة عبارة عن رقم حقيقي يتراوح بين ٠ و١ يلخص تأثير جمع الخصائص الملحوظة على احتمالية المشاركة في البرنامج.

الدلالة. تشير الدلالة الإحصائية إلى احتمالية ارتكاب خطأ *النوع الأول*، أي احتمالية اكتشاف أثر ليس موجودًا بالفعل. عادةً ما يُرمز إلى مستوى الدلالة الإحصائية بالرمز الإغريقي α (ألفا). المستويات الشائعة للدلالة هي ١٠ و ٥ و ١ بالمائة. كلما كان مستوى الدلالة أصغر، زادت الثقة في مدى واقعية الأثر المقدّر. على سبيل المثال، إذا حددت مستوى الدلالة عند ٥ بالمائة، فيمكنك أن تكون واثقًا بنسبة ٩٥٪ في الخلوّص إلى أن البرنامج كان له أثر إذا وجدت أثرًا ملحوظًا بالفعل.

سلسلة النتائج. تحدد منطق البرنامج عن طريق تفسير كيف يمكن تحقيق الهدف الإنمائي. وهي تنص بوضوح على تسلسل المدخلات والأنشطة والمخرجات المتوقع أن تحسن الحاصل.

الصلاحية الخارجية. يكون التقييم ذا صلاحية خارجية إذا كانت عينة التقييم تمثل بصورة دقيقة المجتمع الإحصائي محل الاهتمام المتألف من *الوحدات* المؤهلة. يمكن بعد ذلك تعميم نتائج التقييم على المجتمع الإحصائي المكون من الوحدات المؤهلة. من الناحية الإحصائية، ليكون تقييم الأثر ذا صلاحية خارجية، يجب أن تكون عينة التقييم ممثلة للمجتمع الإحصائي محل الاهتمام. راجع أيضًا *الصلاحية الداخلية*.

الصلاحية الداخلية. يكون التقييم ذا صلاحية داخلية إذا كان يوفر تقديرًا دقيقًا *للواقع المضاد* من خلال *مجموعة مقارنة صالحة*.

طريقة الوحدات الضابطة الاصطناعية. هي طريقة مطابقة خاصة تسمح للإحصائيين بتقدير الأثر في أماكن حيث تكون وحدة (مثل دولة أو شركة أو مستشفى) واحدة تنتفع *بتدخل* أو معرضة لحدث ما. بدلاً من مقارنة هذه الوحدة المعالجة بمجموعة من الوحدات غير المعالجة، تستخدم هذه الطريقة معلومات حول خصائص الوحدة المعالجة والوحدات غير المعالجة لإنشاء وحدة مقارنة اصطناعية أو صناعية عن طريق موازنة كل وحدة غير معالجة بطريقة تجعل وحدة المقارنة الاصطناعية مماثلة للوحدة المعالجة بشكل كبير. ويتطلب ذلك سلسلة طويلة من الملاحظات بمرور الوقت للخصائص الخاصة بكل من وحدة المعالجة والوحدات غير المعالجة. وهذا الجمع بين وحدات المقارنة في وحدة اصطناعية يوفر مقارنة أفضل للوحدة المعالجة عن أي وحدة غير معالجة بشكل فردي.

الطريقة شبه التجريبية. طريقة تقييم الأثر التي لا تعتمد على *التعيين العشوائي* للمعالجة. ومن بين الأمثلة على الطرق شبه التجريبية: *الاختلاف في الاختلافات*، و*تصميم انقطاع الانحدار*، و*المطابقة*.

العامل المتغير بمرور الزمن. العامل الذي يتغير بمرور الوقت.

العامل غير المتغير بمرور الزمن. العامل الذي لا يتغير بمرور الوقت، أي أنه ثابت.

عدم استجابة العناصر. يحدث ذلك عندما تكون البيانات غير مكتملة بالنسبة لبعض *الوحدات* التي تشملها العينة.

عدم استجابة الوحدات. ينشأ عندما لا تكون هناك معلومات متوفرة لمجموعة فرعية من الوحدات، أي عندما تكون العينة الفعلية مختلفة عن العينة المخطط لها.

العينة. في الإحصاء، العينة هي مجموعة فرعية من المجتمع الإحصائي محل الاهتمام. بشكل أساسي، المجتمع عدد أفراد كبير جدًا، والاضطلاع بتعداد سكاني أو تعداد كامل لجميع القيم في المجتمع أمر غير عملي، بل مستحيل. بدلاً من ذلك، يستطيع الباحثون اختيار مجموعة فرعية تمثيلية للمجتمع ككل (باستخدام إطار تحديد العينة) وجمع الإحصائيات حول هذه العينة، التي يمكن استخدامها لطرح استدلالات أو استقراء للمجتمع الإحصائي ككل. ويُشار إلى هذه العملية باسم أخذ العينات. قارن ذلك بمصطلح التعداد السكاني.

عدم الاستجابة. يحدث عندما تكون البيانات ناقصة أو غير مكتملة بالنسبة لبعض الوحدات التي تشملها العينة. ينشأ عدم استجابة الوحدات عندما لا تكون هناك معلومات متوفرة لبعض وحدات العينة؛ أي عندما تكون العينة الفعلية مختلفة عن العينة المخطط لها. ويُعد *التناقص* شكلاً من أشكال عدم استجابة الوحدات. يحدث عدم استجابة العناصر عندما تكون البيانات غير كاملة لبعض الوحدات المشمولة في العينة في مرحلة زمنية معينة. قد يسبب عدم الاستجابة تحيزاً في نتائج التقييم إذا كان مرتبطاً بحالة المعالجة.

العلوم المفتوحة. هي حركة تهدف إلى إضفاء المزيد من الشفافية على الأساليب البحثية، بما يشمل ذلك من خلال تسجيل التجارب، واستخدام خطط التحليل المسبق، وتوثيق البيانات، والتسجيل.

عمليات المحاكاة السابقة. هي تقييمات تستخدم البيانات المتاحة لمحاكاة التأثيرات المتوقعة على *الحصائل* محل الاهتمام عند تنقيح برنامج أو سياسة ما.

العينة الطبقية. يتم الحصول عليها عن طريق قسمة المجتمع الإحصائي محل الاهتمام (*إطار تحديد العينة*) إلى مجموعتين (على سبيل المثال، ذكور وإناث)، ومن ثم سحب عينة عشوائية ضمن كل مجموعة. العينة الطبقية هي عينة احتمالية: كل وحدة في كل مجموعة (أو طبقة) لها احتمال معروف بأن يتم سحبها. بشرط أن تكون كل فئة كبيرة بما يكفي، تتيح طريقة أخذ العينات الطبقية إمكانية اكتساب استدلالات حول *الحصائل*، ليس على مستوى المجتمع الإحصائي فحسب بل أيضاً ضمن كل فئة.

العينة العشوائية. هي عينة يتم سحبها على طريقة جمع *العينات العشوائية* حيث ترتبط كل وحدة في *إطار تحديد العينة* باحتمال معروف بأنه سيتم سحبها. إن اختيار عينة عشوائية هو أفضل طريقة لتجنب اختيار عينة غير ممثلة. ينبغي عدم الخلط بين جمع العينات العشوائية والتعيين العشوائي.

العينة العنقودية. هي عينة تتألف من مجموعات عنقودية.

الفرضية البديلة. هي الفرضية التي تنص على أن *الفرضية الصفرية* غير صحيحة. وفي تقييم *الأثر*، الفرضية البديلة هي عادة الفرضية التي تنص على أن *التدخل* له أثر على *الحصائل*.

الفرضية الصفرية. هي فرضية يمكن بيان عدم صلاحيتها على أساس البيانات الملحوظة. وعادة تقترح الفرضية الصفرية موقفاً عاماً أو افتراضياً. في تقييم *الأثر*، عادةً ما تنص الفرضية الصفرية على أن البرنامج لا يكون له أثر؛ أي أن الفرق بين *الحصائل* في مجموعتي المعالجة والمقارنة هو صفر.

الفرضية. تفسير مقترح لظاهرة قابلة للملاحظة. راجع أيضاً *الفرضية الصفرية* و *الفرضية البديلة*.

الاختلاف في الاختلافات. يُعرف أيضاً باسم *الفرق المزدوج* أو *DD*. يقارن الاختلاف في الاختلافات التغير في *الحصائل* بمرور الوقت بين مجموعة المعالجة ومجموعة المقارنة. وهذا يلغي أي فروق بين هاتين المجموعتين تتسم بالثبات بمرور الوقت.

فريق التقييم. الفريق الذي يجري التقييم. من الضروري عقد شراكة مع مجموعتين: فريق من صانعي السياسات ومديري البرامج (فريق السياسات) وفريق من الباحثين (فريق الأبحاث).

قابلية التعميم. مدى صحة نتائج تقييم مُنفَّذ محليًا في أماكن أخرى وفيما بين مجموعات مجتمع إحصائي آخر.

القوة (أو القوة الإحصائية). احتمالية أن يكتشف تقييم الأثر أثرًا (أي فرق بين مجموعة المعالجة ومجموعة المقارنة) عندما يكون هناك أثر بالفعل. القوة الإحصائية تساوي ١ مطروح منه احتمال وقوع خطأ النوع الثاني، وتتراوح بين ٠ و ١. أشهر مستويين للقدرة الإحصائية هما ٠,٨ و ٠,٩. يدل ارتفاع مستوى القوة على أنها أكثر تحفظًا، وهذا يعني أنه توجد احتمالية منخفضة لعدم اكتشاف آثار حقيقية للبرنامج.

القوة الإحصائية. قوة الاختبار الإحصائي هي احتمال أن يقوم الاختبار بدحض الفرضية الصفرية عندما تصح الفرضية البديلة (أي تلك التي لا تسبب خطأ النوع الثاني). مع تزايد القوة، تقل فرص حدوث خطأ النوع الثاني. إن احتمال حدوث خطأ النوع الثاني يُشار إليه بالمعدل السلبي الكاذب (β). وبالتالي، القوة تساوي $1 - \beta$.

متابعة البيانات. البيانات المستمدة من متابعة برنامج ما توفر معلومات أساسية حول طريقة تقديم التدخل، بما يشمل ذلك تحديد مَنْ المستفيدين وأي فوائد أو مخبرات للبرنامج قد يتلقونها. وبيانات المتابعة هي نوع من أنواع البيانات الإدارية.

المتابعة. العملية المستمرة لجمع المعلومات وتحليلها من أجل تقييم مدى جودة أداء برنامج أو مشروع أو سياسة ما. عادةً تنطوي المتابعة على تتبع المدخلات والأنشطة والمخرجات، وفي أوقات أخرى تشمل النتائج أيضًا. تُستخدم المتابعة لتوفير معلومات تقوم عليها القرارات والإدارة على أساس يومي. ويمكن استخدامها أيضًا لتتبع الأداء مقابل النتائج المتوقعة وعقد مقارنات عبر البرامج وتحليل الاتجاهات بمرور الوقت.

المتغير التابع. عادةً ما يكون المتغير الناتج. وهو المتغير المراد تفسيره على عكس المتغيرات التفسيرية.

المتغير التفسيري. يُعرف أيضًا باسم المتغير المستقل. هو متغير مستخدم على الجانب الأيمن من الانحدار ويساعد على تفسير المتغير التابع الموجود على الجانب الأيسر من الانحدار.

المتغير المساعد. يُعرف أيضًا بـ *الأداة المساعدة*. تعتمد طريقة المتغير المساعد (IV) على مصدر خارجي للتباين من أجل تحديد حالة المعالجة. يؤثر المتغير المساعد على احتمال المشاركة في برنامج ما، ولكنه خارج عن سيطرة المشارك ولا علاقة له بخصائص المشارك.

المتغير. في الاصطلاح الإحصائي، هو رمز يمثل قيمة تتسم بالتغير.

المتغيرات غير الملحوظة. الخصائص غير الملحوظة. وقد تشمل خصائص مثل الدافع أو التفضيلات أو سمات شخصية أخرى يصعب قياسها.

متوسط تأثير المعالجة (ATE). أثر البرنامج تحت افتراض الامتثال الكامل، أي أن جميع الوحدات التي تم تعيينها إلى برنامج هي بالفعل مسجلة فيه ولم تستفد أي وحدة من وحدات المقارنة من البرنامج.

متوسط تأثير المعالجة المكاني. هو أثر البرنامج المقدّر لمجموعة فرعية خاصة للمجتمع الإحصائي، مثل الوحدات التي تمثل لتعيينها إلى مجموعة المعالجة أو المقارنة في وجود الامتثال غير المثالي أو بالقرب من درجة انقطاع الأهلية عند تطبيق تصميم انقطاع الانحدار. وبالتالي، لا يوفر متوسط تأثير المعالجة المكاني (LATE) إلا تقديرًا موضعيًا لأثر البرنامج ويجب عدم تعميمه على المجتمع الإحصائي بالكامل.

المجتمع الإحصائي محل الاهتمام. هو عبارة عن مجموعة شاملة من جميع *الوحدات* (مثل الأفراد والأسر المعيشية والشركات والمرافق) التي تُعد مؤهلة للاستفادة من التدخل أو المعالجة، ويسعى *تقييم الأثر* إلى تقدير آثار البرنامج عليها.

مجلس المراجعة المؤسسية (IRB). هو عبارة عن لجنة تم تعيينها لمراجعة وموافقة ومراقبة الأبحاث التي تنطوي على عناصر بشرية. ويُطلق عليها أيضًا *لجنة الأخلاقيات المستقلة (IEC)* أو *مجلس المراجعة الأخلاقية (ERB)*.

المجموعة الضابطة. تُعرف أيضًا باسم *مجموعة المقارنة* (راجع التعريف).

المجموعة العشوائية. الوحدات التي تنتمي إلى فئة معينة ويمكن أن تشارك في الخصائص. على سبيل المثال، الأطفال الملتحقون بالمدرسة نفسها سينتمون إلى مجموعة عشوائية لأنهم يشاركون في المرافق المدرسية والمعلمين ويعيشون في الحي نفسه.

مجموعة المعالجة. تُعرف أيضًا باسم *المجموعة المعالجة* أو *مجموعة التدخل*. مجموعة المعالجة هي مجموعة من *الوحدات* التي تنتفع *بتدخل* ما مقابل عدم انتفاع *مجموعة المقارنة* به.

مجموعة المقارنة. تُعرف أيضًا باسم *المجموعة الضابطة*. مجموعة المقارنة الصالحة تتماثل في الخصائص - في المتوسط - مع مجموعة المستفيدين من البرنامج (*مجموعة المعالجة*)، فيما عدا حقيقة واحدة، ألا وهي أن الوحدات في مجموعة المقارنة لا تستفيد من البرنامج. وتُستخدم مجموعات المقارنة لتقدير *الواقع المضاد*.

المخرج. المنتجات والسلع والخدمات الملموسة التي يتم إنتاجها (تزويدها) مباشرة عن طريق *أنشطة* البرنامج. يكون تقديم المخرجات خاضعًا للتحكم المباشر من قبل الوكالة المنفذة للبرنامج. إن استخدام المخرجات من قبل المستفيدين يسهم في إحداث تغييرات في *الحصائل*.

المدخلات. الموارد المالية والبشرية والمادية المستخدمة لتمكين *التدخل*.

مسح المتابعة. يُعرف أيضًا باسم *مسح ما بعد التدخل*. وهو دراسة مسحية يتم إجراؤها ميدانيًا بعد بدء البرنامج، بمجرد أن يحقق المستفيدون الاستفادة منه بعد فترة من الوقت. يمكن أن يتضمن *تقييم الأثر* عدة دراسات مسحية للمتابعة يُشار إليها في بعض الأحيان *بالدراسات المسحية لخط المنتصف وخط النهاية*.

مطابقة درجة الميل. هي طريقة مطابقة تعتمد على *درجة الميل* لإيجاد مجموعة مقارنة لمجموعة معالجة معينة

المطابقة. هي طريقة تقييم أثر شبه تجريبية تستخدم مجموعات البيانات الكبيرة والأساليب الإحصائية من أجل إنشاء أفضل مجموعة مقارنة ممكنة لمجموعة معالجة معينة استنادًا إلى الخصائص الملحوظة.

المعالجة. راجع مصطلح *التدخل*.

مقارنات المشتركين وغير المشتركين. تُعرف أيضًا باسم *المقارنات المختارة ذاتيًا*. تقارن هذه الإستراتيجية بين *حصائل الوحدات* التي تختار المشاركة في البرنامج و*حصائل الوحدات* التي تختار عدم المشاركة في البرنامج.

المقارنة القبلية والبعديّة. تُعرف أيضًا باسم *المقارنة السابقة واللاحقة* أو *المقارنة الانعكاسية*. وتعمل هذه الإستراتيجية على تتبع التغييرات في *الحصائل* للمستفيدين من البرنامج بمرور الوقت باستخدام القياسات السابقة واللاحقة لتنفيذ البرنامج أو السياسة من دون استخدام مجموعة مقارنة.

المقدّر. في الإحصاء، المقدّر هو قاعدة مستخدمة لتقدير خاصية مجهولة للمجتمع الإحصائي (مصطلحها الفني هو *المعلمة*) بالاستناد إلى البيانات، ويكون التقدير نتيجة التطبيق الفعلي للقاعدة على عينة محددة من البيانات.

منطقة الدعم المشترك. عند استخدام طريقة *المطابقة*، يأخذ الافتقار إلى الدعم المشترك صورة انعدام التداخل بين *درجات الميل* الخاصة بمجموعة المعالجة (أو المجموعة المشاركة) ودرجات الميل للمجموعة غير المشاركة.

الموافقة المستنيرة. أحد الأركان الأساسية لحماية حقوق العناصر البشرية. في حالة *تقييمات الأثر*، فهي تشترط أن يكون المجيبون على فهم واضح بالغرض والإجراءات والمخاطر والفوائد لعملية جمع البيانات التي يُطلب منهم المشاركة فيها.

مؤشر الأهلية. يُعرف أيضًا باسم *متغير الفرض*. هو *متغير* يصنف المجتمع الإحصائي *محل الاهتمام* على طول خط متصل وله حد أو درجة انقطاع تحدد مَنْ يكون مؤهلاً ومن ليس كذلك.

المؤشر. هو *متغير* يقيس ظاهرة محل اهتمام بالنسبة لفريق التقييم. يمكن أن تكون الظاهرة عبارة عن *مدخل أو مخرج أو حصيلة أو خاصية أو سمة*. راجع أيضًا SMART.

النشاط. الإجراءات المتخذة أو الأعمال المؤداة التي يتم فيها *حشد المدخلات*، مثل الأموال والمساعدة الفنية وأنواع الموارد الأخرى، لإنتاج *مخرجات* محددة، مثل المال المنفق أو الكتب الإرشادية الموزعة أو عدد المشاركين المسجلين في برنامج توظيف.

نظرية التغيير. تفسر القنوات التي من خلالها يمكن للبرامج التأثير في *الحصائل النهائية*. وتصف المنطق السببي لكيفية وأسباب قدرة برنامج معين، أو طريقة البرنامج، أو الابتكار التصميمي على الوصول إلى *الحصائل المستهدفة*. وتُعد نظرية التغيير ركيزة أساسية لأي *تقييم أثر*، وذلك بالنظر إلى تركيز البحث على السبب والنتيجة.

الواقع المضاد. هو *الحصيلة* (Y) التي كانت ستنتج في حالة عدم إشراك المستفيدين في البرنامج (P). وبحسب التعريف، يتعذر ملاحظة الواقع المضاد. ولهذا السبب، يجب تقديره باستخدام *مجموعة مقارنة*.

الوحدة. هي شخص أو أسرة معيشية أو مجتمع أو شركة أو مدرسة أو مستشفى أو وحدة أخرى للملاحظة قد تنتفع ببرامج ما أو تتأثر به.

الحفاظ على البيئة

بيان الفوائد البيئية

تلتزم مجموعة البنك الدولي بالحد من آثارها السلبية على البيئة. وفي سبيل دعم هذا الالتزام، يستعين "قسم النشر والمعرفة" بخيارات النشر الإلكترونية وتقنية الطباعة عند الطلب، الموجودة في المراكز الإقليمية حول العالم. وبذلك، تعمل هذه المبادرات على خفض مرات الطباعة وتقليل مسافات الشحن، ما ينتج عنه انخفاض في استهلاك الورق، واستخدام المواد الكيميائية، وانبعاثات غازات الدفئية، والنفايات.

يتبع "قسم النشر والمعرفة" المعايير الموصى بها لاستخدام الورق الذي حددته "مبادرة الصحافة الخضراء". وأغلبية كتبنا مطبوعة على ورق معتمد من هيئة الإشراف على الغابات (FSC) مع احتوائها جميعها تقريبًا على ٥٠ إلى ١٠٠ بالمائة من المحتوى المعاد تدويره. الألياف المعاد تدويرها في ورق كتبنا إما غير مبيضة أو مبيضة باستخدام عمليات خالية تمامًا من الكلور (TCF) أو المعالجة الخالية من الكلور (PCF) أو العمليات العنصرية المعززة الخالية من الكلور (EECF).

تتوفر المزيد من المعلومات حول الفلسفة البيئية للبنك على الرابط التشعبي <http://www.worldbank.org/corporateresponsibility>



دليل تقييم الأثر في الواقع العملي هو مقدمة شاملة ومتاحة لصانعي السياسات والممارسين لتقييم الأثر في مجال التنمية. وقد تم استخدامه على نطاق واسع عبر المجتمعات الإنمائية والأكاديمية. يتضمن الكتاب أمثلة واقعية لتقديم مبادئ توجيهية لعملية تصميم وتنفيذ تقييمات الأثر. سيكتسب القراء فهماً لتقييمات الأثر وأفضل الطرائق لاستخدامها لتصميم سياسات وبرامج قائمة على الأدلة. تغطي النسخة المحدثه أحدث التقنيات لتقييم البرامج وتتضمن أحدث نصائح التنفيذ، بالإضافة إلى مجموعة موسعة من الأمثلة ودراسات الحالة التي تستند إلى تحديات التنمية الأخيرة. كما يتضمن مواد جديدة عن أخلاقيات البحوث والشراكات لإجراء تقييم الأثر. وينقسم الدليل إلى أربعة أقسام:

- ١- ما يجب تقييمه ولماذا؟
 - ٢- الأساليب الرئيسية لتقييم الأثر .
 - ٣- كيفية إدارة تقييمات الأثر .
 - ٤- أخذ عينات تقييم الأثر وجمع البيانات.
- كما ويحتوي الدليل دراسات حالة لمختلف تطبيقات تقييمات الأثر. سيكون هذا الدليل موردًا قيمًا للمجتمع الإنمائي الدولي والجامعات وصانعي السياسات الذين يتطلعون إلى بناء أدلة أفضل حول ما يصلح في التنمية.