Asian Development Bank

تقييم أثر التدخلات الإنمائية - دليل عملي

"هاورد وايت
ديفيد ايه رايتزر "

تمت الترجمة والمراجعة من خلال

مركز الأثر
IMPACT CENTER
مستقبل أفضل
BETTER FUTURE

تقييم أثر التدخلات الإنمائية

دليل عملي

هوارد وايت

ديفيد إيه رايتزر

بنك التنمية الآسيوي

ⓒ مركز الأثر للبحوث والدراسات المحدودة ، ١٤٤٦هـ

رايتزير ، دايفيد ايه
تقييم أثر التدخلات الإنمائية - دليل عملي. / رايتزير ، دايفيد ايه ؛ وايت ، هاورد ؛ العمودي ، محمد ؛ ترجمان للترجمة .- الرياض ، ١٤٤٦هـ

رقم الإيداع: ١٤٤٦/٩٥٦٧
ردمك: ٩٧٨-٦٠٣-٩٢٢٦٥-٦-٣

هـذا المصنـف متـاح بمقتضـى ترخيـص المشـاع الإبداعـي، نَسـب المُصنَّـف ٣,٠ لفائـدة المنظمـات الحكوميـة الدوليـة (CC BY 3.0 IGO)

حقوق الطبع والنشر © لعام ٢٠١٧ محفوظة لبنك التنمية الآسيوي
٦ شارع بنك التنمية الآسيوي، مدينة ماندالويونغ، ١٥٥٠ مترو مانيلا، الفلبين
هاتف: ٤٤٤٤ ٦٣٢ ٢ ٦٣+؛ فاكس: ٢٤٤٤ ٦٣٦ ٢ ٦٣+
www.adb.org

بعض الحقوق محفوظة. نُشِر في عام ٢٠١٧.
طُبع في الفلبين.

الرقم الدولي الموحد للكتاب (ISBN) ٦-٠٥٨-٩٢٦١-٩٢-٩٧٨ (النسخة المطبوعة)، ٣-٠٥٩-٩٢٦١-٩٢-٩٧٨ (النسخة الإلكترونية)
رقم تعريف الكتاب: TCS179188-2
مُعرّف الكائن الرقمي (DOI): http://dx.doi.org/10.22617/TCS179188-2

تعبر الآراء الواردة في هذا الكتاب عن آراء المؤلفين ولا تعكس بالضرورة وجهات نظر بنك التنمية الآسيوي (ADB) وسياساته أو آراء أعضاء مجلس محافظيه أو الحكومات التي يمثلونها.

ولا يضمن بنك التنمية الآسيوي دقة البيانات الواردة في هذا الكتاب، كما لا يتحمل أي مسؤولية ناجمة عن أي نتيجة لاستخدامها. ولا يعني ذكر شركات أو منتجات محدّدة تصدرها الشركات المصنعة أن تلك الشركات أو المنتجات معتمدة أو موصى بها من جانب بنك التنمية الآسيوي تفضيلًا لها على الشركات أو المنتجات الأخرى ذات الطبيعة المماثلة التي لم تُذكر في هذا الكتاب.

إذا ذُكِرت أي تسمية أو إشارة إلى أي إقليم أو منطقة جغرافية معينة، أو إذا اُستُخدِم مصطلح "دولة/قطري" في هذه الوثيقة، فإن ذلك لا يعني أن بنك التنمية الآسيوي يعتزم إصدار أي أحكام فيما يتعلق بالحالة القانونية أو أي حالة أخرى لأي إقليم أو منطقة.

هذا العمل مُتاح بموجب ترخيص المشاع الإبداعي، نَسب المُصنَّف ٣,٠ لفائدة المنظمات الحكومية الدولية (CC BY 3.0 IGO) المبين بهذا الرابط (https://creativecommons.org/licenses/by/3.0/igo/). وتجدر الإشارة إلى أن استخدامك لمحتوى هذا الكتاب يعني موافقتك على الالتزام بشروط هذا الترخيص. لمعرفة معلومات عن الإسناد والترجمات والتعديلات والأذونات، يُرجى قراءة أحكام الاستخدام وشروطه على هذا الرابط (https://www.adb.org/terms-use#openaccess).

لا يسري ترخيص المشاع الإبداعي (CC) على المواد المحمية بحقوق الطبع والنشر غير المملوكة لبنك التنمية الآسيوي في هذا الكتاب. وإذا كانت تلك المواد منسوبة إلى مصدر آخر، فيُرجى الاتصال بمالك حقوق الطبع والنشر أو ناشر ذلك المصدر للحصول على إذنه بإعادة نسخها. ويُعفى بنك التنمية الآسيوي من تحمل المسؤولية الناجمة عن أي مطالبات تنشأ نتيجة استخدامك لهذه المواد.

يُرجى التواصل من خلال البريد الإلكتروني (pubsmarketing@adb.org) إذا كانت لديك أي أسئلة أو تعليقات فيما يخص المحتوى، أو إذا كنت ترغب في الحصول على إذن مرتبط بحقوق الطبع والنشر كي تتمكن من استخدام هذا المحتوى، وهو ما لا يندرج ضمن هذه الشروط، أو للحصول على إذن لاستخدام شعار بنك التنمية الآسيوي.

تخضع الصور الواردة في هذا الكتاب لملكية بنك التنمية الآسيوي.

ملاحظات:
تشير مفردة "دولار"، في هذا الكتاب، إلى الدولار الأمريكي.
يمكن العثور على تصويبات لإصدارات بنك التنمية الآسيوي على هذا الرابط (http://www.adb.org/publications/corrigenda).

إقرار

"إن دليل تقييم أثر التدخلات الإنمائية - دليل عملي الذي نشره بنك التنمية الآسيوي في الأصل باللغة الإنجليزية تحت عنوان:

Impact Evaluation of Development Interventions A Practical Guide

© 2017 بنك التنمية الآسيوي.

جميع الحقوق محفوظة.

متاح على العنوان التالي:

https://www.adb.org/publications/impact-evaluation-development-interven tions-practical-guide هذا الدليل متاح ومرخص بموجب ترخيص المشاع الإبداعي نسب المصنف - غير موطنة 3.0 دولي، وفقا للشروط التالية: نَسب المُصنَّف – يجب عليك نَسب العمل لصاحبه بطريقة مناسبة، وتوفير رابط للترخيص، وبيان إذا ما قد أُجريت أي تعديلات على العمل. يمكنك القيام بهذا بأي طريقة مناسبة، ولكن على ألا يتم ذلك بطريقة توحي بأن المؤلف أو المرخِّص 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://cre- ativecommons.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. ""الآراء المعرب عنها في هذا المنشور هي آراء المؤلفين ولا تعكس بالضرورة آراء وسياسات بنك التنمية الآسيوي (ADB) أو مجلس محافظيه أو الحكومات التي يمثلونها. ""لا يضمن مصرف التنمية الآسيوي دقة البيانات الواردة في هذا المنشور ولا يقبل أي مسؤولية عن أي نتيجة لاستخدامها. ولا يعني ذكر شركات أو منتجات محددة للمصنعين أن مصرف التنمية الآسيوي يؤيدها أو يوصي بها تفضيلا لشركات أو منتجات أخرى ذات طابع مماثل غير مذكورة. داخلي. وهذه المعلومات متاحة لإدارة مصرف التنمية الآسيوي وموظفيه. يمكن مشاركته خارج مصرف التنمية الآسيوي بإذن مناسب. "إن إدراج مصرف التنمية الآسيوي، أي تسمية أو إشارة إلى إقليم أو منطقة جغرافية معينة، أو باستخدام مصطلح" بلد "في هذه الوثيقة، لا يعتزم إصدار أي أحكام بشأن الوضع القانوني أو أي وضع آخر لأي إقليم أو منطقة".

المحتويات

الجداول والأشكال والمربعات vi
تمهيد viii
شكر وتقدير x
الاختصارات xi

١. مقدمة: تقييم أثر التنمية القائمة على الأدلة ١
١-١ ما أسباب أهمية تقييم الأثر؟ ١
١-٢ أهداف تقييم الأثر ٣
١-٣ ما الأسئلة التي قد يجيب عنها تقييم الأثر؟ ٤
١-٤ تقييم الأثر في دورة المشروع ١١
١-٥ تقييم الأثر والتقييم والتحليل الاقتصادي ١٣
١-٦ نبذة عن هذا الكتاب ١٥

٢. استخدام نظريات التغيير لتحديد أسئلة تقييم الأثر ٢٠
٢-١ مقدمة ٢٠
٢-٢ ما نظرية التغيير؟ ٢١
٢-٣ إشكاليات تطرحها نظرية التغيير ٢٣
٢-٤ بناء نظرية التغيير ٢٦
٢-٥ تطبيق نظرية التغيير ٢٨

٣. المفاهيم الأساسية لتقييم الأثر ٣١
٣-١ ما تقييم الأثر؟ ٣١
٣-٢ تحديد المجموعات الضابطة ومجموعات المقارنة ٣٣
٣-٣ التحيزات والتحديات المتعلقة بالاستدلال السببي ٣٦
٣-٤ البعد الزمني للآثار ٤٠
٣-٥ وحدة التعيين والمعالجة والتحليل ٤٢
٣-٦ القياسات المختلفة للأثر - لمن تُوجَّه الآثار؟ ٤٣
٣-٧ الصلاحية الداخلية والخارجية ٤٥

٤. التجارب العشوائية المضبوطة ٤٧

٤-١ مقدمة ٤٧
٤-٢ لماذا يتم أخذ العينات عشوائيًا؟ ٤٨
٤-٣ أنواع تصميمات التجارب العشوائية المضبوطة ٤٩
٤-٤ الخطوات المتبعة في تنفيذ التجارب العشوائية المضبوطة ٥٤
٤-٥ تقديم تقارير تتناول تفاصيل التجارب العشوائية المضبوطة ٥٩
٤-٦ الاستخدام العملي للتجارب العشوائية المضبوطة ٥٩
٤-٧ الملخص ٦٤

٥. التصاميم غير التجريبية ٦٦

٥-١ مقدمة ٦٦
٥-٢ تقديرات الاختلاف فى الاختلافات (DiD) ٦٨
٥-٣ الوحدات الضابطة الاصطناعية ٧٢
٥-٤ مطابقة درجة الميل ٧٣
٥-٥ موازنة درجة الميل والتقنيات القوية المزدوجة ٧٨
٥-٦ تصميم انقطاع الانحدار والسلاسل الزمنية المتقطعة ٧٩
٥-٧ المتغيرات المساعدة ٨٤
٥-٨ مقاربات المعالجة داخلية النشأة ووظيفة التحكم ٨٧
٥-٩ الانحدارات التبديلية داخلية النشأة ٨٩
٥-١٠ الملخص ٩٠

٦. تحديد ما يجب قياسه وكيفية قياسه: جمع البيانات لتقييم الأثر ٩٦

٦-١ أهمية البيانات عالية الجودة ٩٦
٦-٢ تحديد ماهية البيانات المطلوبة ٩٧
٦-٣ مصادر البيانات ٩٨
٦-٤ تصميم المسح ١٠١
٦-٥ أخذ العينات ١٠٨
٦-٦ إدارة جمع البيانات ١١٠

٧. تحديد حجم العينة لجمع البيانات ١١٦

٧-١ حسابات القوة الإحصائية: مقدمة ١١٦
٧-٢ ماذا نعني بالقوة الإحصائية؟ ١١٧
٧-٣ حسابات القوة الإحصائية لتصاميم الدراسة البسيطة ١٢١
٧-٤ حسابات القوة الإحصائية للتصاميم العنقودية ١٢٣
٧-٥ خطر الدراسات ضعيفة القدرة الإحصائية ١٢٦
٧-٦ وسائل مساعدة لاعتبار حسابات القوة الإحصائية وإجرائها ١٢٧

٨. إدارة عملية تقييم الأثر ١٣١
٨-١ مقدمة إلى إدارة تقييمات الأثر ١٣١
٨-٢ التخطيط لتقييمات الأثر ١٣١
٨-٣ اختيار تصميم تقييم الأثر ١٣٤
٨-٤ التوقيت وإعداد الميزانية لتقييم الأثر ١٣٧
٨-٥ مصادر الدعم لإجراء تقييم الأثر ١٤٣
٨-٦ تفسير استنتاجات تقييم الأثر وتقديمها ١٤٥
٨-٧ الخلاصات: التوجيهات التنموية القائمة على الأدلة ١٤٧

الملحق الأول. تطبيق طرق التقدير على تقييم الأثر ١٥٣
القسم الأول: إطار الحصائل المحتملة ١٥٣
القسم الثاني: التجارب العشوائية المضبوطة ١٥٧
القسم الثالث: نماذج الاختلاف فى الاختلافات (DiD) والتأثيرات الثابتة ١٥٩
القسم الرابع: الوحدات الضابطة الاصطناعية ١٦٣
القسم الخامس: المقارباتالمستندة إلى درجة الميل
(المطابقة، والوزن، وعوامل التقدير القوية المزدوجة) ١٦٥
القسم السادس: المقاربات المستندة إلى المتغيرات المساعدة (مربعات صغرى ثنائية المرحلة،
وانحدارات المعالجة داخلية النشأة، والانحدارات التبديلية داخلية النشأة) ١٧١
القسم السابع: تصميم انقطاع الانحدار ١٧٩

الملحق الثاني: تصميم المسح الميداني ١٨٥
القسم الأول: أنواع البيانات اللازمة لتقييم الآثار ١٨٥
القسم الثاني: أخذ العينات والتمثيل ١٨٦
القسم الثالث: اعتبارات إضافية بخصوص حجم العينة وحسابات القوة الإحصائية ١٩٠
القسم الرابع: تصميم أداة المسح ١٩١
القسم الخامس: تنفيذ المسح ١٩٧
القسم السادس: إدارة البيانات ٢٠٠

الجداول والأشكال والمربعات

الجداول

١-١ تقييم الأثر والتحليل الاقتصادي للاستثمارات..... ١٤
١-٢ توضيح لأحد نماذج التغيير السلوكي في شكل جدول..... ٢٦
١-٣ أمثلة على وحدات تعيين ومعالجة وتحليل ٤٣
١-٤ التشابه بين عينتيْن مأخوذتيْن من المجتمع الإحصائي نفسه ٤٨
١-٥ أثر الطرق الريفية في تنمية الأسواق في فيتنام..... ٧٠
٢-٥ مثال للمتغيرات المستقلة في معادلة المشاركة لمطابقة درجة الميل ٧٥
٣-٥ حساب تقدير أثر درجة الميل: مثال باستخدام بيانات درجة الاختبار..... ٧٧
٤-٥ تصميم انقطاع الانحدار لتقديرات أثر النفقات السنوية للفرد المستمد من تقييم أثر برنامج "بانتاويد باميليا" للتحويل النقدي المشروط في الفلبين، (2013 Philippine Pesos) ٨٢
٥-٥ موجز طرق تقييم الأثر غير التجريبية..... ٩١
١-٧ الأخطاء المحتمل ارتكابها عند تخمين الأثر..... ١١٧
١-٨ اختيار تصميم تقييم الأثر: المقاربة المتبعة لاتخاذ القرار..... ١٣٥
٢-٨ جدول زمني توضيحي لتقييم الأثر ١٣٩
٣-٨ ميزانيات مختارة لتقييمات أثر مدعومة من بنك التنمية الآسيوي ١٤٠

الأشكال

١-١ الإصدار السنوي لتقييمات الأثر ٣
١-٢ تقييم الأثر ودورة المشروع..... ١١
١-٢ نظرية التغيير لأحد مشاريع المياه والصرف الصحي..... ٢٢
٢-٢ مثال على مخطط قُمعي للتناقص..... ٢٣
١-٣ شكل توضيحي لتقييم الأثر..... ٣٢
٢-٣ معدلات الإنتاجية الواقعية والواقعية المضادة الناجمة عن أحد مشاريع الري ٤١
١-٥ شكل توضيحي للاختلاف فى الاختلافات (DiD)..... ٦٩
٢-٥ مثال لتوزيع درجات الميل..... ٧٦
٣-٥ مثال لانقطاع الانحدار حول عتبة الأهلية لمتغير التعيين..... ٨٢

تابع الأشكال

١-٧ خطأ من النوع الأول ١١٩
٢-٧ رسم توضيحي للخطأ من النوع الثاني ١١٩
٣-٧ الحد من خطأ من النوع الثاني عن طريق تقليل مستوى الدلالة ١٢٠
٤-٧ الحد من خطأ من النوع الثاني عن طريق زيادة حجم العينة ١٢٠
٥-٧ كلما كبر التأثير، كان اكتشافه أسهل ١٢١
٦-٧ زيادة حجم العينة تؤثر عكسيًا في الحد الأدنى لحجم التأثير: العلاقة بين الحد الأدنى لحجم التأثير وحجم العينة ١٢٣
٧-٧ زيادة عدد المجموعات العنقودية تأثيرها أكبر على القوة الإحصائية مقارنةً بزيادة عدد الملاحظات في كل مجموعة عنقودية ١٢٥

المربعات

١-١ استخدام الأدلة الناتجة عن عمليات تقييم الأثر في تحديد مدى انتشار التحويلات النقدية المشروطة في أمريكا اللاتينية ٣
٢-١ أمثلة على أسئلة الجيل الأول: مشروع توسعة مترو العاصمة الجورجية تبليسي ٥
٣-١ استنتاجات لدراسات حول تقييم الأثر في مجال البنية التحتية للكهرباء ٦
٤-١ استنتاجات لدراسات حول تقييم الأثر في مجال البنية التحتية للنقل ٧
٥-١ التعرف على تصميم المشروع من تقييم الأثر: المصابيح الكهربائية الموفرة للطاقة في باكستان ٩
٦-١ الاختبار التجريبي في القطاع الخاص ١٠
٧-١ دراسات تقييم الأثر السريعة ١٠
١-٢ تنقيح سياسة التغيير استنادًا إلى مدخلات الأطراف المعنية: القسائم المدرسية في الفلبين ٢٧
١-٣ إظهار التوازن في دراسة تتعلق بالترويج لغسل اليدين في باكستان ٣٤
٢-٣ تحيّز الاختيار في إطار الحصول على المياه المُحسنة في نيبال ٣٨
١-٤ تصميم تشجيعي متوالٍ يرتكز على مجموعة متعددة المعالجة ٥٤
١-٥ استخدام المتغيرات المساعدة لقياس أثر توصيل الكهرباء على الأسر الريفية ٨٥
١-٦ مصطلحات المسح ١٠٠
٢-٦ منصات لإجراء المقابلات الشخصية بمساعدة الحاسوب ١٠٦
٣-٦ مثال لنص الموافقة المستنيرة ١١٢
١-٨ إعداد تقارير بشأن معالجة القضايا الأخلاقية ١٣٧
٢-٨ تحليل فعالية تكلفة التدخلات الرامية لزيادة المواظبة المدرسية ١٤٧

تمهيد

اتجاهان مهمان يشكلان مستقبل المساعدة الإنمائية، ولا سيما في آسيا. يتمثل الاتجاه الأول في أن البلدان النامية أصبحت قادرةً أكثر من أي وقت مضى على الحصول على التمويل من مجموعة أوسع نطاقًا من المصادر، بما في ذلك المصادر الخاصة. في حين يتمثل الاتجاه الثاني في تزايد اهتمام صانعي السياسات بالأدلة التي يمكن أن تجعل البرامج أكثر فعالية. وهذا يعني أن عرض القيمة الخاص بوكالات التنمية، مثل بنك التنمية الآسيوي (ADB)، أصبح يعتمد بشكل متزايد على القدرة على تقديم المعرفة، بدلاً من التمويل وحده. وفي هذا الإطار، هناك عنصر بالغ الأهمية من هذه المعرفة يُستمد من الأدلة المتعلقة بالتأثيرات المقصودة وغير المقصودة للتدخلات. ويعد تقييم الأثر الوسيلة الرئيسية للاختبار التجريبي لما يحدث بالفعل عند تنفيذ التدخلات.

بالتوازي مع هذين الاتجاهين، فقد أضحى الاقتصاد السلوكي مثار اهتمام متزايد للمعنيين بمجال اقتصاديات التنمية. وفي هذا المجال، غالبًا ما كان يسود الاعتقاد خلال العقدين الماضيْن أن السلوك البشري يتبع آليًا الافتراضات الكلاسيكية الجديدة. وخلال الفترة التي تلت ذلك، أدرك الاقتصاديون بشكل متزايد الحاجة إلى مراجعة أفكارهم واختبار مدى صحة تلك الافتراضات، وذلك عبر استخدام تقنيات تقييم الأثر التجريبية وشبه التجريبية، والتي كانت رائدة إلى حد كبير في الطب ومجالات العلوم الأخرى.

إن هذا التقارب بين الحاجة المتزايدة إلى أدلة لتقييم الأثر، بين الخبراء العاملين فيأ ميدان التنمية، والاهتمام المتزايد بذلك، بين الأكاديميين، يتيح فرصة فريدة للبحث والممارسة على حد سواء. ومن الممكن أن يجذب تقييم الأثر بعضًا من المواهب الاقتصادية الرائدة في العالم للمشاركة في مشاريع تنموية محددة. ولن تؤدي مثل هذه المشاركة إلى التوصل فحسب إلى أدلة جديدة دقيقة حول "الوسائل الصالحة والفعالة" في مجال التنمية، ولكنها تعزز أيضًا بشكل مباشر تنفيذ المشروع. وغالبًا ما يكون لدى الباحثين الرائدين الذين عملوا في العديد من البلدان والبرامج رؤى يمكن أن تسهم في تصميم أفضل لعملية التدخل أثناء وضع تصور لعملية تقييم الأثر. كما يتطلب تقييم الأثر توضيح المنطق المرتبط بنتائج المشروع والافتراضات الأساسية قبل اختبارها، مما يتيح تصميمًا أفضل للمشاريع.

يمكن أن يساعد تقييم الأثر في الحصول على أنواع التعقيبات الإيجابية التي يتم استخدامها بشكل منتظم في تطوير المنتجات في القطاع الخاص. ومن الممكن أن تقدم الأدلة الأسباب الجوهرية التي تدعم الاستمرار في المشاريع والبرامج الفعالة أو توسيع نطاقها، بغض النظر عن البيئة السياسية. كما أن تقييم الأثر يمكن أن يختبر طرقًا مختلفة لمعالجة مشكلة ما، فضلاً عن تحديد العوامل التي تؤثر على تحقيق الآثار المرجوة، وكذلك تقديم رؤى حول كيفية تنفيذ التدخلات ودمجها. ويتيح تقييم الأثر أيضًا الوصول إلى برنامج عمل يدعم التوصل إلى إثباتات حول صحة المفاهيم والأفكار المتعلقة بالابتكارات. وفي النهاية، يمكن أن يساعد أيضًا في بناء نظريات أساسية حول السلوك البشري وسبل تنميته، والاستفادة من الآراء الحكيمة السائدة وتحويلها إلى أسس يمكن الارتكاز عليها، وإعادة توجيه التنمية نحو تبني مقاربات أكثر فعالية.

لكي يحدث هذا، يجب نشر تقييم الأثر واعتماده، بحيث يتمكن الخبراء في ميدان التنمية من القيام بانتظام بالتفكير في الدلالات التي يمكن الاستفادة منها عند أخذ أدلة الأثر السابقة في الاعتبار أثناء تقييم أنشطتهم، وكذلك كيف يمكن أن يساعد تقييم الأثر الناجم عن تدخلاتهم في خلق أدلة جديدة لمشاريع أخرى بحاجة إلى التقييم. يهدف هذا الكتاب إلى المساعدة في إثبات هذا التأصيل من خلال إمكانية استخدامه كمرجع يمكن الاستفادة منه من قبل مجموعة من القراء ذوي الصلة، مدعومًا في ذلك بسنوات من الخبرة التي راكمها المؤلفان في مجال تنفيذ دراسات تقييم الأثر. وفيما يتعلق بالقراء غير العاملين في الحقل الاقتصادي، فإنه يوفر أوصافًا عامة للمفاهيم الأساسية ومقدمات للطرق الرئيسية التي يمكن الاستفادة منها في هذا الخصوص، فضلاً عن "القواعد الأساسية" لفهم الموضوعات الفنية، مثل حساب القوة الإحصائية. وبالنسبة للقراء الأكثر تخصصًا على الصعيد الأكاديمي، يقدم الكتاب أوصافًا أكثر تفصيلاً للتقنيات ومقدمات لأوامر برنامج ستاتا (STATA) في الملاحق الفنية. وبالمقارنة مع مؤلفات أخرى سابقة، يقدم هذا الكتاب مجموعة أكبر من المنهجيات، بالإضافة إلى المزيد من الوصف فيما يتعلق بالاعتبارات العملية.

في الوقت الراهن، يعمل بنك التنمية الآسيوي على توسيع نطاق تغطية تقييم الأثر للمشاريع التي يساهم فيها، ويُكرس مزيدًا من الاهتمام والموارد للدراسات الجديدة في مجال تقييم الأثر. وفي إطار دعم هذه العملية، يعد هذا الكتاب مصدرًا قيّمًا يمكن أن يساعد في زيادة الوعي بما تقدمه مقاربة تقييم الأثر وكيف يمكن تطبيقه. إنني أوصي به كمصدر عملي لأولئك الذين لديهم اهتمام باستنباط أو استخدام أدلة دقيقة على "الوسائل الصالحة والفعالة" في مجال التنمية.

ياسويوكي ساوادا
كبير الخبراء الاقتصاديين والمدير العام
قسم البحوث الاقتصادية والتعاون الإقليمي

شكر وتقدير

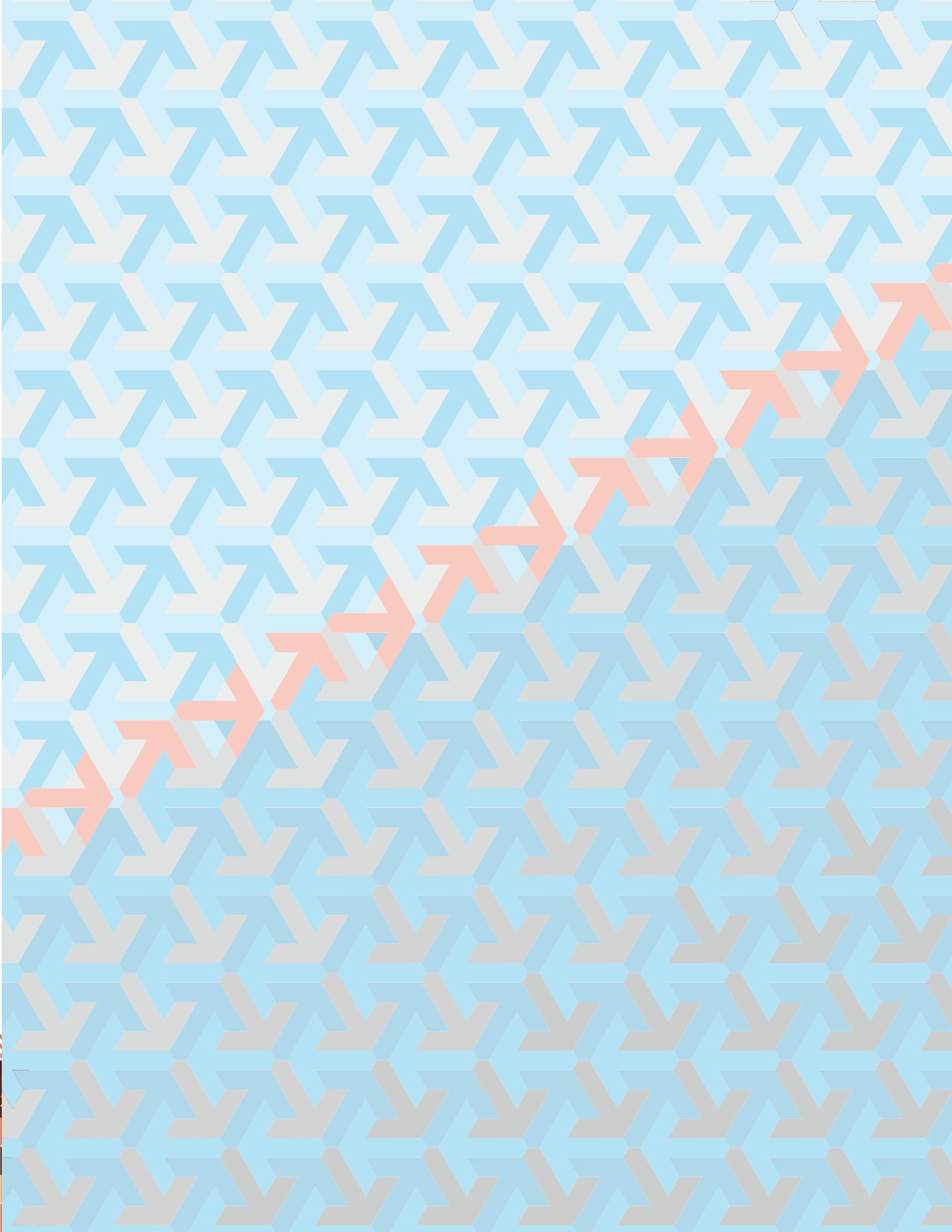
هذا الكتاب هو ثمرة مجهودات ومساهمات العديد من الأفراد داخل وخارج بنك التنمية الآسيوي. وقد تم إنتاجه وفقًا لتوجيه شامل من قبل إديمون جينتينغ، مدير قسم التحليل الاقتصادي ودعم العمليات، وكذلك رنا حسن، مديرة قسم اقتصاديات ومؤشرات التنمية. وقد استفاد هذا الكتاب أيضًا من التوجه العام للجنة تقييم الأثر بين الإدارات المختلفة التابعة لبنك التنمية الآسيوي. وقد قام كل من هوارد وايت، الذي يعمل مستشارًا مستقلاً، وديفيد رايتزر، الذي يعمل خبيرًا اقتصاديًا في قسم التحليل الاقتصادي ودعم العمليات، بتأليف محتوى الكتاب.

على الصعيد ذاته، قدَّم ساكيكو تاناكا رؤى وإسهامات قيِّمة في الإصدارات الأولى لهذا الكتاب. كما قدم سكوت روزيل، أحد الباحثين في جامعة ستانفورد، مواد معرفية أساسية، في حين اعتمدت بعض أقسام الكتاب على مواد أعدتها المستشارة المستقلة نينا بلوندال. وقدَّمت كل من ياسمين سيبال وماري آن كاجاس إسهامات ومدخلات إضافية. وتم تقديم دعم إداري من قِبل كل من ليليبيث بوت، وأماندا مامون، وجي آن بوراك، وريكاسول كالالوان، وروزلين بيريز، وجليني كاستيلو. وقد ساهم زملاء أعضاء في "لجنة تقييم الأثر" بتقديم مراجعات قيِّمة، وهم: آري بيردانا، وأرتور أندريسياك، وأرتورو مارتينيز، وبرنارد وودز، وكريستوفر إدموندز، وإليزابيتا جينتيل، وجواو فاهرينا، وكية. إي. سيثارام، وكيوشي تانيجوتشي، ولاكشمان ناجراج راو، ولارس جوهانس، ومايا فيجاياراجافان. وقد تم وضع التصورات والمفاهيم الأولية لهذا الكتاب وفقًا للتوجيه العام للسيد سين يونغ بارك، المدير السابق لقسم التحليل الاقتصادي ودعم العمليات.

على الصعيد ذاته، قام السيد تيوزداي سوريانو بتنقيح محتوى هذا الكتاب، بينما تولى السيد جو مارك جانابان مهمة إعداد تصميمه. وقد تم إنتاج هذا الكتاب بموجب المساعدة التقنية REG-0012: وضع منهجيات تقييم الأثر ومقارباته وقدراته في عدد محدد من البلدان الأعضاء النامية.

الاختصارات

3ie: International Initiative for Impact Evaluation	المبادرة الدولية لتقييم الأثر
ADB: Asian Development Bank	بنك التنمية الآسيوي
ATE: average treatment effect	متوسط تأثير المعالجة
ATT: average treatment effect on the treated	متوسط تأثير المعالجة على المُعالَج
ATU: average treatment effect on the untreated	متوسط تأثير المعالجة على غير المُعالَج
CCT: conditional cash transfer	التحويل النقدي المشروط
DiD: difference-in-differences	الاختلاف في الاختلافات (DiD)
DMF: design and monitoring framework	إطار التصميم والمتابعة
ICC: intracluster correlation coefficient	معامل الارتباط داخل المجموعات
IE: impact evaluation	تقييم الأثر
IPW: inverse probability weighting	ترجيح الاحتمال العكسي
ITS: interrupted time series	السلاسل الزمنية المتقطعة
ITT: intention to treat	النية للعلاج
IV: instrumental variable	المتغير المساعد
LATE: local average treatment effect	متوسط تأثير المعالجة المحلي
MDE: minimum detectable effect	الحد الأدنى للتأثير القابل للاكتشاف
MES: minimum effect size	الحد الأدنى لحجم التأثير
OLS: ordinary least squares	المربعات الصغرى العادية
PMT: proxy means test	اختبار قياس مستوى الدخل الفعلي
PSM: propensity score matching	مطابقة درجة الميل
RCT: randomized controlled trial	تجربة عشوائية مضبوطة
RDD: regression discontinuity design	تصميم انقطاع الانحدار
ToC: theory of change	نظرية التغيير



الفصل ١
مقدمة: تقييم أثر التنمية القائمة على الأدلة

رسائل أساسية

- يهدف تقييم الأثر إلى التقدير التجريبي للتأثيرات التي يمكن أن تُعزى إلى تدخل ما معين، نظرا للأهمية الإحصائية لتلك التأثيرات.
- يعتمد استنباط المعرفة والأدلة الجديرة بالثقة، من عمليات التنمية، على تقييم الأثر.
- لتقييم الأثر العديد من المزايا. فلا يقتصر دوره على تحديد ما إذا كان التدخل فعّالاً، وإنما يمكن الاستفادة منه أيضًا في المقارنة بين الخيارات المتاحة لجعل عمليات التدخل أكثر فعالية.
- يمكن الاسترشاد بالأدلة المستمدة من تقييم الأثر في وضع الافتراضات التي يقوم عليها التحليل الاقتصادي لاستثمارات محددة، فضلاً عن إستراتيجيات أوسع نطاقًا للقطاعات والمناطق والبلدان.

١-١ ما أسباب أهمية تقييم الأثر؟

تتوفر المؤسسات المعنية بالتنمية على كافة الصلاحيات التي تتيح لها المساهمة في تحقيق أهداف التنمية. فعلى سبيل المثال، تعيد إستراتيجية ٢٠٢٠ الخاصة ببنك التنمية الآسيوي التأكيد على رؤية البنك حيال العمل على القضاء على الفقر في منطقة آسيا والمحيط الهادئ. وتتمثل مهمة بنك التنمية الآسيوي في مساعدة البلدان النامية الأعضاء على تحسين الظروف المعيشية وجودة الحياة لمواطنيها. ولتحقيق هذه الغاية، يتم تخصيص مليارات الدولارات بهدف تمويل برامج تنموية كل عام. والسؤال الذي يطرح نفسه في هذا الخصوص: ما الآثار التي تترتب على البرامج الممولة؟

تتطلب الإجابة عن هذا السؤال دليلاً يتم استنباطه من تقييمات الآثار "المغايرة للواقع". وبدون تقييم الأثر، لا يمكن التأكد من التأثيرات المترتبة عن التدخلات الإنمائية. وفي حالة عدم فهم التأثيرات التي تترتب عن جهود التنمية المبذولة، فإنه لا يمكن المضي قدمًا في أي إجراءات تتعلق بالمساءلة بشأن نفقات التنمية، وكذلك عدم استخلاص المعرفة ذات المغزى من عمليات التنمية بهدف تحسين سياسات هذه الأخيرة.

تتمثل عمليات تقييم الأثر في دراسات تجريبية تهدف إلى تحديد تأثيرات التدخلات على الحصائل المستهدفة. وهذا يختلف كثيرًا عن تقييمات السيرورة التقليدية التي تهتم بتحديد السمات المتعلقة بكيفية تنفيذ المشاريع. وترتكز عمليات تقييم الأثر على تحليل ما ترتب على التدخل الذي تم القيام به، مقارنةً بسيناريو مُقدر تجريبيًا ومغاير للواقع لما كان سيحدث في حالة عدم القيام بهذا التدخل. ويمثل هذا الاختلاف بين الحصائل المتحققة والحصائل المغايرة للواقع المقياس المُستخدم لتقييم الأثر، وهو الاختلاف الذي يمكن أن يُعزى إلى التدخل. وتجدر الإشارة إلى أنه يمكن قياس التأثيرات على أي مستوى. وعلى عكس التصور الشائع في هذا الخصوص، فإن ذلك لا يتطلب الاهتمام فقط بالأهداف طويلة المدى أو "الآثار" وفقًا للمصطلحات الدارجة في الأطر المنطقية. وفي الوقت نفسه، يُعد تقييم الأثر الوسيلة الوحيدة التي يمكن أن تقدم أدلة على تلك التأثيرات طويلة المدى.

تتميز عملية تقييم الأثر بالتفرد من حيث كونها تعتمد على البيانات وتحاول تقليل الافتراضات، التي لا يمكن التحقق منها، عن طريق إسناد التأثيرات والنتائج المترتبة إلى الجهات أو العوامل المُسببة لها. وتتمثل الفكرة الرئيسية هنا في أنه لا يتم تقدير الآثار المحددة من حيث الحجم فقط، ولكن أيضًا من حيث الدلالة الإحصائية. وينبغي عدم الخلط بين هذه المقاربة و "تقدير الأثر"، والذي ينطوي في الغالب على وضع نماذج متأصلة في اتخاذ افتراضات هيكلية وغالبًا ما تكون كلاسيكية جديدة حول السلوك على النحو المُحدد، والتي لا يمكنها التحقق من التأثيرات ذات الدلالة الإحصائية.

إن العمل على أن يكون هناك توجه ترتكز فيه المساعدة الإنمائية على سياسات وتصميم مشاريع قائمة على الأدلة، والإدارة القائمة على النتائج، أمر يعتمد على تعميم مقاربة تقييم الأثر. إذ تتيح هذه المقاربة التحقق من الافتراضات التي تدعم منطق النتائج المترتبة على التدخلات، فضلاً عن الكشف عن العواقب غير المعروفة سابقًا.

يأتي استخدام نتائج الأبحاث في جوهر السياسات المرتكزة على الأدلة، وذلك بهدف وضع الافتراضات واستبدالها عند تصميم البرامج والسياسات (Sanderson 2002). وهذا بدوره يعتمد على استنباط أدلة جديدة حول الفعالية، فضلاً عن إدماج الأدلة في نطاق صياغة التصورات والمفاهيم المتعلقة بالبرامج. ويعد إجراء تحليل اقتصادي للاستثمارات أحد الممارسات التي من شانها أن تكشف عن مدى ارتباط الأدلة بوضع التصاميم والتصورات والحكم على فعالية مشروع ما. وتتيح مقاربة تقييم الأثر التحقق من صحة التأثيرات والنتائج المترتبة على التدخل وقياس حجمها، وتكتسي هذه التأثيرات أهمية بالغة بالنظر إلى دورها في فهم الفوائد والمزايا المتعلقة بالمشروع. ومن الممكن أن تشكل الاستنتاجات، التي يتم استنباطها من الآثار الخاصة بأحد التدخلات، التحليل الاقتصادي لمشروع متابعة يهدف إلى توسيع نطاق الاستثمار القائم أو لاستثمارات مماثلة في أماكن أخرى.

يعتبر نمو التحويلات النقدية المشروطة (CCTs) في أمريكا اللاتينية (المربع ١-١) أحد أفضل الأمثلة المعروفة للسياسة القائمة على الأدلة في مجال التنمية الدولية. وفي نفس السياق، لعبت مقاربة تقييم الأثر، المدعومة من بنك التنمية الآسيوي، لبرنامج البطاقات التموينية في منغوليا دورًا في إقناع الحكومة بتوسيع نطاق هذا البرنامج (بنك التنمية الآسيوي، ٢٠١٤).

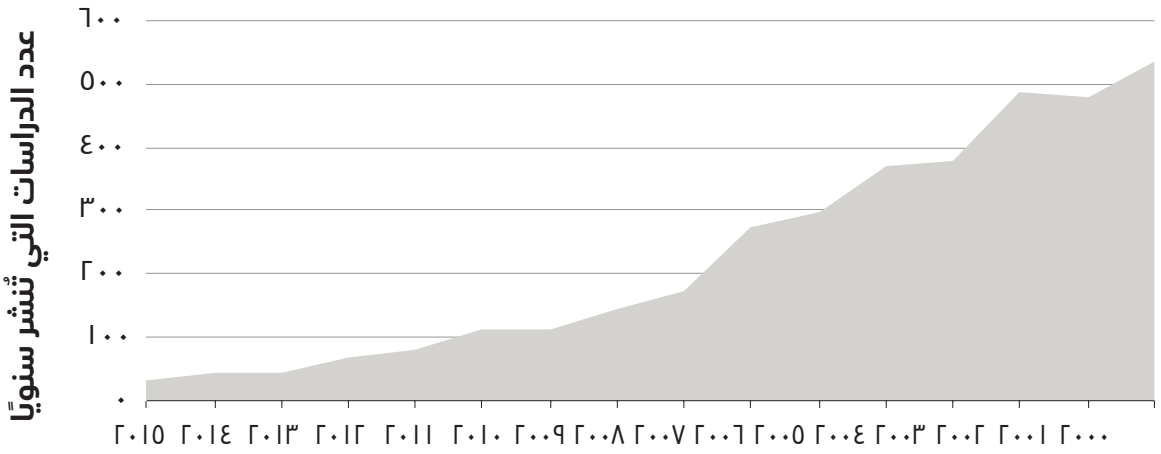
(المربع ١-١): استخدام الأدلة الناتجة عن عمليات تقييم الأثر في تحديد مدى انتشار التحويلات النقدية المشروطة في أمريكا اللاتينية

أطلقت الحكومة المكسيكية برنامج التحويل النقدي المشروط، والمعروف باسم PROGRESSA، في منتصف تسعينيات القرن الماضي. وقد قررت الحكومة إجراء تقييم دقيق باستخدام عينة عشوائية عند تصميم البرنامج. وأظهرت الدراسة مدى الأثر الإيجابي للتحويل النقدي المشروط على مستوى الفقر والصحة والتعليم. وقد أشارت نتائج هته الدراسة إلى أن البرنامج قد تمكن من الصمود أمام تغيير تعرضت له الحكومة من خلال مجرد تغيير في الاسم. ويمكن سرد قصة مماثلة حول برنامج التحويل النقدي المشروط المعروف باسم Familias en Acion في كولومبيا. وفي البرازيل، كلف الرئيس بإجراء تقييم لآثار برنامج Bolsa Familia ليكون قادرًا على مواجهة منتقدي هذا البرنامج، لا سيما أولئك الذين أشاروا إلى أنه حال دون دخول الفقراء سوق العمل. وقد أظهرت الدراسة أن البرنامج لم يتسبب مطلقًا في ذلك، ومن هنا فقد استمر في التوسع حتى وصل إلى أكثر من ١٢ مليون أسرة بحلول عام ٢٠١٢.

المصدر: Behrman (2010).

وقد انتشرت حركة تقييم الأثر في جميع أنحاء العالم وعبر مختلف القطاعات (الشكل ١-١). وتُظهر قاعدة بيانات تضم أكثر من ٤٠٠٠ من عمليات تقييم الأثر في مجال التنمية هذا النمو السريع، وكان من المتوقع إجراء ٥٠٠ دراسة جديدة في هذا الخصوص سنويًا بحلول عام ٢٠١٥. ومعظم هذه الدراسات في القطاعات الاجتماعية، غير أن هناك أعدادًا متزايدة بالنسبة للعديد من الموضوعات الأخرى، مثل الكهربة الريفية والإمداد المائي والنقل.

الشكل (١-١): الإصدار السنوي لتقييمات الأثر



المصادر: كاميرون (Cameron) وميشرا (Mishra) وبراون (Brown)؛ أرقام تقديرية للمؤلفين مبنية على أرشيف عمليات تقييم الأثر الخاص بالمبادرة الدولية لتقييم الأثر (3ie).

١-٢ أهداف تقييم الأثر

على غرار أشكال التقييم الأخرى، ثمة هدفان رئيسيان لمقاربة تقييم الأثر. يتمثل الهدف الأول في *المساءلة*، وذلك لضمان أن تؤدي إجراءات التنمية إلى نتائج حقيقية على صعيد مجال *التنمية*. ويتمثل الهدف الثاني في *التعلُّم*، وذلك لتقديم قاعدة من الأدلة بهدف اختيار وتصميم التدخلات الإنمائية التي يُحتمل أن تكون فعَّالة في تعزيز الحصائل المرجوة.

ويتجلى كلا الهدفين بوضوح في الاتجاهات المهمة، والتي يجب أن تتجاوب معها وكالات التنمية. وقد قامت مجموعة من صانعي السياسات والأطراف المعنية بزيادة الاشتراطات اللازمة للإثبات الدقيق للنتائج المستخلصة من عملية التمويل في مجال التنمية (منظمة التعاون الاقتصادي والتنمية (OECD)، ٢٠١١). وقد أتاح هذا بدء توجيه تخصيص الموارد نحو الوكالات والبرامج التي تبذل جهدًا للتوصل إلى تقدير يتسم بالمصداقية حول ما إذا كانت الحصائل والتأثيرات المتوقعة تحدث بالفعل نتيجة تدخلاتها.

هناك أيضًا طلب متزايد من مجموعة من الأطراف المعنية بأن تعكس مقترحات السياسات وعروض الاستثمار رؤى تستند إلى الاستخدام المنهجي للأدلة (Parkhurst 2017). ومن الممكن أن تستجيب وكالات التنمية لهذه المتطلبات، من خلال كل من (١) تقديم نتائج تقييم الأثر السابقة في تجربة المشروع/القطاع، و(٢) تعزيز المبادرات التجريبية الجديدة التي تشمل استخدام عملية تقييم الأثر كوسيلة منهجية للتحقق من الابتكارات. ومن خلال القيام بذلك، تُقدم الوكالات نفسها باعتبارها مؤسسات يمكن الرجوع إليها بسبب امتلاكها "معرفة" في القطاعات التي تتخصص فيها.

تجدر الإشارة إلى أن بنوك التنمية متعددة الأطراف، ولا سيما البنك الدولي وبنك التنمية للبلدان الأمريكية، قد لعبت دورًا مهمًا في تعزيز عمليات تقييم الأثر. ويمتلك البنك الدولي برامج مختلفة لتقديم الدعم الفني والمالي لعمليات تقييم الأثر، بما في ذلك الصندوق الإستراتيجي لتقييم الأثر. وبحلول عام ٢٠١٣، تضمنت جميع الموافقات الجديدة على القروض لدى بنك التنمية للبلدان الأمريكية عمليات تقييم للأثر في تصميمها. وفي عام ٢٠١٤، وضع بنك التنمية الإفريقي سياسة جديدة تتطلب المزيد من عمليات تقييم الأثر.

هذا وقد انضم بنك التنمية الآسيوي لهذه الحركة عبر أنشطة متعددة. ففي الآونة الأخيرة، أنشأ البنك صناديق مساعدة تقنية كبيرة بهدف توفير الموارد التي تتيح إجراء المزيد من عمليات تقييم الأثر. ويعد هذا الكتاب بمثابة مصدر يمكن الاستفادة منه من جانب موظفي المشاريع والشركاء الحكوميين وخبراء التنمية الآخرين الذين قد يهتمون بإدراج عمليات تقييم الأثر في مشاريعهم، أو الحصول على أدلة من التدخلات الأخرى ذات الصلة، أو فهم كيفية الاستفادة من نتائج تقييمات الأثر.

١-٣ ما الأسئلة التي قد يجيب عنها تقييم الأثر؟

يجيب تقييم الأثر عن أسئلة مثل: (١) ما الاختلاف الذي تحدثه سياسة أو برنامج ما؟ (٢) ما تصميمات البرامج الأكثر فعالية للحصول على حصيلة واحدة أو أكثر من الحصائل المحددة القابلة للقياس الكمي؟ كما أنه يمكن أن يتيح فهمًا لمدى اختلاف تلك الحصائل بين المجموعات المستهدفة والعوامل التي تحدد تلك الحصائل.

الدور المركزي الذي يلعبه تحليل الواقع المضاد

تم تصميم عمليات تقييم الأثر لتناول السؤال التالي الذي يحدد على إثره سبب الفعالية أو إلى ما تعزى: هل أحدث التدخل فرقًا ذا أهمية إحصائية بالنسبة لحصائل محددة؟ تتطلب الإجابة عن هذا السؤال إجراء تحليل للواقع المضاد فيما يتعلق بسيناريو بديل لم يحدث فيه التدخل، حيث قد لا يكون هذا البديل تدخلاً، أو تدخلاً بديلاً (يُطلق عليه اسم "التصميم أ/ب" (A/B design)، حيث يقارن التدخل "أ" بالتدخل "ب"). ويمثل إنشاء الواقع المضاد التحدي الأساسي في عملية تقييم الأثر. ويُعزى هذا التحدي إلى أنه بينما يتم رصد السيناريو الفعلي بشكل مباشر، فإن ذلك لا يتم عادة مع سيناريو الواقع المضاد. وعلى الرغم من هذا التحدي، فإن تحليل الواقع المضاد يعد ضروريًا لتحديد البرامج الأكثر فعالية، أو ما إذا كان البرنامج ذو الصلة يُحدث حقًا أي فرق على الإطلاق.

أسئلة تقييم الأثر

يعد تقييم الأثر الوسيلة الوحيدة للتحقق عمليًا من مدى إحداث مبادرات المشروع والسياسة ذات الصلة فروقًا قابلة للقياس في الحصائل المتحققة، مقارنة بالحصائل التقديرية للواقع المضاد (أي في السيناريو الذي لا ينطوي على تدخل). وعادة ما يتطلب التقييم الدقيق للأثر إجراء مسوحات خط الأساس ومسوحات خط النهاية مُصمَّمة بعناية للتوصل إلى التخمينات الأكثر مصداقية وتعبيرًا عن الواقع المضاد. ويمكن لأنظمة المتابعة أو بيانات عملية المعالجة تتبع الحصائل المتحققة في مجال الرفاهية، والتي تشير إلى ما تحقق في المجالات التي يغطيها المشروع. وتعد هذه بيانات واقعية، لكنها لا تُجيب عن السؤالين السببيين التاليين: "ما الفرق الذي أحدثه المشروع؟" و"إلى أي مدى تُعزى التغييرات المرصودة إلى التدخل الذي تم القيام به؟". وفي هذا الإطار، يمكن لتقييم الأثر الإجابة عن مثل هذه الأسئلة.

إن السؤال المركزي حول "عناصر النجاح"، والذي يستفسر بشأن ما إذا كانت حصائل التنمية المرجوة تُعزى إلى المشروع ذي الصلة، يعد محور تركيز معظم عمليات تقييم الأثر، ويمكن تسميته "سؤال الجيل الأول (السؤال الأساسي)". ويقدم المربع (١-٢) أمثلة على هذه الأنواع من الأسئلة من عملية تقييم للأثر حظيت بدعم بنك التنموية الآسيوي.

المربع (١-٢): أمثلة على أسئلة الجيل الأول: مشروع توسعة مترو العاصمة الجورجية تبليسي

سوف يتناول التقييم الأسئلة المتعلقة بأثر المشروع على تعزيز مستويات الرفاهية المتاحة للمواطنين:

١. إلى أي مدى ستؤدي توسعة المترو إلى تعزيز التنمية الاقتصادية المحلية، بما في ذلك زيادة النشاط التجاري والإيرادات وخلق فرص عمل؟
٢. إلى أي مدى ستؤثر محطة المترو على طلاب الجامعات الذين يتنقلون من جامعة تبليسي الحكومية وإليها؟ وعلى وجه الخصوص، كيف سيؤثر ذلك على استخدام الوقت وأنماط الإنفاق ومعدلات الحضور ودرجات الاختبار؟
٣. إلى أي مدى ستساهم توسعة المترو في تحسين جودة الهواء/تقليل التلوث؟

المصدر: بنك التنمية الآسيوي (٢٠١٢)

يمكن أن تكشف عمليات تقييم الأثر عن قدر كبير من الأدلة حول مجموعة واسعة من الآثار، والتي ربما لم يتم أخذ بعضها في الاعتبار من قبل منفذي المشروع. ولا شك أن الأدلة، التي يتم الحصول عليها من تقييم الأثر حول مدى ملاءمة تدخل ما معين لعملية تنمية أو تطوير أوسع، والدور الذي تلعبه عمليات التدخل التكميلية، والأجواء التي يمكن في ظلها تحقيق أفضل أشكال الفعالية على صعيد التنمية، قد تساعد في تحسين الكيفية التي يتم من خلالها تصميم المشاريع وتنفيذها (المربعان "١-٣" و"١-٤" يقدمان عينة من الرؤى والتحليلات المتعلقة بمجالَي الطاقة والنقل).

المربع (١-٣): استنتاجات لدراسات حول تقييم الأثر في مجال البنية التحتية للكهرباء

على الرغم من أن عدد عمليات تقييم الأثر فيما يتعلق بالتدخلات التي يتم القيام بها في مجال الطاقة قد شهد نموًا أبطأ مقارنةً بالقطاعات الأخرى، فإن الدراسات التي أُجريت حتى الآن في هذا الصدد تقدم أدلة مثيرة للاهتمام فيما يخص تأثيرات الحصول على الكهرباء. وتظهر الاستنتاجات التي تم الوصول إليها من هذه الدراسات وجود تأثيرات على مستوى مجموعة من الحصائل، في مجالات مختلفة تتنوع بين التعليم والصحة والدخل والمساواة بين الجنسين. وفي الوقت نفسه، تُستخلص معظم هذه النتائج من مواقف وتدخلات محددة. لذلك، تظهر الحاجة لإجراء دراسات إضافية للتحقق من مدى قابلية تعميم الاستنتاجات.

١. من الممكن أن يؤدي توصيل الكهرباء إلى *تغييرات في استخدام الوقت*، لا سيما فيما يتعلق بزيادة وقت الدراسة للأطفال، فضلاً عن زيادة ساعات العمل، وزيادة الوقت الذي يقضيه الكبار في الأنشطة غير الزراعية المُدرة للدخل (Barron and Torero 2015 وGrimm et al. 2013, Dasso and Fernandez 2015 وArraiz and Calero 2015).

٢. من الممكن أن تؤدي زيادة وقت الدراسة بفضل الحصول على الكهرباء إلى *تحسين الحصائل التعليمية للأطفال* (Arraiz and Calero 2015 وKhandker et al. 2013). ومع ذلك، فقد يؤدي هذا أيضًا إلى زيادة عمالة الأطفال على حساب تعليمهم (Squires 2015).

٣. قد تؤدي التغييرات في استخدام الوقت الناتجة عن الربط بشبكة الكهرباء إلى *نشوء أعمال صغيرة* (Dinkelman 2011 وKhandker et al. 2013, Rao 2013 وDasso and Fernandez 2015).

٤. من الممكن أن يترتب على زيادة فرص العمالة والتوظيف *زيادة في الدخل والاستهلاك والإنفاق* (Dinkelman 2011 وKhandker et al. 2013 وRao 2013 وDasso and Fernandez 2015).

٥. قد يؤدي الحصول على الكهرباء إلى تحسين الأحوال الصحية، ويمكن قياس ذلك من خلال انخفاض الحالات التي يتم الإبلاغ عنها جرّاء التهابات الجهاز التنفسي والأمراض الأخرى المرتبطة بالتدخين. ويبدو أن هذا يترتب على التحسن في جودة الهواء الداخلي بفضل استخدام المنازل للكهرباء بديلاً عن الكيروسين (ADB 2010 وBarron and Torero 2015).

٦. لقد أشارت بعض النتائج إلى أن الحصول على الكهرباء يمكنه أن *يعزز تنظيم الأسرة*. لوحظ أن زيادة مشاهدة التلفاز بسبب مد المنازل بالكهرباء تقلل من معدلات الخصوبة، وهذا ويرجع جزئيًا إلى زيادة تلقي المعلومات المتعلقة بتنظيم الأسرة، والتي تساعد على زيادة استخدام وسائل منع الحمل (Grimm et al. 2015).

٧. تشير النتائج الأخرى إلى أن الكهرباء يمكن أن تؤدي إلى *تعزيز المساواة بين الجنسين*. وقد تم التوصل إلى حقيقة أن التأثيرات والنتائج المتعلقة بالتعليم تتسم بأنها أكثر إيجابية بالنسبة للفتيات مقارنة بالأولاد (van de Walle et al. 2013)، والأمر ذاته ينطبق على التأثيرات المتعلقة بالعمالة والتوظيف (Grogan and Barron and Torero 2015 وSadanand 2012).

المصدر: المؤلفان.

المربع (١-٤): استنتاجات لدراسات حول تقييم الأثر في مجال البنية التحتية للنقل

ثمة مجموعة صغيرة ولكنها سريعة التزايد من الأدلة الدقيقة المتعلقة بالآثار المترتبة على الاستثمارات في مجال النقل. وتُظهر الدراسات التالية عددًا من التأثيرات المتعلقة بمجموعة واسعة من حصائل التنمية، وذلك على الرغم من أن هناك حاجة إلى إجراء مزيد من الدراسات لاستكشاف كيفية حدوث هذه التأثيرات، وما إذا كانت تحدث خارج نطاق السياقات المحددة الخاضعة للدراسة.

١. من الممكن أن تؤثر التدخلات المتعلقة بالنقل على *أسواق العقارات*. وقد توصلت تجربة عشوائية مضبوطة لرصف شوارع في مناطق حضرية إلى حدوث تأثيرات كبيرة على قِيم العقارات والأراضي (Gonzalez-Navarro and Quintana-Domeque 2016).

٢. من الممكن أن تؤدي التدخلات التي تهم البنية التحتية للنقل إلى *التقليل من الهجرة*. وقد تم التوصل إلى أن تحسين الطرق الريفية يساهم في التقليل من الهجرة إلى خارج المناطق الريفية الأقل حظًا ورعاية، وذلك بفضل استفادتها من تنمية اقتصادية أفضل (Gachassin 2013 وAkee 2006).

٣. قد يؤدي تحسين البنية التحتية للطرق إلى *زيادة الانتظام في المدارس*، لا سيما في المرحلة الثانوية، وذلك نتيجة لانخفاض تكلفة السفر (Sengupta et al. 2007 وKhandker et al. 2009). وفي بعض الحالات، يكون هذا التأثير أكثر وضوحًا بالنسبة للفتيات مقارنةً بالأولاد (Iimi et al. 2015).

٤. يكون للتدخلات المتعلقة بالنقل *تأثيرات مهمة على الصحة*. وقد تم التوصل إلى أن وجود طرق أفضل يساهم في تسهيل الوصول بشكل أفضل إلى المرافق الصحية واستخدامها (Lokshin and Yemtsov 2003). كما تم التوصل إلى أن القيام بتدخلات بسيطة لتعزيز السلامة الطرقية يؤدي إلى الحد بشكل كبير من الحوادث والإصابات (Banerjee et al. 2014 وHabyarimana and Jack 2012 وHabyarimana and Jack 2009).

٥. من الممكن أن يؤدي توفر بنية تحتية أفضل في مجال النقل إلى *تحفيز تنمية السوق*. وقد تم التوصل إلى أن تحسين شبكات الطرق يؤدي إلى خفض تكاليف المدخلات، والسماح بمزيد من المرونة في إمدادات المدخلات الثابتة، وزيادة التجارة المحلية، والسماح بمتابعة أسواق المخرجات الجديدة (Lokshin and Yemtsov وMu and van de Walle 2011 2003). كما تم التوصل إلى ظهور تأثيرات طبيعية على تطوير المشاريع، فضلاً عن تحسن مستمر في الإنتاجية (Datta 2011 وGhani et al. 2016).

٦. *يزداد النشاط الاقتصادي* بفضل التحسينات التي يشهدها مجال النقل. وقد تكون هناك زيادات كبيرة في الناتج المحلي الإجمالي في المناطق التي تستفيد من تدخلات تهدف إلى تحسين شبكات النقل (Faber وBanerjee et al. 2012 Yoshino and Abidhadjaev 2015 وWang and Wu 2015 و2014).

٧. من الممكن أن يؤدي تحسين النقل إلى *زيادة الطلب على العمالة وكذلك زيادة الأجور*. وإلى جانب تحسين القدرة على الحركة والتنقل للوصول إلى الفرص المتاحة في سوق العمل، من الممكن أن يؤدي تحسين البنية التحتية لقطاع النقل أيضًا إلى تأثيرات كبيرة على التوظيف والدخل (Rand 2011 وGertler et al. 2015 وAkee 2006).

٨. من الممكن أن يكون للتدخلات المتعلقة بالنقل تأثيرات كبيرة فيما يتصل *بالحد من الفقر* (Sengupta et al. 2007 وDercon et al. 2009). علاوة على ذلك، وُجِد أن تطوير الطرق له تأثير قوي على أداء الشركة ومعدلات التوظيف في المناطق الأكثر فقرًا (Mu and van de Walle 2011 وGibson and Rozelle 2002).

المصدر: المؤلفان.

من الممكن أيضًا أن تجيب دراسات تقييم الأثر عن أسئلة تتعلق بشكل مباشر بتنفيذ المشروع. ففي كثير من الأحيان، يكون هناك شيء من عدم اليقين لدى من يتولون إعداد المشاريع حول أفضل تصميم للبرنامج. وفي حالة كان من الممكن تقديم برنامج بتصميميْن متنافسيْن أو أكثر في مجالات مختلفة، فيمكن عندئذٍ استخدام الطرق المتبعة في دراسات تقييم الأثر للتحقق والوصول إلى تصميم البرنامج الأكثر فعالية، والذي يتيح إحداث التغيير المطلوب في الحصائل.

يمكن تسمية هذه الأسئلة المتعلقة بالتصميم بأسئلة "الجيل الثاني". وتكتسي كل من أسئلة الجيل الأول والثاني أهمية بالغة. كما تكتسي الأدلة المتعلقة بالفعالية أهمية بسبب دورها في تحقيق المساءلة ومساعدة كبار المسؤولين عن وضع السياسات على اتخاذ قرارات تخصيص الموارد. وفي أغلب الأحوال، يهتم مديرو المشاريع بالأسئلة المتعلقة بـ "الكيفية"، والتي تجيب عنها دراسات الجيل الثاني. ثمة تقدم طبيعي في دراسات تقييم الأثر التي تتناول موضوعًا معينًا، حيث يبدأ ذلك بالتأكد أولاً من أن نوع التدخل يمكن أن يكون فعالاً (أسئلة الجيل الأول)، وبمجرد إثبات هذا، يمكن بعد ذلك تعزيز درجة الاستيعاب حول كيفية الارتقاء بمستوى الفعالية (أسئلة الجيل الثاني).

يمكن الحصول على أمثلة حول الأدلة المتعلقة بتصميم البرامج من العديد من دراسات تقييم الأثر التي تُجرى للتعرف على أثر التحويلات النقدية المشروطة على الحصائل التعليمية. ومن خلال تجميع الأدلة من هذه الدراسات، تم التوصل إلى العديد من الدروس المهمة المتعلقة بالتصميم الفعال لعمليات التحويل النقدي المشروط (CCTs) بهدف تحفيز الحضور إلى المدارس (Baird et al. 2014):

- تكون التحويلات النقدية المشروطة أكثر فعالية على مستوى المدارس الثانوية مقارنة بمستوى المدارس الابتدائية.
- تكون التحويلات النقدية المشروطة أكثر فعالية مع دفعات المبالغ الأقل والأكبر حجمًا مقارنة بالدفعات الصغيرة المتكررة.
- المال مهم: كلما زاد حجم المال المدفوع، زاد التأثير.
- تتميز البرامج التي تتحلى بمستوى أكبر من المتابعة وإنفاذ الشروط بتأثير أكبر مقارنةً بتلك التي تتسم بمستوى ضعيف من المتابعة والإنفاذ. وتكون احتمالية حضور الأطفال، الذين يعيشون في مجتمعات تطبق التحويل النقدي المشروط مع مستوى قوي من المتابعة والإنفاذ، إلى المدرسة أكبر بنسبة ٦٠٪ مقارنةً بأولئك الذين يعيشون في مجتمعات تطبق برنامج تحويل بدون آلية للمتابعة أو الإنفاذ.

الجمع بين التحليل السياقي وتحليل الواقع المضاد

بالرغم من أن تحليل الواقع المضاد يمثل جوهر دراسات تقييم الأثر، يلعب التحليل السياقي دورًا مهمًا في مثل هذه الدراسات للمساعدة في فهم سبب نجاح التدخل أو عدم نجاحه في سياقات مختلفة ولمجموعات مختلفة. ويعد كل من تقييم الأثر وتقييم السيرورة مقاربتين متكامليْن وليستا بديلتيْن.

في هذا الإطار، يعد الاستهداف أحد المجالات المهمة للتحليل السياقي. ويمكن استخدام البيانات الكمية لتقدير أخطاء الاستهداف، كما هو الحال عندما يتم فقدان أعضاء المجموعة المستهدفة، أو عندما يكون هناك مستفيدون ليسوا في المجموعة المستهدفة. وقد تكون البيانات النوعية مفيدة بشكل عام في تحديد العوائق التي تحول دون الاعتماد.

إذا أشار التحليل السياقي إلى أسباب محتملة لعدم المشاركة، فيمكن استخدام التحليل الكمي للتحقق من هذه الأسباب. فعلى سبيل المثال، توصلت دراسة يدعمها بنك التنمية الآسيوي حول استخدام مصابيح الفلورسنت الكهربائية صغيرة الحجم في باكستان أن الغالبية العظمى من المستهدفين قد قللوا من شأن توفير الطاقة الناتج عن استخدام هذه المصابيح مقارنةً بالمصابيح المتوهجة التقليدية (المربع ١-٥).

المربع (١-٥): التعرف على تصميم المشروع من تقييم الأثر: المصابيح الكهربائية الموفرة للطاقة في باكستان

أطلقت الحكومة الباكستانية برنامجًا وطنيًا بتكلفة ٦٠ مليون دولار يهدف إلى استبدال ٣٠ مليونًا من المصابيح المتوهجة بمصابيح فلورسنت كهربائية صغيرة الحجم (CFLs) في القطاع السكني. وقد حظي هذا البرنامج بدعم بنك التنمية الآسيوي من خلال قرض بقيمة ٤٠ مليون دولار. وتوصلت دراسة لتقييم الأثر تم إجراؤها أثناء إعداد القرض إلى النتائج التالية:

- هناك أقلية لا يُستهان بها (١١٪) من الأسر ليست على دراية بمصابيح الفلورسنت الكهربائية صغيرة الحجم.
- لدى الغالبية العظمى من الأسر معلومات مغلوطة عن مستوى الكفاءة العالية لمصابيح الفلورسنت الكهربائية صغيرة الحجم مقارنة بأنواع أخرى من المصابيح. وعلى سبيل المثال، تدوم هذه المصابيح لفترة أطول بمعدل ١٠ مرات على الأقل مقارنة بالمصابيح المتوهجة. ومع ذلك، أجاب ثلث المشاركين في الدراسة بأنهم لم يسبق لهم معرفة هذا الفرق، بينما قال ربعهم إن مصابيح الفلورسنت الكهربائية صغيرة الحجم تدوم ضعف نظيرتها المتوهجة. وأجاب أقل من ١٠٪ من المشاركين في الدراسة بأن مصابيح الفلورسنت تدوم بمعدل ١٠ مرات مقارنة بمصابيح أخرى.
- تتم المبالغة في تقدير المزايا إذا تم الاستناد إلى القرار، المتخذ بخصوص استبدال المصابيح المتوهجة بمصابيح فلورسنت الكهربائية صغيرة الحجم، وحده. و في نفس السياق، يتم اغفال تقدير بعض العيوب الناجمة عن "التأثير الارتدادي" الذي ينشأ عن استهلاك الأسر ضوءًا أكثر عند استخدام مصابيح الفلورسنت الكهربائية صغيرة الحجم بدلاً من المصابيح المتوهجة.

يُظهر الاستنتاجان الأول والثاني أهمية تضمين عنصر طلب (معرفة المستهلك) في البرنامج، في حين يتناول الاستنتاج الثالث التحليل الاقتصادي.

المصدر: Chun and Jiang (2013).

الاستفادة من دراسات تقييم الأثر في الاختبار التجريبي للتدخلات المبتكرة

من الممكن دمج دراسات تقييم الأثر في مشاريع مبتكرة للاستفادة منها في التوصل إلى تصميمات مشاريع أخرى، من خلال تقييم اثر الاختبار التجريبي. فعلى سبيل المثال، يمكن تقدير طرق مختلفة للترويج لاعتماد تقنية أو خدمة ما، أو تشجيع الاستخدام الفعال للمياه أو الطاقة، من خلال مقارنة ببعضها البعض باتباع مقاربة التصميم أ/ب؛ أي دراسة طريقتيْ المعالجة المُستخدمتيْن والمقارنة بينهما. يُعدّ دمج هذه المقاربات في عمليات التدخل التي يمكن الاسترشاد بها كنماذج تجريبية يمكن محاكاتها، في القطاع العام، نوعا من عمليات "التعلم" التي أصبحت "عادة مكتسبة" بالنسبة للعديد من أنشطة القطاع الخاص (المربع ١-٦). وستكون هذه الدراسات

التجريبية بمثابة دراسات سريعة لتقييم الأثر (المربع ١-٧)، مع وجود حلقات تعقيبات واضحة تهدف إلى التوصل إلى تصميم المشروع وتنفيذه.

المربع (١-٦): الاختبار التجريبي في القطاع الخاص

يتم استخدام التجارب العشوائية المضبوطة (RCTs) على نطاق واسع في القطاع الخاص، بهدف التوصل إلى المقاربة الإدارية وتصميم المنتج اللذين يمكن استخدامهما. وتُجري شركات تقنية رائدة، مثل مايكروسوفت (Microsoft) وجوجل (Google) وأمازون (Amazon) ومواقع بيع بالتجزئة مثل بوكينغ.كوم (Booking.com)، بشكل معتاد أكثر من ١٠٠٠٠ تجربة عشوائية مضبوطة سنويًا. وعادة ما تكون هذه الدراسات تقييمات سريعة للأثر مع تصميمات أ/ب للمقارنة، على سبيل المثال، بين عروض المنتجات أو أساليب تسويقها المختلفة.

وعلى نحو أكثر تحديدًا، تقوم شركة ياهو (Yahoo) المتخصصة في محركات البحث بتجارب تتعلق بإعادة تصميم صفحتها الرئيسية بهدف زيادة النقرات التي تنقل إلى مواقع أخرى (والتي تجني منها المال) عن طريق الإحالة العشوائية لـ ١٠٠٠٠٠ من زوارها خلال ساعة إلى الموقع المعاد تصميمه (المجموعة الخاضعة لنفس المُعَالجَة "أ"). ويتم توجيه الملايين الآخرين من الزوار إلى الموقع الحالي ((المجموعة الخاضعة لنفس المُعَالجَة "ب"، المعالجة الحالية). وبعد مرور ساعة واحدة فقط، تكتمل الدراسة عن طريق مقارنة معدلات النقرات بين مجموعتيْ المعالجة "أ" و"ب".
المصدر: Kohavi and Thomke (2017).

المربع (١-٧): دراسات تقييم الأثر السريعة

تشير دراسات تقييم الأثر السريعة إلى التجارب العشوائية المضبوطة التي تتسم بإطارات زمنية أقصر وميزانيات أقل مقارنةً بدراسات تقييم الأثر التقليدية.

ويمكن أن تكون التجربة العشوائية مضبوطة سريعة (أي يتم إجراؤها في فترة تتراوح من ١٢ إلى ١٨ شهرًا) وفقًا للشروط التالية:

- أن تكون التجربة العشوائية مضبوطة بسيطة، بمعنى أن التعيين العشوائي يكون على مستوى الفرد أو الشركة أو الأسرة.
- أن تكون الحصائل التي يتم قياسها هي تلك التي يمكن، على نحو معقول، توقع تأثير كبير كافٍ لها داخل مجموعة كبيرة بما فيه الكفاية أثناء وقت التقييم. وغالبًا ما تركز دراسات تقييم الأثر السريعة على الاعتماد أو الاستخدام بدلاً من الحصائل النهائية المتعلقة بمستوى الرفاهية المنشود والمتحقق، والتي يستغرق تحقيقها وقتًا أطول وتتطلب عينة أكبر ليتم قياسها.

من الممكن أن تكون دراسات تقييم الأثر المتعلقة بالاختبار التجريبي مناسبة تمامًا لمقاربة تقييم الأثر السريعة.

المصدر: Kohavi and Thomke (2017).

في أغلب الأحوال، يُمكن أن يكون الاختبار التجريبي جزءًا من التقييم التكويني (أو التقييم الذي يهدف إلى التوصل إلى بنية البرنامج)، حيث يتم تقييم البرنامج على نطاق صغير قبل تطبيقه على نطاق أوسع. وعادة ما تركز التقييمات التكوينية على التأثيرات التقريبية، مثل معدلات الاعتماد (الاستخدام).

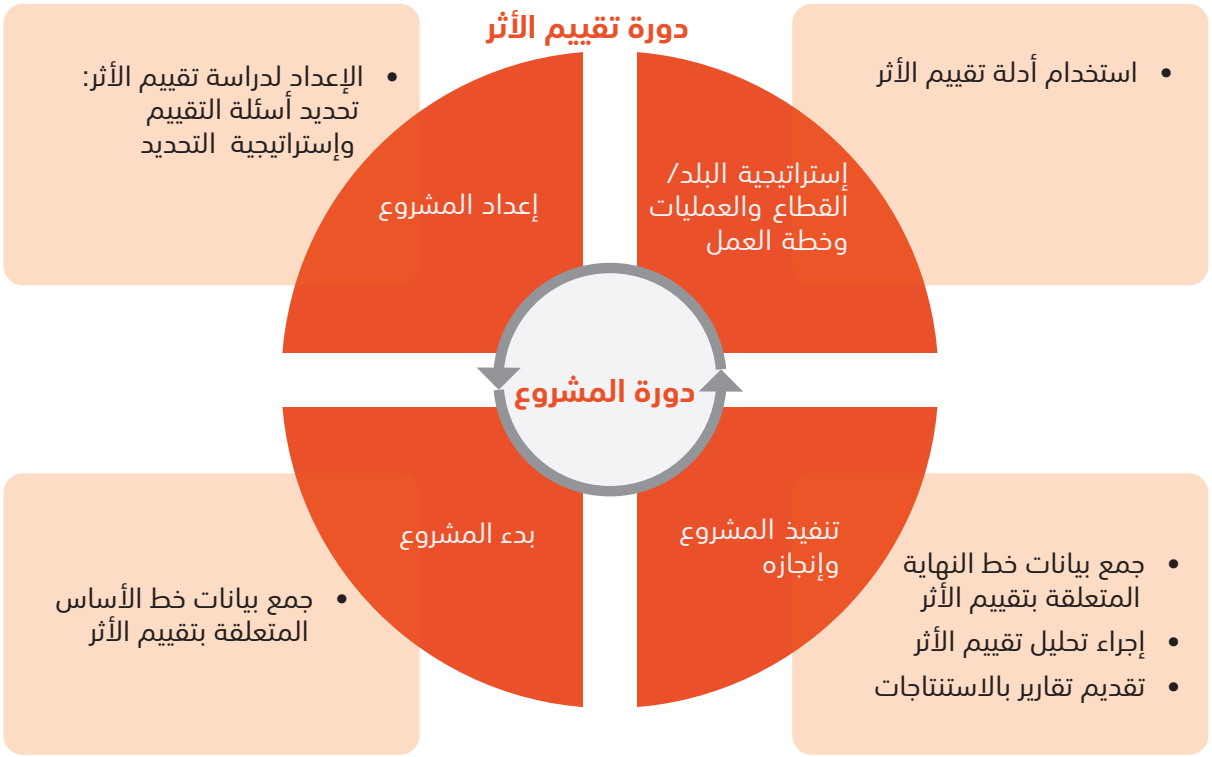
يظل هناك العديد من الإشكاليات السياقية المهمة التي لا يتم تناولها بواسطة دراسات تقييم الأثر، مثل دقة التنفيذ. لذلك، قد يكون من المفيد تضمين التحليل السببي لتقييم الأثر في إطار تقييم أوسع نطاقًا،

بحيث يمكن للتحليل أن يتناول بشكل أفضل "تحديد عوامل النجاح وأسبابه". (سيتم تناول ذلك في الفصل الثاني).

١-٤ تقييم الأثر في دورة المشروع

يوضح الشكل (١-٢) أين سيكون استخدام دراسات تقييم الأثر مناسبًا في دورة مشاريع مؤسسات التنمية، مثل بنك التنمية الآسيوي. ومن خلال هذا الشكل المتراكب، تظهر بعض التحديات (الاقتصاد السياسي) التي يجب التغلب عليها في العديد من الجهود المتعلقة بدراسات تقييم الأثر. وبالإضافة إلى الجهود المبذولة لتعزيز الوعي بمنهجيات دراسات تقييم الأثر، ثمة إشكاليات أخرى تتعلق بالتوقيت وتحفيز الأطراف المعنية تتطلب هي الأخرى الاهتمام.

الشكل (١-٢): تقييم الأثر ودورة المشروع



المصدر: المؤلفان.

التحفيز والتوقيت

قد تكون الاعتبارات المتعلقة بدراسات تقييم الأثر معقدة، ومن الأفضل النظر إليها بعين الاعتبار بشكل أولي في وقت مبكر من دورة المشروع، ويُفضل أن يكون ذلك أثناء تصميم المشروع ومعالجته (بحيث يتم إدراج أنشطة تنفيذ دراسات تقييم الأثر، مثل مسوحات خط الأساس، في نطاق المشروع). وقد لا يكون العديد من موظفي المشروع على دراية بمنهجيات دراسات تقييم الأثر، وقد يفضلون لهذا السبب تجنب القيام بنشاط إضافي أثناء إعداد المشروع، لا سيما إذا كان تصميم تقييم الأثر يتطلب جهودًا لا يُعتقد أنها تلبي المتطلبات الفورية.

علاوة على ذلك، إذا كان الموظفون المعنيون بعمليات التشغيل يتغيرون بشكل متكرر، فقد لا يظلون يشغلون نفس الدور الوظيفي الذي يتيح لهم إما الاستفادة من استنتاجات دراسات تقييم الأثر أو تلقي أي إقرار محتمل يتعلق بإجراء مثل هذه الدراسات.

تجدر الإشارة إلى أن ثمة ما قد لا يحظى بالتقدير المناسب في سياق المواعيد النهائية للإعداد للمشروع، ويتمثل في أن جهود دراسات تقييم الأثر تؤدي مباشرة إلى مشاريع أفضل تتميز بسلاسة أكثر في التنفيذ وإنفاق أنسب للوقت، نظرًا لأنه تتم مراعاة الافتراضات بالغة الأهمية بشكل أفضل (Legovini et al. 2015). بالإضافة إلى ذلك، يمكن أن توفر بيانات خط الأساس التي يتم الحصول عليها عند بداية المشروع مدخلات قيِّمة لتنفيذ المشروع، وذلك في حالة تم إشراك موظفي المشروع لضمان تزويدهم بالمعلومات التي يحتاجونها. وقد يحتاج الموظفون المعنيون بالتشغيل إلى أن يكونوا على دراية بهذه الفوائد.

البدء

تبدأ دراسة تقييم الأثر عادة في المراحل الأولى لتنفيذ المشروع. وبعد الموافقة على المشروع، غالبًا ما تكون هناك فترة تأخير كبيرة تنجم عن الشروع في توقيع اتفاقيات المشروع، ومنح عقود المشتريات، فضلاً عن وضع الأساس لجعل مخرجات المشروع حقيقةً ملموسة. وخلال هذه الفترة، غالبًا ما يصبح التوصل إلى كيفية التنفيذ أمرًا أكثر قابلية للتنبؤ، غير أن المُخرجات لا تكون فعالة أو متاحة بعد، ومن ثم تظل هناك إمكانية لإجراء مسوحات خط الأساس.

- إذا تضمنت دراسة تقييم الأثر تعيينًا عشوائيًا، فسيؤثر ذلك على تصميم المشروع. لذا، يجب تضمين آلية التعيين العشوائي في تصميم المشروع.

- دائمًا ما يتم تعزيز تقديرات دراسات تقييم الأثر من خلال إتاحة بيانات خط الأساس، والتي يمكن الحصول عليها بشكل أفضل عن طريق التخطيط المبكر.

- عند إعداد المشروع، قد يكون من الممكن دمج تقييم الأثر مع عملية جمع بيانات متابعة والتقييم الخاصة بالمشروع، وذلك لتجنب المسوحات المتكررة، وتحسين عملية متابعة المشروع وتقييمه، فضلاً عن الاستفادة من الموارد القليلة المخصصة لإجراء المسح بشكل أكثر فعالية.

التنفيذ

سيتم إجراء دراسات تقييم أثر للاختبار التجريبي خلال السنوات الأولى من تنفيذ المشروع. وقد يتم تضمين مسح إحصائي يُجرى في منتصف مدة المشروع في تقييم الأثر، والذي قد يكون أكثر تركيزًا على تناول المشكلات المتعلقة بسيرورة المشروع لتحديد التعديلات التي يمكن إجراؤها في منتصف المدة. ومع ذلك، فإنه اعتمادًا على توقيت نهاية المشروع، قد يكون من الضروري أيضًا التركيز في منتصف المدة على اتخاذ القرارات المتعلقة بأي مشروع متابعة.

الإنجاز

في مرحلة الإنجاز أو عند اقترابها، يتم إجراء مسح خط النهاية للتمكن من تقييم الأثر. وهناك اعتباران رئيسيان قد يؤثران على التوقيت، وهما:

- يمكن استخدام الأموال المخصصة للمشروع لتمويل عملية جمع البيانات وربما تحليلها، بحيث يتعين إكمال الدراسة بينما لا يزال المشروع جاريًا. وفي هذا الإطار، قد يلزم توخي الحيطة إزاء مشاركة التكاليف مع الحكومة أو مصادر التمويل الأخرى، بحيث يمكن إجراء التحليل بعد الإغلاق المالي.

- من الممكن أن تساعد نتائج تقييم الأثر في صياغة القرارات المتعلقة بتمويل وتصميم عمليات التدخل التي تتم في إطار المتابعة. ومع ذلك، قد يتم التوصل إلى تصميم مشاريع المتابعة الفردية قبل توقيت دراسة تقييم الأثر وفقًا لمسح يتم إجراؤه عند اكتمال المشروع. ولهذا السبب، قد تُحدث دراسات تقييم الأثر تأثيرًا غير مباشر من خلال تقديم أدلة تسهم في وضع إستراتيجيات أوسع نطاقًا للبلد أو القطاع.

١-٥ تقييم الأثر والتقييم والتحليل الاقتصادي

في بعض الأحيان، قد يشعر غير الاقتصاديين وغير العاملين في حقل التقييم بالارتباك بسبب الاختلافات بين دراسات تقييم الأثر والتقييم عمومًا والتحليل الاقتصادي. وغالبًا ما يكون التقييم لدى بنك التنمية الآسيوي والعديد من الوكالات الأخرى *تقييمًا للسيرورة* بُغية معرفة كيفية تنفيذ المشاريع والبرامج. ويتعلق هذا في الغالب بكيفية استخدام المدخلات في الأنشطة لتوليد المخرجات، وغالبًا ما يكون مجال التركيز في أي مجموعة بيانات أصلية نوعيًا ووصفيًا. إضافة إلى ذلك، غالبًا ما يكمن الهدف في تقديم تصنيف معياري لتنفيذ المشروع/البرنامج. ويختلف هذا اختلافًا جوهريًا عن تقييم الأثر، والذي يعد شكلاً إيجابيًا من أشكال البحث التطبيقي، ويركز على التأثيرات السببية للتدخلات، وذلك من أجل استخلاص استنتاجات أوسع نطاقًا ذات صلة بالسياسة المتبعة.

عادة ما تكون الوحدات التنظيمية المُكلفة بالتقييم لدى بنك التنمية الآسيوي والمؤسسات المالية الدولية الأخرى مستقلة أيضًا عن تنفيذ المشروع، ولا يمكنها في الغالب إجراء الأنشطة سوى في وقت *لاحق*. ومع ذلك، تتطلب طرق تقييم الأثر الأكثر دقة التفاعل الوثيق أثناء تنفيذ المشروع، وذلك لضمان الجمع المناسب للبيانات، وربما التعيين العشوائي للتدخل قبل التنفيذ واسع النطاق. وهذا غالبًا ما يضع دراسات تقييم الأثر أيضًا ضمن اختصاصات جهات البحث والتنفيذ، وليس أقسام التقييم.

قد يحدث الالتباس أيضًا لأن كلاً من التحليل الاقتصادي وتقييم الأثر يعتمدان على على خبراء اقتصاديين يتولون مهمة تحليل تأثيرات التدخلات الإنمائية. ومع ذلك، يركز التحليل الاقتصادي لدى بنك التنمية الآسيوي والعديد من الوكالات الأخرى في الغالب على التحليل *السابق* أثناء الإعداد للمشروع، وذلك عندما يتعين التنبؤ التأثيرات الناجمة عن المشروع أو افتراضها. هذا في حين يركز تقييم الأثر على تقديم تقديرات دقيقة قائمة على الأدلة لتلك التأثيرات (الجدول ١-١). ويكون تقييم الأثر بمعناه الدقيق ممكنًا فقط بمجرد تنفيذ قدر معين من التدخل، على الرغم من أن هذا قد يكون في مرحلة مبكرة (المربع ١-٢). وعند إعداد المشروع، من الممكن الاعتماد على أدلة مستقاة من دراسات تقييم أثر موجودة بالفعل لمشاريع مماثلة للمساعدة في تبرير القيام بالاستثمار ذي الصلة والتوصل إلى تصميم المشروع. كما تنطوي عمليات المراجعة المنهجية على تجميع جميع الأدلة الدقيقة الملائمة حول موضوع ما، ويمكن أن تكون مفتاحًا للمساعدة في إعداد التقييم *السابق*.

الجدول (١-١): تقييم الأثر والتحليل الاقتصادي للاستثمارات

مجالات التحليل الاقتصادي	تحليلات محددة	مساهمة الأدلة المستقاة من تقييم الأثر وعمليات المراجعة المنهجية
التحقق من صلاحية المبرر الاقتصادي	السياق الكلي وتحليل القطاع	الأدلة المتعلقة بأهم الاستثمارات لحصائل القطاع
	المبرر الاقتصادي	الأدلة المنطقية الاقتصادية المتعلقة بتأثيرات تخفيف القيود في التدخلات السابقة
	تحليل الطلب	إبراز الاستعداد للدفع من خلال تجارب في مشاريع أو برامج أخرى
	إطار التصميم والمتابعة	الأدلة على الافتراضات بالغة الأهمية والخطوات المتبعة في سلسلة العلاقات السببية التي يتعين متابعتها
إجراء تحليل اقتصادي سابق	تحليل البديل والأقل تكلفة	الأدلة المتعلقة بالفعالية النسبية لمقاربات التدخل البديلة المتبعة في مشاريع أو برامج الأخرى
	التحليل السابق للتكلفة والمنفعة	القياس الكمي لحجم التأثيرات الناتجة عن تدخلات سابقة
	استدامة استثمارات المشروع	الأدلة المتعلقة بالعوامل التي تؤثر على (١) الاعتماد (الاستخدام) (٢) والاستدامة على نطاق واسع
	تحليلات المخاطر والحساسية	الأدلة المتعلقة بأسباب فشل تدخلات مماثلة في إحداث آثار متوقعة
	تحليل التوزيع	الأدلة المتعلقة بالاستيعاب/المشاركة وكيف يؤثر السلوك على الآثار التوزيعية في تدخلات مماثلة
إجراء تحليل اقتصادي لاحق	التحليل اللاحق للتكلفة والمنفعة	تقديرات التأثيرات التي تُعزى إلى التدخلات

ملاحظة: موضوعات ومجالات من بنك التنمية الآسيوي (٢٠١٣)، باستثناء التحليل الاقتصادي اللاحق الذي لم يتم تناوله في هذه الوثيقة.
المصدر: المؤلفان.

١-٦ نبذة عن هذا الكتاب

هذا الكتاب موجه بالأساس للخبراء الممارسين في مجال التنمية القائمة على الأدلة والباحثين في البلدان النامية وفي وكالات معنية بالتنمية، مثل بنك التنمية الآسيوي. وبالمقارنة مع كتب إرشادية أخرى، يقدم هذا المجلد نطاقًا أوسع من الخيارات العملية والمنهجية، والتي يتم تقديمها بلُغة مُبسّطة يسهل استيعابها.

من الممكن أن تساعد نظريات التغيير في وضع تصميمات دراسات تقييم الأثر، ويتم تناولها في الفصل الثاني. كما يتم تناول أساسيات تقييمات الأثر في الفصل الثالث، مع تسليط الضوء على المزيد من التفاصيل المتعلقة بتصميمات دراسات تقييم الأثر المختلفة في الفصلين الرابع والخامس. في حين يتناول الفصلان السادس والسابع عملية جمع البيانات والقضايا المتعلقة بأخذ العينات. ويعالج الفصل الثامن القضايا الرئيسية المتعلقة بإدارة دراسات تقييم الأثر، ويوفر قائمة بوكالات البحث الرئيسية المعنية بإنتاج دراسات تقييم الأثر.

تم تأليف هذا الكتاب للقراء الذين لا يزالون في بداية عهدهم بمجال التقييم والمتابعة، لذا فهو بمثابة تمهيد لهم ولا يتطلب منهم معرفة مسبقة بالموضوع. وعلى الرغم من ذلك، يتضمن أيضًا محتوى يتناسب مع جمهور القراء الأكثر خبرة في المجال الذي يتناوله. وقد تمت كتابته ليكون دليلًا عمليًا. فهو يركز على الدروس المستفادة من التجارب والنصائح الرئيسية ذات الصلة، والتي ستكون مفيدة لأولئك الذين يفكرون في القيام بدراسات تقييم الأثر ويخططون لها.

يقدم الملحقان الأول والثاني مزيدًا من العروض التقنية التي تتناول طرق التقدير وجمع البيانات لأولئك الذين سيجرون دراسات تقييم الأثر. كما يتضمن الملحق الأول أيضًا مقدمة للأوامر والحزم اللازمة لتنفيذ الطرق في برنامج STATA، وهو برنامج يُستخدم كثيرًا في التحليل الاقتصادي القياسي.

المراجع*

Asian Development Bank (ADB). 2010. Impact Evaluation Study: ADB's *Assistance for Rural Electrification in Bhutan – Does Electrification Improve Quality of Rural Life?. Manila.*

ADB. 2012. Impact Evaluation Study: *Tbilisi Metro Extension Project in Georgia – Evaluation Design and Baseline Survey Report.* Manila.

ADB. 2013. *Key Areas of Economic Analysis of Investment Projects: An Overview. Manila.*

ADB. 2014. *Food Stamps and Medicard – Impact Evaluation Report: Final Report.*
Manila.

Akee, R. 2006. The Babeldaob Road: The Impact of Road Construction on Rural Labor Force Outcomes in the Republic of Palau. *The Institute for the Study of Labor (IZA) Discussion Paper*. No. 2452. Bonn.

Arraiz, I. and C. Calero. 2015. From Candles to Light: The Impact of Rural Electrification. *IDB Working Paper Series*. No. IDB-WP-599.

Baird, S., F. H. G. Ferreira, B. Özler, and M. Woolcock. 2014. Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes. *Journal of Development Effectiveness*. 6 (1). pp. 1–43.

Banerjee, B., E. Duflo, and N. Qian. 2012. On the Road: Access to Transportation Infrastructure and Economic Growth in China. *NBER Working Paper Series*. No. 17897.

Banerjee, A., E. Duflo, D. Keniston, and N. Singh. 2014. The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India. Massachusetts Institute of Technology Working Paper.

Barron, M. and M. Torero. 2015. Household Electrification and Indoor Air Pollution. *Munich Personal RePEc Archive*. MPRA Paper No. 61424.

Behrman, J. R. 2010. The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash Transfer Program. *World Development*. 38 (10). pp.14148501–43.

Cameron, D. B., A. Mishra, and A. N. Brown. 2016. The Growth of Impact Evaluation for International Development: How Much Have We Learned? *Journal of Development Effectiveness*. 8 (1). pp. 1–21.

CAREC Institute Research Program. 2010. Final Report: Retrospective Impact Evaluation of the Korday–Almaty Road Project (Zhambyl Oblast). http://www.carecprogram.org/uploads/docs/Research/Impact-Evaluation/2010/IE-Report-Korday-Almaty-Road-Project-Final.pdf.

* يشير بنك التنمية الآسيوي (ADB) إلى "الصين" باعتبارها جمهورية الصين الشعبية، و"فيتنام" باعتبارها "فييت نام".

Chun, N. and Y. Jang. 2013. How Households in Pakistan Take on Energy Efficiency Lighting Technology. *Energy Economics*. 40. pp. 277–284.

Cody, S. and A. Asher. 2014. Proposal 14: Smarter, Better, Faster: The Potential for Predictive Analytics and Rapid-Cycle Evaluation to Improve Program Development and Outcomes. In Section 4: Improving Safety Net and Work Support. The Hamilton Project. Washington, DC: The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2016/06/predictive_analytics_rapid_cycle_evaluation_cody_asher.pdf.

Dasso, R. and F. Fernandez. 2015. The Effects of Electrification on Employment in Rural Peru. *IZA Journal of Labor & Development*. 4 (6).

Datta, S. 2011. The Impact of Improved Highways on Indian firms. *Journal of Development Economics*. 99 461). pp. 5721.

Dercon, S., D. Gilligan, J. Hoddinott, and T. Woldehanna. 2009. The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages. *American Journal of Agricultural Economics*. 91 (1). pp.1010211–43.

Dinkelman, T. 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review*. 10. pp. 3078– 3108.

Faber, B. 2014. Trade Integration, Market Size and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*. 81. pp. 1046-10704.

Gachassin, M. 2013. Should I Stay or Should I Go? The Role of Roads in Migration Decisions. *Journal of African Economies*. 22 (1). pp.79826–43.

Gertler, P., M. Gonzalez-Navarro, T. Gracner, and A. Rothenberg. 2015. Road Quality, Local Economic Activity, and Welfare: Evidence from Indonesia's Highways. Preliminary Draft October 2015.

Ghani, E., A. Goswami, and W. Kerr. 2016. Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*. 126 (1). pp313571 1–43.

Gibson, J. and S. Rozelle. 2002. Poverty and Access to Infrastructure in Papua New Guinea. Department of Agricultural and Resource Economics, University of California Davis Working Paper No. 02-008.

Gonzalez-Navarro, M. and C. Quintana-Domeque. 2016. Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects. *Review of Economics and Statistics*. 98 (1). pp. 25267–21.

Grimm, M., J. Peters, and M. Sievert. 2013. Impacts of Pico-PV Systems Usage using a Randomized Controlled Trial and Qualitative Methods. Evaluation report commissioned by the Operations Evaluation Department (IOB) of The Netherlands Ministry of Foreign Affairs.

Grimm, M., R. Sparrow, and L. Tasciotti. 2015. Does Electrification Spur the Fertility Transition? Evidence from Indonesia. *Demography*. 52 (1). pp.1717961–43.

Grogan, L. and A. Sadanand. 2012. Rural Electrification and Employment in Poor Countries: Evidence from Nicaragua. *World Development*. 43. pp. 252–265. Habyarimana, J. and W. Jack. 2009. Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya. Center for *Global Development*. Working Paper No. 169.

Habyarimana, J. and W. Jack. 2012. State vs. Consumer Regulation: An Evaluation of Two Road Safety Interventions in Kenya. *NBER Working Paper Series*. No. 18378.

Iimi, A., E. Lancelot, I. Manelici, and S. Ogita. 2015. Social and Economic Impacts of Rural Road Improvements in the State of Tocantins, Brazil. *World Bank Policy Research Working Paper*. No. 7249.

Khandker, S., Z. Bakht, and G. Koolwal. 2009. The Poverty Impact of Rural Roads: Evidence from Bangladesh. *Economic Development and Cultural Change*. 57 (1). pp.68722–43.

Khandker, S., D. Barnes, and H. Samad. 2013. Welfare Impacts of Rural Electrification: A Panel Data Analysis from Viet Nam. *Economic Development and Cultural Change*. 61 (1). pp. 65692–21.

Kohavi, R. and S. Thomke. 2017. The Surprising Power of Online Experiments. *Harvard Business Review*. September–October 2017 issue. pp. 74–82. https://hbr.org/2017/09/the-surprising-power-of-online-experiments

Legovini, A., V. Di Maro, and C. Piza. 2015. Impact Evaluation Helps Deliver Development Projects. *World Bank Policy Research Working Paper*. No. WPS 7157. http://documents.worldbank.org/curated/en/676351468320935363/ Impact-evaluation-helps-deliver-development-projects.

Lokshin, M. and R. Yemtsov. 2003. Evaluating the Impact of Infrastructure Rehabilitation Projects on Household Welfare in Rural Georgia. *World Bank Policy Research Working Paper*. No. 3155.

Mu, R. and D. van de Walle. 2011. Rural Roads and Local Market Development in Vietnam. The Journal of Development Studies. 47 (1). pp. 70734–21.

Organisation for Economic Co-operation and Development (OECD). 2011. *Aid Effectiveness 2005–10*: Progress in Implementing the Paris Declaration. Paris: OECD Publishing.

Parkhurst, J. 2017. The *Politics of Evidence*. London: Routledge.

Rand, J. 2011. Evaluating the Employment-Generating Impact of Rural Roads in Nicaragua. *Journal of Development Effectiveness*. 3 (28). pp. 1431.

Rao, N. 2013. Does (Better) Electricity Supply Increase Household Enterprise Income in India?. Energy Policy. 57. pp. 532–541.

Sanderson, I. 2002. Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration*. 80 (1). pp. 1–22. doi:10.1111/1467-9299.00292.

Sengupta, R., D. Coondoo, and B. Rout. 2007. Impact of a Highway on the Socio-Economic Well-Being of Rural Households Living in Proximity.*Contemporary Issues and Ideas in Social Sciences.* 3 (3).

Squires, T. 2015. The Impact of Access to Electricity on Education: Evidence from Honduras. Job Market Paper.

Van de Walle, D., M. Ravallion, V. Mendiratta, and G. Koolwal. 2013. Long-Term Impacts of Household Electrification in Rural India. *World Bank Policy Research Working Paper.* No. 6257.

Wang, Y. and B. Wu. 2015. Railways and the Local Economy: Evidence from Qingzang Railway. *Economic Development and Cultural Change.* 63 (1). pp.55588–43.

Yoshino, N. and U. Abidhadjaev. 2015. An Impact Evaluation of Investment in Infrastructure: The Case of the Railway Connection in Uzbekistan. *ADBI Working Paper Series.* No. 548. Tokyo: Asian Development Bank Institute.

مراجع إضافية للاطلاع

Gertler, P., S. Martinez, L. B. Rawlings, P. Premand, and C. M. J. Vermeersch. 2016. *Impact Evaluation in Practice: Second Edition.* Washington, DC: World Bank and Inter-American Development Bank. https:// openknowledge.worldbank.org/handle/10986/25030.

Khandker, S., G. Koolwal, and H. Samad. 2009. *Handbook on Impact* Evaluation: Quantitative Methods and Practices. Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstream/hand le/10986/2693/520990PUB0EPI1101Official0Use0Only1.pdf.

WorldBank.ImpactEvaluationToolkit.http://web.worldbank.org/WBSITE/ EXTERNAL/TOPICS/ EXTHEALTHNUTRITIONANDPOPULATION/ EXTHSD/EXTIMPEVALTK/0,,contentMDK:23262154~pagePK:6416842
7~piPK:64168435~theSitePK:8811876,00.html.

الفصل ٢
استخدام نظريات التغيير لتحديد أسئلة تقييم الأثر

رسائل أساسية

- تعد نظرية التغيير أداة إرشادية يمكن استخدامها لتحديد الحصائل والعلاقات المحتملة التي يمكن اختبارها وقياسها من خلال تقييم الأثر.
- تهدف نظرية التغيير إلى إظهار الروابط الموجودة في سلسلة النتائج التي تربط بين الأنشطة والمخرجات والحصائل والأهداف طويلة المدى، فضلاً عن الافتراضات الشرطية.
- يمكن أن تقدم نظرية التغيير قائمة من الخيارات المتعلقة بتقييم أثر التدخل. ويمكنها أن تساعد في تحديد اعتبارات التناقص التي تؤثر في عملية أخذ العينات وجمع البيانات، وكذلك الافتراضات السلوكية، والتي يمكن أن يساعد تقييم الأثر في التحقق منها.

٢-١ مقدمة

يهدف تقييم الأثر إلى تحديد التأثيرات الناجمة عن تدخل ما، بالنسبة لمجتمع إحصائي محدد. ومع ذلك، يمكن أن يكون لأي تدخل العديد من النتائج المحتملة. إن وصف النتائج التي يمكن توقعها، ومن سيتأثر بها، وعبر أي آلية، وفي أي أُطر زمنية، من شأنه أن يتيح للقائم بعملية تقييم الأثر استيعاب ماهية الخيارات الموجودة بالنسبة لدراسة تقييم الأثر. ويمكن دراسة هذه الخيارات على أساس الاستيعاب الحالي للقطاع ذي الصلة، وذلك بهدف تضييق نطاق الفرضيات محل الاهتمام التي ستخضع للاختبار عبر مقاربات تقييم الأثر.

تعتمد معظم وكالات التنمية على استخدام الأطر المنطقية التي تُظهر العلاقات القائمة بين أنشطة المشروع والمخرجات والحصائل والأهداف طويلة المدى، أو ما يُعرف باسم "سلسلة النتائج". وفي بنك التنمية الآسيوي (ADB)، يأخذ هذا شكل إطار تصميم ومتابعة (ADB 2006). وتعبر سلسلة النتائج عن نظرية تتعلق بكيف يمكن للتدخل أن يُفضي إلى تحقيق التأثيرات المرجوة. وعندما يتم التعبير عن هذه الافتراضات بشكل أكثر وضوحًا، فإن ذلك يشير إلى "نظرية التغيير" (ToC)، والتي تتناول كيف يمكن تحقيق المرجو من العلاقات السببية التي تنشأ بين الأنشطة والمخرجات والحصائل والآثار. إن التعبير عن النظرية بشكل واضح من شأنه أن يساعد في تحديد الوضع الحالي، والوضع المُخطط له أو المرجو، وما يجب القيام به للانتقال من وضع إلى آخر.

لا شك أن ارتكاز تصميم دراسة تقييم الأثر على نظرية تغيير واضحة أمر من شأنه أن يساعد في تحديد البيانات التي سيتم جمعها، وكيف ينبغي تحليل هذه البيانات وإعداد تقارير بشأنها. كما يمكن أن يتيح ذلك للدراسة الانتقال إلى ما هو أبعد من مجرد الإجابة عن السؤال "ما الذي يصلح؟" إلى الإجابة عن "ما الذي يصلح ولماذا؟". يقدم هذا الفصل لمحة عامة عن ماهية نظرية التغيير وكيفية إعدادها.

٢-٢ ما نظرية التغيير؟

تعتبر نظرية التغيير (ToC) بمثابة بيان يوضح كيف تؤدي المدخلات التي يتم توفيرها (الأموال والأشخاص والتغييرات في البيئة التنظيمية أو المتعلقة بالسياسات) إلى الحصائل والآثار المرجوة. وتحدد النظرية الخطوات المتبعة في السلسلة السببية (إطار التصميم والمتابعة في مصطلحات بنك التنمية الآسيوي) والافتراضات الأساسية التي يجب دعمها حتى تعمل النظرية على النحو المتوقع. كما تساعد نظرية التغيير في تحديد المؤشرات أو المتغيرات التي يجب جمع البيانات بشأنها، وقد تحدد "النظريات المضادة" التي يعمل فيها التدخل ذو الصلة بطرق أخرى غير تلك التي تم التخطيط لها، مما يؤدي إلى حصائل غير مقصودة (Carvalho and White 2004). ومن الممكن أن تكون أشجار المشكلات (مقاربة تعتمد على التحليل الشجري للمشكلات)، المُستخدمة لدى بنك التنمية الآسيوي وفي أماكن أخرى لتصميم التدخل، بمثابة مدخلات مفيدة لتحديد نظريات التغيير (ToCs).

لمحة عن نظريات التغيير

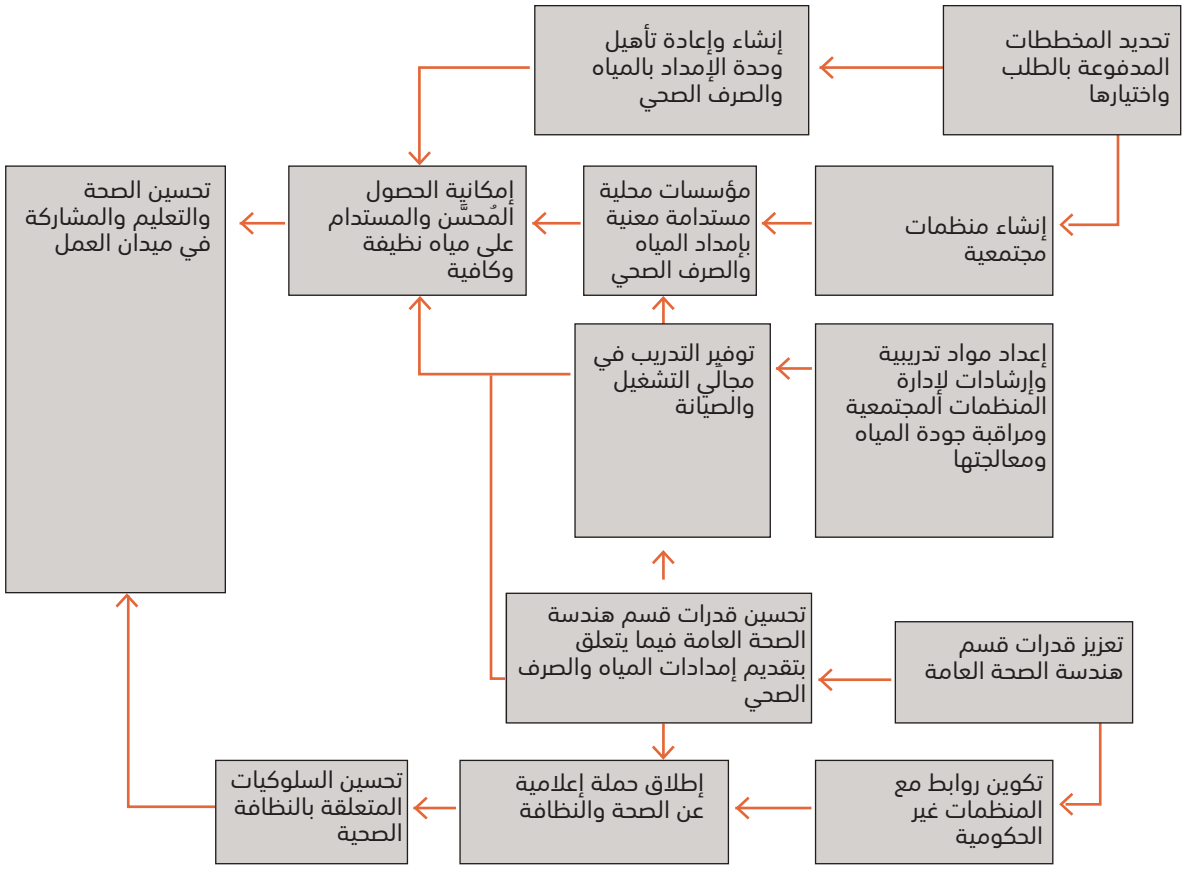
وحسب فوجل، فقد تتخذ نظريات التغيير أشكالاً عديدة، لكنها تهدف جميعًا إلى إنشاء روابط وافتراضات سببية واضحة (Vogel 2012). وقد يكون من المفيد تمثيل نظرية التغيير بشكل تصويري، كما هو الحال في الشكل (٢-١)، والذي يتناول أحد برامج إمدادات المياه والصرف الصحي. ويمكن أيضًا تقديم نظرية التغيير في شكل نص أو جدول، إلا أن التمثيل التصويري يكون أوضح بشكل عام لقدرته على تحديد العلاقات السببية.

لجعل الروابط التي تنشأ عن نظرية التغيير واضحةً قدر الإمكان، يجب تمييز كل مجال من مجالات النشاط. وللقيام بذلك، يجب أن تتجاوز نظرية التغيير "الأنظمة المنعزلة" التي تجمع جميع المدخلات وجميع المخرجات وجميع الحصائل معًا. بدلاً من ذلك، يجب ربط أنشطة محددة بمخرجات محددة، كما يجب أن ترتبط المخرجات المحددة بحصائل محددة. ويختلف هذا العرض عن ذلك الخاص بالإطارات المنطقية التقليدية، مثل إطار التصميم والمتابعة لدى بنك التنمية الآسيوي، وذلك نظرًا لعدم وجود سلسلة سببية واحدة بل العديد منها. ولهذا، يُعد هذا التحديد الأكثر تفصيلاً للعلاقات السببية مفيدًا في تصميم التقييم.

عادةً ما يكون هناك تسلسل من اليسار إلى اليمين أو من الأعلى إلى الأسفل، بدءًا من المدخلات وانتهاءً بالحصائل أو الآثار. ويمكن دمج حلقات التعقيبات بسهولة في أي من التمثيلات التصويرية لنظرية التغيير. كما يمكن تضمين كل من العواقب غير المقصودة والتداعيات ذات الصلة بالمجموعات غير المرتبطة بالتدخل.

تتوقف جميع نظريات التغيير على الافتراضات، ويتعين أن تجعل تلك الافتراضات واضحة. على سبيل المثال، في الشكل (٢-١)، تمت كتابة الافتراضات أسفل الشكل التوضيحي، بحيث يتم توضيح الخصائص السلوكية والشروط اللازمة في كل خطوة.

الشكل (٢-١): نظرية التغيير لأحد مشاريع المياه والصرف الصحي



الافتراضات المحددة

اختيار مخططات مناسبة

تكوين شراكات فعالة بين المنظمات المجتمعية والمنظمات غير الحكومية

وعي الناس بالأنشطة المتعلقة بالنظافة الصحية والمشاركة فيها

اتباع ممارسات صحية تتيح تحسين المياه للحد من الأمراض

إعداد إرشادات ومواد تدريبية مناسبة

استبدال مصادر المياه الحالية وزيادة استخدام المياه

استخدام أساليب توفير الوقت على نحو يعزز الإنتاجية

NGO = منظمة غير حكومية، O&M = التشغيل والصيانة، PHED = قسم هندسة الصحة العامة، WSS = إمداد المياه والصرف الصحي.

المصدر: مقتبس من خطة تحليل تقييمي لدراسة تتعلق بتقييم أثر مشاريع قطاع الإمداد بالمياه والصرف الصحي في ريف البنجاب/المجتمع المحلي في باكستان.

يمكن إضافة بُعد زمني إلى نظرية التغيير، بُناءً على الجدول الزمني للتدخل والوقت اللازم لتحقيق الحصائل. ويمكن لهذا التحليل تحديد توقيت جمع البيانات (سيتم مناقشة ذلك بمزيد من التفصيل في الفصل الثامن).

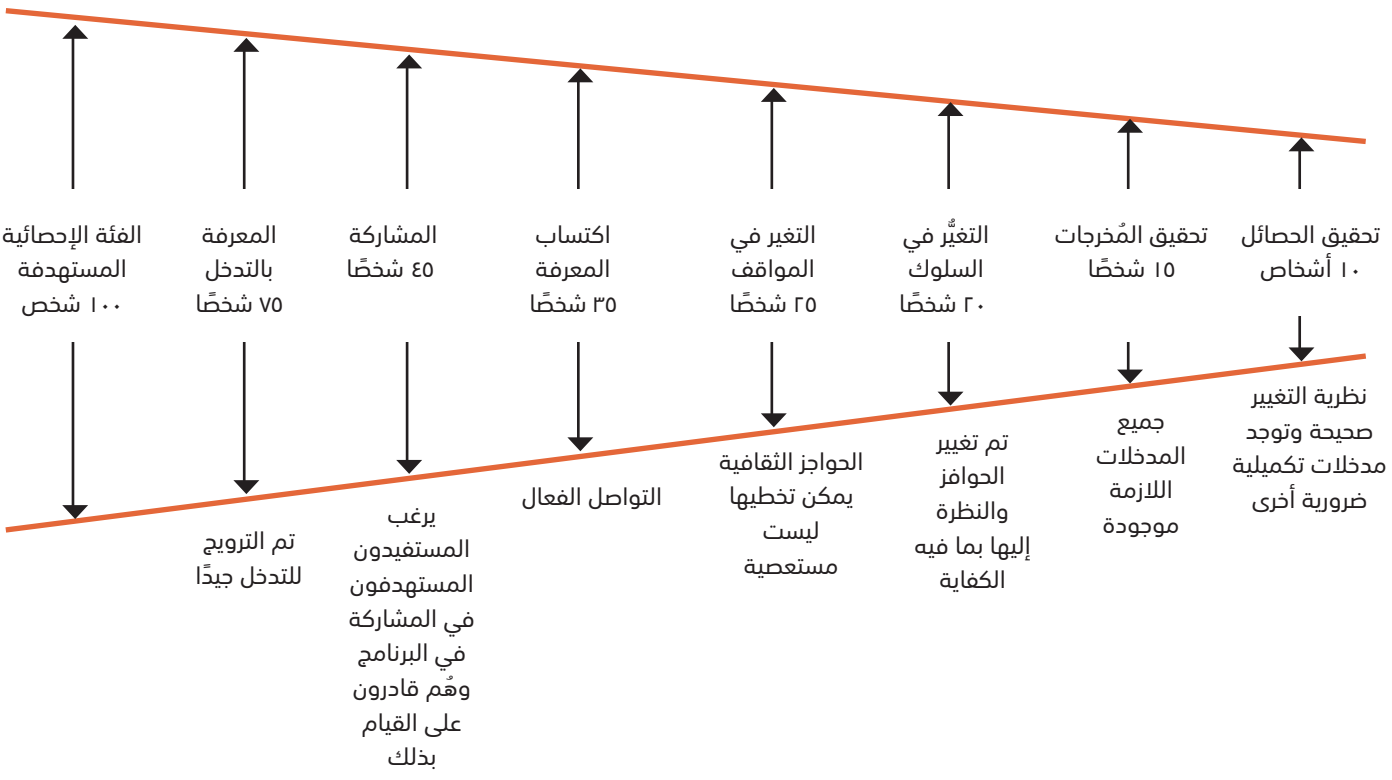
٢-٣ قضايا تطرحها نظرية التغيير

العوائق والميسرات التي تظهر بامتداد السلسلة السببية: المخطط القُمعي للتناقص

وفقًا لهوارد وايت، يعد المخطط القُمعي للتناقص (الشكل ٢-٢) أداة إرشادية لعرض وتصور الافتراضات الأساسية للسلسلة السببية (White 2013). ويكمن الدافع وراء استعمال المخطط القُمعي في أن معدلات المشاركة وأحجام تأثيرها تتناقص بامتداد السلسلة السببية، لذا فإن التأثيرات النهائية لا تكون كبيرة كما يتصورها مصممو المشروع في كثير من الأحيان. وهذا يشكل أهمية لدراسات تقييم الأثر، نظرًا لأن المبالغة في تقدير معدلات المشاركة المتوقعة وأحجام التأثير يمكن أن تؤدي إلى دراسات تعتمد على عينات ذات أحجام صغيرة جدًا يتعذر معها الحصول على نتائج مهمة.

يكمن أحد استخدامات المخطط القُمعي للتناقص، والمُوضح في الشكل (٢-٢)، في الإجابة عن السؤال "من بين ١٠٠ مستفيد مقصود، كم عدد المستفيدين بالفعل من البرنامج؟". وقد يكون هناك تناقص كبير لأن التعرض لتأثيرات البرنامج لا يكون شاملاً، وقد تكون المشاركة جزئية، وقد لا يحدث تغيير في السلوك دائمًا، وقد لا تكون الشروط والظروف المتعلقة بحدوث التأثيرات الكاملة موجودة دائمًا.

الشكل (٢-٢): مثال على مخطط قُمعي للتناقص



المصدر: White 2013.

تتضمن اعتبارات التناقص الشائعة ما يلي:

- **هل المستفيدون المستهدفون على دراية بالبرنامج؟** هل يتم النشر في وسائل الإعلام المناسبة (عبر آليات تتيح الوصول إلى أولئك الذين لديهم معرفة محدودة بالقراءة والكتابة، أو عبر التلفاز في المواقع النائية التي يمتلك فيها القليلون تلفازًا أو يشاهدونه) أو باللغة المحلية الملائمة؟ فإذا لم يكن المستفيدون المستهدفون على دراية بالبرنامج، فمن غير المرجح أن يشاركوا به.

- **هل يرغب المستفيدون المستهدفون في المشاركة في البرنامج؟** هناك العديد من الأسباب التي قد تدفع نحو احتمالية عدم حدوث ذلك. وفي الغالب، ثمة سببان مهمان يتمثلان في: (١) أن تقديرهم الخاص للتكاليف والفوائد الخاصة يقودهم إلى استنتاج أن البرنامج ليس مجديًا، و (٢) أن لديهم شكوكًا حول جدوى التدخل.

- **هل يستطيع المستفيدون المشاركة؟** هل تتم أنشطة المشروع في وقت ومكان وبتكلفة (بما في ذلك تكلفة الفرصة) تتيح لهم إمكانية القيام بذلك؟ هل سيشعر المستفيدون المقصودون بأنهم مستبعدون وغير مرحب بهم اجتماعيًا، أم سيشعرون بالخوف من الإعداد لأنشطة المشروع؟

- **هل تتم عملية نقل المعرفة على النحو الفعال؟** تتضمن معظم مشاريع التنمية نقل المعرفة في مجال معرفي ما إلى موظفي المشروع أو الوكالة، وكذلك إلى الوسطاء والمستفيدين المستهدفين. وهناك رابط ضعيف شائع في السلسلة السببية يكمن في افتراض فعالية عملية نقل المعرفة، أو أن المشاركين يتعلمون المرجو منهم.

- **هل يحدث التغيُّر في السلوك على النحو المتوقع؟** تتطلب معظم التدخلات الإنمائية تغييرًا في السلوك بين المسؤولين الحكوميين أو الوسطاء أو المستفيدين المستهدفين. وفي كثير من الأحيان، قد يختلف التغيُّر الفعلي في السلوك للمشاركين بشكل كبير عما يتوقعه مديرو المشروع.

- **هل تظل القيود الأخرى على الفعالية دون معالجة؟** قد يتمثل أحد القيود التي تحول دون القدرة على تغيير السلوك في المدخلات التكميلية. على سبيل المثال، فحتى بعد تخفيض تكاليف النقل بفضل أحد مشاريع الطرق، قد تظل الشركات تواجه قيودًا أو عقبات ائتمانية تتعلق بتحسين قدراتها وتوسيع نطاقها.

- **هل يكون للتدخل تأثير دائم؟** غالبًا ما يكون التأثير غير متكرر، حتى لو توفرت جميع الشروط والظروف المتعلقة بالفعالية. فعلى سبيل المثال، إذا أدى تحسين إمداد المياه للمناطق السكنية إلى تقليل متوسط تكرار الإصابة بأحد الأمراض المنقولة عن طريق المياه، فقد يستمر انتقال هذا المرض خارج المنزل أو عبر قنوات أخرى، وقد لا تتم ملاحظة التأثير الناجم على الحالة الصحية إلا بشكل غير متكرر. وبالمثل، فإنه قد لا تتم ملاحظة التأثير المباشر لأي من التدخلات التأمينية على آليات المكافحة إلا في السنوات التي يقع فيها حادث ما تترتب عنه خسارة.

تطور التأثيرات

يتم إجراء دراسة تقييم الأثر في وقت ما معين، وغالبًا ما يكون ذلك في نهاية المشروع أو البرنامج. ومن المفترض استمرار التأثيرات التي تم قياسها عند مستوى مماثل بعد ذلك. ومن الممكن أن تلعب نظرية التغيير دورًا مهمًا في توضيح هذا الافتراض المتعلق بالاستدامة.

تعتمد استدامة المزايا على الاستخدام السليم للمخرجات وصيانتها، مثل البنية التحتية. وقد يحدث التناقص عندما لا تكون الصيانة كافية. وقد تصبح المدخلات التكميلية الضرورية لإحداث الأثر أكثر أو أقل تكرارًا، وذلك نتيجة للتغيرات السياقية.

يتوقف البعد الزمني للتناقص على منحنى الانتشار لأي تغيير يتم تعزيزه من خلال التدخل. وتتنبأ نظرية الانتشار الكلاسيكية بنمط انتشار على شكل الحرف الإنجليزي S، مع تقبل بطيء أولي من قبل المشاركين الأوائل (Rogers 2003). وبمجرد إثبات التغيير، يكون هناك تقبل سريع، والذي يستمر بعد ذلك على هذا النحو ليتحول إلى معدل اعتماد أبطأ بين المشاركين المتأخرين. في الوقت نفسه، قد ينشأ عدم إقبال على التغييرات مما يؤدي إلى تقلص في حجم الانتشار المتوقع بمرور الوقت. وقد يكون لدى المعتمدين الأوائل أيضًا اختلافات مهمة عن المعتمدين اللاحقين، مثل السن الأصغر والترابط الأفضل والتعليم العالي، مما يعني تعرضهم لتأثيرات لا تُعزى إلى التدخل.

تحديد العناصر السلوكية للتغيير المرجو

يعد تقييم الأثر مقاربة في محلها عندما تعتمد التأثيرات الناجمة عن التدخلات على السلوك البشري والردود أو الاستجابات ذات الصلة، والتي لا يمكن التنبؤ بها آليًا. ولهذا السبب، يمكن أن يشكل اختبار ما إذا كان التغيُّر السلوكي يحدث على النحو المرجو من التدخل أحد العناصر المهمة لمقاربة تقييم الأثر. وقد يكون نموذج التغيير السلوكي أحد الأساليب الإرشادية المهمة لتحقيق هذه الغاية (الجدول ٢-١). ويتيح هذا النوع من النماذج بأن تكون الافتراضات السلوكية المتعلقة بالتدخل واضحة جدًا، فضلاً عن إمكانية ترجمتها إلى متغيرات يمكن أن يساعد تقييم الأثر في قياسها. ويمكن استخدام هذه المتغيرات كحصائل يُتوقع حدوثها أو كمتغيرات وسيطة توضح ما إذا كان سيسفر الأمر عن آثار طويلة المدى ووقت حدوث ذلك.

الجدول (٢-١): توضيح لأحد نماذج التغيير السلوكي في وسيلة تأخذ شكل جدول

الحصيلة المرجوة	شخص فاعل (أو مجموعة من الأشخاص الفاعلين الذين من المتوقع أن يتغيروا بنفس الطريقة) يتعين عليه أن يخضع لتغيُّر في السلوك	يلزم أن يؤدي التغيير في السلوك إلى تحقيق الحصيلة المرجوة	المؤشرات المتوقعة لحدوث تغيير في السلوك

المصدر: المؤلفان.

تحديد العوامل الخارجية والعواقب غير المقصودة

قد يكون للتدخلات فوائد وعواقب سلبية على حد سواء. فعلى سبيل المثال، يمكن للطرق الجديدة أن تؤدي إلى زيادة في معدلات الحوادث، ونزوح شركات وأعمال محلية، فضلاً عن إحداث تأثيرات اجتماعية غير مرغوب فيها. ومن الممكن أن تساعد مشاركة الأطراف المعنية في إعداد نظرية التغيير (سيتم مناقشة ذلك فيما يلي) في تحديد العواقب المحتملة غير المقصودة.

من الممكن أن تفيد نظرية التغيير أيضًا في تحليل التداعيات أو التأثيرات التي تتجاوز الفئة السكانية أو الإحصائية المُعالَجة، إلى جانب المساعدة في تحديد العواقب غير المقصودة. ويعد الانتشار عبر التواصل الشفهي أو رصد التأثيرات مثالاً على التداعيات الإيجابية، والذي قد يكون جزءًا واضحًا من نظرية التغيير، كما هو الحال في مدارس تدريب المزارعين. وهناك أمثلة أخرى على عوامل خارجية إيجابية تتمثل في تقليل التلوث أو الازدحام أو حالات الأمراض المعدية. في حين قد تشمل التداعيات السلبية عكس ذلك، أو انخفاض الطلب على العمالة غير الماهرة.

٢-٤ بناء نظرية التغيير

خطوات بناء نظرية التغيير

تمثل الخطوات السبع التالية وسيلة مفيدة لتحديد سمات نظرية التغيير.

(١) **إجراء تحليل سياقي:** يجب أن تبدأ نظرية التغيير بتحديد المشكلة الجذرية التي يجب أن يعالجها التدخل، وعواقب المشكلة، والأسباب التي تؤدي إلى تفاقم المشكلة. وتنبغي دراسة فرص معالجة المشكلة، وذلك في ضوء المبادرات الأخرى القائمة والمخطط لها.

(٢) **تحديد التدخل والأهداف والحصائل:** تتطلب نظرية التغيير تحديدًا واضحًا لماهية التدخلات، والمخرجات التي تقدمها، وأنشطتها ذات الصلة. ويجب أن يكون ذلك محددًا قدر الإمكان. ويتعين أن تتماشى الحصائل المحددة مع أهداف التدخل.

(٣) **وضع الخطوات الرئيسية في السلسلة السببية:** تربط نظرية التغيير المدخلات بالحصائل المقصودة من خلال الأنشطة والمخرجات والحصائل متوسطة المدى. ويجب تحديد كيفية ربط السلاسل السببية المختلفة عبر هذه العناصر.

(٤) **تصور المؤشرات بامتداد السلسلة السببية:** يمكن استخدام هذه المؤشرات لتتبع الروابط السببية وتحديد العوائق التي تحول دون تحقق الأثر المطلوب. فعلى سبيل المثال، تتناول دراسة تقييم الأثر المدعومة من بنك التنمية الآسيوي حول توسيع مترو تبليسي الأثر الواقع على طلاب الجامعات من خلال جمع البيانات المتعلقة باستخدام الوقت وأنماط النقل وتكاليف السفر وبالآثار ذات الصلة المسجلة على صعيد أنماط الاستهلاك، فضلاً عن معدلات الحضور ودرجات الاختبار (المربع ٢-١).

(٥) **تحديد الافتراضات الأساسية:** بالإضافة إلى الروابط التي تنشأ في السلسلة السببية، ستكون هناك افتراضات أساسية يجب أخذها بعين الاعتبار حتى تؤدي السلسلة السببية الدور المنشود منها. وينبغي عدم أخذ أي افتراضات كأمور مُسلَّم بها.

(٦) **التمييز بين قنوات الحصائل:** قد يكون من المفيد أن تستخدم نظرية التغيير محور عرض تقديمي واحدًا للتمييز بين المسارات المؤدية إلى حصائل محددة لمجموعات محددة من المستفيدين أو فئات مختلفة من الحصائل.

(٧) **التحقق من الصلاحية والمراجعة:** ينبغي التحقق من صلاحية نظرية التغيير من خلال إجراء مناقشات مع الأطراف المعنية الرئيسية، مثل موظفي البرنامج ومديريه، والمستفيدين المستهدفين والوكالات التي لديها برامج مماثلة. وقد تؤدي هذه المشاورات إلى إجراء مراجعات وتنقيحات تهم كلاّ من نظرية التغيير وأسئلة التقييم (المربع ٢-١).

المربع (٢-١): تنقيح سياسة التغيير استنادًا إلى مدخلات الأطراف المعنية: القسائم المدرسية في الفلبين

تكمن نظرية التغيير الشائعة للقسائم المدرسية في أن الأطفال الذين يتلقون قسائم يحصلون على تعليم أفضل في القطاع الخاص، وبالتالي تكون الحصيلة النهائية تحسنًا في نتائج التعليم. وعندما تم عرض نظرية التغيير هذه على موظفي وزارة التعليم في الفلبين، اعترضوا على أنه لم يكن من المفترض أن تكون هذه هي الطريقة التي تعمل بها القسائم المدرسية في الفلبين. فقد كانت القسائم أحد البرامج التي تهدف إلى معالجة الاكتظاظ في المدارس الحكومية. يعمل نقل التلاميذ الذين يتلقون قسائم على خفض نسبة التلاميذ لكل معلم، ومن ثم تحسين نتائج التعلم لأولئك الذين يبقون في المدرسة الحكومية. وبدون هذه الاستشارة، والتي أدت إلى تحديد سلسلة سببية جديدة، لم يكن لتقييم الأثر أن يأخذ في الاعتبار أثر البرنامج على غير المستفيدين من القسائم.

المصدر: المؤلفان

عملية بناء نظرية التغيير

يتم وضع تصور مبدئي لنظرية التغيير عندما يتم تصميم المشاريع والسياسات والبرامج (المشار إليها فيما يلي بـ "التدخلات") بشكل أولي. لكن، غالبًا ما يتم التخطيط لتقييم الأثر بعد هذه الخطوة، عندما يكون هناك قدر أكبر من اليقين بشأن ما إذا كان التدخل سيحدث أم لا. وهذا يعني أن الخطوة الأولى تصبح مراجعة الأدوات المستعملة حاليًا في نظرية التغيير، مثل الأطر المنطقية (إطارات التصميم والمتابعة)، والتشاور مع الأطراف المعنية المناسبة بشأن التحديثات والتنقيحات المطلوبة. وفي معظم الحالات، سوف يلزم إضافة مستوى أعلى من التفاصيل مقارنةً بما تم وضعه سابقًا لإعداد المشروع بهدف التوصل إلى تصميم تقييم الأثر.

يعد إعداد نظرية التغيير عملية استشارية تتم صياغتها على أفضل نحو من خلال إشراك الأطراف المعنية الرئيسية بانتظام، والذين يمكن التماس آرائهم أثناء تطوير نظرية التغيير. وتكمن إحدى طرق القيام بذلك في عقد ورش عمل تحضيرية مع موظفي البرنامج وإدارته خلال مرحلة تصميم دراسة تقييم الأثر. ومن شأن الحصول على موافقة موظفي البرنامج، لتحديد نظرية التغيير التي تدعم أسئلة التقييم، أن يساعد في دعمهم لتصميمات التقييم المقترحة.

قد يتم دعم الروابط السببية في نظرية التغيير، عبر الأبحاث القائمة، بهدف توفير إطار عمل كأساس لروابط محددة. وتعتمد العديد من المشاريع على افتراضات اقتصادية جزئية تقليدية عند الاشتغال على نظرية التغيير. وعندما يتعلق الأمر بتغيير السلوك، يمكن استخدام العديد من النماذج المستقاة من علم النفس السلوكي والتخصصات المعرفية الأخرى. وفي الغالب، ثمة إشكاليات جديرة هي الأخرى بالوضع في الاعتبار مثل المعلومات غير المتماثلة ودور تكاليف المعاملات.

٢-٥ تطبيق نظرية التغيير

تتيح نظرية التغيير تشكيل مجموعة مختارة من أسئلة التقييم والمتغيرات التي يتعين استخدامها كأداة من أدوات المسح الإحصائي. يتمثل الدور الأساسي لنظرية التغيير كأداة إرشادية في أنها تساعد في تحفيز المناقشة حول الحصائل المتوقعة، وكيف ينبغي تناولها في تصميم تقييم الأثر. ونادرًا ما يتم تناول جميع الحصائل في دراسة تقييم أثر واحدة. ومع ذلك، فإن توضيح العلاقات السببية التي يمكن التحقق منها يتيح مقارنة تلك العلاقات مع الأدلة المتاحة، وذلك لتحديد الفرضيات التي يمكن أن يتولد من خلالها فهم جديد. ويمكن أيضًا التحقق من هذه العلاقات على أساس المخطط القُمعي لتقدير مدى احتمالية أن تكون أحجام العينة والتأثيرات كافية لإجراء التحليل في فترات زمنية معينة. كما يمكن أن تساعد دراسة منحنيات الاعتماد (التقبل) وسمات المعتمدين الأوائل في فهم الفئة الإحصائية (السكانية) التي يمكن قياس التأثيرات التي طرأت عليها في فترات زمنية مختلفة.

يمكن استخدام الفروع الأولية لنظرية التغيير لإنشاء مؤشرات تكشف كيف يؤثر التنفيذ أو العوامل الشرطية على الحصائل. ومن الممكن أن يؤدي ذلك إلى إثراء الفرضيات المتعلقة بالجيل الثاني من أسئلة تقييم الأثر. ويجب أن تعمل المؤشرات المرتبطة بسلاسل النتائج المختارة لإجراء عملية التحقق على توجيه المتغيرات التي تم الحصول عليها خلال تصميم عملية جمع البيانات وإستراتيجية التحديد الخاصة بدراسة تقييم الأثر. ومن الممكن أن يوفر نموذج التغيير السلوكي الذي تم تطويره قائمة بالحصائل المحتملة والمتغيرات المتوسطة التي يمكن استخدامها في تحليل تقييم الأثر.

المراجع

ADB. 2006. *Guidelines for Preparing a Design and Monitoring Framework*. Manila.

Carvalho, S. and H. White. 2004. Theory-Based Evaluation: The Case of Social Funds. *American Journal of Evaluation*. 25 (2). pp. 141–60.

Rogers, E. 2003. *The Diffusion of Innovations*. Fifth Edition. New York: The Free Press.

Vogel, I. 2012. Review of the Use of "Theory of Change" in International Development. Review Report for the Department for International Development of the United Kingdom. http://www.dfid.gov.uk/r4d/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf.

White, H. 2013. Using the Causal Chain to Make Sense of the Numbers. Evidence Matters. Blog. 12 February. http://blogs.3ieimpact.org/using-the-causal-chain-to-make-sense-of-the-numbers/.

مراجع إضافية للاطلاع

Taplin, D. and H. Clark. 2012. *Theory of Change Basics: A Primer on Theory of Change*. New York: Actknowledge. http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.

Rogers, P. 2014. *Theory of Change: Methodological Briefs – Impact Evaluation No. 2*. Florence: UNICEF Office of Research. https://www.unicef-irc.org/publications/747/.

Stein, D. and C. Valters. 2012. Understanding Theory of Change in International Development. JSRP Paper 1. London: Justice and Security Research Programme, London School of Economics. http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/UNDERSTANDINGTHEORYOFChangeSteinValtersPN.pdf.

الفصل ٣
المفاهيم الأساسية لتقييم الأثر

رسائل أساسية

- تعتمد دراسات تقييم الأثر على استخدام تقنيات تجريبية لتحديد أو نمذجة سيناريو واقع مضاد دون التدخل محل الاهتمام بهدف المقارنة مع الوضع الفعلي.
- تنطوي تقنيات تقييم الأثر على تصميمات تجريبية وطرق شبه تجريبية ومقاربات قائمة على الانحدار.
- تعتمد جميع طرق تقييم الأثر على رصد ملاحظات حول ما يطرأ على كل من المجتمعات الإحصائية التي تتأثر بالتدخل وتلك التي لم تتأثر به.
- تكمن إحدى الإشكاليات الرئيسية التي يجب أن يعالجها تقييم الأثر في تحيّز الاختيار، والذي يتمثل في أن تلك المجتمعات الإحصائية التي تخضع لتدخلات ليست مثل تلك التي لا تخضع لها.
- من الأهمية استيعاب وحدة التعيين ووحدة العلاج ووحدة التحليل ذات الصلة بالتدخل.
- يترتب على تقنيات تقييم الأثر المختلفة تقديرات متعلقة بالتأثيرات تكون صالحة للمجتمعات الإحصائية المختلفة.

٣-١ ما تقييم الأثر؟

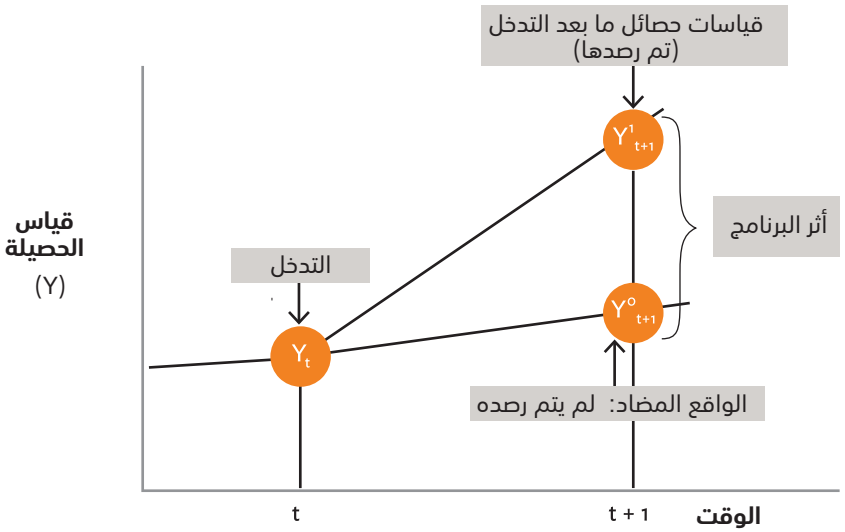
تقيس تقييمات الأثر *تأثيرات المعالجة*. وفي هذا السياق، نعني بالمعالجة التعرض لتدخل ما، مثل سياسة أو مشروع جديد. أما التأثيرات، فالمقصود بها الفروق التي يحدثها التعرض لهذا التدخل على حصائل، مثل الدخل والإنتاجية والفقر والصحة والعديد من الجوانب الأخرى.

يعتمد تقييم الأثر على *تحليل الواقع المضاد* الذي يقارن بين ما كان سيحدث في حالة عدم وجود تدخل والحصائل الفعلية المترتبة على التدخل.[1] ويُمكن لتقييم الأثر أيضًا مقارنة نتائج تدخل معين بالنتائج المتعلقة بتدخل مختلف.

[1] يتنوع التدخل بدءًا من أنشطة محددة، مثل تقديم المنح النقدية للأسر الفقيرة أو وجبات التغذية لأطفال المدارس وكذلك توفير البنية التحتية مثل الطرق الريفية أو معالجة مياه الصرف الصحي في المناطق الحضرية، وصولاً إلى سياسات أكثر عمومية، مثل إصلاح التعليم.

يتناول الشكل (٣-١) تقييم الأثر بشكل بصري. يحدث التدخل في الوقت t، عندما يكون مستوى حصيلتنا محل الاهتمام Y_t. وبعد التدخل، تصبح الحصيلة محل الاهتمام Y^1_{t+1}، بينما كانت ستصبح Y^0_{t+1} فقط بدون التدخل. وتشير الحصيلة الأخيرة إلى قيمة Y الخاصة بالواقع المضاد.

الشكل (٣-١): شكل توضيحي لتقييم الأثر



المصدر: المؤلفان.

يمكن تمثيل تقييم الأثر، كما هو موضح في الشكل (٣-١)، جبريًا على النحو المُبيَّن في المعادلة (٣-١):

$$\text{الأثر} = Y^1_{t+1} - Y^0_{t+1} \qquad (٣-١)$$

حيث يشير Y إلى الحصيلة محل الاهتمام مثل عدد الفقراء أو استخدام الوقت أو الإصابة بالأمراض. يشير الرمز السفلي $t+1$ إلى نقطة زمنية (فترة زمنية) بعد التدخل، أو إلى نقطة زمنية (فترة زمنية) بدرجة كافية في نطاق التدخل ليتسنى التوقع بشكل معقول أن هناك تأثيرًا على الحصيلة. في حين يشير الرمز السفلي 1 إلى الحصيلة المُسجلة عند المشاركة في التدخل، أي الحصيلة الواقعية. ويشير الرمز السفلي 0 إلى الحصيلة نفسها للمجموعة نفسها من الأفراد، وذلك في نفس النقطة الزمنية (الفترة الزمنية) في حالة عدم مشاركتهم في التدخل، أي الواقع المضاد.

يتمثل التحدي الرئيسي لتقييم الأثر في تحديد وقياس التقديرات الصحيحة المرتبطة بالواقع المضاد للحصائل المرجوة. بينما يمكن ملاحظة ما يحدث لأولئك الذين يتلقون دعما من، على سبيل المثال، برنامج التمويل متناهي الصغر (Y^1)، فإنه لا يمكن ملاحظة الحصائل المتعلقة بالمجموعة نفسها في حالة عدم مشاركتها (Y^0).

وبعيدًا عن التقييم الدقيق للأثر، غالبًا ما تُستخدم مقارنات بسيطة لاستخلاص استنتاجات حول التأثيرات الناجمة عن التدخلات. ومع ذلك، تتسم هذه المقارنات بعيوب بالغة.

المقارنة الأحادية: يتمثل أحد الاحتمالات في مقارنة الحصائل قبل التدخل وبعده، وهو ما يُعرف باسم المقارنة الأحادية. فعلى سبيل المثال، سيتم إجراء مسح لاستقصاء مستوى دخل متلقي التمويل متناهي الصغر. وبعد عام، سيتم إجراء مسح آخر لاستقصاء مستوى الدخل، ثم تتم مقارنة متوسط الدخل قبل البرنامج وبعده. ومن النادر جدًا أن يكون هذا فعالاً، لأن التغيير في الدخل، سواء أكان إيجابيًا أم سلبيًا، يمكن أن يكون بسبب العديد من العوامل الأخرى بخلاف البرنامج.

عادة ما تتغير الحصائل بمرور الوقت بغض النظر عن التدخل، لذلك فإن متابعة مدى تغير الحصائل بمرور الوقت يوفر معلومات حول تأثيرات جميع العوامل التي تتغير بمرور الوقت، بدلاً من التدخل محل الاهتمام. وهذا يعني أن هذه التقنية البسيطة تنطوي على مشكلات تتعلق بالسببية التي تتسم بالخلط.

المقارنة المقطعية: يتمثل أحد الاحتمالات الأخرى في مقارنة الحصائل المتعلقة بمتلقي الدعم من برنامج التمويل متناهي الصغر بأشخاص آخرين لم يشاركوا في هذا البرنامج داخل البلدة. ومع ذلك، قد يكون لدى أولئك الذين شاركوا في برامج التمويل متناهي الصغر خصائص مختلفة عن أولئك الذين لم يشاركوا، وبالتالي فإن الزيادة في الدخل قد تكون بسبب الخصائص الأخرى وليس البرنامج نفسه. مرة أخرى، تخلط معظم تقنيات المقطع العرضي البسيطة بين التدخل وعوامل أخرى.

٣-٢ تحديد المجموعات الضابطة ومجموعات المقارنة

للحصول على قياس صالح للآثار الناجمة عن تدخل ما، يلزم استخدام تقنيات تتيح عدم تشويش عملية المقارنة بعوامل أخرى. وفي هذا الإطار، تتمثل أحد المقاربات الرئيسية في استخدام مجموعة ضابطة أو مقارنة، وهي مجموعة من الأشخاص الذين يتمتعون بخصائص تشبه إلى حد كبير خصائص المجموعة التي حظيت بالتدخل، ولكنها بالفعل لم تحظَ بالتدخل.[٢] وتشير مجموعة المقارنة إلى عينة من الأفراد أو الأسر أو الشركات أو بعض الوحدات الأخرى المنتمية إلى نفس المنطقة الجغرافية الممتدة التي شملها التدخل. وفي المتوسط، يجب أن تتمتع مجموعة المقارنة بنفس خصائص مجموعة المعالجة عند خط الأساس، وهو ما يُسمى التوازن (يقدم المربع ٣-١ مثالاً في هذا الخصوص). وحتى يتسنى تحديد مجموعة مقارنة، يتعين على المحلل فهم *آلية التعيين*، وهي العملية التي يتم من خلالها اختيار أولئك الذين يحظون بالتدخل، ويمكن أن يختار أفراد تلك العينة بأنفسهم المشاركة.

[٢] في بعض الأحيان، تُستخدم "مجموعة المقارنة" و"المجموعة الضابطة" بشكل تبادلي. وقد جرت العادة على استخدام مصطلح "مجموعة المقارنة" في التصاميم غير التجريبية، بينما يُستخدم مصطلح "المجموعة الضابطة" في التصاميم التجريبية. ويتم تناول الاختلافات بين كل تصميم في القسم التالي.

المربع (٣-١): إظهار التوازن في دراسة تتعلق بالترويج لغسل اليدين في باكستان

تناول الباحث بوين وآخرون (Bowen et al.) في تقرير لهم (في عام ٢٠١٣) دراسة متابعة امتدت لخمسة أعوام حول أحد مشاريع الترويج لغسل اليدين في كراتشي. وعند المتابعة، تم إعادة تسجيل ٨٤٪ من الأسر التي شاركت في الدراسة الأصلية في هذه الدراسة الاستقصائية. ويوضح الجدول التالي متوسط الخصائص المتعلقة بالأسر المشاركة في المعالجة والمقارنة لعدد من المتغيرات. وكما يتضح، فإن الأرقام متشابهة جدًا. ويشير العمود الأخير في الجدول إلى احتمالية أن تكون المتوسطات ناجمة عن مجتمعات إحصائية مختلفة عن بعضها البعض، وهذا هو الحال إذا كانت قيمة p أقل من ٠,١. وفي جميع الحالات، تكون قيمة p أعلى بكثير، مما يدل على أن المجموعة المُعالجة والمجموعة الضابطة متماثلتان.

المتغير	المعالجة	المقارنة	قيمة p
حجم الأسرة	٨,٤	٨,٥	٠,٨
إلمام الأم بالقراءة والكتابة	٣٣٪	٣١٪	٠,٦
إلمام الأب بالقراءة والكتابة	٦٤٪	٥٥٪	٠,٦
تحدث اللغة الأوردية داخل المنزل	٩٦٪	٩٦٪	٠,٩
امتلاك جهاز راديو	١٤٪	١٣٪	١,٠
امتلاك جهاز تلفزيون	٩١٪	٩٢٪	٠,٩
امتلاك ثلاجة	٥٦٪	٥٦٪	١,٠
حصول المنزل على إمدادات مياه البلدية	٣١٪	٣٥٪	٠,٩
حجم العينة (n)	٣٠١	١٦٠	

المصدر: تناول الباحث بوين وآخرون (Bowen et al.) في تقرير (٢٠١٣).

في التجارب المعملية، حيث يتحكم الباحثون في التدخل محل الاهتمام، يمكن استيعاب عملية التعيين استيعابًا كاملاً. ومع ذلك، يكون هذا أكثر صعوبة عندما يتعلق الأمر بوضع برامج التنمية. وتتمثل المشكلة التي تظهر غالبًا عند محاولة تحديد مجموعة مقارنة في أنه يتم القيام بالتدخل في أماكن محددة أو أنه يستهدف أشخاصًا معينين لسبب ما. بالإضافة إلى ذلك، فإنه عادةً ما يكون الأشخاص الذين يختارون المشاركة في البرنامج مختلفين عن أولئك الذين لا يختارون ذلك. ذلك أنهم قد يكونون أكثر دراية أو تعليمًا، أو أكثر استعدادًا للمخاطرة، أو أكثر نشاطًا، أو لديهم اختلافات سلوكية أخرى عن أولئك الذين لا يشاركون. وغالبًا ما يكون تكرار خيارات التنسيب والاختيار للحصول على مجموعة مقارنة صالحة أمرًا ليس باليسير.

لتحديد مجموعة ضابطة أو مجموعة مقارنة، ثمة تصميمات تجريبية وغير تجريبية على حد سواء لدراسات تقييم الأثر.

١. *التصميم التجريبي* هو المصطلح المُستخدم عندما يتم تعيين المجموعات المُعالَجة وغير المُعالَجة بشكل عشوائي بواسطة عامل خارجي لا صلة له بالتدخل.

- *التجارب العشوائية المضبوطة (RCTs):* في بعض الأحيان، يُطلق على هذه التجارب اسم *التجارب العشوائية* أو *التقييمات العشوائية*. وتنطوي هذه المقاربة على التعيين العشوائي لمن سيتلقون المعالجة أو مكان ذلك. وهذا يعني أنه يتم اختيار المجموعات الموجودة ضمن مجموعة التدخل وتلك الموجودة ضمن المجموعة الضابطة عشوائيًا من قائمة تتضمن المجموعات المؤهلة للبرنامج. وهذا يختلف عن أخذ عينات عشوائية من أولئك المندرجين بالفعل ضمن المجموعة المُعالجة والمجموعة غير المُعالجة. ويتم تناول التجارب العشوائية المضبوطة بمزيد من التفصيل في الفصل الرابع.

٢. *التصاميم غير التجريبية* تُستخدم عندما لا يكون التصميم التجريبي ممكنًا بسبب التكاليف أو اعتبارات أخرى حالت دون التعيين العشوائي. وغالبًا ما يتم النظر بعين الاعتبار إلى التصاميم غير التجريبية في حالة تم اتخاذ قرار بإجراء تقييم للأثر بعد حدوث التدخل.

- *التجارب الطبيعية*: عندما تفصل حالة طبيعية أو مواقع جغرافية أو سياسة حكومية ما فئة سكانية متجانسة بطبيعتها إلى مجموعتين أو أكثر يتم تمييزهما فقط من خلال التعرض لتدخل ما، فيمكن اعتبار ذلك تجربة طبيعية. وتتمثل بعض الأمثلة على ذلك فيما يلي: (١) نهر (أو طريق سريع مُشيَّد حديثًا) يقسم القرية إلى نصفيْن، حيث تقع عيادة صحية في أحد النصفيْن، مما يتيح للقرويين المقيمين في هذا النصف الوصول إلى هذا المرفق الصحي على نحو أكثر سهولة من نظرائهم المقيمين في النصف الآخر؛ (٢) منطقتان سكنيتان يوجد فيهما سكان من نفس الفئة الاجتماعية والاقتصادية يعيشون مع بعضهم البعض، لكن هاتين المنطقتين تنتميان إلى مقاطعتيْن مختلفتيْن، ومن ثم فإن أي تغيير في السياسة الضريبية لإحدى المقاطعتيْن سيؤثر على نصف هؤلاء السكان. وعلى الرغم من أن التجارب الطبيعية تتلاءم بشكل فريد مع دراسات تقييم الأثر، فإنه يتم تناولها عمليًا من خلال التصاميم شبه التجريبية أو المقاربات القائمة على الانحدار.

- *التصاميم شبه التجريبية:* تستخدم هذه التصاميم طرقًا إحصائية لإنشاء مجموعة مقارنة، والتي تتمتع بنفس خصائص مجموعة المعالجة، بغض النظر عن المعالجة. وتتمثل المقاربات شبه التجريبية الرئيسية في التباين المزدوج أو الاختلاف فى الاختلافات (DiD)، فضلاً عن مطابقة درجة الميل وتصميم انقطاع الانحدار، والتي سيتم تناولها في الفصل الخامس. وتعد هذه المقاربات الأكثر شيوعًا في الدراسات التي يدعمها بنك التنمية الآسيوي حتى الآن. ويمكن استخدام تصميم استمرارية الانحدار عندما تكون هناك قاعدة تستلزم استيفاء شرط الحد الأدنى للأهلية.

- *المقاربات القائمة على الانحدار:* تتيح هذه المقاربات تقدير نماذج الانحدار التي يتم فيها عادةً تناول المشاركة من خلال متغير صوري. وعلى نحو أقل شيوعًا، تكون المعالجة متغيرًا مستمرًا مثل حجم القرض أو مدة التدريب. وتشمل هذه المقاربات نماذج المعالجة داخلية النشأة والمتغيرات المساعدة والانحدارات التبديلية والانحدار القوي المزدوج. ويتم تناول هذه المقاربات في الفصل الخامس.

في حالة استخدام جولة واحدة فقط من الملاحظات لإظهار صلاحية مجموعة المقارنة، أي أن التحليل يقارن فعلاً بين شيئين متشابهين، فحينها يجب أن تقدم الدراسات شبه التجريبية جدول مقارنة إحصائية بين الاختلافات. ويتعين أن يُظهر هذا الجدول أنه تم التحقق من أن مجموعتي المعالجة والمقارنة/ الضابطة كانتا متماثلتيْن قبل التدخل (كما هو موضح في المربع ٣-١). ومع ذلك، يمكن لجداول المقارنة الإحصائية بين الاختلافات التحقق من هذا التوازن (الاتساق) عبر خصائص يمكن ملاحظتها فقط، ومن ثم قد يظل هناك تحيز ناجم عن عدم التوازن بشأن خصائص لم يتم ملاحظتها.

٣-٣ التحيزات والتحديات المتعلقة بالاستدلال السببي

ثمة العديد من التحديات التي تواجه تحديد مجموعة مقارنة صالحة للاستدلال السببي، وتكمن فيما يلي:

١. *تحيّز الاختيار:* من المُرجح أن يختلف أولئك الذين لا يشاركون في البرنامج في نواحٍ مهمة عن أولئك الذين يشاركون فيه. وينشأ تحيّز الاختيار من تنسيب البرنامج (*تحيز التنسيب*) والاختيار الذاتي في البرامج على حد سواء. إذا كانت محددات الاختيار في البرنامج مرتبطة بالحصائل محل الاهتمام، فإن "التقدير المبسط للأثر"، الذي يقارن الحصائل الناتجة بين عينات عشوائية للمشاركين وغير المشاركين، سينتج عنه تقدير متحيز فيما يتعلق بأثر البرنامج (يقدم المربع ٣-٢ مثالاً في هذا الخصوص). فعلى سبيل المثال، قد يستهدف مشروع ما فقراء؛ لذا فإن المستفيدين من المشروع يكونون أكثر فقرًا من غير المستفيدين، وهذا ليس بسبب فشل المشروع في حد ذاته ولكن لأنه نجح في استهدافه. ولمعرفة ما إذا كان للمشروع تأثيرات، فمن الضروري مقارنة ما حدث بحصائل تتعلق بمجموعة مماثلة من الأسر الفقيرة التي لم تشارك في المشروع. ويُطلق على المقاربة المُتبعة للتعامل مع تحيز الاختيار والعوامل الممتزجة الأخرى اسم *إستراتيجية التحديد*. ويعد امتلاك إستراتيجية تحديد قوية الفكرة الرئيسية التي ترتكز عليها الطرق ذات الصلة التي يتم تناولها في هذا الكتاب. وقد يتعلق الاختيار *بعناصر قابلة للملاحظة* أو متغيرات يتم ملاحظتها، وفي هذه الحالة يكون هناك العديد من الخيارات التي يمكن استخدامها للتحكم في تحيّز الاختيار. وقد يتعلق الاختيار أيضًا بعناصر غير قابلة للملاحظة، والتي تتمثل في خصائص (سلوكية غالبًا) لا تتوفر بشأنها بيانات. وإذا كانت الحالة الأخيرة هي الواقعة، فقد تكون هناك صعوبة أكبر في تنفيذ إستراتيجية تحديد صالحة.

٢. *التداخل أو الانتشار:* يُفترض أن تكون مجموعة المقارنة مشابهة لمجموعة المعالجة، باستثناء أن الأولى لا تتعرض للبرنامج المُنفَذِ. ومع ذلك، لا يتمتع القائمون بالتقييم في الواقع بالقدرة على التحكم الكامل فيما يحدث في مجموعة المقارنة. فقد تكون هناك وكالة أخرى تنفذ برنامجًا آخر في المنطقة الخاضعة للمقارنة، مما يؤثر على الحصائل محل الاهتمام. ومن هنا، يتعين فهم *التداخل* الحادث في مجموعة المقارنة حتى يتسنى معالجته. وللمساعدة في التخفيف من هذه المخاطر، من المفيد تضمين عملية جمع بيانات عن التدخلات السابقة والجارية الأخرى في مجال الدراسة. بالإضافة إلى ذلك، قد تنتشر بعض التدخلات، وهو أمر ينجم مثلاً عن توفير المعلومات، على نحو مستقل خارج موقع التدخل. وإذا كان من المحتمل حدوث مثل هذا الانتشار داخل أي من المواقع، فيجب وضع ذلك في الاعتبار عند أخذ عينات.

٣. *تأثير التداعيات:* يعد افتراض قيمة معالجة الوحدة المستقرة، والذي يشار إليه غالبًا بالاختصار الإنجليزي SUTVA، عنصرًا بالغ الأهمية في دراسات تقييم الأثر (Rubin 1980). ويعني هذا الافتراض بشكل أساسي أنه يمكن تمييز مجموعة المعالجة والمجموعة الضابطة، وأن هاتين المجموعتين لا تسجلان تأثيرات تفاعلية غير معروفة ناجمة عن التدخل. ولا يمكن اعتبار هذا أمرًا مُسلّمًا به، حيث قد يؤثر البرنامج بشكل غير مباشر على الأشخاص خارج المجموعة المستهدفة المقصودة من خلال تأثير التداعيات الإيجابية أو السلبية. فعلى سبيل المثال، قد يتسبب التدخل في إحداث تغيير على طلب السوق والأسعار لموقع أوسع نطاقًا، أو قد يؤدي إلى عوامل خارجية اجتماعية أخرى. وإذا كان من المتوقع أن تكون هذه التداعيات مهمة، فحينها يجب استيعابها بشكل واضح. وللقيام بذلك، قد يشتمل تصميم التقييم على مجموعة مُعالَجة، ومجموعة غير مُعالَجة ولكن مُعرضة للتداعيات، ومجموعة غير مُعالَجة وغير مُعرضة لتداعيات. وتعتمد طريقة تحديد هذه المجموعات على النطاق الجغرافي للتداعيات، وتحديدًا ما إذا كانت داخل أو بين مجموعات عنقودية أو كليهما. وإذا أثرت التداعيات على مجموعة المقارنة التي لم تتعرض في حد ذاتها للتداعيات، فإن هذا يُطلق عليه *التداخل الذاتي*. هذا التداخل الذاتي من شأنه أن يؤدي إلى حدوث تحيّز أو انحراف في تقدير الأثر. ولتجنب التداخل الذاتي، من الأفضل أن يكون هناك فصل جغرافي بين منطقتي المعالجة والمراقبة، بدلاً من أن تكونا متجاورتيْن جغرافيًا. وفي الوقت نفسه، ينبغي ألا تكونا بعيدتيْن عن بعضهما بشكل كبير للغاية حتى لا يترتب على ذلك عدم إمكانية إجراء المقارنة.

المربع ٣-٢: تحيّز الاختيار في إطار الحصول على المياه المُحسنة في نيبال

يعرض أول عمودين في الجدول متوسط خصائص الأسر التي تتمتع أو لا تتمتع بإمكانية الحصول على إمدادات مياه مُحسَّنة. وتتسم مجموعتا الأسر بأنهما مختلفتان للغاية. إذ تتمتع المجموعة التي تحصل على مياه محسنة بأنها أكثر ثراءً وأفضل تعليمًا ومن المرجح أنها تعيش في مناطق حضرية. لذا، فإن أي اختلاف يتم رصده في الحصائل، مثل إصابة الأطفال بمرض الإسهال، قد لا ينتج عن الماء النظيف، ولكن من الخصائص الأخرى التي تؤثر على كل من الحصول على الماء والحصيلة.

يعرض العمودان الأخيران متوسط الخصائص بعد إجراء المطابقة باستخدام طريقة مطابقة درجة الميل (طريقة يتم تناولها بالتوضيح في الفصل الخامس). ومن الممكن ملاحظة أن المجموعتين متشابهتان الآن إلى حد كبير، لذلك يمكن الآن أن يُعزى الاختلاف في الحصيلة إلى التدخل.

المتغير	مقارنة بسيطة		بعد المطابقة	
	تتمتع بصرف صحي مُحسّن	بدون صرف صحي مُحسّن	تتمتع بصرف صحي مُحسّن	بدون صرف صحي مُحسّن
توصيل المياه إلى المنزل عبر مواسير	٢٣٪	٥٪	٢٣٪	١٥٪
المنطقة الريفية	٥٢٪	٨٤٪	٥٣٪	٥٨٪
ربُّ الأسرة حاصل على شهادة تعليم ثانوي	٥٦٪	٣٠٪	٤٥٪	٤١٪
أرضية منزل إسمنتية	٢٩٪	٣٪	٣٠٪	٣٣٪
الشريحة السكانية الأكثر ثراءً	٥٤٪	٤٪	٥٢٪	٥٢٪

المصدر: Bose (2009).

طرق تقييم الأثر في سياق التحيّزات والتحديات

إذا كانت عملية التعيين والمشاركة في البرنامج تتم "عشوائيًا"، بحيث يكون ذلك مشابهًا لتجربة معملية، فحينها يمكن القضاء تمامًا على التحيّز الناجم عن اختيار المشاركين ذاتيًا وأماكن تنفيذ البرنامج. كما يمكن أن يضمن التعيين العشوائي الدقيق التحكم في التداخل، فضلاً عن الاستيعاب المناسب لتأثيرات التداعيات. ولهذه الأسباب، ينظر العديد من الأكاديميين إلى التجارب العشوائية المضبوطة (RCTs) باعتبارها "المعيار الذهبي" في دراسات تقييم الأثر (Athey and Imbens 2015). هناك اعتقاد سائد أن التجارب الطبيعية تقترب من تحقيق درجة المثالية التي تتصف بها التجارب العلمية، وذلك عندما يكون التعيين خارجيًا ويتم التحكم في التداخل ويمكن معالجة تأثيرات التداعيات.

يمكن أيضًا تجنب التحيز في الاختيار من خلال اللجوء إلى التصاميم غير التجريبية. وتعتمد جميع التصاميم غير التجريبية تقريبًا على فهم كيفية تعيين التدخل وكذلك على نمذجة محددات الاختيار. واستنادًا إلى نموذج الاختيار، يمكن التحكم في الاختلافات المرتبطة بالاختيار والموجودة قبل المعالجة أو المستقلة عنها، أو التخلص تمامًا من هذه الاختلافات. وبالتالي، يتمثل أحد الجوانب الرئيسية التي تتحكم في مدى الصلاحية أو الدقة فيما إذا كان نموذج الاختيار صالحًا. وفي ظل افتراضات معينة، تتحكم بعض التصاميم غير التجريبية في عملية الاختيار المتعلقة بالعناصر القابلة للملاحظة وتلك غير القابلة للملاحظة. ومع ذلك، فإنه بالنسبة لآخرين ليس من الممكن دائمًا تحديد ما إذا كان الاختيار المتعلق بعناصر غير قابلة للملاحظة موجودًا ويؤخذ في الاعتبار بشكل مناسب.

تجدر الإشارة إلى أنه لا توجد تقريبًا تقنية ممكنة دون رصد ملاحظات، ذات مستويات مختلفة، عن التعرض للتدخل (على سبيل المثال، مع أو بدون تدخل) خلال نفس الفترة الزمنية، وذلك بموجب افتراض قيمة معالجة الوحدة المستقرة. وتعتمد المنهجيات الأكثر دقة بشكل عام على رصد ملاحظات خاصة بالأفراد في ظل وجود التدخل وبدونه خلال الفترة التي سبقت التدخل وبعده. ويجب أن تعرض هذه الملاحظات لنفس الأفراد في كلتا الفترتين، وذلك على شكل بيانات طولية مُجمعة.

مقارنة تقييمات الأثر المعتمدة على العينة ذات الحجم الكبير بنظيرتها ذات الحجم الصغير

تتسم المقاربات الرئيسية لتقييم الأثر بأنها جميعًا عبارة عن تصميمات إحصائية لـ "عينة ذات حجم كبير" (large-n). وهذا يعني أنه يكون هناك عدد كبير من الملاحظات التي يمكن من خلالها إجراء اختبارات ذات دلالة إحصائية حول الاختلاف في الحصائل بين مجموعتيْ المعالجة والمقارنة. ويتم تحديد ما إذا كان حجم العينة كبيرًا بما يكفي لهذا الغرض من خلال حسابات القوة الإحصائية، والتي سيتم تناولها في الفصل السابع. وكما سيتم التناول في هذا الفصل، فإن عدد وحدات التعيين يمثل العنصر الأكثر أهمية بالنسبة للقوة الإحصائية مقارنةً بالعدد الإجمالي للملاحظات. وإذا لم يكن حجم العينة كبيرًا بما يكفي، فسيكون تقييم الأثر ضعيفًا، مما يعني أن احتمالية عدم العثور على تأثير له دلالة إحصائية ستكون كبيرة على الرغم من أن البرنامج يكون له حقًا أثر فعلي.

يدعم بنك التنمية الآسيوي وجِهات مانحة أخرى الأنشطة المختلفة التي من المحتمل أن تتطلب تصميمات قائمة على "عينة ذات حجم صغير إلى متوسط"، حيث لا يوجد العديد من المجتمعات المماثلة التي تتعرض للتدخل أو لا تتعرض له في نفس الوقت. تكمن أكثر مكونات المشروع وضوحًا في هذه الفئة في دعم إصلاح السياسات على المستوى الوطني أو داخل وكالة واحدة. وقد تمثل البنية التحتية الكبيرة، مثل الموانئ والطرق السريعة، أيضًا هذه الحالة، وذلك على الرغم من أنه قد تظل التصاميم القائمة على عينة ذات حجم كبير ممكنة كما هو موضح أدناه.

عندما تتوفر ملاحظات كافية بمرور الوقت، يمكن التحايل على إشكالية "صِغَرِ حجم العينة" (small-n) من خلال طريقتين:

- *الوحدات الضابطة الاصطناعية* تتيح إمكانية التوصل إلى تخمينات للآثار المرتبطة بالواقع المضاد عندما يكون هناك عدد أكبر من فترات الملاحظة مقارنةً بالوحدات المُعالجة (يتم تناول ذلك في الفصل الخامس).
- *مقاربات السلاسل الزمنية المتقطعة* قد تتيح انقطاعات في الاتجاهات بمرور الوقت عند القيام بالتدخلات لتحديد آثار البرنامج (يتم تناول ذلك أيضًا في الفصل الخامس).

في حالة عدم إمكانية استخدام أي من هاتيْن الطريقتيْن أو تصميم يعتمد على عينة كبيرة الحجم، فحينها يجب اتخاذ خيار حول ما إذا كان سيتم التخلي عن تقييم الأثر أو المضي قدمًا في مقاربة تعتمد على عينة ذات حجم صغير أو متوسط. ولا يتم تناول هذه المقاربات في هذا الكتاب. لكن، يمكن العثور على مزيد من المعلومات في المصادر التالية:

- من الممكن أن تستخدم التدخلات المبنية على عينة متوسطة الحجم، مثل تقديم الدعم لمجموعة من الوسطاء الماليين، التحليل *المُقارن النوعي*، وهي مقاربة كمية غير مُختبَرة نسبيًا للتحليل السببي. من خلال هذا التحليل، يتم البحث عن أنماط معينة في البيانات لتحديد الشروط الضرورية والكافية للسببية دون إجراء اختبارات ذات دلالة إحصائية (Ragin 2000).
- تعتمد المقاربات النوعية عينة صغيرة الحجم للاستدلال السببي على التحليلات المنهجية لنظرية التغيير، والتي تُسمى أحيانًا *الحصائل المرتبطة بالآلية والسياق* (White and Phillips, 2012).
- عادة ما يكون تقييم الأثر تجريبيًا ويهدف عادةً إلى تقليل الافتراضات الهيكلية. وتعد نمذجة المحاكاة بديلاً نظريًا وليس تجريبيًا لتقييم الأثر. فعلى سبيل المثال، يمكن نمذجة آثار إصلاح الاقتصاد الكلي والبنية التحتية الكبرى باستخدام تحليل التوازن العام القابل للحساب. وقد يتم طرح برامج جديدة كتغييرات في المتغيرات الخارجية ("الصدمات")، في حين قد تتطلب السياسات تغييرات في المعلمات أو حتى مواصفات النموذج (Bourguignon and de Silva 2003).

٤-٣ البعد الزمني للآثار

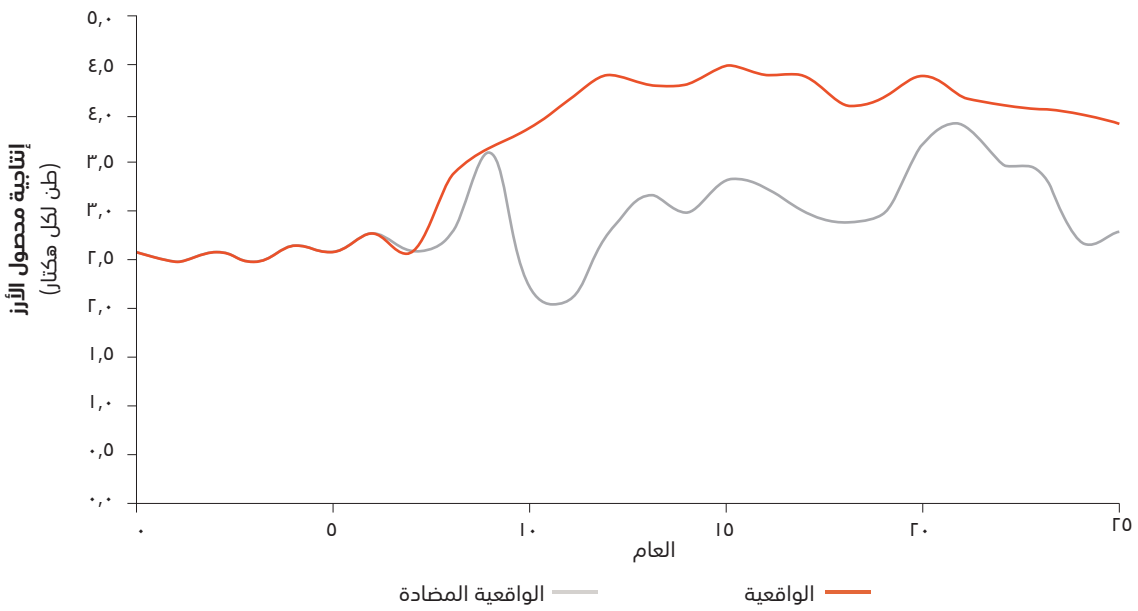
غالبًا ما تستغرق آثار التدخل وقتًا طويلاً حتى تتحقق، وهذا يفرض معوقات تتعلق بتحديد وقت محاولة تقييم تلك الآثار بالإضافة إلى ذلك، تتغير الآثار بمرور الوقت. وقد يكون هذا بسبب الفترة المستغرقة للإعداد للمشروع بُغية التعرف على كيفية تطبيق التدخل، أو قد يكون بسبب اختلاف المشاركين الأوائل عن أولئك الذين ينضمون إلى البرامج لاحقًا. فعلى سبيل المثال، يعرض الشكل (٣-٢) إنتاجية محصول الأرز في منطقة تستفيد من أحد مشاريع الري. ويوضح الخط الأحمر حجم الإنتاجية الفعلية للمحصول. خلال الأعوام السبعة الأولى، بينما كان الإنشاء جاريًا، بلغ متوسط إنتاجية المحصول ٢,٦ طن للهكتار. وعلى مدى الأعوام الخمسة التالية، تم الانتهاء من القنوات الثانوية والقنوات من الدرجة الثالثة، وبالتالي فقد ازداد متوسط إنتاجية المحصول في منطقة مستجمعات المياه مع اتصال المزيد من المزارعين بالنظام، لتصل إلى ٤,٣ أطنان للهكتار بحلول العام الثاني عشر.[٣]

يوضح الخط الرمادي الواقع المضاد لما كان سيحدث في حالة عدم التدخل. ويُرجى ملاحظة أن هذا هو الواقع المضاد الحقيقي. وإذا تم تخمين معدلات الإنتاجية في حالة الواقع المضاد باستخدام مجموعة مقارنة، فيجب أن تكون متشابهة ولكن ليست متطابقة في فترة ما قبل المشروع.

[٣] تمثل إنتاجية محصول الأرز مجرد حصيلة واحدة يجب أخذها في الاعتبار، حيث إن جزءًا كبيرًا من الأثر الذي يُحدثه الري يكون من خلال تأثيره على أنماط زراعة المحصول (التركيب المحصولي). ويتعين أيضًا إدراج صافي دخل المزرعة ودخل الأسرة كحصائل.

يتطلب تحديد توقيت جمع البيانات دراسة متأنية، حيث يمكن أن يترتب عليه تحديد نطاق تقييم الأثر ونتائجه. وقد تم إكمال مشروع الري الوارد في الشكل (٣-٢) في العام الثامن، مع استمرار العمل الممول محليًا بهدف توسيع التوصيلات في منطقة مستجمعات المياه. ومن شأن القيام بقياس الأثر في العام الثامن أن يقلل من شأنه، نظرًا لعدم قيام العديد من المزارعين بالتوصيل بعد. و بالتالي، فإن تقييم الأثر في عام انتهاء المشروع سيكون سابقًا لأوانه. يتطلب تصميم دراسة لتقييم الأثر فهم الجوانب الزمنية لنظرية التغيير، ومتى يمكن توقع التأثيرات على الحصائل محل الاهتمام بشكل واقعي. فقد يسفر برنامج تدخل لعلاج حالات إصابة بالإسهال عن الوصول إلى أثر سريع على معدلات الإصابة، ولكن إجراء إصلاح في نظام الحوكمة قد يستغرق عدة دورات انتخابية قبل أن يصبح الأثر قابلاً للقياس على الإطلاق.[٤]

الشكل (٣-٢): معدلات الإنتاجية الواقعية والواقعية المضادة الناجمة عن أحد مشاريع الري



المصدر: محاكاة للمؤلف.

مع ذلك، في واقع الحال، لا يمكن معرفة الأثر الفعلي قبل إجراء المسوحات التي تهدف إلى تقييم الأثر. وبالتالي، فإن ذلك يتطلب مزيجًا من الأحكام القائمة على النظريات والخبرة المهنية. وتتمثل المسألة بالغة الأهمية التي يجب مراعاتها في هذا الخصوص فيما إذا كان قد مضى وقت كافٍ على التدخل ذي الصلة حتى تتجلى تأثيراته محل الاهتمام. فقد تؤدي محاولة قياس الأثر قبل الأوان إما إلى استنتاجات خاطئة مفادها أن الأثر محدود، أو قد تحد من إمكانية تقييم تأثيرات ذات مغزى. وحتى عند مرور وقت كافٍ بما يتيح لأن تكون التأثيرات قابلة للقياس، فقد يكون هناك تباين في التأثير الذي لحق بالمشاركين الأوائل والجمهور المستهدف بشكل عام.

[٤] ردًا على سؤال حول أثر الثورة الفرنسية في عام ١٧٨٩، أجاب تشو إنلاي، رئيس وزراء جمهورية الصين الشعبية آنذاك قائلاً: "من السابق لأوانه الحديث عن هذا الأمر".

الوقت المناسب للتخطيط لتقييم الأثر

عادةً ما يتم إجراء تقييمات الأثر في نهاية المشروع أو بالقرب من نهايته. وهذا لا يعني أنه يمكن تأجيل كل الأفكار المتعلقة بتقييم الأثر حتى سنوات الانتهاء هذه. وتكون خطط تقييم *الأثر السابقة*، والتي يتم إجراؤها قبل الوصول إلى الفعالية الميدانية للتدخل، أقوى لأنها (١) تتيح جمع بيانات خط الأساس المناسبة من كل من المشروع/التدخل (العلاج) ومجالات المقارنة؛ (٢) تأخذ التعيين العشوائي في عين الاعتبار. ويعد التوزيع العشوائي مقاربة ضرورية يتم إجراؤها في وقت سابق للمشروع، حيث يكون للتعيين العشوائي إنعكاسات على كيفية تنفيذ المشروع، ومن ثم لا يمكن إجراؤه في وقت لاحق له.

يشير الواقع السائد في هذا الخصوص إلى أن الطلب على تقييم الأثر قد ينشأ فقط بعد بدء المشروع أو البرنامج، وربما حتى بعد فوات فرصة جمع بيانات خط الأساس ذات صلة. وبالرغم من أن هذه الظروف بعيدة عن المثالية، فإن تصميمات تقييمات *الأثر اللاحقة* تكون ممكنة، ومن ثم يجب اختيار أفضل تصميم ممكن في ظل هذه الظروف.

٣-٥ وحدة التعيين والمعالجة والتحليل

يتطلب التخطيط لتصميم تقييم الأثر فهمًا واضحًا لوحدة التعيين والمعالجة والتحليل:

- *وحدة التعيين* هي أقل وحدة (جغرافية عادةً) يتم فيها اتخاذ قرار التدخل، والتي عادة ما ستتمثل في بلدية أو منطقة أو منطقة فرعية أو وكالة أو مجتمع أو مجموعة أو شركة. ونادرًا ما يتم تحديد أنواع التدخل التي يدعمها بنك التنمية الآسيوي أو جهات مانحة أخرى على مستوى الأسرة.
- *وحدة المعالجة* هي الوحدة التي يتم تقديم المعالجة لها، والتي تكون أقل مستوى من وحدة التعيين. فعلى سبيل المثال، سيوفر برنامج التمويل متناهي الصغر الدعم من خلال مؤسسات التمويل متناهي الصغر (وحدة التعيين)، غير أن وحدة المعالجة تتمثل في الشركات التي تحصل على قروض.
- *وحدة التحليل* هي الوحدة التي يتم قياس الحصائل بها. ومرة أخرى، قد تكون هذه الوحدة أقل مستوى من وحدة التعيين أو المعالجة.

وفي مثال مشروع التمويل متناهي الصغر، قد تكون الحصيلة هي الأجور المدفوعة لموظفي المؤسسة، ومن ثم، فإن العاملين في المؤسسة يمثلون وحدة التحليل.

ليس من المعتاد أن تكون وحدة التعيين هي نفسها وحدة المعالجة ووحدة التحليل. وإذا لم تكن هذه الوحدات متطابقة، فسيلزم وضع تصميم عنقودي، مع استخدام طريقة أخذ عينات عنقودية لجمع البيانات. ولهذه الوسيلة الواقعية إنعكاسات ملموسة على جمع البيانات. يتناول الجدول (٣-١) بعض الأمثلة حول مدى الاختلاف الذي قد يوجد بين وحدات التعيين والمعالجة والتحليل.

الجدول (٣-١): أمثلة على وحدات تعيين ومعالجة وتحليل

التدخل	التعيين	المعالجة	التحليل
تحسين الصرف الصحي	القرية	القرية والمدارس والأسر	الأسر والأطفال
تطوير الطرق السريعة الرئيسية	المدينة	المدينة	الشركة
التدريب المهني	المدرسة	المدرسة	الطالب

المصدر: المؤلفان.

يبدو أن مشاريع البنية التحتية الكبيرة تمثل صعوبة خاصة للتصميمات الإحصائية التي تعتمد على عينة ذات حجم كبير (large-n)، لأنه في بعض الحالات مثل إعادة تأهيل الموانئ أو بناء طريق سريع وطني، يكون حجم العينة المُعالَجة (n) مساويًا لـ ١. ومع ذلك، من الممكن إنشاء مجموعات معالجة ومقارنة عبر اعتبار تلك المجتمعات (القرى والمناطق الفرعية والبلدات وغيرها) بمثابة وحدة التعيين. وفي المثال الوارد في الجدول (٣-١) الخاص بتحسينات الطرق السريعة، يتم اعتبار المدن الواقعة على طول الطرق السريعة الأربعة بمثابة مجموعة المعالجة، مع اتخاذ المدن غير الواقعة على الطرق السريعة والبالغ عددها ١٨ مدينة كمجموعة مقارنة. وعلى النقيض من ذلك، غالبًا ما تكون التدخلات التعليمية أكثر قابلية لاستخدام تصميمات تقييمات الأثر المعتمدة على عينة كبيرة الحجم، حيث تمثل المدرسة وحدة شائعة للتعيين، كما في حالة التدريب المهني في الجدول (٣-١).

يمكن إيجاز ما سبق من خلال: (١) عادة ما تكون وحدة المعالجة في نفس المستوى أو في مستوى أقل من وحدة التعيين، وعادة ما تكون وحدة التحليل في نفس المستوى أو في مستوى أقل من وحدة المعالجة؛ و(٢) قد تكون هناك وحدات متعددة في كل مستوى، مثل المعالجات المختلفة للقرية والمدارس والأسر كما في مثال الصرف الصحي في الجدول (٣-١).

٣-٦ القياسات المختلفة للأثر - لمن تُوجَّه الآثار؟

يمكن أن تقدم دراسات تقييم الأثر قياسات مختلفة لتأثيرات التدخل، وذلك اعتمادًا على المجموعة السكانية (الإحصائية) التي تخضع للتقييم. ومن المهم فهم ذلك، حيث لا يمكن لجميع المنهجيات تقدير جميع القياسات.

- *متوسط تأثير المعالجة (ATE):* متوسط تأثير المشاركة في البرنامج على المجموعة السكانية المؤهلة بالكامل.
- *تأثير النية للعلاج (ITT):* متوسط تأثير التعرض للبرنامج، على سبيل المثال، على جميع أولئك الذين يعيشون في منطقة مؤهلة للمشاركة في البرنامج.
- *متوسط تأثير المعالجة على المعالَج (ATT):* متوسط التأثير على أولئك الذين يشاركون بالفعل في التدخل (يختارون الامتثال له أو استخدامه).
- *متوسط تأثير المعالجة على غير المُعالَج (ATU):* متوسط التأثير المحتمل على أولئك الذين لم يشاركوا في المعالجة في حالة تمت معالجتهم. ويعد هذا مقياسًا مناسبًا لفهم الآثار المحتملة لتوسيع نطاق البرنامج.
- *متوسط تأثير المعالجة المحلي (LATE):* متوسط التأثير على مجموعة فرعية من المجموعة المستفيدة، وعادة ما يكونون على عتبة الأهلية. ويترتب على بعض تصميمات تقييمات الأثر متوسط تأثير معالجة محلي بدلاً من متوسط تأثير معالجة بصفة عامة.

ترتبط النية للعلاج (ITT) ومتوسط تأثير المعالجة على المُعالَج (ATT) بمعدل المشاركة:

$$ITT = \frac{\text{إجمالي التأثير}}{\text{عدد المستفيدين المستهدفين}} \qquad (٢-٣)$$

$$= \frac{\text{إجمالي التأثير}}{\text{عدد المشاركين}} \; x \; \frac{\text{عدد المشاركين}}{\text{عدد المستفيدين المستهدفين}}$$

= متوسط تأثير المعالجة على المعالَج x (ATT) معدل المشاركة (PR)

ومن ثم، يتضح أن

$$ITT \leq ATT \qquad (٣-٣)$$

من هنا يمكن استنتاج أن متوسط التأثير على المجموعة السكانية المستهدفة لن يكون بالضرورة أكبر من متوسط التأثير على أولئك الذين يشاركون بالفعل. وتؤدي معدلات المشاركة المنخفضة إلى إحداث فجوة بين الاثنين. وقد يكون للتدخل أثر كبير جدًا على أولئك الذين يشاركون فعليًا، لكن القليل فقط هم من يشاركون بالفعل، ومن ثم يكون تأثير النية للعلاج (ITT) منخفضًا جدًا.

يتم قياس متوسط الأثر الذي تم تحقيقه من خلال التعرض لتأثيرات البرنامج حتى تاريخه عبر مقياس النية للعلاج (ITT)، في حين أن متوسط تأثير المعالجة المحلي غالبًا ما يوفر تأثيرًا هامشيًا عبر توسيع نطاق التدخل. ويوفر متوسط تأثير المعالجة على المُعالَج قياسًا للأثر لأولئك الذين يشاركون بالفعل، وغالبًا ما يكون هو المقياس المُفضل للباحثين المهتمين بقياس الأثر حتى تاريخه. على الجانب الآخر، قد يكون متوسط تأثير المعالجة على غير المُعالَج أكثر ملاءمة لصناعة قرارات بشأن السياسات ذات الصلة بتوسيع نطاق البرنامج.

ومع افتراض وجود أشخاص منطقيين، إلى جانب وضع برنامج هادف للمواقع والمجموعات السكانية التي تتجاوب بشكل أفضل مع التدخلات، وأن أولئك الذين يختارون أنفسهم في التدخلات يحققون مزايا أكثر من أولئك الذين لا يفعلون ذلك، فحينها يجب ملاحظة العلاقة التالية بين مؤشرات الأثر:

$$ATT > ATE > ATU \qquad (٤-٣)$$

إذا لم ينشأ هذا النمط، فقد يشير ذلك إما إلى وجود خلل في إستراتيجية التحديد المتعلقة بتقييم الأثر أو أن البرنامج به عيوب كبيرة في الاستهداف أو التنسيب.

٣-٧ الصلاحية الداخلية والخارجية

تتمتع أي دراسة *بالصلاحية الداخلية* إذا كانت تقديرات هذه الدراسة صالحة بالنسبة للعينة التي تم تقديرها. وتعد الصلاحية الداخلية بمثابة دالَّة تشير إلى دقة تصميم التقييم. وبالنسبة لتقييم الأثر، فإن أحد الجوانب بالغة الأهمية للصلاحية الداخلية يكمن في قوة إستراتيجية التحديد، أي المقاربة المُستخدمة لمعالجة مشكلة السببية، بما في ذلك التحيّز في الاختيار. وهناك عوامل أخرى تؤثر أيضًا على الصلاحية الداخلية، مثل إستراتيجية أخذ العينات المُستخدمة لجمع البيانات في مناطق المعالجة والمراقبة، فضلاً عن جودة البيانات التي تم جمعها.

بينما تتعامل الصلاحية الداخلية مع كيف يمكن للتقييم أن يتخلص من التحيزات المحتملة، تمثل *الصلاحية الخارجية* المدى الذي يمكن من خلاله تعميم نتائج الدراسة. ولا تكون دراسة واحدة كافية لتقديم أحكام عامة من على شاكلة: "التأمين ضد الأضرار الناشئة عن أحوال جوية والمرتبط بمؤشرات لا يصلح" أو "الشراكة بين القطاعين العام والخاص تعد من أكثر الوسائل المتميزة بفعالية التكلفة المخصصة لإنشاء بنية تحتية ضخمة". ولإصدار مثل هذه الأحكام، يلزم الاعتماد على مجموعة أوسع من الأدلة التي تشمل العديد من الدراسات. ومع ذلك، فإن هذا لا يعني أنه لا يمكن استخلاص دروس عامة من دراسة واحدة. ففي كثير من الأحيان، يمكن نقل الدروس المُستفادة على نحو معقول وتطبيقها في سياق مماثل.

تجدر الإشارة إلى أنه يتم تعزيز الصلاحية الخارجية (١) من خلال وجود نظرية قوية للتغيير مع إجراء تحليل بامتداد السلسلة السببية؛ (٢) عندما تكون عينة الدراسة ممثلة للمجتمع الإحصائي الذي يستهدفه المشروع/البرنامج (والتعميمات التي أُجريت على مجتمعات إحصائية مماثلة)؛ (٣) عندما يكون التدخل الذي تم تقديره إما مطبقًا على نطاق واسع حاليًا أو يتمتع بإمكانية واسعة للتطبيق؛ و(٤) عندما يكون التحليل وثيق الصلة بالمنطق الاقتصادي الذي ينطبق على نطاق أوسع من سياق أو محيط التدخل. وبشكل عام، غالبًا ما يمثل إقرار الصلاحية الخارجية تحديًا أكبر بالنسبة للمقاربات التي ترتكز على التجارب العشوائية المضبوطة، حيث يؤدي التطبيق التجريبي إلى تدخلات قد يكون لها صلة محدودة بالعالم الواقعي، في حين أن الصلاحية الداخلية تمثل تحديًا أكبر للمقاربات غير التجريبية، حيث تكون عناصر التحكم الخاصة بالعوامل المثيرة للبس غير مكتملة.

المراجع

Athey, S. and G. W. Imbens. 2017. Chapter 3: The Econometrics of Randomized Experiments. In E. Duflo and A. Banerjee, eds. *Handbook of Field Experiments*. Volume 1, 1st Edition. Amsterdam: North Holland.

Bourguignon, F. and L. A. Pereira da Silva, eds. 2003. *The Impact of Economic Policies on Poverty and Income Distribution: Economic Techniques and Tools*. Washington, DC: World Bank and Oxford University Press.

Bose, R. 2009. The Impact of Water Supply and Sanitation Interventions on Child Health: Evidence from DHS Surveys. Paper prepared for the Bi-Annual Conference on Impact Evaluation. Colombo, Sri Lanka. 22–23 April 2009.

Bowen et al. 2013. Sustained Improvements in Handwashing Indicators More than 5 Years after a Cluster-Randomised, Community-Based Trial of Handwashing Promotion in Karachi, Pakistan. *Tropical Medicine and International Health*. 18 (3). pp. 259–267.

Ragin, C. C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Rubin, D. B. 1980. Comment on: Randomization Analysis of Experimental Data in the Fisher Randomization Test by D. Basu. *Journal of the American Statistical Association*. 75. pp. 591–593.

White, H. and D. Phillips. 2012. Addressing Attribution of Cause and Effect in Small-n Impact Evaluations: Towards an Integrated Framework. 3ie Working Paper 15. Delhi.

مراجع إضافية للاطلاع

ADB. 2011. *A Review of Recent Developments in Impact Evaluation*. Manila. https://www.adb.org/sites/default/files/publication/28622/developments-impact-evaluation.pdf.

Caliendo, M. and R. Hujer. 2005. The Microeconometric Estimation of Treatment Effects – An Overview. Discussion Paper No. 1653. IZA. Bonn, Germany. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.421.6476&rep=rep1&type=pdf.

Ravallion, M. 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. *The World Bank Economic Review*. 15 (1). pp. 115–140. http://documents.worldbank.org/curated/en/410881468180273314/The-mystery-of-the-vanishing-benefits-an-introduction-to-impact-evaluation.

الفصل ٤
التجارب العشوائية المضبوطة
Randomized Controlled Trials

رسائل أساسية

- تضمن التجارب العشوائية المضبوطة (RCTs) التي تُجرى بشكل صحيح توفر أولئك الذين خضعوا لتدخلات وأولئك الذين لم يخضعوا لها على سمات وخصائص متوازنة، بحيث تكون الاختلافات ناتجة عن التدخل فقط.
- تمثل معظم التجارب العشوائية المضبوطة كتصميمات عنقودية، حيث تحتوي وحدة التعيين على وحدات مُعالَجة متعددة.
- يمكن إجراء التعيين العشوائي عبر عدة طرق، بعضها يترتب عليه فقط تغيير التسلسل أو عتبة الأهلية أو حوافز الاستفادة من التدخل، بدلاً من التطبيق على المشروع بشكل عام.
- يمكن أن تحتوي التجارب العشوائية المضبوطة على مجموعات مختلفة ومتبدلة من التدخلات للتحقق من التأثيرات الناجمة عن التفاعل وإجراء المقارنات.
- يلزم الإشراف الدقيق لضمان دقة التجارب العشوائية المضبوطة التي تُجرى ميدانيًا.

٤-١ مقدمة

تنطوي التجربة العشوائية المضبوطة (RCT)، والتي تُسمى أحيانًا التقييم العشوائي أو التصميم التجريبي، على التعيين العشوائي لأعضاء من المجتمع الإحصائي (السكاني) المؤهل للانضمام إلى واحدة أو أكثر من "مجموعات المعالجة" التي تحظى بالتدخل، وكذلك إلى "المجموعة الضابطة" التي لا تحظى بالتدخل أو تدخل مرجعي مقارن أو تدخل علاجي وهمي، في بعض الحالات.[1]

يضمن التعيين العشوائي، من خلال توفر الحجم الكافي للعينة، إحداث توازن. وهذا يعني أن الخصائص المتوسطة بين المجموعة المعالجة والمجموعة الضابطة تكون متماثلة في المتوسط عند خط الأساس. وسوف ينطبق هذا على كل من الخصائص القابلة للملاحظة وتلك غير القابلة للملاحظة على حد سواء.

[1] في التجارب السريرية، قد تتلقى مجموعة المعالجة دواءً وهميًا. صولأسباب أخلاقية وعملية، قد لا يكون هذا خيارًا في العديد من التدخلات الاجتماعية والاقتصادية.

ولن يكون هناك أي تحيز في الاختيار، نظرًا لأن الاختيار في عملية التعيين تم إجراؤه عشوائيًا ولم يكن مقصودًا ومتعمدًا. ويمكن حساب الأثر على أنه الاختلاف أو الفرق في الحصائل بين المجموعة المعالجة والمجموعة الضابطة عند خط النهاية. ويُفضل استخدام تقدير الاختلاف فى الاختلافات (DiD) (يتم تناوله في الفصل الخامس)، والذي يقارن الفرق أو الاختلاف بين التغييرات التي تطرأ بمرور الوقت بين المجموعة المعالجة والمجموعة الضابطة، وذلك لأنه يقلل الفروق الناتجة عن أخطاء أخذ العينات.

ينبغي عدم الخلط بين التعيين العشوائي وأخذ العينات العشوائية. إذ يشير أخذ العينات العشوائية إلى كيفية الحصول على عينة من مجتمع إحصائي واحد أو أكثر. في حين يشير التعيين العشوائي إلى كيفية تعيين الأفراد أو المجموعات إما لمجموعة معالجة أو مجموعة ضابطة. وعادة ما تستخدم التجارب العشوائية المضبوطة كلاً من التعيين العشوائي وأخذ العينات العشوائية، وذلك نظرًا لعدم الحاجة إلى مجتمع المعالجة بالكامل لتحليل تقييم الأثر.

٤-٢ لماذا يتم أخذ العينات عشوائيًا؟

ثمة أسباب منطقية واضحة وراء استخدام التعيين العشوائي. فإذا تم أخذ عينة ممثلة من مجتمع إحصائي، فإن القيمة المتوقعة لمتوسط أي خاصية لهذه العينة هي متوسط المجتمع الإحصائي الحقيقي. وكلما زاد حجم العينة التي يتم الحصول عليها، زادت احتمالية أن يكون متوسط العينة قريبًا من متوسط المجتمع الإحصائي التي تم أخذها منه.

يترتب على ذلك أنه إذا تم أخذ عينتين أو أكثر من نفس المجتمع الإحصائي، فيجب أن يكون متوسط الخصائص لكل عينة متماثلاً بشكل أساسي، حيث من المتوقع أن يكون لكل منهما متوسط الخصائص للمجتمع الإحصائي ككل. وهذا يتضح في الجدول (٤-١)، الذي يوضح متوسط خصائص الأسر في عينتين، إحداهما تسمى مجموعة المعالجة والأخرى تُسمى المجموعة الضابطة. ويزداد تشابه العينتيْن مع زيادة حجم العينة، مع عدم وجود فروق ذات دلالة إحصائية بمجرد أن تصبح العينتان كبيرتيْن بدرجة كافية. وهذا يعني وجود توازن بين العينتيْن.

الجدول (٤-١): التشابه بين عينتيْن مأخوذتيْن من المجتمع الإحصائي نفسه

حجم العينة	مناطق ريفية (٪)			سنوات التعليم			حجم الأسرة		
	المعالجة	الضابطة		المعالجة	الضابطة		المعالجة	الضابطة	
١	١٠٠	٠	*	١٢,٠	٩,٠	*	٩,٠	٥,٠	*
١٠	٧٠	٨٠	*	٦,٤	٥,٨	*	٦,٤	٦,٧	
٢٥	٧٢	٦٠	*	٥,٨	٥,٣	*	٦,٤	٦,٥	
١٠٠	٦٤	٦١		٦,٠	٥,٥		٦,٧	٦,٥	
١٠٠٠	٦٦	٦٤		٥,٢	٥,٤		٦,٥	٦,٥	

ملاحظات: حجم العينة لكل عينة. * يشير إلى أن الفرق كبير عند مستوى ٥٪.

المصدر: المؤلفان.

يوضح الجدول (٤-١) أن العينتين متوازنتان فيما يتعلق بالخصائص التي يمكن ملاحظتها. ويتعين أن يترتب على الاختيار العشوائي المناسب أيضًا توازن فيما يتعلق بالخصائص غير القابلة للملاحظة أيضًا، على الرغم من أنه لا يمكن التحقق من ذلك بشكل مباشر. وبمجرد إحداث التوازن عن طريق التعيين العشوائي في مجموعة المعالجة والمجموعة الضابطة، فإنه يترتب على ذلك ما يلي: بما أن العينتيْن تتمتعان بنفس الخصائص، فإن أي فرق (أو اختلاف) في الحصائل بعد التدخل يجب أن يُعزى إلى التدخل.

ثمة نتيجتان مهمتان تترتبان على هذه المناقشة:

(١) يجب أن تكون مجموعة المعالجة والمجموعة الضابطة من نفس المجتمع الإحصائي؛ على ألا يتمثل ذلك المجتمع في المجتمع الإحصائي ككل، ولكن ذلك الذي يتحدد من خلال معايير الأهلية الخاصة بالتدخل أو المجتمع الإحصائي في منطقة الاستجماع الخاصة بالتدخل.

(٢) يجب أن يكون حجم العينة كبيرًا بما يكفي ليتم التيقن على نحو معقول من أن العينتين ستتمتعان بمتوسط الخصائص نفسه.

يضمن التعيين العشوائي أن يتمتع جميع المستفيدين المحتملين بفرصة متساوية في أن يتم تضمينهم في التدخل. وحيث إن ندرة الموارد تعني أنه لا يمكن تقديم الخدمات لكافة أفراد المجتمع الإحصائي، فإن الاختيار العشوائي يعد وسيلة شفافة وعادلة لتخصيص هذه الموارد.

تجدر الإشارة إلى أن الاختيار العشوائي يتمتع بميزة ثانوية أخرى ربما تكون في الغالب بنفس أهمية تفوقها على الصعيد النظري. وحتى يتسنى تنفيذ التدخل اعتمادًا على الاختيار العشوائي الفعال للعينات، يجب أن يشارك القائمون على تقييم الأثر عن كثب في تنفيذ هذا التدخل، وذلك نظرًا لأنهم يشاركون في تعيين المجموعة المعالجة. ولهذه المشاركة فوائد لكل من موظفي البرنامج والقائمين على التقييم على حد سواء. وفي هذا الصدد، يرى جلينيرستر وآخرون أنها تعزز من أهمية التقييم ومدى ملاءمته، كما توفر فرصًا لعملية تقييم الأثر للمساعدة في تحسين تصميم المشروع وإدارته (Glennerster et al. 2017).

٤-٣ أنواع تصميمات التجارب العشوائية المضبوطة

هناك العديد من التصاميم المختلفة للتجارب العشوائية المضبوطة. وترتبط الاختلافات بـ (١) مستوى التعيين، (٢) المقاربات المختلفة المتبعة في التعيين العشوائي، و(٣) نوع مجموعات العلاج التي يتم تقديرها. وتعتمد الاختيارات التي يتم إجراؤها من بين هذه التصاميم المختلفة على تصميم التدخل، وخاصةً قواعده التشغيلية المتعلقة بتحديد من المؤهل للبرنامج أو كيفية تحديد منطقة الاستجماع.

بالنسبة لجميع المقاربات، من المهم التأكد من تطبيق آلية مناسبة لإخفاء التعيين أو التوقيت. بعبارة أخرى، يرى الكاتبان كيم وشين أنه يجب ألا يكون لدى المشاركين في تنفيذ المعالجة علم مسبق بعملية التعيين، وهو ما قد يسمح بوجود تحيز أو تلاعب في عملية التعيين هته (Kim and Shin 2014).

مستوى التعيين: التجارب العشوائية المضبوطة البسيطة مقارنةً بنظيرتها العنقودية

تشير وحدة التعيين في التجربة العشوائية المضبوطة إلى الوحدة المُستخدمة لاختيار من يتم علاجه. في التجربة العشوائية المضبوطة البسيطة، تكون وحدة التعيين هي نفس وحدة المعالجة والقياس. وقد يكمن أحد الأمثلة على ذلك في برنامج تطوير أعمال لمؤسسات صغيرة ومتوسطة الحجم، حيث يتم تعيين أو اختيار المؤسسات المؤهلة عشوائيًا للمجموعة المعالجة والمجموعة الضابطة. ويمكن أن تشتمل الحصائل، على مستوى الشركة، رقم المبيعات والربحية والتوظيف.

لأسباب عملية وأخلاقية، غالبًا ما يُستخدم تصميم التجربة العشوائية المضبوطة العنقودية، حيث تحتوي وحدة التعيين على العديد من وحدات المعالجة. ومن الناحية العملية، ستكون هناك جدوى أكبر عندما يتم التعيين العشوائي لخدمة ما تتعلق ببنية تحتية مجتمعية مشتركة، مثل الكهرباء أو إمدادات المياه على مستوى المجتمع أو الكتلة السكنية، وليس على مستوى الأسرة. فعلى سبيل المثال، تخدم كهرباء الريف والطرق المجتمعات وتوفر علاجًا للعديد من مشكلاتهم، بالرغم من أنه يمكن قياس الفوائد ذات الصلة على مستوى الشركة أو الأسرة. وقد تنشأ أسئلة لوجستية وأخلاقية من ظهور أي من أشكال عدم الإنصاف في حالة كان التعيين على مستوى الأسرة أو الفرد، مثل أن يتم تعيين بعض أعضاء صف مدرسي ولا يتم تعيين آخرين.

تساعد تصميمات التجارب العشوائية المضبوطة العنقودية أيضًا في احتواء تأثيرات التداعيات والتداخل. وغالبًا ما تنتشر المعرفة بسبل المعالجة داخل المجتمع. وإذا كان التعيين على مستوى الأسرة أو الفرد، فإن هذه المعرفة يترتب عنها تداعيات وتخلق تأثيرًا غير مباشر قد يغير سلوك غير المُعالَج، مما يؤدي إلى تحيز أو انحراف في التجربة إذا كانت المجموعة الضابطة مأخوذة من جيران غير معالَجين. ومن الممكن أن تساعد التجارب العشوائية المضبوطة العنقودية في إنشاء وحدات تعيين كبيرة ومميزة بشكل كافٍ بحيث يتم تقليل هذه التداعيات.

في التجربة العشوائية المضبوطة العنقودية، تكون وحدة التعيين أعلى من وحدة المعالجة أو القياس. فعلى سبيل المثال، قد يصلح برنامج تطوير أعمال في مدن معينة، حيث تكون جميع الشركات في تلك المدن مؤهلة لهذا البرنامج. وسيتم اختيار هذه المدن بشكل عشوائي من قائمة المدن المؤهلة، مع اعتبار تلك التي لم يتم اختيارها (أو عينة عشوائية من تلك التي لم يتم اختيارها) مجموعة ضابطة.

تحديد القوة الإحصائية للتصميم مرهون، إلى حد كبير، بعدد المجموعات في الدراسة بدلاً من عدد الوحدات المُعالَجة. وهذا يعني أن البرنامج النموذجي يجب أن يغطي عددًا كبيرًا، وبشكل معقول، من المدن التي ستخضع للمعالجة للحصول على دراسة قوية بشكل كافٍ (يقدم الفصل السابع مزيدًا من التفاصيل حول حسابات القوة الإحصائية لتحديد أحجام العينة).

مقاربات مختلفة للتعيين العشوائي

ثمة العديد من المقاربات المتبعة في التعيين العشوائي. ويتعين تحديد المقاربة المناسبة بناءً على القواعد التشغيلية للبرنامج.

- عندما تكون هناك طلب مفرط (زيادة وإفراط في حجم الاشتراك) على برنامج ما أو عندما يتجاوز المجتمع الإحصائي المؤهل ما يمكن تقديمه عبر الموارد المتاحة، فحينها يمكن استخدام أي من وسائل الاختيار العشوائي، مثل القرعة، لتحديد أي من المتقدمين المؤهلين سيتم تضمينه وأيهم سيتم إدراجه في المجموعة الضابطة. ونظرًا لأن البرنامج لن يكون متاحًا لجميع المؤهلين، يمكن أن يكون الاختيار العشوائي في البرنامج الوسيلة الأكثر عدالة وشفافية لتحديد من سيتم إدراجه. وفي هذا الإطار، يمكن استخدام مُولد أرقام عشوائية، غير أنه يكثر بشكل متزايد إقامة فعاليات عامة للاختيار العشوائي بهدف زيادة مستوى الشفافية.

- يتيح الاختيار العشوائي المرتكز على تغيير عتبة الأهلية التعيين العشوائي، عبر إحداث تغيير طفيف في عتبة الأهلية. ومن خلال تخفيف عتبة الأهلية هته، من الممكن تحديد عدد أكبر من المجتمع الإحصائي المؤهل الذي يمكن معالجته، حيث يتم في نطاقه تعيين المجموعة المعالجة بشكل عشوائي. فعلى سبيل المثال، إذا كان معيار الأهلية لبرنامج تغذية ما هو الأسر التي لديها أطفال تصل أعمارهم إلى ٢٤ شهرًا، فحينها يمكن رفع هذا الحد إلى ٣٠ شهرًا. ويمكن استخدام مقاربة مشابهة جغرافيًا. ويمكن لبرنامج يخطط للعمل في ٥٠ مجتمعًا القيام أولاً بتحديد ١٠٠ مجتمع، ومن ثم اختيار ٥٠ مجتمعًا عشوائيًا من هذا العدد الإجمالي للانضمام إلى البرنامج. وفي هذه الحالة الأخيرة، ستعمل تقنية الاختيار العشوائي للأزواج المتطابقة (التي يتم تناولها أدناه) على زيادة القوة الإحصائية للتصميم.

- تعمل التصاميم المتوالية أو المتدرجة على اختيار الترتيب المتعلق بالمعالجة عشوائيًا، بدلاً من المعالجة نفسها. ويعني الاختيار العشوائي المتوالي أن جميع وحدات التعيين ستلتحق بالبرنامج تدريجيًا ومع مرور الوقت. ويتم التعيين العشوائي عندما يحين وقت الالتحاق بالبرنامج. وغالبًا ما تطرح الوكالات المعنية بالتنفيذ البرنامج على مراحل، مما يتيح إمكانية الاختيار العشوائي للترتيب الذي يلتحق فيه المشاركون بالبرنامج. فعلى سبيل المثال، إذا حالت قيود تتعلق بالميزانية والخدمات اللوجستية دون البدء الفوري للبرنامج على الصعيد الوطني، فقد يكون من الممكن اختيار الوحدات التي ستشارك في البرنامج عشوائيًا خلال المرحلة الأولى. ويكمن أحد الأمثلة على هذه المقاربة في برنامج "بانتاويد باميليا" (Pantawid Pamilya) للتحويل النقدي المشروط في الفلبين. ففي مرحلته الأولية، كان البرنامج تجريبيًا على ١٤٠ مجتمعًا، حيث تلقى نصف المجتمعات البرنامج في بادئ الأمر وعمل النصف الآخر كمجموعة ضابطة لمدة عامين (البنك الدولي، ٢٠١٤). ومن ثم، فقد تم تقسيم هذه المجتمعات بشكل عشوائي إلى مجموعتين للمشاركة في البرنامج، وذلك إما في العام الأول أو العام الثالث (أي أن أولئك الذين شاركوا في البرنامج في العام الثالث عملوا كمجموعة ضابطة لمدة عامين).

- *يمكن استخدام التصاميم التشجيعية* مع البرامج والسياسات المتاحة ولكن غير المعتمدة بشكل عام. ويتم تشجيع مجموعة المعالجة لتقبل التدخل، ولكن ينبغي ألا يكون هذا التشجيع شيئًا يؤثر على الحصائل المرجوة. ويكمن أحد الأمثلة على التشجيع المناسب في القيام بحملة إعلامية تشجيعية لأحد البرامج الجارية. ويتم اختيار القرى التي ستُجرى فيها الحملات بشكل عشوائي من بين جميع القرى التي يتم تنفيذ البرنامج فيها (والتي قد تكون جميع القرى داخل الدولة في إطار برنامج وطني). ويُقاس الأثر الذي يُحدثه البرنامج على الحصائل المرجوة من خلال مقارنة الحصيلة ذات الصلة بين القرى المدرجة ضمن المجموعة المعالجة والمجموعة الضابطة. وفي هذه الحالة، تتمثل في القرى التي تتعرض للحملة الإعلامية. وتتيح هذه المقاربة إجراء تخمين لمتوسط تأثير المعالجة المحلي بسبب معدلات القبول المتفاوتة بين القرى التي تتعرض للحملة الإعلامية وتلك التي لا تتعرض لها.

- *يمكن استخدام التقسيم الطبقي أو المطابقة المسبقة لضمان* إحداث التوازن مع حجم عينة أصغر. ويعمل الاختيار العشوائي للأزواج المتطابقة على المطابقة بين الوحدات (على سبيل المثال، المجتمعات)، ومن ثم تصنيفها في أزواج بناءً على الخصائص المرصودة، مع التعيين العشوائي لأحد المُجتمعيْن من كل زوج في مجموعة المعالجة والآخر في المجموعة الضابطة. لذلك إذا كان هناك، على سبيل المثال، مجتمعان بعيدان أو كبيران أو بهما أقليات عرقية على وجه الخصوص، فإن المطابقة المسبقة تضمن أن يتم إدراج أحدهما في مجموعة المعالجة والآخر في المجموعة الضابطة. وبدون المطابقة المُسبقة، قد ينتهي الأمر بكليهما في إحدى هاتين المجموعتيْن، مما يؤدي إلى عدم توازن العينة. وعندما يكون ذلك ممكنًا، فإن الاختيار العشوائي للأزواج المتطابقة يتيح إمكانية تحسين قوة العينة وكفاءتها وفعاليتها، ومن ثم تعزيز القوة الإحصائية (Imai et al. 2009). ووِفقاً للتقسيم الطبقي، يتم تقسيم المشاركين إلى مجموعات (طبقات). على سبيل المثل، يمكن تقسيم المشاركين إلى ذوي الدخل المنخفض والمتوسط والمرتفع. وبعد ذلك، يتم إجراء الاختيار (التوزيع) العشوائي لكل مجموعة. ويرى كوسوك إيماي وآخرون أن التقسيم الطبقي يضمن أن يكون لكل من مجموعة المعالجة والمجموعة الضابطة نفس النسبة من الوحدات المتعلقة بالمتغيرات المستخدمة في التقسيم الطبقي (على سبيل المثال، ذوو الدخل المنخفض/المتوسط/المرتفع، مناطق ريفية/حضرية، فقراء/غير فقراء)، كما يمكن أن يساعد في تعزيز القوة الإحصائية وتسهيل تحليل المجموعات الفرعية (Imai et al. 2008).

مجموعات المعالجة المتعددة والمجموعات الضابطة المُعالَجة

لا يلزم أن تكون المجموعة الضابطة غير مُعالَجة (يُطلق على هذا مصطلح مجموعة ضابطة سالبة). وتوفر المجموعة الضابطة غير المُعالَجة (أو مجموعة الدراسة) الواقع المضاد لعدم التدخل، وبالتالي تجيب عن السؤال التالي: "كيف يمكن مقارنة هذا التدخل بعدم القيام بأي شيء؟" ومع ذلك، نادرًا ما يكون هذا هو السؤال الوحيد محل الاهتمام المتعلق بالسياسة المتبعة.

وغالبًا ما يهتم صانعو السياسات بمعرفة أوجه الشبه والاختلاف بين عمل "أ" وعمل "ب". إذ يقارن تصميمي أ/ب (A/B designs) بين مجموعتيْ (ذراعيْ) معالجة تتمثلان في مجموعة المعالجة "أ" ومجموعة المعالجة "ب". وفي المجال الطبي، غالبًا ما تُعرف مجموعة المعالجة "ب" باسم "معيار الرعاية الحالي". وهذا يعني أن الواقع المضاد يمثل السياسة الحالية المتبعة، وبالتالي سيكون سؤال التقييم: "كيف يمكن مقارنة هذا التدخل بما يتم القيام به حاليًا؟" وفي هذه الحالة، تُعرف "أ" باسم "مجموعة مقارنة نشطة". وبعبارة أخرى، يمكننا القول أنه، قد يكون كل من "أ" و"ب" مقاربتين مختلفتين لتنفيذ مشروع ما. ويُعَد تصميمي أ/ب وسيلة مناسبة للتعلم التكيفي الذي يهدف إلى تحسين تصميم المشروع. ويمكن إضافة المزيد من مجموعات (أذرع) المعالجة (على سبيل المثال، تصميمات أ/ب/ج)، غير أن كل مجموعة تعمل على زيادة متطلبات حجم العينة.

تتمثل إشكالية عدم وجود مجموعة ضابطة لا تخضع للمعالجة في أنه لا يمكن حساب الفعالية المطلقة. وقد يبدو أن التدخليْن، اللذين يتم التحقق من فعاليتهما، يعملان بشكل جيد على حد سواء. لكن في واقع الأمر قد لا يكون لأي منهما أي أثر يُذكر. وإذا ثبُت من خلال دراسات دقيقة سابقة أن المعالجة الحالية تعمل على نحو جيد، فلن تحدث هذه المشكلة. ومع ذلك، لا يزال من المفيد في أغلب الأحيان تخمين مدى الفعالية المطلقة لأغراض تتعلق بتحليل فعالية التكلفة أو تحليل التكلفة والمنفعة.

تعد التصاميم العاملية نوعًا خاصًا من الدراسات متعددة المجموعات (الأذرع) التي تتلقى فيها إحدى المجموعات تدخلات متعددة. فعلى سبيل المثال، تحصل إحدى المجموعات على التدخل "أ"، في حين تحصل أخرى على التدخل "ب"، وتحصل الثالثة على التدخليْن "أ" و"ب"، بينما لا تتم معالجة الرابعة. وتتيح التصاميم العاملية اختبار ما إذا كانت التدخلات المختلفة مكملات أو بدائل لبعضها البعض. وكثيرًا ما يُقال إن هناك تكاملاً بين التدخلات المختلفة (على سبيل المثال، التمويل الأصغر وتطوير الأعمال، ودعم المدخلات والخدمات الإرشادية، وتحسين التثقيف في مجال المياه والنظافة الصحية). وتتيح التصاميم العاملية التحقق من هذا الأمر.

ترتبط التصاميم المتقاطعة بالتصاميم العاملية، غير أن عمليات المعالجة تكون متسلسلة وليست متزامنة. وهذا يعني أن المجموعة الثالثة تحصل على التدخل "ب" متبوعًا بالتدخل "أ"، وقد تحصل مجموعة إضافية على التدخل "أ" متبوعًا بالتدخل "ب"، بخلاف التصميم العاملي الذي يتم من خلاله تقديم التدخليْن "أ" و"ب" معًا. ويمكن أن يتيح هذا التحقق مما إذا كان تسلسل التدخلات مهمًا، ولكنه يتطلب مجموعات معالجة أكثر مقارنة بالتصميم العاملي.

يمكن الجمع بين المقاربات المختلفة في دراسة واحدة، كما في حالة دراسات تقييم الأثر التي يُجريها بنك التنمية الآسيوي على تزويد المناطق الريفية في ولاية ماديا براديش الهندية بالكهرباء (المربع ٤-١).

المربع (٤-١): تصميم تشجيعي متوالٍ يرتكز على مجموعة متعددة المعالجة

تجمع دراسة تقييم أثر تزويد الريف بالكهرباء على جودة حياة المرأة وتمكينها في ولاية ماديا براديش الهندية، والتي يدعمها بنك التنمية الآسيوي، بين ثلاثة تصميمات: مجموعات معالجة متعددة مع وسائل تشجيعية يتم طرحها تدريجيًا ومع مرور الوقت (التصميم المتوالي).

تتمثل عمليات المعالجة المتعددة في تكاليف التوصيل المدعوم والتدريب على استخدام الكهرباء. وتتلقى مجموعة المعالجة الأولى التوصيل المدعوم فحسب، في حين تتلقى مجموعة المعالجة الثانية التوصيل المدعوم والتدريب. وهناك مجموعة ضابطة غير مُعالجة لن تتلقى الكهرباء على الإطلاق.

لم يتم اختيار القرى التي سيتم توصيل الكهرباء إليها عشوائيًا. ومع ذلك، فإن الاختيار ضمن مجموعتيْ المعالجة تم بُناءً على عملية تعيين عشوائي غطت ٢٤٠ قرية. ويعد دعم عملية التوصيل أحد الأساليب التشجيعية، والذي يتعين أن يؤدي إلى اعتماد أو قبول أعلى في قرى المعالجة المدرجة ضمن المجموعة الضابطة وهو ما يُفضي إلى إجراء تحليل بشأن أثر التزويد بالكهرباء. وفي قرى المعالجة البالغة ١٦٠، تم توفير دعم التوصيل في الفترة من ٢٠١٥ إلى ٢٠١٧، وكان من المقرر توفيره لاحقًا للقرى المُدرجة في المجموعة الضابطة.

ملاحظة: تم تبسيط التصميم لأغراض العرض والتوضيح.

المصدر: Impact Evaluation Study Proposal, 21 November 2014 (اقتراح دراسة تقييم أثر، ٢١ نوفمبر، ٢٠١٤).

٤-٤ الخطوات المتبعة في تنفيذ التجارب العشوائية المضبوطة

يلزم التخطيط للتجربة العشوائية المضبوطة قبل التنفيذ الميداني الواسع للتدخل. ولا يمكن إجراء التجارب العشوائية المضبوطة بأثر سابق.[٢] ومن الممكن القيام بدراسات تقييم الأثر الاختبارية التجريبية في المراحل الأولى من المشروع بحيث يمكن تحسين تصميمات المشروع من خلال الأدلة التي يتم جمعها من هذه الدراسات. ووفقا لهوارد وايت، يقتضي تصميم التجارب العشوائية المضبوطة اتخاذ الخطوات التالية (White 2013):

(١) تحديد الأسئلة محل الاهتمام

ينطوي التصميم الشامل للتقييم على اختيار أسئلة التقييم التي يتم استنباطها من نظرية التغيير. ويتعين على موظفي الوكالة المعنية بالتنفيذ، بدءًا من الإدارة العليا وصولاً إلى الموظفين الميدانيين، الاقتناع بالتصميم المبني على أخذ عينات عشوائية حتى تكون عملية الاختيار العشوائي ممكنة في المقام الأول، فضلاً عن الحفاظ على نزاهة التصميم خلال الممارسة العملية.

(٢) الفصل بين عناصر المعالجة (عمليات المعالجة) محل الاهتمام

ستمثل عمليات المعالجة محل الاهتمام تجسيدًا للأسئلة التي يجب أن تساعد الدراسة في الإجابة عنها. وفيما يتعلق بالتجربة العشوائية المضبوطة، يتعين أن تكون المعالجة التي يجري تقييمها محددةً بوضوح حتى يتسنى تطبيقها على نحو متجانس. ويُمثل هذا أهمية أيضًا للصلاحية الخارجية: يلزم أن يكون ما تم تقييمه واضحًا.

٢ يكمن الاستثناء المحتمل لهذا في "التصميم التشجيعي"، الذي لا يُعيِّن المشاركين عشوائيًا في التدخل ولكن يتيح الحصول على حوافز تشجعهم على المشاركة.

وبصفة عامة، يجب أن يكون التدخل مجزأً قدر الإمكان، مع تقييم عناصر التدخل المجمعة فقط في التصاميم العاملية أو التصاميم المتقاطعة. وبخلاف ذلك، لن يكون من الممكن تحديد مكونات التدخلات التي أدت إلى تأثيرات محددة، وستكون الصلاحية الخارجية محدودة.

بالإضافة إلى ذلك، يجب أن يحدد القائم بالتقييم ما إذا كان ينبغي استخدام وسيلة معالجة وهمية لأي من المجموعات الضابطة في التجربة. إذ يمكن أن تتيح وسائل المعالجة الوهمية القضاء على التحيز الناجم عن معرفة المعالجة، فضلاً عن تحسين مستوى الدقة. ومع ذلك، غالبًا ما تكون وسائل المعالجة الوهمية غير مجدية للعديد من التدخلات التي تحدث في العالم الحقيقي، كما أنها تثير قضايا أخلاقية حول تمثيل الدراسة والموافقة المستنيرة. وهذا يعني أن وسائل المعالجة الوهمية نادرة عند الحديث عن الممارسة العملية بالنسبة لدراسات تقييم الأثر المتعلقة بالتدخلات الإنمائية.

(٣) مراعاة تأثيرات التداعيات

إذا كانت نظرية التغيير تشير إلى أنه ستكون هناك تأثيرات للتداعيات، فيجب مراعاة هذه التداعيات في تصميم الدراسة. ويجب تجميع وحدة التعيين على نحو يكفي لعدم وجود أي تداعيات بين الوحدات المُعالَجة وتلك غير المعالَجة. وعندما يُتوقع أن تكون تأثيرات التداعيات كبيرة، فقد ينطبق التعيين العشوائي على ثلاث مجموعات: المُعالجَة، وغير المُعالجَة ولكنها عرضة لتأثيرات التداعيات، وغير المُعالَجة ولكنها ليست عرضة للتداعيات.

(٤) تحديد مستويات التعيين والمعالجة والتحليل

يعتمد تصميم عملية الاختيار العشوائي على التحديد الواضح لوحدة التعيين والمعالجة والتحليل. ويكون الاختيار العشوائي في المقام الأول عند مستوى التعيين. ويجب أن يحدث التعيين العشوائي عبر نطاق يتيح الحصول على قائمة كافية يمكن من خلالها الاستخلاص العشوائي لوحدات المعالجة والضابطة المكافئة. ويجب إجراء حسابات القوة الإحصائية للعينة (التي يتم تناولها في الفصل السابع) في هذه المرحلة للتأكد من أنه يمكن تضمين وحدات كافية تتيح التمتع باحتمالية عالية لاكتشاف التأثيرات.

يجب أن يقرر فريق الدراسة أيضًا المجموعات الفرعية المحتملة محل الاهتمام في بداية الدراسة، وذلك للتأكد من تمتع الدراسة بالقوة الكافية لإجراء تحليلات المجموعات الفرعية.

(٥) تحديد نوع التجربة العشوائية المضبوطة (تم تناول ذلك في القسم ٤-٣)

يتوقف اختيار نوع التجربة العشوائية المضبوطة على تصميم التدخل وأسئلة التقييم وما هو ممكن. وعادةً ما تتطلب طبيعة التدخلات الإنمائية إجراء تجارب عشوائية مضبوطة عنقودية نظرًا لأنه يتم تعيين المشاريع حسب الوحدة الجغرافية، وغالبًا ما تحدث التداعيات داخل المجتمعات.

ستتطلب أسئلة الجيل الأول (هل ينجح التدخل؟) مجموعة ضابطة غير مُعالَجة وعينة كبيرة، نظرًا لأن أحجام التأثير المتعلقة بالحصائل النهائية عادةً ما تكون صغيرة نسبيًا. ومن المرجح أن تستخدم أسئلة الجيل الثاني (ما مقاربات التدخل الأكثر فعالية؟) تصميمي أ/ب (A/B designs). وفي بعض الحالات، يتطلب ذلك أطرًا زمنية أقصر وعينات أصغر حجمًا، لأن الحصيلة المرجوة تكون أقرب للتحقق (عادة ما ينعكس ذلك في مستوى الاعتماد أو القبول)، حيث تكون أحجام التأثير أكبر وتحدث بسرعة أكبر من الحصائل النهائية.

عادة ما تُصمم التجارب العشوائية المضبوطة في وقت سابق. ولكن إذا كان التدخل جاريًا، فقد يكون من الممكن (١) استخدام التعيين العشوائي في بدء البرنامج، أو (٢) إدخال تغييرات في تنفيذ البرنامج لأغراض التعلُّم التكيفي، أو (٣) استخدام تصميم تشجيعي.

(٦) تحديد المجتمع الإحصائي المؤهل (نطاق التعيين العشوائي)

حتى يتسنى لمعدل المشاركة أن يكون كافيًا لأحجام العينات المُعالَجة من أجل إتاحة الحصول على قوة العينة الإحصائية اللازمة، يجب أن تركز التجربة العشوائية المضبوطة على المجتمع الإحصائي المستهدف الذي من المحتمل أن يشارك و/أو أن يُستهدَف من خلال التدخل المُقيَّم. ومن ثم، يتم الاختيار العشوائي في نطاق هذا المجتمع الإحصائي السكاني، وليس السكان ككل. وللقيام بذلك، يجب تحديد معايير الأهلية أو مناطق الاستجماع المخطط لها لوحدات التعيين بوضوح. وقد تكون المعايير جغرافية، مثل جميع المناطق الفرعية في ١٠ مقاطعات محددة مسبقًا، أو مرتكزة على خصائص معينة، مثل المؤسسات الصغيرة التي ترأسها سيدات. وقد يتم الجمع بين معايير عديدة، على سبيل المثال، المؤسسات الصغيرة التي ترأسها سيدات في المناطق الفرعية التي تنتمي إلى ١٠ مقاطعات محددة مسبقًا، حيث يكون مؤشر الفقر في المناطق الفرعية أقل من حد مُعين.

يمكن تعيين المجتمع الإحصائي المؤهل بإحدى طريقتين أو كليهما: التحديد (identification) أو الاختيار الذاتي (self-selection). فيما يتعلق بالتحديد، يتم وضع قائمة بالمجتمع الإحصائي المؤهل. وقد لا يتمثل هذا "المجتمع" في أفراد أو أسر أو شركات، ولكن يمكن أن يتألف من قرى أو بلدات أو مناطق فرعية أو مناطق. وغالبًا ما تُتاح هذه القوائم من خلال مصادر إدارية أو يتم وضعها كجزء من تصميم المشروع.

كأحد أشكال الاختيار الذاتي، يمكن الإعلان عن البرنامج ويمكن للمهتمين به التقديم للمشاركة به. ويتم فحص المتقدمين للتحقق من مدى أهليتهم للمشاركة في البرنامج. ومع افتراض وجود زيادة في عدد المشتركين، يتم إجراء التعيين العشوائي بين المتقدمين المؤهلين.

(٧) استخلاص عينة للتحليل.

يتم استخلاص عينة ممثلة من المجتمع الإحصائي المؤهل أو المستهدف، ما لم يتم استخدام هذا المجتمع الإحصائي بالكامل للتحليل. وفي حالة التجربة العشوائية المضبوطة العنقودية، ستتجسد عملية أخذ العينات في تصميم يتألف من مرحلتين، ويشار إليه بمجموعات أخذ العينات العنقودية الأولى، وذلك إذا كانت هناك مجموعات عنقودية للمشروع أكثر مما هو مطلوب للدراسة. وسيكون هذا هو الحال بشكل دائم تقريبًا مع التجربة العشوائية المضبوطة العنقودية التي يلزم إجراؤها في نطاق عملية أخذ عينات من مجموعات عنقودية.

(٨) التعيين إلى المجموعة المعالجة والمجموعة الضابطة.

يتوجب تعيين الوحدات المؤهلة إلى مجموعة معالجة ومجموعة ضابطة قبل تنفيذ التدخل في المواقع المُعالَجة. ويتم تحديد الحد الأقصى الممكن لحجم مجموعة المعالجة من خلال التغطية المستهدفة للمشروع. وليس بالضرورة أن يتم تضمين مجتمع المعالجة الإحصائي بأكمله في العينة لإجراء دراسة تقييم الأثر، وهو أمر يتوقف على متطلبات حجم العينة كما يتحدد من خلال حساب القوة الإحصائية.

يمكن أن يقتصر التعيين العشوائي على مجموعة فرعية من مشروع أو برنامج ، إذا كان ذلك التعيين يوفر عينة ذات حجم كافٍ . فعلى سبيل المثال، لنفترض أن مشروعًا لترقية أحياء فقيرة يُخطط لترقية ٢٠٠ مستوطنة على مدى فترة زمنية تبلغ ٥ سنوات، على أن تتم ترقية ٤٠ مستوطنة جديدة كل عام. وتُظهر حسابات القوة الإحصائية أنه يجب تضمين ٤٠ مستوطنة في العينة، والتي تتألف من ٢٠ في مجموعة المعالجة و ٢٠ في المجموعة الضابطة، مع أخذ عينة تتألف من ١٥ أسرة لكل مستوطنة. وسيتمثل التصميم المحتمل في التعيين العشوائي للمستوطنات الـ ٢٠ للعام ١ والعام ٥، حيث ستعمل الأخيرة كمجموعة ضابطة للأعوام ١ إلى ٥. ويتمتع موظفو البرنامج بالحرية في تعيين المستوطنات الـ ١٦٠ المتبقية عبر الأعوام الأخرى حسبما يرغبون.[٣]

(٩) جمع بيانات خط الأساس والتحقق من التوازن.

بالرغم من أن بيانات خط الأساس لا تكون مطلوبة من الناحية النظرية للتجربة العشوائية المضبوطة، فإنه من الناحية العملية عادة ما يتم جمعها لضمان التوازن بين المجموعة المعالجة والمجموعة الضابطة. وعادة ما يلزم جمع بيانات خط الأساس قبل أن يكون للتدخل أي تأثير على المستوى الميداني. ويمكن جمعها قبل تعيين المعالجة لتجنب أي تحيز أو انحراف في جمع البيانات، حيثما أمكن ذلك.

بمجرد جمع البيانات وتنظيفها (يتناول الفصل السادس جمع البيانات)، وكذلك إجراء التعيين في مجموعة المعالجة، يجب إجراء عملية التحقق من التوازن. وإذا لم يكن هناك توازن، فقد يتم الإخلال بالتنفيذ الميداني لبروتوكول الاختيار العشوائي، ومن ثم، يجب إجراء مزيد من التحقيق حول كيفية تنفيذ هذا الاختيار العشوائي.

(١٠) التأكد من نزاهة التصميم ورصد التداخل والتناقص.

يحتاج فريق الدراسة إلى البقاء متفاعلاً عن كثب أثناء تقدم عملية التنفيذ لضمان نزاهة التصميم، مع اتباع بروتوكولات الاختيار العشوائي على نحو صارم بحيث يتم علاج مجموعات العلاج وعدم علاج المجموعات الضابطة. وغالبًا ما يعني الاختيار العشوائي مزيدًا من العمل للفرق الميدانية للوصول إلى مجتمعات يصعب الوصول إليها، أو التعامل مع مستفيدين محتملين أقل انفتاحًا وتقبلاً، أو القيام بالأشياء بترتيب لا يتسم بالكفاءة. وبالتالي، قد يكون لدى الفرق الميدانية حوافز تشجعهم على اختصار عملية الاختيار العشوائي إذا لم يكن هناك مراقبة وثيقة للغاية لضمان اتباع بروتوكول الاختيار العشوائي.

٣ يمكن افتراض حالة تتعلق بالقيام بالتعيين العشوائي في العامين ٢ و٥. ومن المحتمل أن يكون هناك تعلُّم في العام الأول مما يؤدي إلى تكييف المشروع وتهيئته، لذلك من المنطقي تقييم المشروع على نحو ما تم تنفيذه في العام الثاني عندما يكون التصميم "قد استقر".

حتى إذا تم ضمان نزاهة التصميم عند البدء، فقد يتعرض لخلل لاحقًا بسبب التداخل (تم تناوله في الفصل الثالث). ومن المهم التحقق من أن أولئك المدرجين في المجموعة الضابطة لا يعانون من أي تداخل، إما من خلال تدخل مماثل يتم تنفيذه في المناطق التي تنتمي إليها المجموعة الضابطة أو من خلال التداخل الذاتي، حيث ينتقل المشاركون في الدراسة من إحدى مجموعاتها إلى الأخرى، وبالتالي يحدث تداخل أو تشوش في عملية الاختيار العشوائي الأولية.

كمثال على عدم الالتزام الذي يؤدي إلى التداخل، تناولت تجربة عشوائية مضبوطة في جمهورية الصين الشعبية بالدراسة مشروعًا قدم نظّارات طبية لطلاب المدارس الثانوية، وتوصلت إلى أن استخدام النظارات الطبية قد ارتفع أيضًا في بعض المجموعات الضابطة (Glewwe et al. 2009). وقد كشف المزيد من البحث والاستعلام أن الأطباء الذين كانوا يجرون فحوصات العين قد توفرت لديهم نظّارات طبية متبقية من مجموعة المعالجة، وقد أعطوها للطلاب في المناطق التي تنتمي إليها المجموعة الضابطة، وهو ما يجسد مثالاً على التداخل الذاتي. وقد استخدمت الدراسة تصميمًا يرتكز على الأزواج المتطابقة، وبالتالي تمكنت من استبعاد الأزواج التي حدث بها تداخل في المجموعة الضابطة.

من المهم أيضًا التحقق من التناقص بين المجموعات، والذي يحدث عندما ينسحب مشاركون في الدراسة من العينة بين جولة جمع بيانات وأخرى، حيث يمكن أن يؤدي ذلك إلى نتائج مضللة. فعلى سبيل المثال، إذا تم تقديم بيانات الحصائل من قِبل عدد من الأشخاص في المجموعة الضابطة أقل من مجموعة المعالجة، فإن هذا من شأنه أن يؤدي إلى انحراف النتائج، حيث يتم استبعاد المشاركين الذين انسحبوا من التحليل.

بالإضافة إلى ذلك، من الأهمية بمكان إنشاء بروتوكولات للتحقق والمتابعة الميدانية عند تنفيذ تجربة عشوائية مضبوطة، وذلك لفهم أي خطأ محتمل في الإبلاغ ناجم عن التدخل. وقد تكون هناك توقعات متباينة بين أولئك الذين تعرضوا للمعالجة والذين لم يتعرضوا لها بشأن الانعكاسات المحتملة للإجابات التي يقدمونها. وذلك عندما يكون هناك تصور بأن القائمين على تقييم الأثر يمكنهم تقديم تدخلات أو التأثير عليها. فعلى سبيل المثال، قد تعتقد المجتمعات الإحصائية غير المُعالَجة أنها تستطيع المبالغة بشأن المشكلات التي تعاني منها من أجل الاستفادة من عمليات معالجة في المستقبل، أو قد يعتقد أولئك الذين عولجوا أنه يمكنهم الحصول على تعويض إذا أبلغوا عن أن التدخل لم يُسفر عن النتائج المرجوة على النحو المحدد. ومن الممكن أن يساعد تعدد أدوات جمع الردود مع المتابعة والملاحظة المباشرة على التخفيف من هذه الأنواع من التحيزات.

٤-٥ تقديم تقارير تتناول تفاصيل التجارب العشوائية المضبوطة

تتطلب التجارب العشوائية المضبوطة ضمان أن تتسم آلية التعيين بالملاءمة فضلاً عن دقة الاختيار العشوائي، من أجل ضمان جودة الدراسة. ولهذا السبب، من المهم توفير تفاصيل كافية عند تقديم المنهجية والاستنتاجات. ومن المهم أيضًا تقديم وصف كافٍ للتدخل الذي يجري تقييمه، فضلاً عن إستراتيجية أخذ العينات. وفي هذا الوصف، من الضروري تحديد ووصف كل من عدد المجموعات وعدد الأسر و/أو الأفراد في مجموعة المعالجة والمجموعة الضابطة بمرور الوقت. ويجب أن يتضمن التقرير أيضًا جداول حول عمليات التحقق الخاصة بالتوازن (التي تم تناولها أعلاه)، ومناقشة ما إذا كان هناك أي تحيزات محتملة في البيانات التي تم جمعها.

عند توفر بيانات خط الأساس، يمكن الإبلاغ عن تقديرات الأثر باستخدام تحليل الاختلاف فى الاختلافات (DiD). وبالرغم من أنه ينبغي ألا يكون هذا ضروريًا مع التعيين العشوائي، فإن الاقتصاديين غالبًا ما يقدرون الأثر باستخدام نموذج انحدار، والذي يتحكم في التباين المتبقي في المتغيرات المشتركة المرصودة. ويمكن الإبلاغ عن نتائج التحليل للعينة بأكملها، وكذلك بالنسبة للمجموعات الفرعية، وذلك بهدف تحليل التأثيرات غير المتجانسة.

كما هو الحال مع جميع دراسات تقييم الأثر، يجب الانتباه إلى حجم التأثير، وليس فقط دلالته وأهميته. ويمكن استخدام حجم التأثير لمزيد من التحليل الاقتصادي، مثل تحليل فعالية التكلفة أو تحليل التكلفة والمنفعة.

٤-٦ الاستخدام العملي للتجارب العشوائية المضبوطة

التغلب على المقاومة التي قد تعترض الاختيار العشوائي

غالبًا ما يواجه الاختيار العشوائي معارضة أولية من الفرق المعنية بالتنفيذ. ومع ذلك، هناك العديد من الأسباب التي تجعل الاختيار العشوائي فكرة جيدة وليست مرهقة للمشاريع على نحو ما قد يُتصور.

١. تكمن النقطة الأكثر تعزيزًا للفكرة سالفة الذكر في أن **التجربة العشوائية المضبوطة المُصممة جيدًا تعكس الروابط السببية بأكبر قدر ممكن من الإيضاح.** وحسب التصميم، لا يكون هناك أي صلة بين خصائص المستفيدين وتعيين التدخل. ويجب أن يكون أي اختلاف في الحصائل بسبب التدخل، وليس بسبب أي اختلاف أساسي بين المجموعة المعالجة والمجموعة الضابطة.

٢. **سهولة تحليل التجارب العشوائية المضبوطة،** حيث يتم الحصول على النتائج عبر حساب الفرق (أو الاختلاف) في متوسط الحصائل بين المجموعة المعالجة والمجموعة الضابطة. وتتسم هذه النتائج بسهولة الحساب والعرض. لذلك، ليست هناك حاجة لإجراء تحليلات إحصائية معقدة.(عادةً ما يحسب الاقتصاديون متوسط الفرق باستخدام الانحدار مع إضافة متغيرات المقارنة، ولكن يمكن تقديمه بسهولة على أنه المتوسط البسيط للفرق.)

٣. **تعد التجارب العشوائية المضبوطة وسيلة عادلة وشفافة لتعيين البرامج.** ففي برنامج تنمية نموذجي، قد يكون لدى المستفيدين المستهدفين وحتى موظفي الوكالة فكرة محدودة عن كيفية أو أسباب اختيار مجتمعات بعينها للاستفادة من البرنامج. ومن الممكن أن يعمل الاختيار العشوائي على تعزيز الإنصاف في الاختيار، مقارنةً بتنسيب البرنامج الذي قد يتأثر باعتبارات سياسية أو يستند إلى شبكات محسوبية. وقد تحظى العمليات الشفافة، مثل القرعات العامة، أيضًا بتقدير المستفيدين كوسيلة مُحسَّنة لاختيار المشاركين.

٤. **لا يلزم أن تكون هناك مجموعة ضابطة غير مُعالَجة.** وقد يكون للتجربة العشوائية المضبوطة مجموعات (أذرع) معالجة متعددة في أبسط صورها، حيث تتم مقارنة التدخل "أ" بالتدخل "ب". كما هو الحال مع العديد من التجارب السريرية، قد تكون المعالجة "ب" ما يتم القيام به بالفعل. ويضيف التصميم العاملي مجموعة معالجة إضافية تتلقى كلا التدخليْن "أ" و"ب". وهذا يساعد في الإجابة عن السؤال حول ما إذا كان التدخلان يعملان بشكل أفضل معًا أو عندما يكونان منفصليْن.

٥. **يمكن أن تؤدي التجارب العشوائية المضبوطة إلى استهداف أفضل.** ويتم إجراء الاختيار العشوائي عبر المجتمع الإحصائي المؤهل، لذلك يظل التدخل مستهدفًا كما هو مخطط له. ونظرًا لأن التجربة العشوائية المضبوطة تتطلب من موظفي المشروع تحديد المجتمع الإحصائي المؤهل وإدراجه بكل وضوح، فقد يؤدي ذلك إلى استهداف أفضل مما يمكن تحقيقه بدون هذا النظام.

٦. **يجب ألا يؤثر الاختيار العشوائي على مشروع أو برنامج بأكمله.** فبمجرد إجراء حسابات القوة الإحصائية، سيتم توقع حجم العينة اللازمة لإجراء الاختيار العشوائي. وبالنسبة للبرامج الكبرى، من المحتمل أن تكون هناك حاجة إلى مجموعة فرعية فقط من المجتمع الإحصائي المستفيد والمستهدف. ويمكن لمديري البرامج أن يقوموا بما يحلو لهم مع باقي هذا المجتمع، والذي قد يمثل الغالبية.

٧. **لا يعني الاختيار العشوائي بالضرورة تغييرات كبيرة في تنفيذ البرنامج:**

(أ) يمكن إجراء تعديلات طفيفة على معايير الأهلية ("تصميم عتبة مرتفعة للأهلية") للحصول على مجموعة ضابطة صالحة بطريقة غير تدخلية. وقد تنشأ زيادة في عدد المشتركين من خلال تعديل هذه العتبة. ويتم بعد ذلك اختيار المشاركين بشكل عشوائي، أي عن طريق قرعة.

(ب) يتيح التصميم التشجيعي، وليس البرنامج نفسه، التعيين العشوائي لوسيلة تشجيعية تحفز على المشاركة في البرنامج. ولن يكون لهذا أي تأثير على كيفية تشغيل البرنامج، بل سيؤدي في الواقع أيضًا إلى توفير معلومات مفيدة حول مدى زيادة الإقبال على المشاركة في البرنامج.

(ج) أخيرًا، تستفيد التجربة العشوائية المضبوطة ذات التصميم المتوالي من حقيقة أن البرنامج يتم تنفيذه بمرور الوقت (تدريجيًا)، وأنه من شبه المؤكد أن يكون هناك أعضاء غير مُعالَجين من المجتمع الإحصائي يمكنهم تشكيل مجموعة ضابطةصالحة. لذلك، تعمل التجربة العشوائية المضبوطة ببساطة على ترتيب عملية المعالجة عشوائيًا.

٨. **يمكن للتجارب العشوائية المضبوطة جيدة التصميم أن تكشف النقاب عن كيفية حدوث التأثير.** ولا تحتاج التجارب العشوائية المضبوطة إلى التركيز فقط على الاستفسار بشأن ما إذا كان "التدخل فعالاً". إذ يمكنها أيضًا النظر في التباينات الموجودة في التصميم الخاص بالتدخل لتحديد الطرق التي يمكن من خلالها أن يعمل بشكل أفضل. وحتى عندما تكون السلسلة السببية للبرنامج معقدة للغاية بحيث لا يمكن كشفها، فإنه سيظل بإمكان التجارب العشوائية المضبوطة تقديم رؤى حول الظروف التي تتحقق في ظلها الآثار.

عشرة أشياء يمكن أن تخفق مع التجارب العشوائية المضبوطة

على الرغم من أن استخدام التعيين العشوائي يمكن أن يسمح بإجراء تقديرات صحيحة للأثر، فمن الضروري الاحتراس من إجراء تجارب باهظة التكلفة وغير ضرورية. فيما يلي ١٠ أشياء مهمة قد تخفق مع التجارب العشوائية المضبوطة. وتنطبق العديد من هذه النقاط على دراسات تقييم الأثر بشكل عام:

١. **تقييم التدخلات عندما يكون مستوى الاعتماد/المشاركة أقل بكثير مما هو متوقع:** تفشل العديد من دراسات تقييم الأثر بسبب ضعف الإقبال على التدخل أو عدم قبوله. وإذا كانت هناك نسبة صغيرة فقط من المستفيدين المستهدفين مهتمة بالتدخل، فقد لا يكون تقييم الأثر ضروريًا. ويعد المُخطط القمعي للتناقص (الذي تم تناوله في الفصل الثاني) أداة يمكن أن تساعد في تحديد مستوى القبول المنخفض وتقييم ما إذا كان يمكن إصلاحه. ومن الممكن أن تعطي الدراسة التشخيصية قبل إجراء تقييم الأثر معلومات عن الميسرات والعوائق التي تحول دون المشاركة.

٢. **الاهتمام فقط بالجوانب البحثية المجردة (الاستيلاء البحثي):** قد يكون الباحثون أكثر اهتمامًا بإجراء دراسة ينتج عنها منشور أكاديمي، بدلاً من الإجابة عن أسئلة التقييم التي تهم منفذي البرنامج. وعادة ما تتيح عملية التحقق من صحة النظرية الاقتصادية السلوكية أفضل الفرص للباحثين باعتبارها موضوعًا دسمًا للنشر. ومع ذلك، يمكن لمثل هذا التركيز أن يأخذ التقييم بعيدًا عن الصلاحية الخارجية أو المتطلبات المعلوماتية العملية لصانعي السياسات. وفي حالة تصميم الدراسة التقييمية من قبل أكاديميين لا تتوفر لديهم المعرفة بسياق المجال ذي الصلة، فقد تحتوي مثل هذه الدراسات أيضًا على عيوب خطيرة في التصميم. ويتناول الفصل الثامن هذا الأمر بالمزيد من المناقشة.

٣. **إجراء تقييمات منخفضة القدرة الإحصائية:** تُصمم الدراسات بشكل عام بحيث تتمتع بقوة إحصائية تبلغ ٨٠٪، مما يعني أن أي دراسة ستخفق في قياس أو تقييم خُمس (٢٠٪) الفترة الزمنية التي يتم فيها استخدام التدخل ذي الصلة. وفي الواقع، تبلغ القوة الإحصائية الفعلية للعديد من التجارب العشوائية المضبوطة حوالي ٥٠٪ فقط. وفي هذه الحالات، لن تكون التجربة العشوائية المضبوطة وما يترتب عليها أفضل مما سينتج عن الاقتراع بقطعة نقود معدنية لمعرفة ما إذا كان التدخل ناجحًا أم لا. وحتى عندما يتم إجراء حسابات القوة الإحصائية بشكل صحيح، قد تكون الدراسات أيضًا ضعيفة. ويرجع ذلك في أغلب الأحيان إلى افتراض أن المشروع سيكون له أثر أكبر بكثير مما يتحقق بالفعل، أو أن معدلات المشاركة ستكون أعلى مما يحدث بالفعل. ويتم تناول حسابات القوة الإحصائية في الفصل ٧ والملحق الفني ٢.

٤. **عدم استيعاب الأخطاء المعيارية بشكل صحيح:** تعد معظم التجارب العشوائية المضبوطة تجارب عنقودية يكون فيها التعيين العشوائي عند مستوى أعلى من الوحدة التي تُقاس بها الحصائل. فعلى سبيل المثال، يتم تعيين تدخل ما عشوائيًا للمدارس، ولكن تكون الحصيلة محل الاهتمام هي تعلم الأطفال. أو يتم تعيين التدخل عشوائيًا للمناطق، ولكن يتم تقييم الحصائل على مستوى القرى. وفي هذه الحالات، يجب تعديل الأخطاء المعيارية لهذه المجموعات العنقودية، حتى لا تتفاقم الأخطاء وتجد الدراسات صعوبة في تحقيق الدلالة أو الأهمية الإحصائية. وقد تستنتج الدراسات التي لا تُعدل الأخطاء المعيارية للمجموعات العنقودية، على نحو غير صحيح، أن الأثر مهم. وإذا لم تؤخذ المجموعات العنقودية في الاعتبار في حسابات حجم العينة، فقد يترتب على ذلك دراسة ضعيفة أو منخفضة القوة الإحصائية مع عدد قليل جدًا من المجموعات العنقودية (يتم تناول ذلك في الفصل السابع).

٥. **عدم الحصول على الموافقة أو الإشراف الكافي فيما يتعلق بالاختيار العشوائي:** تظل فكرة التعيين (التخصيص) العشوائي للبرنامج تمثل إشكالية كبيرة بالنسبة للعديد من منفذي البرامج، حيث يمكن أن تأتي على حساب الكفاءة قصيرة المدى. وقد يفشل التعيين العشوائي إذا لم يتمكن الباحثون من الحصول على موافقة الوكالة الرئيسية ذات الصلة بهدف التنفيذ الميداني. بالإضافة إلى ذلك، بالنسبة للموظفين الميدانيين، قد يبدو التعيين العشوائي غير متوقع وقد يؤدي إلى عمل إضافي شاق. على سبيل المثال، قد يقتضي التعيين العشوائي ضمنًا أن يتم تجاوز موقع يسهل الوصول إليه لصالح منطقة نائية، أو أن يتم تجاهل مجتمع إحصائي مهتم ويسهل إشراكه لصالح مجتمع آخر غير مهتم ويصعب تفاعله. وبدون إشراف أو مساءلة كافيَيْن، سيكون للموظفين الميدانيين حوافز قوية للانحراف أو التلاعب ببروتوكول التعيين العشوائي.

٦. **التداخل الذاتي:** يحدث التداخل عندما تتعرض المجموعة الضابطة لنفس التدخل أو تدخل آخر يؤثر على نفس الحصائل. كما يحدث التداخل الذاتي عندما يتسبب التدخل أو المشروع نفسه في حدوث التداخل. وقد يحدث هذا التداخل من خلال التداعيات، مثل ما يترتب على التواصل الشفهي في حالة التدخلات المعلوماتية. وقد يحدث إذا استخدم الأشخاص في المجموعة الضابطة الخدمات المُقدمة في منطقة المعالجة. بالإضافة إلى ذلك، قد يحدث ذلك عندما يُترك لموظفين من الوكالة المعنية بالتنفيذ موارد غير مستخدمة من منطقة المشروع، والتي يستخدمونها بهدف النشر الإضافي في المناطق التي تنتمي إليها المجموعة الضابطة.

٧. **قياس الحصائل الخاطئة:** قد يتم إجراء الدراسة على نحو جيد، غير أنها تفشل في التأثير على الجماهير المستهدفة إذا لم تقِس الأثر المتعلق بالحصائل التي تهمهم، أو تلك التي تهم المستفيدين أكثر. ويكمن أحد الأسباب الشائعة لعدم قياس الحصائل المهمة في تجاهل العواقب غير المقصودة، والتي كان من المفترض استيعابها بشكل مثالي في نظرية التغيير. ومن الممكن أن يقلل العمل النوعي السابق في مرحلة تصميم التقييم والتفاعل مع صانعي السياسات والمستفيدين المستهدفين وغيرهم من الأطراف المعنية الرئيسية من المخاطر التي قد تترتب على هذا الأمر.

٨. **التوقعات الخاطئة للتأثيرات الحقيقية للتدخل:** قد يكون للتوقعات المرجوة من التدخل تأثيرات بخلاف التدخل نفسه، وهذا قد يتمثل في استثمارات أو تغييرات في أسعار العقارات تحسبًا لإنشاء بنية تحتية جديدة. وفي حالة عدم وجود معالجة وهمية ذات صلة، فلا يمكن في الغالب فصل التأثيرات المتعلقة بهذه التوقعات بشكل كامل.

٩. **رفع سقف التوقعات والاهتمام بالدلالات الإحصائية العالية للأثر:** يهتم العاملون في حقل الاقتصاد بشدة بمدى "أهمية ما يتحقق ودلالته"، حيث يتم أحيانًا إيلاء الكثير جدًا من الاهتمام للدلالة أو الأهمية الإحصائية، في حين يكون الاهتمام قليلاً جدًا بحجم معامل تأثير المعالجة وأهميته. ومن ثم، قد يفوت الباحثون حقيقة أن الأثر ذا الدلالة الكبيرة جدًا هو في واقع الأمر صغير جدًا من حيث القيمة المطلقة وضئيل للغاية بحيث يتعذر معه إثارة اهتمام صانعي السياسات. وعندما تكون هناك حصيلة واحدة واضحة ناجمة عن التدخل، فإن تحقيق فعالية التكلفة تعد وسيلة جيدة للإبلاغ عن الأثر، ويُفضَّل أن يكون ذلك في جدول يشتمل على أوجه المقارنة مع التدخلات الأخرى التي تؤثر على الحصيلة نفسها. وعندما يتحقق الباحثون من مدى فعالية التكلفة وجدواها الاقتصادية، فقد يؤدي إلى تغيير الاستنتاجات المرتبطة بالسياسة المتبعة، والمستمدة من التركيز على الدلالة أو الأهمية الإحصائية وحدها.

١٠. **الإبلاغ عن نتائج متحيزة:** يجب أن تتناول الدراسات جميع الحصائل المتوقعة وتركز عليها. ويُفضَّل أن تكون هذه الحصائل قد تم تحديدها في مرحلة تصميم التقييم. ويتعين أيضًا تسجيل التصميم

في سجل تقييمات الآثار الإنمائية على الصعيد الدولي التابع للمبادرة الدولية لتقييم الأثر (http://www.3ieimpact.org/en/evaluation/ridie/). وتركز العديد من الدراسات بشكل غير ملائم على المُعامِلات المهمة ذات الدلالة الإحصائية، والتي غالبًا ما تكون إيجابية، مع استبعاد النتائج "الضارة" (السلبية) وغير المهمة. وهناك تحيز ضمني مفاده أن الفشل في تحديد الأثر لا يعكس استنتاجًا حتميًا بأنه لا يوجد أثر، وبالتالي فهو غير قاطع. ويصعب نشر النتائج غير القاطعة في الأعمال التي يراجعها الأقران، وقد تواجه شكوكًا كبيرة. ومن الممكن أن يؤدي التناقص التفاضلي بين المجموعة المُعالَجة والمجموعة الضابطة أيضًا إلى تحيز غير مقصود.

٧-٤ الملخص

توفر التجارب العشوائية المضبوطة أساسًا يتيح التكامل بين العمل التنموي والعمل البحثي، فضلاً عن التحقق من فعالية الابتكارات بشكل منهجي. ولا توجد مقاربة آخرى يمكنها أن تحقق مثل هذا التكامل الوثيق بين البحث والتنمية، وهو أيضًا تكامل يعمل على تحسين التدخل الخاضع للتقييم وكذلك دراسات تقييم الأثر.

تكمن الميزة الرئيسية للتجارب العشوائية المضبوطة في قوة إستراتيجية التحديد، والتي تضمن التوازن بين العناصر التي يمكن ملاحظتها وتلك التي لا يمكن ملاحظتها. وهذا يجعل من السهولة توضيحها لصانعي السياسات. وقد تتمثل إحدى الصعوبات الرئيسية في الاعتراض على الاختيار العشوائي، والذي يمكن مواجهته. وعلى الرغم من أنه يسهل في واقع الأمر التصدي لمثل هذه الاعتراضات، فإن هذا قد يتسبب في الإخلال بنزاهة التصميم. كما يعد ضمان دقة التنفيذ تحديًا رئيسيًا، لذلك قد يتطلب تنفيذ المشروع في المناطق ذات التعيين العشوائي مزيدًا من الإشراف.

ثمة العديد من التصاميم الممكنة للتجارب العشوائية المضبوطة، لذلك سيكون من الممكن غالبًا العثور على تصميم يناسب على الأقل بعض مكونات التدخل. وبالنسبة للتدخلات المتعلقة بالبنية التحتية، سوف تطرح العديد من الاحتمالات التي ستسلط الضوء على أسئلة الجيل الثاني، والتي تدور، على سبيل المثال، حول الصيانة أو الفحص، بدلاً من سؤال الجيل الأول: "هل يعد التصميم أو التدخل فعالاً؟".

المراجع*

Glennerster, R. 2017. The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency. In E. Duflo and A. Banerjee, eds. *Handbook of Field Experiments*. Volume 1, 1st Edition. Amsterdam: North Holland.

Glewwe, P., A. Park, and M. Zhao. 2009. Visualizing Development: Eyeglasses and Academic Performance in Rural Primary Schools in China. Working Paper No. 12–2. Center for International Food and Agricultural Policy, University of Minnesota. http://purl.umn.edu/120032.

Imai, K., G. King, and C. Nall. 2009. The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*. 24. pp. 29–53.

Imai, K., G. King, and E. Stuart. 2008. Misunderstandings among Experimentalists and Observationalists about Causal Inference. *Journal of the Royal Statistical Society*: Series A. 171. pp. 481–502.

Kim, J. and W. Shin. 2014. How to Do Random Allocation (Randomization). Clinics in Orthopedic Surgery. 6 (1). pp. 103–109. https://doi.org/10.4055/ cios.2014.6.1.103.

White, H. 2013. An Introduction to the Use of Randomised Control Trials to Evaluate Development Interventions. Journalof *Development Effectiveness*. 5 (1). pp. 30–249

World Bank. 2014. Philippines Conditional Cash Transfer Program: Impact Evaluation 2012. Report Number 75533-PH. Washington, DC.

مراجع إضافية للاطلاع

Bloom, H. S. 2006. The Core Analytics of Randomized Experiments for Social Research. MDRC Working Papers on Research Methodology. New York. http://www.mdrc.org/sites/default/files/full_533.pdf.

Duflo, E., R. Glennerster, and M. Kremer. 2006. Using Randomization in Development Economics Research: A Toolkit. NBER Technical Working Paper Series. No. 333. http://www.nber.org/papers/t0333.pdf.

King, G., R. Nielsen, C. Coberley, J. E. Pope, and A. Wells. 2011. Avoiding Randomization Failure in Program Evaluation. *Population Health Management*. 14 (1). pp. S11–S22. http://j.mp/2oTEZI6.

Riecken, H. W., R. F. Boruch, D. T. Campbell, N. Caplan, T. K. Glennan, Jr., J. W. Pratt, A. Rees, and W. Williams. 1974. The Purposes of Social Experimentation. *Educational Researcher*. 3 (11). pp. 5–9. https:// www.jstor.org/stable/1175978.

* يشير بنك التنمية الآسيوي (ADB) إلى "الصين" باعتبارها جمهورية الصين الشعبية.

الفصل ٥
التصاميم غير التجريبية
Nonexperimental Designs

رسائل أساسية

- في أغلب الأحيان، يظل من الممكن إجراء تقييم أثر صارم عندما يتعذر إجراء تجارب عشوائية مضبوطة.
- تتعزز تصاميم تقييم الأثر بشكل كبير بامتلاك بيانات تم جمعها، قبل التدخل وبعده، عن المجتمعات الإحصائية الخاضعة للمعالجة والأخرى غير الخاضعة للمعالجة.
- تعتمد الطرق غير التجريبية على تكوين تصور مفاهيمي يفسر التدخلات التي تم تعيينها وعلى فهم ماهية المتغيرات التي تكون بمثابة مؤشرات للمعالجة.
- علاوة على ذلك، تتطلب معظم الطرق غير التجريبية بيانات حول المحددات الأخرى للحصائل بخلاف التدخل المعني.
- تعتمد كل التصاميم غير التجريبية على الافتراضات الشرطية. يمكن التقليل من شأن هذه الافتراضات في الحالات التي توجد فيها قاعدة واضحة أو عامل يحدد شرط الأهلية بشرط قاطع وكانت بيانات خط الأساس متوفرة.
- نظرًا إلى أن الطرق المختلفة تعطي تقديرات صالحة للمجتمعات الإحصائية المختلفة، فإن من الواجب تفسير الاستنتاجات غير التجريبية بعناية.

٥-١ مقدمة

لا يعتبر التعيين العشوائي خيارًا مقبولاً في الغالب لعدة أسباب عملية. فمثلاً، قد لا ترغب بعض الوكالات المعنية بالتطبيق في قبول سمة "العشوائية" في التعيين. أو قد لا تتوفر بنية تحتية شبكية لنشر تسلسل عشوائي بشكل فعال. أو قد يتعذر إجراء تعديلات على مشاريع الاستثمار الكبرى من حيث المكان أو التسلسل. أو قد لا يصبح تقييم الأثر محل اهتمام إلا بعد أن يكون البرنامج قيد التنفيذ، بل قد يكون اكتمل بالفعل. عندما يتعذر استخدام سمة "العشوائية"، يظل بالإمكان غالبًا تقدير الأثر من خلال مجموعة متنوعة من التصاميم غير التجريبية.

يمكن تقسيم الطرق غير التجريبية بشكل عام إلى فئتين:

- *الطرق شبه التجريبية* تشكل مجموعة مقارنة عن طريق الطرق الإحصائية بدلاً من التعيين العشوائي. ومن بين هذه المقاربات مطابقة درجة الميل (Propensity Score Matching) التي يرمز لها اختصارًا بـ(PSM)، والاختلاف في الاختلافات (difference-in-differences) المعروف اختصارًا بـ (DiD)، والوحدات الضابطة الاصطناعية (Synthetic Controls)، وتصاميم انقطاع الانحدار (Regression Discontinuity Designs). تسعى كل هذه المقاربات إلى تأسيس مجموعة مقارنة تكون مماثلة لمجموعات المعالَجة قدر الإمكان. أي أن تكون مجموعتا المعالَجة والمقارنة متوازنتين، بمعنى أن يكون هناك تماثل تقريبي في متوسط قيم الخصائص الملحوظة. ومن ثم يتم حساب الأثر إما باعتباره الفرق في الحصائل بعد التدخل (الفرق الفردي المسجل (بعد التدخل))، أو الاختلاف في الاختلافات (DiD) في الحصائل بين خط الأساس وخط النهاية. وللتحكم بشكل أفضل في تحيز الاختيار، يجوز الجمع بين الفروق وفقًا لأحد أشكال المطابقة.

- *المقاربات القائمة على الانحدار* لا تؤسس مجموعة مقارنة صريحة، رغم أن البيانات يجب أن تشمل ملاحظات حول الوحدات غير المعالَجة أو تلك المعالَجة بدرجة أقل. وتتضمن هذه المقاربات المتغيرات المساعدة (instrumental variables) (تكون في الغالب مُقدَّرة كمربعات صغرى ذات مرحلتين)، ونماذج المعالجة داخلية النشأة (endogenous treatment)، والانحدارات التبديلية داخلية النشأة (endogenous switching regressions)، ونماذج التاثيرات الثابتة (fixed effects)، التي تجمع بين الاختلاف في الاختلافات (DiD) والنموذج الهيكلي. تستند المقاربات القائمة على الانحدار إلى تحديد النموذج الهيكلي الأساسي، أي مجموعة العلاقات السلوكية التي تؤدي إلى إحداث أثر البرنامج. ويجسد هذا النموذج الهيكلي نظرية التغيير. ولهذا، يمكن أن يكون تقدير معاملات النموذج جزءًا مفيدًا من التحليل التسلسلي الاعتيادي. عادةً ما تمثل المقاربات القائمة على الانحدار الخيار الوحيد إذا كانت المعالجة مستمرة (أي تطرأ تغيرات في الكمية/المستوى) بدلاً من المقاربات الثنائية (الوحدات المعالَجة مقابل غير المعالَجة).[1]

يتطلب اختيار هذه المقاربات ووضع تصميم لها فهمًا جيدًا لتصميم التدخل، وبالأخص تنسيب البرنامج. تشمل أسئلة تنسيب البرنامج الأساسية ما يلي:

١. ما الإجراءات والمعايير التي تم استخدامها لاختيار مجالات المشروع/التدخل؟

٢. إذا كان البرنامج يسمح باختيار المشاركين لأنفسهم (الاختيار الذاتي) فما الخصائص المرجحة التي تحدد المشاركة؟

٣. هل من المرجح أن يؤدي تنسيب البرنامج أو الاختيار الذاتي وظيفة متغيرات تتصف بكونها ملحوظة ومترابطة مع الحصائل المعنية؟

١ العديد من التدخلات التي تبدو في البداية ثنائية التفرع قد تكون في الواقع مستمرة، لأن درجة المعالجة تختلف ضمن المجتمع الإحصائي المعالَج. مثلاً، البنية التحتية للطرق ستكون لها تأثيرات مختلفة على التنقلات من حيث الوقت والتكلفة، ويعتمد ذلك على بُعد المنازل عن الطريق.

تشترط المقاربات غير التجريبية توافر بيانات مستقاة من كلا المجتمعين الإحصائيين: المعالَج وغير المعالَج. ويتم استخدام البيانات المستقاة من المجتمع الإحصائي غير المعالَج لعمل مقارنات. وبما أن الملاحظات غير المعالَجة لن تكون جميعها مناسبة للمقارنة، فإن أحجام العينات الأكبر تكون مطلوبة بشكل عام أكثر مما هو الحال مع التجارب العشوائية المضبوطة (RCT). عادةً ما تكون هناك فرصة لإنشاء تصاميم أقوى تؤدي إلى توقعات أكثر موثوقية للأثر في حال توافرت بيانات خط الأساس، بما أن هذه البيانات (١) تسمح بحساب توقعات الاختلاف فى الاختلافات (DiD) بدلاً من الفرق الفردي المسجل بعد التدخل، و(٢) ولأن المتغيرات المستخدمة للمطابقة يجب ألا تكون متأثرة بالتدخل وهذا طبقًا لتعريف الحالة بالنسبة للبيانات التي يتم جمعها قبل التدخل.[٢] تستعرض الأقسام التالية أهم التقنيات، ولكن سيتم تقديمها ووصفها بشكل أكثر توسعًا واتسامًا بالطابع التقني في قسم STATA الوارد بالملحق الأول.

٢-٥ تقديرات الاختلاف في الاختلافات (Difference-in-Differences)

وصف موجز للطريقة

تستند تقديرات الاختلاف فى الاختلافات المعروف اختصارًا بـ (DiD) إلى الفرق في التغييرات التي تشهدها الحصيلة فيما بين مجموعتي المعالَجة والمقارنة بمرور الوقت. تجمع نماذج التأثيرات الثابتة الفروق مع النماذج متعددة المتغيرات من أجل تفسير الفروق بين المتغيرات الملحوظة بمرور الوقت.

وصف الطريقة

تسلك هذه الطريقة مسار مجموعة المقارنة باعتباره مسار الواقع المضاد لمجموعة المعالجة. بمعنى أن التغير في الحصيلة الذي يقع في مجموعة المقارنة يؤخذ على أنه ما كان سيحدث لمجموعة المعالجة في غياب التدخل. لذلك، نحصل على قياس للأثر من خلال طرح هذا التغير في الحصيلة الملحوظة في مجموعة المقارنة (المجموعة الضابطة) من الحصيلة الملحوظة في مجموعة المعالجة. وعند تقدير الأثر، يتم استثناء تأثيرات كل العوامل التي تتغير بمرور الوقت أو تلك التي لا تؤثر على التغييرات بمرور الوقت. بإمكاننا أن نتوقع أن تكون العديد من محددات تنسيب البرنامج والمشاركة ثابتة على مر الزمن، وهذا سبب يجعلنا نتقبل تلك المقاربة بشكل أكبر.

[٢] هناك نقطة ضعف تتمثل في أنه إذا كانت هناك تأثيرات متوقعة للتدخل، فإن هذه التأثيرات إذًا تؤثر على قيم خط الأساس. على سبيل المثال، قد يتغير سلوك الأشخاص نحو الادخار أو الاستثمار إذا نما إلى علمهم أنهم سيحصلون على تمويل أصغر في المستقبل القريب.

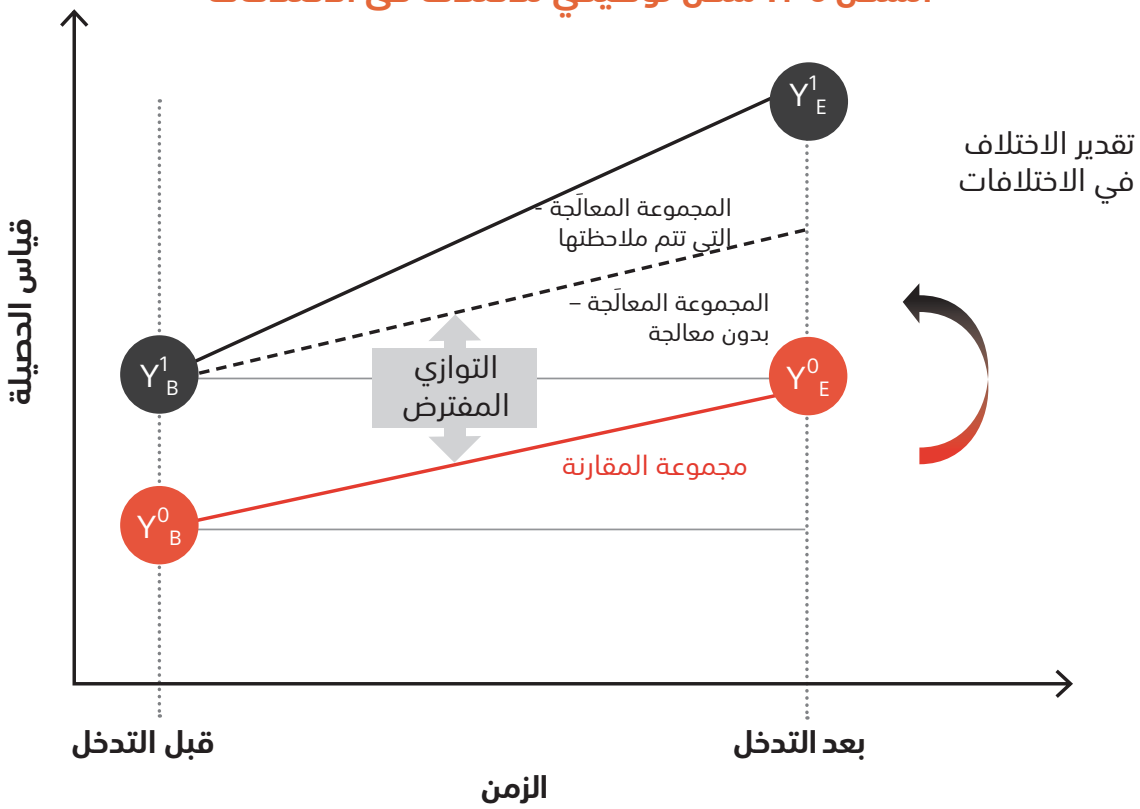
إن مُقدِرات أثر الفرق الفردي (SD) المسجل بعد التدخل عبارة عن الفرق بين المؤشر (Y) (عادةً ما يرمز إلى حصيلة) بعد التدخل (أي، عند خط النهاية = وقت الانتهاء [ويرمز له بالحرف E]) لمجموعة في وجود تدخل (Y^1)، بالمقارنة مع المؤشر الخاص بمجموعة مقارنة في غياب التدخل (Y^0)، الذي له صيغة على هيئة معادلة (٥-١). ويتضح ذلك فيما يلي:

$$SD = Y_E^1 - Y_E^0 \quad (٥-١)$$

تقدير أثر الاختلاف في الاختلافات (DiD) (أو التباين المزدوج) هو عبارة عن الفرق بين التغييرات بمرور الوقت للمجموعتين (الشكل ٥-١). بعبارة أخرى، يستند تقدير الأثر هنا إلى ناتج طرح التغييرات التي طرأت من خط الأساس (قبل التدخل) إلى خط النهاية (بعد التدخل) لمجموعة المقارنة من التغييرات التي طرأت من خط الأساس إلى خط النهاية لمجموعة المعالجة:

$$DiD = (Y_E^1 - Y_B^1) - (Y_E^0 - Y_B^0) \quad (٥-٢)$$

الشكل ٥-١: شكل توضيحي للاختلاف في الاختلافات



المصدر: المؤلفان

يلغي الاختلاف في الاختلافات (DiD) أي فرق في المؤشر بين مجموعتي المعالجة والمقارنة كان موجودًا عند خط الأساس، وهذا أمر مفيد لأن هذه الفروق من الواضح أنها ليست نتيجة للتدخل. وهو يلغي أيضًا تأثيرات الاتجاهات العامة المؤثرة في كل من ملاحظات مجموعتي المعالجة والمقارنة.

يبدو جليًا، ودون تمحيص، أنه إذا كانت قيم مجموعتي المعالجة والمقارنة للمؤشر هي نفسها عند خط الأساس ($Y_B^1 = Y_B^0$) فسيكون تقدير كل من الفرق الفردي والمزدوج متكافئين إذًا، حيث إن الحدين Y_B^1 و Y_B^0 يلغيان بعضهما البعض في المعادلة (٥-٢). ومع ذلك، حتى إذا كانت تقديرات مجموعتي المعالجة والمقارنة لقيمة خط الأساس مماثلة لتخمينات عينة من المجتمع الإحصائي الأساسي نفسه، فقد لا تكون الحالة تمامًا متمثلة في أن $Y_B^1 = Y_B^0$ بسبب خطأ أخذ العينة.

يلزم توافر بيانات خط الأساس للتمكن من تقدير الاختلاف في الاختلافات (DiD). ويجب جمع بيانات خط الأساس قبل التدخل مباشرة. ولكن في حالة كان يجري استخدام التمويل لاطلاق المشروع، فغالبًا ما سيتعذر جمع بيانات خط الأساس. بسبب قلة التمويل. عمليًا، لا يهم ذلك الأمر كثيرًا، حيث إن المطلوب هو بيانات خط الأساس قبل أن يؤثر البرنامج في المجتمع الإحصائي المستفيد المستهدف، وفي الغالب تكون هناك فترة مبدئية للمشروع تتراوح من ٦ إلى ١٢ شهرًا تسبق أي أنشطة على المستوى الميداني، وبالتالي يمكن إجراء مسوحات خط الأساس في هذه الفترة.

حساب الأثر: مثال

يُظهر الجدول ٥-١ بيانات من تقييم أثر "مشروع النقل الريفي في فيتنام" الذي أنشأ وأعاد تأهيل ٥ آلاف كيلومتر من الطرق الريفية ما بين عامي ١٩٩٧ و٢٠٠١ في بلديات تقع ضمن ١٨ مقاطعة عبر ربوع فيتنام. يوضح الجدول قياسات تنمية الأسواق في فيتنام. وتم أخذ هذه القياسات حسب المتوسطات على مستوى البلديات، بالنسبة للبلديات التي تنتمي إلى مجموعتي المعالجة والمقارنة قبل وبعد إنشاء الطرق أو إعادة تأهيلها.

الجدول ٥-١: أثر الطرق الريفية في تنمية الأسواق في فيتنام
(المتوسطات على مستوى البلديات)

	١٩٩٧		٢٠٠٣		الاختلاف في الاختلافات (DiD)
	المشروع	غير المشروع	المشروع	غير المشروع	
الأسواق	٠,٥١	٠,٤٤	٠,٦٢	٠,٤٦	٠,٠٩
رواج الأسواق	١,١٣	١,٠٥	١,٤٣	١,١٦	٠,١٩
المتاجر	٠,٦٣	٠,٥٩	٠,٨٤	٠,٧٧	٠,٠٣
الصيدليات	٠,٦٢	٠,٥٨	٠,٦٩	٠,٥٢	٠,١٣

المصدر: Mu and Van de Walle (2011).

تقدير الفرق الفردي (SD) هو الفرق في المتوسطات على مستوى البلديات لمتغيرات تنمية الأسواق بين البلديات ذات الطرق المحسَّنة وتلك التي تفتقر إلى الطرق المحسَّنة. على سبيل المثال، نسبة البلديات التي بجوارها طريق ريفي وبها سوقها الخاصة ارتفعت إلى ١١٪ (من ٠,٥١ إلى ٠,٦٢) مقارنةً بارتفاع ٢٪ في القرى التي تخلو من الطرق المحسَّنة (من ٠,٤٤ إلى ٠,٤٦). وهذا يعني أن تقدير أثر الاختلاف في الاختلافات (DiD) هو (٠,٦٢-٠,٥١) – (٠,٤٦-٠,٤٤) = ٠,١١-٠,٠٢ = ٠,٠٩ (٩٪). يلغي الاختلاف في الاختلافات (DiD) الفرق المبدئي في تنمية الأسواق بين مجموعتي المعالجة والمقارنة. ولا يفعل ذلك *الفرق* الفردي المسجل بعد التدخل، وبالتالي قد يبالغ في توقع أثر البرنامج (٠,٦٢ – ٠,٤٦ = ٠,١٦ أو ١٦٪).

يوفر الاختلاف فى الاختلافات (DiD) تقديراً غير متحيز لأثر البرنامج إذا ثبت افتراض "الاتجاهات المتوازية" وهو الافتراض بأن المتغير الناتج يتبع المسار نفسه بمرور الوقت في مجموعتي المعالجة والمقارنة في غياب التدخل. ويمكن اختبار افتراض الاتجاهات المتوازية باستخدام بيانات ما قبل التدخل إن توافرت، على الرغم من أن هذا الاختبار لا يقدم شيئًا سوى الدعم للافتراض فحسب ولا يظهر مدى صلاحيته من عدمها. على الأرجح سيثبت الافتراض إذا استخدمنا إحدى طرق المطابقة، مثل مطابقة درجة الميل (المذكورة في القسم ٥-٤)، لضبط الأسباب الملحوظة للفرق في المسار.

تجمع نماذج التأثيرات الثابتة بين مقاربات الاختلاف فى الاختلافات (DiD) والانحدار متعدد المتغيرات. وهذا يسمح بضبط عوامل أخرى قد تؤثر في الحصيلة، من أجل الحصول على تقدير أدق من الاختلاف فى الاختلافات (DiD) البسيط لمتوسط الحصائل. مع تضمين التغييرات الملحوظة في المتغيرات المشتركة بمرور الوقت في الانحدار، لن يغطي افتراض الاتجاهات المتوازية إلا مجموعة من العوامل غير الملحوظة فقط.

ما المطلوب لتطبيق طريقة الاختلاف فى الاختلافات (DiD)؟

يتطلب تطبيق هذه الطريقة بيانات حول الحصائل من مجموعتي المعالجة والمقارنة عند خط الأساس وخط النهاية. من المحبذ توافر بيانات حول الحصائل السابقة للتدخل لاختبار افتراض الاتجاهات المتوازية. إذا كان من المقرر استخدام مقاربة المطابقة، فإذًا يلزم توافر بيانات لمطابقتها أيضًا، وإذا كان سيتم تطبيق نموذج التأثيرات الثابتة، فيجب تصنيف المحددات الأخرى للحصائل حسب البيانات كذلك. قد تكون نماذج الاختلاف فى الاختلافات (DiD) البسيطة محددة في حزمة برامج ستاتا (STATA) باستخدام الأمر "regress" مع متغيرات صورية للدلالة على الزمن، والمعالجة، والمعالجة حسب حدود التفاعل الزمنية، كما هو موضح في الملحق الأول، القسم ٣-٢. قد تكون نماذج التأثيرات الثابتة محددة في حزمة برامج STATA باستخدام الأمر "xtreg"، كما هو موضح في الملحق الأول، القسم ٣-٣.

مزايا وعيوب الاختلاف فى الاختلافات (DiD)

يتسم الاختلاف فى الاختلافات (DiD) بسهولة التطبيق والفهم. إلا أن البيانات لا تتوافر عادةً لاختبار مدى صلاحية النموذج. ولذلك، ستكون حصيلته أدق إذا ما تم استخدامه جنبًا إلى جنب مع تقنية مطابقة أو مع تطبيق نموذج تأثيرات ثابتة، وسيؤدي ذلك إلى ضبط وحدات الخلط الملحوظة على نحو أفضل. ينشئ هذا النموذج متوسط تأثير المعالجة على المُعالَج (ATT) الذي يعزز من فهم التأثيرات على أولئك المشاركين بالفعل ولكنه ليس مقياسًا لتأثيرات التدخل على المجتمع الإحصائي بشكل عام.

٣-٥ الوحدات الضابطة الاصطناعية (Synthetic Controls)

وصف موجز للطريقة

يرى ألبيرتو أبادي وآخرون أن الوحدات الضابطة الاصطناعية تقوم على مفاهيم مقاربات الاختلاف في الاختلافات (DiD). وفي هذا السياق، نحصل على تخمين للأثر بواسطة قياس الفرق في الاتجاهات بين الحصيلة وملاحظات مجموعة المقارنة (Abadie et al. 2010). إلا أن هذه الوحدات الضابطة الاصطناعية لا تعتني كثيرًا بافتراض الاتجاهات المتوازية، إنما تعزز الوحدات الضابطة عن طريق موازنة ملاحظات مجموعة المقارنة بحيث تتطابق تلك الاتجاهات التي تظهر على المتغيرات المشتركة وحصائل الوحدات الضابطة الاصطناعية مع تلك الخاصة بالمعالجة المسبقة للتدخل.

وصف الطريقة

بموجب هذه الطريقة، يتم إعداد لوحة الانحدار (نموذج الانحدار للبيانات المقطعية عبر الزمن) للحصائل حول المتغيرات المشتركة (باستثناء المعالجة)، ويُحدد متغير ثنائي يشير إلى حالة المعالجة للملاحظات الفردية. ويُستخدم إجراء تحسيني من أجل تحديد أوزان لملاحظات مجموعة المقارنة الفردية، بحيث تطابق اتجاهات الوحدات الضابطة الاصطناعية الموزونة في العوامل المشتركة والحصائل مع تلك الخاصة بالوحدات المعالَجة قبل المعالجة في أقرب وقت ممكن. من شأن تطبيق هذه الأوزان على ملاحظات مجموعة المقارنة في أثناء فترة المعالجة أن يسمح بنشوء وحدة ضبط اصطناعي أو واقع مضاد يمكن مقارنته مع الاتجاه الفعلي للمجموعات المعالَجة.

لا تُختبر الدلالة الإحصائية بموجب هذه الطريقة بالطريقة التقليدية. بدلاً من ذلك، تُجرى عمليات محاكاة تجريبية وهمية استنادًا إلى توزيع الملاحظات للمتغيرات في وحدة الضبط الاصطناعي، ويُستخدم التوزيع الاحتمالي للحصائل الممثلة بالنماذج من أجل استنتاج القيمة الاحتمالية (قيمة p).

ما المطلوب للحصول على الوحدات الضابطة الاصطناعية؟

لا تنطبق الوحدات الضابطة الاصطناعية إلا على البيانات الطولية المُجمعة التي تكون فيها المعالجة ثنائية، ولا تنطبق المعالجة إلا على الفترات الزمنية التي تلي الملاحظات. يسمح الأمر "synth_runner" في حزمة برامج STATA بالتطبيق التلقائي (Quistorff and Galiani 2017)، كما ستتم المناقشة في الملحق ١، القسم ٤-١. لا تنطبق التقنيات الحالية إلا على البيانات الطولية المُجمعة المتوازنة بالكامل، ولا تزال الطرق المخصصة للبيانات الطولية المُجمعة غير المتوازنة قيد التطوير. كما هي الحال مع الاختلاف في الاختلافات (DiD)، يجب توافر وحدات غير معالَجة في فترات الملاحظة النهائية. على عكس المقاربات التقليدية، لا تحتاج وحدات الضبط الاصطناعية الكثير من الملاحظات المعالَجة، لذا يمكن تطبيقها على التدخلات صغيرة العدد (small-n)، مثل البنية التحتية الكبرى. ومع ذلك، فهي تحتاج قطعًا إلى تسلسل من الملاحظات السابقة للمعالجة، ويهدف ذلك إلى حث الموازنة على مطابقة الديناميات الزمنية للوحدات الضابطة الاصطناعية مع الوحدات المعالَجة.

مزايا وعيوب الوحدات الضابطة الاصطناعية

المزايا الرئيسية للوحدات الضابطة الاصطناعية هي كما يلي: (١) إنها تسمح بتقدير تأثير المعالجة حتى عندما يكون عدد وحدات المعالَجة صغيرًا، و(٢) تحد من التحيز عند عدم ثبوت افتراض "الاتجاهات المتوازية" الداعم للاختلاف في الاختلافات (DiD). أما من حيث العيوب، لهذه التقنية فعالية أقل من الاختلاف في الاختلافات (DiD) عندما يكون افتراض الاتجاه المتوازي صالحًا. وأيضًا لا تنطبق هذه التقنية على الحزم الإحصائية الحالية عند استيفاء متطلبات البيانات المحددة نسبيًا (البيانات الطولية المُجمعة المتوازنة مع فترات طويلة من الملاحظات السابقة للمعالجة والمعالجة الثنائية). العناية أيضًا مطلوبة لضمان ألا تتضمن وحدات المقارنة المأخوذة بعين الاعتبار ملاحظات تخضع لتطورات مسببة للخلط بمرور الوقت. قد يسفر أيضًا الافتقار إلى الاختبارات التقليدية للدلالة الإحصائية عن تفسير النتائج بشكل مربك في أوساط المعنيين بوضع السياسات. على غرار طريقة الاختلاف في الاختلافات (DiD)، تنشئ هذه الطريقة متوسط تأثير المعالجة على المُعالَج (ATT).

٤-٥ مطابقة درجة الميل (Propensity Score Matching)

وصف موجز للطريقة

يرى كل من روزينباوم وروبين أن درجة الميل هي الاحتمالية المقدَّرة لانتماء فرد أو وحدة إلى مجموعة المعالجة علمًا بالخصائص الملحوظة المستقاة من نموذج انحدار المشاركة (Rosenbaum and Rubin 1983). يمكن استخدام درجات الميل المقدَّرة لدعم مجموعة واسعة من طرق تقييم الأثر، بما يشمل مطابقة درجة الميل (PSM)، وأكثر التقنيات شيوعًا: المقدرات الموزونة، والانحدارات الموزونة. تنشئ مطابقة درجة الميل مجموعة مقارنة من ملاحظات غير معالَجة عن طريق مطابقة ملاحظات المعالجة مع واحدة أو أكثر من الملاحظات المستقاة من العينة غير المعالجة، استنادًا إلى الخصائص الملحوظة. وفي هذا السياق، تتم مطابقة الوحدات المعالّجة مع الوحدات غير المعالّجة المماثلة لها من حيث درجة الميل. في مقاربات درجة الميل، يتعذر تضمين الاختيار من الخصائص غير الملحوظة، لذا قد تعطي تقديرات متحيزة إذا كانت لهذه الخصائص دلالة.

وصف الطريقة

للحصول على تطابق مثالي، يتعين مطابقة كل فرد أو وحدة في مجموعة المعالجة مع شخص أو وحدة في مجموعة المقارنة بحيث يكونان متطابقين في جميع الخصائص الملحوظة ذات الصلة، مثل العمر والتعليم والدين والمهنة والثراء والتصرف تجاه المخاطر وما إلى ذلك. من الواضح أن من المستحيل تحقيق ذلك. ولكنه غير ضروري أيضًا. تتطلب الموازنة أن يكون متوسط الخصائص لمجموعتي المعالجة والمقارنة متماثلاً قبل التدخل. على الرغم من أن المطابقة الفردية ستحقق ذلك، فهي ليست عملية. ومع ذلك، توجد طرق مطابقة أخرى تُعد عملية وتضمن الموازنة، ومن بين هذه المقاربات الأكثر شيوعًا مطابقة درجة الميل. ففي مطابقة درجة الميل هذه، لا تكون المطابقة على أساس كل خاصية منفردة ولكن على أساس كل رقم فردي: درجة الميل.

درجة الميل هي عبارة عن احتمال شرطي. بشكل أكثر تحديدًا، هي مدى احتمال مشاركة شخص/وحدة ما في التدخل نظرًا إلى الخصائص الملحوظة له/لها. نحصل على هذا الاحتمال من "معادلة المشاركة": انحدار احتمالي أو لوغاريتمي يكون فيه المتغير التابع ثنائي التفرع، آخذًا القيمة ١ لكل مَن شارك في التدخل، و٠ لكل مَن لم يشارك. تتضمن المتغيرات المفسِّرة كل المتغيرات الملحوظة (الفرد، والأسرة أو الشركة، والمجتمع أو السوق) التي قد تؤثر على المشاركة، ولكنها لا تتأثر بالتدخل. لا تتأثر قيم خط الأساس لجميع المتغيرات (بما في ذلك الحصائل) بالتدخل. لذلك، يساعد توافر بيانات خط الأساس على تحقيق مطابقة أدق. هناك ابتكار حديث وهو *درجة ميل موازنة المتغيرات المشتركة (covariate balancing propensity score)*، وهو يطبق موازنة تقديرات درجة الميل لضمان أن المتغيرات المشتركة متوازنة بشكل أكثر كمالاً قبل تطبيق درجات الميل في عملية المطابقة (Imai and Ratkovic 2014).

بما أن التحليل لا يعتني بدلالة المعامِلات الفردية، فإن جميع المتغيرات التي تتوافر لها بيانات يمكن تضمينها في معادلة المشاركة بشرط استيفاء معيار عدم تأثرها بالتدخل. وبعبارة أخرى، تعدد العلاقات الخطية ليس شاغلاً في معادلة المشاركة، إذا كانت مقدَرة باستخدام درجة الميل التقليدية. مثلاً، في تحليل المياه المنقولة بالأنابيب في الهند (Jalan and Ravallion, 2001)، يوجد ٩٠ متغيرًا مستقلاً في معادلة المشاركة: ١٥ متغيرًا صوريًا على مستوى الولاية، و٢٠ متغيرًا على مستوى القرية، و٥٥ متغيرًا على مستوى المنازل (الجدول ٥-٢). ومع ذلك، تحتاج درجات ميل موازنة المتغيرات المشتركة إلى أن تكون أكثر انتقائية في المتغيرات المستخدمة بحيث يمكن تحقيق الموازنة.

درجة ميل الفرد هي القيمة الملائمة المستقاة من معادلة المشاركة. بعد حساب درجات الميل لكل الملاحظات، تُحدد منطقة الدعم المشترك. في الملاحظات غير المعالَجة، إذا كانت الملاحظات ذات درجة ميل أقل من أدنى قيمة ملحوظة في مجموعة المعالجة، فهي تكون خارج الدعم المشترك ولا تُستخدم. على نحو مماثل، الملاحظات في مجموعة المعالجة التي لها درجة ميل أعلى من أعلى قيمة ملحوظة في المجموعة غير المعالَجة لا تُستخدم. وتلك الملاحظات التي نستبقي عليها من المجموعة غير المعالجة تشكل مجموعة المقارنة.

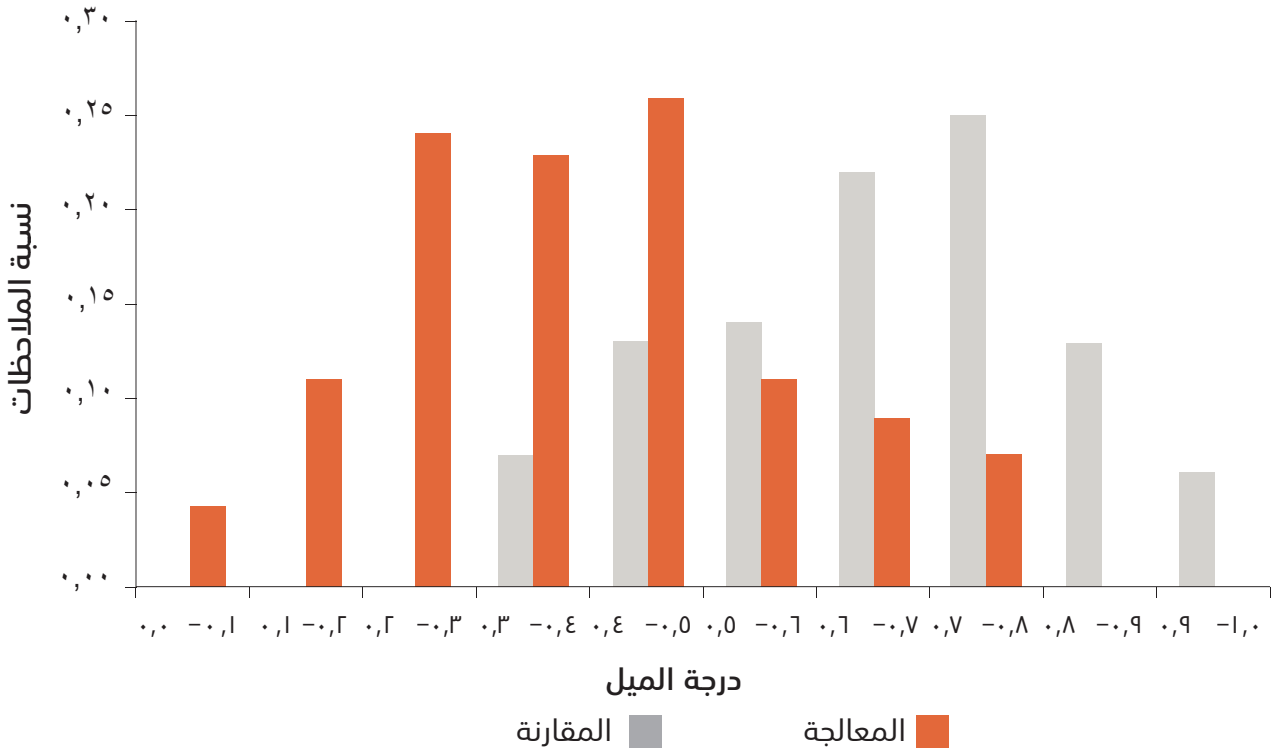
الجدول ٥-٢: مثال للمتغيرات المستقلة في معادلة المشاركة لمطابقة درجة الميل

المستوى	المتغيرات
الولاية	متغير صوري على مستوى الولاية
القرية	حجم القرية (لوغاريتم)، نسبة المساحة المزروعة الإجمالية التي تخضع للري، ما إذا كان يوجد مركز للرعاية النهارية في القرية، ما إذا كانت توجد مدرسة ابتدائية في القرية، ما إذا كانت توجد مدرسة إعدادية في القرية، ما إذا كانت توجد مدرسة ثانوية في القرية، نسبة الطالبات إلى الطلاب في القرية، نسبة الطالبات إلى الطلاب بالنسبة للأقليات، طريق ممهد رئيسي مؤدي إلى القرية، ما إذا كانت توجد محطة حافلات داخل القرية، ما إذا كانت توجد محطة سكة حديد داخل القرية، ما إذا كان يوجد مكتب بريد داخل القرية، ما إذا كانت توجد مرافق خطوط هاتفية في القرية، ما إذا كان يوجد مركز تلفزيوني مجتمعي في القرية، ما إذا كانت توجد مكتبة في القرية، ما إذا كان يوجد بنك في القرية، ما إذا كانت توجد سوق في القرية، نسبة الطلاب إلى المعلمين في القرية
الأسرة	ما إذا كانت الأسرة تنتمي إلى قبيلة مجدولة، ما إذا كانت الأسرة تنتمي إلى طائفة مجدولة، ما إذا كانت الأسرة هندوسية، ما إذا كانت الأسرة مسلمة، ما إذا كانت الأسرة مسيحية، ما إذا كانت الأسرة سيخية، حجم الأسرة، استغلال الحيازات الأرضية المستخدمة في الزراعة، ما إذا كانت الأسرة تمتلك المنزل، ما إذا كانت الأسرة تحوز ممتلكات أخرى، ما إذا كانت الأسرة تمتلك دراجة، ما إذا كانت الأسرة تمتلك ماكينة خياطة، ما إذا كانت الأسرة تمتلك ماكينة درس (درَّاسة)، ما إذا كانت الأسرة تمتلك غربالاً هوائيًا، ما إذا كانت الأسرة تمتلك عربة ثيران، ما إذا كانت تمتلك الأسرة جهاز راديو، ما إذا كانت الأسرة تمتلك جهاز تلفزيون، ما إذا كانت الأسرة تمتلك مروحة، ما إذا كانت الأسرة تمتلك ماشية، طبيعة المنزل، حالة المنزل، عدد الغرف في المنزل، ما إذا كانت الأسرة تمتلك مطبخًا منفصلاً، ما إذا كان المطبخ جيد التهوية، ما إذا كان المنزل به كهرباء، مهنة رب الأسرة، ما إذا كان أفراد الأسرة من الذكور يستمعون إلى الراديو، ما إذا كان أفراد الأسرة من الإناث يستمعن إلى الراديو، ما إذا كان أفراد الأسرة من الذكور يشاهدون التلفزيون، ما إذا كان أفراد الأسرة من الإناث يشاهدن التلفزيون، ما إذا كان أفراد الأسرة من الذكور يقرؤون الصحف، ما إذا كان أفراد الأسرة من الإناث يقرأن الصحف، نسبة أفراد الأسرة الذين تزيد أعمارهم عن ٦٠ عامًا، نسبة الإناث بين البالغين، نسبة الذكور بين الأطفال، نسبة الإناث بين الأطفال، جنس رب الأسرة، الحالة الاجتماعية لرب الأسرة، تعليم رب الأسرة، ما إذا كان رب الأسرة حاصلاً على شهادة المرحلة الثانوية العليا، المساحة المزروعة الإجمالية، المساحة المروية الإجمالية، حجم القطعة الأرضية التي تحوزها الأسرة

المصدر: Jalan and Ravallion (2001).

يوضح الشكل ٥-٢ التوزيع النموذجي لدرجات الميل. التوزيع لمجموعة المعالجة يكون على يسار توزيع المجموعة غير المعالَجة، أي أفراد مجموعة المعالجة الذين يميلون إلى امتلاك درجات ميل أعلى من أولئك في المجموعة غير المعالَجة. لا يوجد فرد في مجموعة المعالجة لديه درجة ميل أقل من ٠,٣، ولا يوجد فرد في المجموعة غير المعالَجة لديه درجة ميل أكثر من ٠,٨. لذا، لتأسيس منطقة الدعم المشترك، لا تُستخدم الـ ٣٩٪ من ملاحظات المجموعة غير المعالَجة التي تتراوح درجة ميلها بين ٠ و٠,٣، والـ ١٩٪ من ملاحظات مجموعة المعالجة التي تتراوح درجة ميلها بين ٠,٨ و١,٠. (عمليًا، سيستخدم ذلك نقطة انقطاع أكثر دقة، بدلاً من تلك الموضحة حسب التصنيف الفئوي للبيانات.)

الشكل ٥-٢: مثال لتوزيع درجات الميل



المصدر: المؤلفان

يتطابق كل عضو في مجموعة المعالجة مع واحد أو أكثر من أعضاء مجموعة المقارنة. ويُستخدم نموذج "مطابقة المجاور الأقرب" لمطابقة فرد من مجموعة المعالجة مع أكثر فرد من مجموعة المقارنة يقاربه في درجة الميل. ويشيع إجراء المطابقة مع أكثر من مجاور واحد، مثل أقرب خمسة أفراد. يُستخدم نموذج "مطابقة الفرجار" لمطابقة كل ملاحظات مجموعة المقارنة في نطاق "مسافة" معينة (أي ما يصل إلى أقصى فرق في درجات الميل)، ويُستخدم نموذج "المطابقة الجوهرية" لمطابقة كل ملاحظات مجموعة المقارنة في منطقة الدعم المشترك التي تحمل موازنة متناسبة عكسيًا مع المسافة. بالإمكان مطابقة ملاحظة فردية في مجموعة المقارنة مع عدة ملاحظات مختلفة في مجموعة المعالجة. ويُستبعد أولئك الأفراد الذين ينتمون إلى مجموعة المقارنة ولا يطابقون أولئك الذين ينتمون إلى مجموعة المعالَجة.

بمجرد إتمام المطابقة، يُجرى اختبار للموازنة عن طريق مقارنة متوسط خصائص مجموعتي المعالجة والمقارنة. من المفترض ألا تنشأ فروق كبيرة في متوسط الخصائص بين المجموعتين.

تقل الفروق بين المجموعتين بعد إجراء المطابقة. هذا التقارب في الخصائص يُلاحظ أيضًا على مستوى درجة الميل، حيث يمثل هذا العدد متوسطًا موزونًا لتلك الخصائص. في تحليل مطابقة درجة الميل للمياه المنقولة بالأنابيب في الهند ((Jalan and Ravallion (2001)، كان متوسط درجة الميل في مجموعتي المعالجة والمقارنة ٠,٥٤٩٥ و٠,١٩٣٣ على التوالي قبل المطابقة، و٠,٣٧٤٣ و٠,٣٧٤٢ بعد المطابقة.

وأخيرًا وليس آخرًا، يتم تقدير الأثر (سواء الذي نحصل عليه من الفرق الفردي أو المزدوج) باتباع الخطوتين التاليتين: (١) حساب الفرق بين المؤشر لفرد من مجموعة المعالجة ومتوسط قيمة الأفراد المطابقين من مجموعة المقارنة. و(٢) عن طريق إيجاد متوسط كل هذه الفروق. على سبيل المثال، إذا كانت المطابقة لأقرب خمسة مجاورات (تضع المعادلة التالية الفرق الفردي في الاعتبار):

$$IMPACT\ (\text{الأثر}) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{1,t+1,i} - \frac{1}{5}\sum_{j=1}^{5}Y_{0,t+1,i,j}\right) \quad (٣-٥)$$

يوضح الجدول ٥-٣ مثالاً عدديًا باستخدام البيانات الواردة في حصائل التعلم لطلاب الصف السادس في اختبار معياري. يُظهر العمود (١) درجة الاختبار لفرد (ف) من مجموعة المعالجة، أما الأعمدة من ٤ إلى ٨، فهي تخص أقرب خمسة مجاورين. ويظهر متوسط درجة المجاورين الخمسة في العمود (٢)، والفرق في درجات الاختبار للفرد من مجموعة المعالجة وأقرب درجة المجاورين يظهر في العمود (٣).

الجدول ٥-٣: حساب تقدير أثر درجة الميل: مثال باستخدام بيانات درجة الاختبار

الملاحظات (ف)	Y_{1i} (١)	Y_{0i}(متوسط) (٢)	Y_{0i}-Y_{1i}(متوسط) (٣)	$Y_{0i(١)}$ (٤)	$Y_{0i(٢)}$ (٥)	$Y_{0i(٣)}$ (٦)	$Y_{0i(٤)}$ (٧)	$Y_{0i(٥)}$ (٨)
١	٤٨,٢	٤٢,٤	٥,٨	٤٤,١	٤٥,١	٤٣,٨	٤٣,٢	٣٥,٨
٢	٥٠,٢	٤٢,٦	٧,٦	٤٢,١	٤٥,٢	٤٨,١	٣٨,٤	٣٩,٣
٣	٥٠,٦	٤٣,١	٧,٥	٤٠,٨	٤٣,٧	٤٥,٣	٤٤,١	٤١,٨
٤	٤٨,١	٣٨,٩	٩,١	٤٣,٦	٣٥,٦	٣٦,٩	٤١,٤	٣٧,٢
٥	٦٩,٠	٥٩,٧	٩,٣	٥٥,٦	٥٧,٦	٥٧,١	٦٢,٤	٦٥,٨
...	...	...	...	...	...	...	...	...
١٩٩	٥٨,٦	٥٢,٢	٦,٤	٥٥,٥	٤٨,٢	٥٤,٧	٥٣,٤	٤٩,١
٢٠٠	٤٥,٤	٣٩,٣	٦,١	٤١,٢	٣٩,١	٣٨,٧	٤٠,١	٣٧,٥
المتوسط	٥٢,٩	٤٥,٥	٧,٤					

... = حُذفت الصفوف المتوسطة لأغراض العرض التقديمي.
المصدر: المؤلفان

ما المطلوب لإجراء مطابقة درجة الميل؟

تتطلب مطابقة درجة الميل بيانات تخص كل من مجموعتي المعالجة وغير المعالَجة، التي تنشأ منها مجموعة المقارنة. يجب أن تتضمن البيانات خصائص المجتمع والأسرة والفرد، تلك التي تحدد مشاركة البرنامج من حيث تنسيب البرنامج والاختيار الذاتي. يجب أن تكون كلتا العينتين أكبر من حجم العينة المقترحة عن طريق حسابات القوة الإحصائية البسيطة، بما أننا نتجاهل الملاحظات التي تقع خارج منطقة الدعم المشترك. عمليًا، لا يحتاج الباحث إلى أداء الخطوات أعلاه يدويًا. تسمح حزم البرامج الإحصائية بإجراء التحليل بتنفيذ أمر واحد فقط، مثل "teffects" المتبوع بالأمر "psm" في حزمة برامج ستاتا (STATA)، كما نوضح في الملحق الأول، القسم ٥-٢.

مزايا وعيوب مطابقة درجة الميل

هناك ميزتان رئيسيتان لمطابقة درجة الميل، وهما: (١) إمكانية إجراء معالجة ثنائية التفرع دائمًا عند توافر كم كافٍ من البيانات (ومن الممكن اعتبارها "طريقة الملاذ الأخير"). و(٢) إمكانية إتمام المطابقة باستخدام القيم الوقائعية (أي الفعلية) حتى في غياب بيانات خط الأساس. وفي حالة عدم توافر بيانات خط الأساس، فلا يزال من الممكن إتمام المطابقة على أساس خصائص ثابتة على مر الزمن، مثل الجنس والدين، واستدعاء الخصائص القبلية للتدخل التي يسهل استدعاؤها، مثل مستوى تعليم رب الأسرة وملكية الأصول الكبيرة. يمكن استخدام مطابقة درجة الميل (PSM) أيضًا للحصول على متوسط تأثير المعالجة على المُعالَج (ATT) ومتوسط تأثير معالجة (ATE) يصح تعميمه على المجتمع الإحصائي برمته.

ما يعيب طريقة مطابقة درجة الميل (PSM) هو اعتماده على المطابقة على أساس الوحدات الملحوظة. فإذا افترضنا تأثر الاختيار (المشاركة) بوحدات غير ملحوظة، فإذًا سنستخرج من مطابقة درجة الميل (PSM) تقديرات أثر متحيزة لتقديرات الفرق الفردي الوقائعي. يرى بعض التجريبيين ككينغ ونيلسن أن تطبيق مطابقة درجة الميل (PSM) قد يفاقم تأثيرات الوحدات غير الملحوظة، حيث إن ملاحظات مجموعة المقارنة المستخدمة تتسم بسلوك مشاركة ينافي بشكل أكبر ما تقترحه الوحدات الملحوظة (King and Nielsen 2016).

٥-٥ موازنة درجة الميل والتقنيات القوية المزدوجة (Propensity Score Weighting and Double Robust Techniques)

وصف موجز للطريقة

إن المطابقة ما هي إلا شكل واحد من أشكال تطبيق درجة الميل. يمكن تطبيق درجة الميل أيضًا في موازنة الملاحظات، وذلك من أجل تحقيق توازن في المتغيرات المشتركة بين ملاحظات المجموعتين المعالَجة وغير المعالَجة (Lunceford and Davidian 2004). وأبسط مقاربة للمطابقة هي "ترجيح الاحتمال العكسي" (IPW) الذي يحسب الفرق في المتوسط الموزون لقيم الفرد الممثلة بالمتغيرات التابعة، وذلك بالنسبة لملاحظات المجموعتين المعالَجة وغير المعالَجة. يمكن العثور على المزيد من التفاصيل حول هذه الطريقة في الملحق الأول، القسم ٥-٣.

يجوز الجمع بين الموازنة وتقنيات الانحدار عندما تكون هناك علاقة معروفة بين المتغيرات المشتركة والحصائل. علاوة على ذلك، هذا الجمع بين انحدار الموازنة والحصيلة يمكن إتمامه بأسلوب "مزدوج القوة" بحيث يتم تخمين متوسط تأثير المعالجة (ATE) غير المتحيز فقط في حالة تحديد انحدار درجة الميل أو انحدار النتائج تحديدًا صحيحًا. هناك شكل من الأشكال الرائدة للتقنيات مزدوجة القوة (double robust)، الذي يُطلق عليه *الانحدار الموزون لترجيح الاحتمال العكسي المعزز (augmented inverse probability weighted regression)*، *بحسب المصطلح التقني*.

وصف الطريقة

يتألف الانحدار مزدوج القوة من نموذج درجة ميل وانحدار نتائج تبديلي للمتغير التابع مقابل المتغيرات المستقلة التي تقيده بشروطها. يتميز انحدار النتائج بسمة "التبديل"، أي أنه يسمح بتقدير ملاحظات مجموعتي المعالَجة وغير المعالَجة بشكل منفصل. وتُحسب موازنات الاحتمال العكسي استنادًا إلى درجة الميل، وتُستخدم بالاقتران مع القيم المتنبأ بها استنباطًا من المعادلتين الناتجتين لإيجاد متوسط موزون يمثل متوسط تأثير المعالجة (ATE).

ما المطلوب لإجراء انحدار مزدوج القوة؟

تمتثل التقنية القوية المزدوجة للمتطلبات العامة نفسها التي تمتثل مطابقة درجة الميل (PSM) لها، بل ويمكن استخدامها مع أنواع البيانات نفسها أيضًا. ويُطلق عليها اسم "teffects" في حزمة برامج STATA، مع الخيار "aipw" كما هو موضح في الملحق الأول، القسم ٥-٤. ومع ذلك، تستوجب هذه التقنية تحديد المتغيرات التي تقيد النتائج، في إطار علاقة هيكلية. بوجه عام، يجب أن يظهر في نموذج درجة الميل جميع المتغيرات الممثلة في المعادلة الناتجة، ولكن يمكن أن يشتمل نموذج درجة الميل (وليس هذا شرطًا) على محددات إضافية للمشاركة لا تؤثر بشكل مباشر على الحصائل.

مزايا وعيوب الانحدار مزدوج القوة

يكون الانحدار مزدوج القوة أقل عرضة لخطأ المواصفات أو التحيز مقارنةً بمقاربة مطابقة درجة الميل (PSM)، حيث إنه يوفر فرصتين لتحديد العلاقات بشكل ملائم. إلا أن العيب في هذه الطريقة يكمن في أنها لا تسمح سوى بتخمين متوسط تأثير المعالجة (ATE). بإمكان ترجيح الاحتمال العكسي (IPW) أن يساعدنا على تخمين متوسط تأثير المعالجة على المُعالَج (ATT)، إلا أنه لا يضاهي خاصية التقنية القوية المزدوجة من حيث الحساسية الأقل تجاه المواصفات. وكلا التقنيتين، مثل مطابقة درجة الميل (PSM)، لا يستطيعان تفسير الاختيار على أساس الوحدات غير الملحوظة. إلا أنه عندما يكون متوسط تأثير المعالجة (ATE) محل الاهتمام ويكون من الممكن تحديد النتائج الشرطية للمتغيرات في نموذج هيكلي، يسمح الانحدار القوي المزدوج باستيعاب احتمالات متعددة لتجنب التحيز.

٥-٦ تصميم انقطاع الانحدار والسلاسل الزمنية المتقطعة (Regression Discontinuity Design and Interrupted Time Series)

وصف موجز للطريقة

يُستخدم تصميم انقطاع الانحدار (RDD) عندما تكون هناك قاعدة تحدد عتبة الأهلية للمشاركة في البرنامج، مثل خط الفقر، أو القرى الواقعة على أي من جانبي الحدود الإدارية، أو الدرجة المستخدمة لتصنيف المشاريع الفرعية المحتملة (Thistlewaite and Campbell 1960). الافتراض الذي يخضع للاختبار كجزء من هذه العملية، يشير إلى أن الوحدات القريبة من أي من جانبي الحدود تكون متماثلة بما يكفي، ما يؤهلها للاستبعاد من البرنامج من أجل تكوين مجموعة مقارنة صالحة.

الفرق في الحصيلة بين هذه الوحدات القريبة من أي من جانبي الحدود، كما هي مقاسة حسب الانقطاع في خط الانحدار عند هذه النقطة، يمكن إسناده إلى البرنامج، وكذلك الحال بالنسبة لقياس الأثر.

يرى كل من سوين وكامبل أن السلاسل الزمنية المتقطعة (ITS) هي تطبيق خاص لتصميم انقطاع الانحدار (RDD) تكون فيه العتبة (الحد) عبارة عن نقطة زمنية يتجلى عندها تأثير البرنامج (Sween and Campbell 1965). وتكون هذه الطريقة ملائمة لا سيما عندما تكون فعالية التدخل فجائية، لا تدريجية، مثل إكمال جسر أو وصلة نقل طاقة كبيرة.

وصف الطريقة

يمكن استخدام تصميم انقطاع الانحدار (RDD) عندما تكون هناك قاعدة تحدد عتبة الأهلية للمشاركة في البرنامج، وحيث عتبة الأهلية مستندة إلى متغير تعيين مستمر مقدَّر لجميع الوحدات المؤهلة بشكل محتمل للتعيين (الأفراد، الأسر، الشركات، إلخ). مثلاً، الأسر التي تكون أعلى خط الفقر أو أدناه، أو الشركات التي تكون أعلى تصنيف ائتماني معين أو أدناه، أو الطلاب الذين إذا تجاوزوا درجة اختبار معينة سيحصلون على منحة دراسية، أو النساء اللائي أعلى سن معين أو أدناه لبرنامج صحي ما. إذا طُبقت قاعدة عتبة الأهلية بشيء من النقص، يصبح من الممكن استخدام الاختلاف على أساس هذه المقاربة، يُسمى *تصميم انقطاع الانحدار* المشوّش (fuzzy RDD) يجب ألا يكون متغير التعيين متغيرًا يمكن معالجته ليصبح مؤهلاً للبرنامج، حيث إن هذا سيفتح بابًا للتحيز في الاختيار.

في حالة السلاسل الزمنية المتقطعة (ITS)، تكون العتبة عبارة عن النقطة الزمنية التي تُنفذ عندها السياسة أو البرنامج. مثلاً في حالة تطبيق سياسة تنص على رفع القيود على الاتصالات، تكون هذه النقطة الزمنية مشتركة بالنسبة لكل العتبات. ولكن هناك تدخلات أخرى، مثل إدخال الكهرباء أو وصلة إلى نظام صرف صحي، قد تؤثر على المجتمعات المختلفة عند نقاط زمنية مختلفة.

يجب أن تكون عتبة الأهلية فريدة بالنسبة للبرنامج. في الهند، مثلاً، تُستخدم حالة "تحت خط الفقر" الرسمية كمعيار للأهلية في العديد من البرامج. في مثل هذه الحالة، يتعذر على تصميم انقطاع الانحدار (RDD) فصل تأثير برنامج محدد باستخدام عتبة "تحت خط الفقر" في مجالات حيث تستخدم برامج أخرى العتبة نفسها. على نحو مماثل، يصبح الأشخاص مؤهلين للحصول على المعاش بمجرد وصولهم إلى سن التقاعد. ومع ذلك، يتوقف الأشخاص أيضًا عن العمل، ولهذا يتعذر على تصميم انقطاع الانحدار (RDD) فصل تأثيرات التقاعد والمعاش من تلك الحصائل العديدة (مثل الحصائل الصحية) التي تتأثر بشكل كبير بكلا الحدثين.

طرأ هذا السياق في تقييم "برنامج الطوابع الغذائية في منغوليا" (Mongolia Food Stamps Program)، المدعوم من قِبل البنك الآسيوي للتنمية (ADB 2014). ففي اختبار قياس مستوى الدخل الفعلي (PMT) نفسه حيث كانت العتبة مشتركة، استُخدم معيار الأهلية نفسه لكل من برنامج الطوابع الغذائية وبرنامج Medicard. لذلك، كما ذكرنا في التقرير، "بما أن كلا البرنامجين تم تقديمهما وفقًا لمعيار أهلية متطابق (تقدير اختبار قياس

مستوى الدخل الفعلي) ويوجد بالتالي تداخل مثالي في اختيار المعالجة، يكون من المستبعد من الناحية الفنية أن يفصل تقييم الأثر بين آثار البرنامجين. نتيجة لذلك، أي أثر يتم التوصل إليه يجب اعتباره "أثرًا مركبًا" لكلا البرنامجين".

لا شك في أننا نستطيع لمس بعض الفروق بين الوحدات المجاورة للعتبة من أعلاها وأسفلها. بالإضافة إلى ذلك، قد يكون معيار العتبة مرتبطًا بالحصيلة، ولهذا ينشأ تحيز في الاختيار عند إجراء مقارنات بسيطة. على سبيل المثال، تُقدم المنح الدراسية على سبيل تحسين حصائل التعلم، إلا أن أولئك الذين لديهم حصائل تعلم أفضل يتم انتقاؤهم لتضمينهم في البرنامج. كمثال، يزداد ترجيح إصابة النساء الأكبر سنًا بسرطان الثدي، ولهذا يتم انتقاء النساء الأكبر سنًا لإخضاعهن للفحص.

على الرغم مما سبق، أولئك المقتربون من جانبي العتبة يكونون على درجة أكبر من التماثل. مثلاً، الطلاب الذين تتراوح درجاتهم في الاختبارات بين ٥٨,٠ و٥٩,٩ ولم يتم اختيارهم للاستفادة من منحة دراسية لوجودهم دون عتبة التأهيل البالغ ٦٠,٠، ليس هناك فرق كبير بينهم وبين أولئك الذين حصلوا على درجات تتراوح بين ٦٠,٠ و٦٠,٩ وتم قبولهم. في العديد من الحالات، يجوز إسناد الفروق بين هاتين المجموعتين إلى خطأ في القياس بشكل أقرب من أي عوامل أخرى. يستند انقطاع الانحدار إلى مقارنة الفرق في متوسط الحصائل لهاتين المجموعتين.

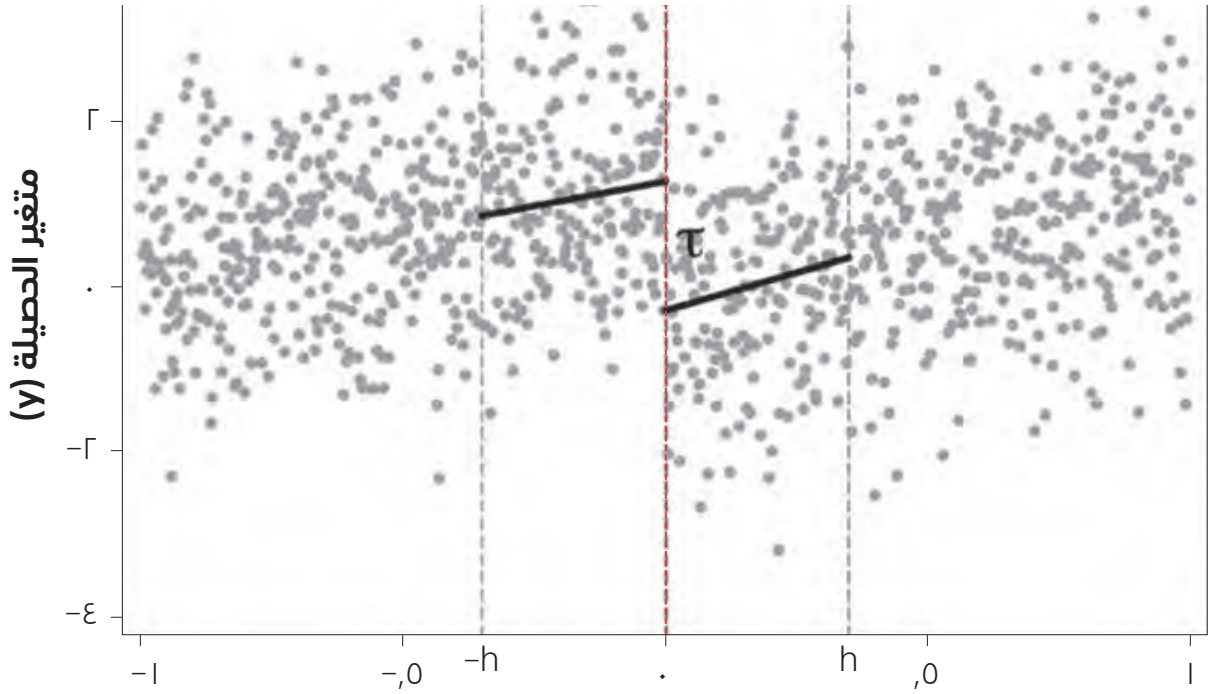
في حالة السلاسل الزمنية المتقطعة (ITS)، من المقبول أن توجد عوامل أخرى مؤثرة في الحصائل محط الاهتمام. على الرغم من ذلك، لن تلعب هذه العوامل الأخرى دورًا كبيرًا مباشرًا على مدار فترات زمنية بالنسبة لكلا جانبي النقطة الزمنية التي يتم إجراء التدخل فيها، لأنها لن تغير الكثير في مثل هذه الفترة الزمنية القصيرة.

تستخدم مقاربة تكرارية لتحديد الهامش المحيط بعتبة الأهلية. في البداية، نحدد هامشًا صغيرًا، ونتحقق من موازنة مجموعتي المعالجة والمقارنة الناتجتين. إذا كانت المطابقة قريبة، يجوز حينها توسيع الهامش قليلاً مع التحقق من الموازنة مجددًا. يمكن تكرار ذلك حتى تبدأ العينات في التباعد. على الرغم من أن الموازنة تقوم على الخصائص الملحوظة، فإن معيار الأهلية إذا كان مطبقًا وكانت المشاركة للمجتمع الإحصائي المؤهل واسعة، لا يكون هناك سبب لتوقع عدم توازن الخصائص غير الملحوظة.

بمجرد تشكيل العينة، نطبق خط الانحدار على العينة القريبة من العتبة (الشكل ٥-٣).[٣] وتكون عينة الانحدار مقيدة بالملاحظات المجاورة لكلا جانبي العتبة، أو عبر عرض نطاق واحد أو أكثر (كما هو موضح في الجدول ٥-٤ كمثال).

٣ لتبسيط هذا المثال التوضيحي، نفترض وجود علاقة خطية. عمليًا، يتم اختبار الشكل غير الخطي، بما أن افتراض الخطية قد يعطي انطباعًا للانقطاع، وهذا لا ينطبق هنا في الواقع. ومع ذلك، اختبار الأشكال الوظيفية الأكثر تعقيدًا قد يتطلب استخدام العينة الكاملة بهدف موازنة الملاحظات البعيدة عن نقطة الانقطاع.

الشكل ٥-٣: مثال لانقطاع الانحدار حول عتبة الأهلية لمتغير التعيين



τ = متوسط تأثير المعالجة المحلي.

المصدر: Orbeta et al. 2014.

الجدول ٥-٤: تصميم انقطاع الانحدار لتقديرات أثر النفقات السنوية للفرد المستمد من تقييم أثر برنامج "بانتاويد باميليا" للتحويل النقدي المشروط في الفلبين، (Philippine Pesos 2013)

الحصائل		عرض النطاق		
		CCT	IK	أخذ العينات
التعليم (لكل طفل في سن التمدرس من ٣ إلى ٢٠ عامًا)	الأثر	٢٠٦,٦١**	٢٠٠,٥٦**	٧٧,٦٧
	الخطأ المعياري	٧٠,٥١	٥٥,٣٢	٥٠,٥٠
	المتوسط بخلاف برنامج بانتويد	٢٥١,٨٢	٢٣٠,٤٨	٢٥٢,٤٩
	الملاحظات	١٤٠٢	٢٠١٨	٢٩٣٩
الأغراض الطبية	الأثر	١٤,٦٧	١٤,٤٢*	١٤,٦٠**
	الخطأ المعياري	٨,١٣	٦,٩١٢	٥,٥٠
	المتوسط بخلاف برنامج بانتاويد	٣٥,٣٤	٣٤,٥٦	٣٤,٣٧
	الملاحظات	١,٧٨٩	٢,١٠٠	٣,١٠٧
الملابس والأحذية.	الأثر	٧٥,٢٨**	٧٣,٤١**	٤٤,٢٧**
	الخطأ المعياري	٢٥,٦٣	٢٤,٩٣	١٧,١٢
	المتوسط بخلاف برنامج بانتاويد	٩١,٥٢	٩٥,٩	١٠٧,٠١
	الملاحظات	١٣٥١	١٤٥٣	٣١٠٨

* ترمز إلى الدلالة عند مستوى ١٠٪، ** ترمز إلى الدلالة عند مستوى ٥٪، CCT = التحويل النقدي الشرطي.

ملاحظة: يشير عرض النطاق إلى عرض النطاق المثالي كما هو مقترح في Imbens and Kalyanaraman (2012) (المشار إليه بالاختصار IK)، وCattaneo, and Titiunik (2014a) (المشار إلي بالاختصار CCT)، وعرض نطاق أخذ العينات كما هو مقدَّر في Grover (2013).

المصدر: Orbeta et al. (2014).

يُجرى تحليل انحدار لكل من مؤشر النتيجة ومتغير التعيين، مثل نتائج اختبار قياس مستوى الدخل الفعلي (PMT) ونقطة تقاطع صورية. نقطة التقاطع الصورية هي متغير ثنائي التفرع يأخذ القيمة ٠ للملاحظات دون العتبة والقيمة ١ للملاحظات عند العتبة وأعلاها. يجوز تضمين ميل صوري أيضًا، وهو عتبة تفاعلية تنتج عن نقطة التقاطع الصورية ومتغير التعيين.

إذا طُبقت العتبة بشيء من النقص (تصميم انقطاع الانحدار المشوش)، تُستخدم مقاربة المتغير المساعد ثنائي المرحلة (مذكورة في القسم ٥-٧). في المرحلة الأولى، ننشئ تصميم انحدار للمشاركة ومتغير التعيين، ونحسب القيم الملائمة. أما في المرحلة الثانية، فنستخدم هذه القيم الملائمة التي حصلنا عليها من المرحلة الأولى، في الانحدار الناتج ليحل محل متغير التعيين.

ما المطلوب لإجراء تصميم انقطاع انحدار؟

يلزم توافر بيانات بشأن متغير التعيين ومؤشر الحصيلة لأعداد كافية من هؤلاء الذين يتم النظر في تضمينهم في البرنامج، بما في ذلك أولئك الذين تم قبولهم ورفضهم. ومن المثمر أيضًا توافر بيانات بشأن متغيرات أخرى للتحقق من صلاحية التوازن عبر العتبة، ويمكن تعزيز عملية التحقق هذه عن طريق إجراء مسح خط الأساس لضمان عدم تأثر المتغيرات المشتركة بالبرنامج. لا تحتفظ العديد من البرامج بمعلومات حول هؤلاء الذين لم يتم قبولهم، ما يجعل من الأصعب تطبيق تصميم انقطاع الانحدار (RDD). ويجب توفير معلومات حول الكيفية التي تم بها تطبيق قاعدة عتبة الأهلية بصرامة، ولكن يجب أن تكون هذه الحقيقة واضحة من خلال الاستدلال بالبيانات. يرى كالونيكو وآخرون أنه يمكن تطبيق هذه التقنية من خلال "rdrobust" في حزمة برامج STATA (Calonico et al. 2014b)، كما هو موضح في الملحق الأول، القسم ٧-١.

مزايا وعيوب تصميم انقطاع الانحدار

يضبط تصميم انقطاع الانحدار (RDD) بشكل أكثر كمالاً الوحدات غير الملحوظة مقارنةً بطرق المطابقة شبه التجريبية الأخرى. تجدر الإشارة إلى أنه من الممكن أيضًا الاستفادة من البيانات الإدارية إلى أكبر حد، وبالتالي نقلل من الحاجة إلى جمع البيانات، إلا أنه قد تكون هناك حاجة لجمع بيانات الحصائل المستقاة من الطلبات المرفوضة.

ما يقيد هذه التقنية هو اشتراط وجود معيار تعيين واضح وعينات كافية لغرض التحليل. غالبًا ما تكون هناك صعوبة في إجراء تصميم انقطاع الانحدار (RDD) بسبب عدم كفاية الملاحظات على كلا جانبي العتبة. جمعت دراسة "برنامج الطوابع الغذائية في منغوليا" بيانات من عينة محددة الغرض حول عتبة اختبار قياس مستوى الدخل الفعلي (PMT) لتفادي هذه المشكلة.

يوجد قيد آخر على هذه التقنية يتمثل في أن تقدير الأثر يعتني بالمجتمع الإحصائي القريب من العتبة فحسب. ويُطلق على هذا التقدير اسم "متوسط تأثير المعالجة المحلي" (LATE)، بدلاً من متوسط تأثير المعالجة للمجتمع الإحصائي المعالَج بالكامل. من حيث المبدأ، يقيد ذلك الأمر الصلاحية الخارجية لهذه المقاربة. على الرغم من ذلك، يمكن القول إن متوسط تأثير المعالجة المحلي (LATE) يعطي معلومات حول التأثير عند هامش الأهلية، ومن ثم فهو ممثل جيد لما هو متوقع إذا تم تمديد البرنامج.

٧-٥ المتغيرات المساعدة (Instrumental Variables)

وصف موجز للطريقة

في انحدار المربعات الصغرى العادية (OLS) التقليدية، قد يحدث الانحدار في الحصيلة بناءً على تدخلٍ صوري من ناحية أو قياس المشاركة في التدخل من ناحية أخرى، مثل مدة التدريب الذي تم حضورها أو المسافة إلى طريق ما. تكون المربعات الصغرى العادية غير متحيزة في ظل ظروف معينة، بما في ذلك أن تكون المشاركة مدفوعة بالخصائص الملحوظة التي تم قياسها. ولكن مع إمكانية اختيار الوحدات غير الملحوظة، يكون من المتوقع حدوث نشأة داخلية للتقديرات الناتجة عن مثل هذه المقاربة. لتصحيح ذلك الأمر، تُستخدم متغيرات مساعدة (IV) للتوصل إلى تقديرات متسقة عن طريق استخدام واحد أو أكثر من المتغيرات المؤثرة على المعالجة، لا الحصائل، باعتبارها ممثلات للتدخل (Reiersol 1945). التجارب الطبيعية (خضعت للمناقشة في الفصل ٣) هي الحالات المثالية لتطبيق طرق المتغيرات المساعدة (IV)، حيث إن الشرط الخارجي النشأة الذي يحدد صلاحية الوصول إلى التدخل يصبح الأداة المساعدة.

وصف الطريقة

في مقاربة انحدار المربعات الصغرى العادية (OLS) البسيطة التقليدية للمعالجات ثنائية التفرع، يتم إنشاء انحدار لكل من الحصيلة ومتغير صوري للدلالة على المشاركة W = 1 (W لمجموعة المعالجة وW = ٠ لمجموعة المقارنة) إلى جانب المتغيرات الأخرى التي تؤثر على الحصيلة. يعبر معامل W عن مقياس الأثر.

المشكلة في هذه المقاربة هو أن التحيز في الاختيار يمكن أن ينعكس على تقدير معامل الأثر جاعلاً إياه متحيزًا. إذا كان الاختيار يعتمد بالكامل على الوحدات الملحوظة، واشتمل الانحدار على متغيرات للدلالة على كل هذه الوحدات الملحوظة، فإن انحدار المربعات الصغرى العادية (OLS) سينتج عنه تقدير أثر صالح بكل تأكيد. ومن النادر افتراض وقوع مثل هذه الحالة.

إذا كانت الوحدات غير الملحوظة ثابتة على مر الزمن، فإن الفروق تزيل تأثيرها، ومن ثم يكون تقدير معادلة الأثر باستخدام الفروق غير متحيز. وعلى العكس، إذا كانت الوحدات غير الملحوظة متغيرة على مر الزمن، فإن مقاربة الاختلاف في الاختلافات (DiD) ستؤدي إلى تقديرات أثر متحيزة. ومن ثم يمكن استخدام تقدير المتغير المساعد كطريقة لإزالة سمة التحيز.

يمكن وصف تقدير المتغير المساعد بأنه انحدار يتم فيه إحلال المتغير الذي يُعد مصدرًا لمشكلة داخلية النشأة (أي W بسبب التحيز في الاختيار) *بوحدة* مساعدة (Z). يجب أن تفي هذه الوحدة المساعدة بشرطين:

١. أن تكون مترابطة مع W (المشار إليها بـ "الملاءمة").

٢. أن تكون غير مترابطة مع الحصيلة (Y) عدا من خلال تأثيرها على W: أي لا تكون هناك علاقة مباشرة بين Z وY (ويُشار إلى ذلك بـ "قيد الاستثناء").

ربما نجد مثالاً بسيطًا في الوحدة المساعدة لتأثير التدخين على سرطان الرئة. أولئك الذين يدخنون قد تكون لديهم خصائص أخرى تختلف عن أولئك الذين لا يدخنون، مثل ممارسة الرياضة أو أي نوع من الممارسات التي تنطوي على مخاطر، وعليه يكون الانحدار المباشر القائم على سمة "التدخين" وحدها متحيزًا. غير أن الضرائب المفروضة على السجائر تؤثر على التدخين ولا تؤثر على الإصابة بسرطان الرئة، على العكس من آثار التدخين، لذا يمكن استخدامها كوحدة مساعدة لتأثير التدخين. ولنضرب مثلاً آخر تكون فيه الأداة المساعدة ممثلة في أثر وصول الكهرباء على الأسر الذي تم تقديره باستخدام بُعد المسافة عن أعمدة الكهرباء. تُحدد رسوم الوصلة الكهربائية في بنجلاديش حسب القرب من أعمدة الكهرباء، ولكنها لا تفرض شرطًا على الحصائل مباشرة، حيث إن الأسر الأكثر فقرًا تميل إلى أن تكون أقرب إلى الأعمدة (المربع ٥-١).

المربع ٥-١: استخدام المتغيرات المساعدة لقياس أثر توصيل الكهرباء على الأسر الريفية

يعد توليد الكهرباء ونقلها وتوزيعها واحدًا من أكثر المجالات التي تستقطب التمويلات المخصصة للتنمية. إلا أنه توجد أدلة ضئيلة نسبيًا على فوائد مثل هذه التدخلات بالنسبة للأسر الريفية الفقيرة. يتمثل أحد أكبر التحديات في أن الأسر الأكثر ثراءً فقط هي التي تقدر وحدها على تحمل تكلفة وصلات الكهرباء، لذلك العلاقات السببية تتسم بالخلط.

استكشف خاندكير وآخرون (Khandker et al. (2012)) المعايير المعمول بها لتسعير وصلات الكهرباء في بنجلاديش، وكشفوا عن ارتفاع كبير في تكلفة الوصلة بالنسبة للأسر التي تبعد عن أعمدة الكهرباء بمسافة أكبر من ١٠٠ قدم. لذا، فإن هذه المسافة تمثل شرطًا لأهلية الحصول على إعانة لتمديد الوصلات. ومع ذلك، قبل توصيل الكهرباء، لا توجد علاقة قوية إيجابية بين البُعد عن أعمدة الكهرباء والحصائل المعنية، مثل الدخل أو النفقات للمزارع وغير المزارع. وهذا يقترح أن البُعد عن أماكن أعمدة الكهرباء غير مثير للخلط، بل يتنبأ بالمعالجة. وعلى هذا الأساس، استُخدم بُعد مسافة الأسرة عن أعمدة الكهرباء بما يقل عن ١٠٠ قدم كوحدة مساعدة في انحدار التأثيرات الثابتة للقرية.

تمثلت المرحلة الأولى في تقدير القيمة المتنبأ بها لتوصيل الكهرباء إلى الأسرة (متغير ثنائي) كدالة للوحدة المساعدة وخصائص الأسرة وسمات القرية. وفي المرحلة الثانية، تم تقدير الحصائل (مثل الدخل والنفقات وإكمال التعليم ووقت الدراسة) على أساس المتغير المتنبأ به وخصائص الأسرة والقرية. بهذه الطريقة، وجدت الدراسة تأثيرات ذات دلالة كبيرة لتوصيل الكهرباء على جميع الحصائل الخاضعة للتقدير تقريبًا.

المصدر: Khandker et al. (2012).

بصفة عامة، كان هناك تحدٍ متمثل في إيجاد وحدة مساعدة صالحة تفي بكلا الشرطين. يمكننا النظر إلى كلتا الطريقتين اللذين ناقشناهما كمثالين للمتغير المساعد (IV) الذي يعبر عن وحدات مساعدة صالحة: التجربة العشوائية المضبوطة (RCT)، وتصميم انقطاع الانحدار (RDD).

يشيع تقدير الأثر استنادًا إلى دراسات التجربة العشوائية المضبوطة (RCT) باستخدام الانحدار بدلاً من إجراء مقارنة بسيطة بين وسائل المعالجة والضبط. في هذه الحالة، يُستخدم التعيين العشوائي باعتباره وحدة مساعدة. يرتبط التعيين العشوائي بالمشاركة (ولكن يختلف المتغير في حالة وجود تداخلات تقاطعية)، ولكنه لا يرتبط بالحصيلة الناتجة عن التصميم. ويُعد تصميم انقطاع الانحدار المشوش حالة خاصة لمتغير مساعد (IV) تكون الوحدة المساعدة فيه عبارة عن متغير التعيين.

يتم اختيار وحدات المساعدة على أفضل نحو عن طريق تحديد النموذج الهيكلي الأساسي، الذي يتم اشتقاقه من نظرية التغيير. وجرت العادة أن يتم تحديد أكثر من وحدة مساعدة واحدة. عندما تتوافر أكثر من وحدة مساعدة واحدة، يتم تطبيق المتغيرات المساعدة غالبًا في صورة مربعات صغرى ثنائية المرحلة. (١) في المرحلة الأولى، أنشئ انحدارًا لكل من المتغير الداخلي النشأة (الذي يقيس مشاركة البرنامج) والوحدات المساعدة، ثم احسب القيمة الملائمة.
(٢) في المرحلة الثانية، قم بتقدير المعادلة الناتجة، مستبدلاً المتغير الداخلي النشأة بالقيم الملائمة التي حصلت عليها من المرحلة الأولى. ويتمثل تقدير الأثر في معامل هذه القيم الملائمة.

ما المطلوب عمله مع المتغيرات المساعدة؟

عمليًا، لا تُنفذ هاتان المرحلتان يدويًا: ستتولى الحزم البرمجية تنفيذ العمليات الحسابية، التي ستعطي أيضًا أخطاء معيارية صحيحة (وهو ما لن تفعله تقديرات انحدار المرحلة الثانية إذا تم إجراء الانحدار يدويًا). على سبيل المثال، في حزمة برامج STATA، الأمر الخاص بالمتغيرات المساعدة هو "ivregress"، ويوفر "ivreg2" تشخيصات إضافية مفيدة (Baum et al. 2010)، كما هو موضح في الملحق الأول، القسم ٦-٢. من أجل تقدير المتغير المساعد (IV)، يلزم توافر بيانات حول الملاحظات المعالَجة وغير المعالَجة، بما يشمل ذلك الحصيلة والوحدات المساعدة وكذلك متغيرات الخلط الأخرى. إذا كان يتم جمع البيانات للدراسة، فمن الضروري تحديد الوحدات المساعدة بشكل مسبق، لذلك يتم تضمين أسئلة ذات صلة في الوحدات المساعدة الخاضعة للمسح.

مزايا وعيوب المتغيرات المساعدة

إذا كان المتغير المساعد (IV) يعبِّر عن وحدة مساعدة صالحة، فإن مصادر كل من الوحدات الملحوظة وغير الملحوظة للتحيز في الاختيار تكون خاضعة للضبط. أما العيب الرئيسي، فهو يكمن في الصعوبة النسبية في إيجاد وحدة مساعدة صالحة، حيث إن العديد من العوامل التي تؤثر على المعالجة تؤثر أيضًا على الحصائل بشكل ما. وينتج عن هذه المقاربة أيضًا متوسط تأثير المعالجة المحلي (LATE)، الذي قد يصعب على المعنيين بوضع السياسات فهمه.

٨-٥ مقاربات المعالجة داخلية النشأة ووظيفة التحكم (Endogenous Treatment and Control Function Approaches)

وصف موجز للطريقة

طبَّق هيكمان (Heckman (1976)) أحد النماذج الأولية في الاقتصاد القياسي للتعامل مع التحيز في الاختيار. كان الغرض في الأصل من تطوير هذا النموذج هو تقدير معادلات الأجور، وكانت هناك مشكلة، وهي أن الأجور لا تكون ملحوظة إلا لهؤلاء الأشخاص الموظفين (على نحو مماثل، لا يكون أثر البرنامج ملحوظاً إلا لهؤلاء الأشخاص المشاركين في البرنامج). تُستخدم معادلات الأجور لتقدير معدل العائد على التعليم، وأصبح نموذج هيكمان معياريًا في مثل هذه التحاليل. تشكل مقاربة وظيفة التحكم هذه الأساس الذي تستند إليه نماذج الدالة الضابطة للمعالجة داخلية النشأة. المفهوم الأساسي الذي نرغب في توضيحه هو أن الاختلاف بين احتمالات المعالجة المتنبأ بها استنادًا إلى الخصائص الملحوظة والمعالجة الفعلية، يمكن تمثيله في هيئة متغيرات إضافية مستخدمة في انحدار ناتج/حصيلة، وتقضي هذه المتغيرات على تأثيرات محددات المعالجة غير الملحوظة، بحيث يمكن تخمين تأثيرات المعالجة الخالية من التحيز.

وصف الطريقة

تقوم هده المقاربة على اتخاذ خطوتين: الخطوة الأولى هي تمثيل متغير ثنائي يقيس المشاركة في التدخل المعني، كدالة لمتغيرات مستقلة. أما الخطوة الثانية، فنستخرج القيم المتنبأ بها من هذا الانحدار لاستخدامها في انحدار ثانٍ للحاصل المعني من أجل تمثيل كل من معدل المشاركة في البرنامج، وبقايا الوحدات المعممة من الخطوة الأولى، والعوامل الأخرى المؤثرة على الحصائل.

تقوم هذه المقاربة على إجراء ثنائي العناصر:

١. تقدير معادلة احتمالية للمشاركة في البرنامج (تمامًا مثل مطابقة درجة الميل باستثناء أن هذا الاحتمال يجب أن يحتوي على واحدة أو أكثر من وحدات المساعدة للمشاركة). حساب القيم الملائمة لـ "نسبة معكوس الاخفاق" و"نسبة الخطر" من الاحتمال.

٢. استخدم متغير المشاركة والقيم الملائمة كعناصر للانحدار في تقدير المربعات الصغرى العادية (OLS) للمعادلة الناتجة. في حالة الملاحظات المعالَجة، سيتم تضمين نسبة معكوس الاخفاق، أما في حالة الملاحظات غير المعالَجة، سيتم تضمين نسبة الخطر. إن معامل متغير المشاركة له دور في قياس الأثر، حيث تقضي معامِلات الانحدار الملائمة الإضافية على التحيز في الاختيار.عمليًا، قد تسمح برامج القياس الاقتصادي بتقدير النموذج في خطوة واحدة. من أجل تعريف النموذج، يجب توافر متغير واحد على الأقل في خطوة الاحتمال يؤثر على اختيار المشاركة في البرنامج، بشرط ألا يؤثر على الحصيلة، وهو بشكل أساسي الشرط نفسه كما لو أنه متغير مساعد. ومع ذلك، يمكن استخدام المتغيرات المشتركة التي تؤثر على كل من الاختيار والحصيلة، في كلتا المرحلتين.

ما المطلوب لإتمام انحدارات معالجة داخلية النشأة؟

يتطلب استخدام مقاربة هيكمان تحديد نموذج مشاركة يتضمن وحدة مساعدة واحدة على الأقل. ويتطلب تقدير نموذج هيكمان توافر بيانات حول كل من مجموعتي المعالجة والمقارنة إلى جانب ملاحظات بشأن العوامل المؤثرة على المشاركة، وكذلك العوامل المؤثرة على الحصائل بالنسبة لمعادلة الحصيلة. قد تتضمن الحزم الإحصائية أمرًا للتقدير (مثل "etregress" في STATA). يوفر القسم ٦-٣ في الملحق الأول مزيدًا من المعلومات حول هذه التقنية وتطبيقها.

مزايا وعيوب انحدارات المعالجة داخلية النشأة

ما يميز نموذج الاختيار من هيكمان هو أنه يستند إلى مقاربة يمكن تطبيقها في أغلب الحالات بشرط توافر البيانات المناسبة. هناك ميزة ثانية يشترك فيها مع مطابقة درجة الميل (PSM)، وهي أن المرحلة الأولى تنطوي على تقدير معادلة مشاركة، وهذا جزء مفيد في عملية التحليل الداعمة للتقييم. تسمح هذه المقاربة أيضًا بالكشف عن وجود أي تحيز في الاختيار عن طريق اختبار دلالة لامدا (lambda) في انحدار المرحلة الثانية. على عكس مطابقة درجة الميل (PSM)، تستطيع مقاربة هيكمان التعامل مع الاختيار من الوحدات غير الملحوظة بموجب افتراضات معينة، وعلى عكس المتغير المساعد (IV) التقليدي، يُستخدم لتقدير متوسط تأثير معالجة (ATE) وليس متوسط تأثير معالجة محلي (LATE). من منظور عام أشمل، قد يكون متوسط تأثير المعالجة (ATE) ذا أهمية أكبر بالنسبة للمعنيين بوضع السياسات.

تحتاج هذه الطريقة إلى متغير واحد على الأقل يمثل وحدة مساعدة، كما هي الحال مع المتغير المساعد (IV). لتحديد المتغير، يتطلب الأمر بعض الافتراضات القوية بشأن وسطية حدود الخطأ وهيكل الارتباط بين متغيرات الوحدات غير الملحوظة التي تحدد تعيين المعالجة والوحدات غير الملحوظة التي تؤثر على الحصيلة، وفي حالة عدم ثبوت هذه الافتراضات، فهذا يعني أن التقديرات ليست خالية من التحيز.

٩-٥ الانحدارات التبديلية داخلية النشأة (Endogenous Switching Regressions)

وصف موجز للطريقة

الانحدارات التبديلية داخلية النشأة هي عبارة عن طرق انحدارية تمثل معادلتين للحصيلة (نظامين)، واحدة للمعالجة والأخرى للمقارنة، ما يسمح بتأثير الاختيار على المعالجة (Maddala and Nelson 1975). يمكن اعتبار هذه المقاربة حالة خاصة من نموذج هيكمان، حيث إن المعادلة الناتجة في المرحلة الثانية هي انحدار تبديلي.[٤] أهم ميزتين لهذه المقاربة هما أنها تسمح بالتأثيرات التفاعلية بين المعالجة والمتغيرات المؤثرة على الحصائل، وأنها تستطيع الإتيان بتخمينات فريدة لكل من متوسط تأثير المعالجة على المعالَج (ATT) ومتوسط تأثير المعالجة على غير المُعالَج (ATU)، ومتوسط تأثير المعالجة (ATE).

وصف الطريقة

تطبيق الانحدارات التبديلية داخلية النشأة تبدأ بتوصيف النموذج:

- بالنسبة لنماذج المعالجة داخلية النشأة، يتم تقدير معادلة المشاركة التي تمثل ما إذا كانت وحدة الملاحظة (الأسرة، الشركة، إلخ) تنتمي إلى مجموعة المعالجة أو مجموعة المقارنة في صورة دالة للخصائص الملحوظة، التي تُستخرج منها القيم الملائمة.
- معادلتان للحصيلة، إحداهما للحصيلة بالنسبة للوحدات المنتمية إلى مجموعة المعالجة، والأخرى للوحدات المنتمية إلى مجموعة المقارنة. تشترك المعادلتان في معامِلات الانحدار نفسها بصرف النظر عن نسبة معكوس الاخفاق ونسبة الخطر.

لا تعطينا هذه المقاربة تقدير أثر واحدًا. بما أن العوامل المشتركة تختلف حسب كل نظام من النظامين، فإن الحصيلة المتوقعة للملاحظتين (واحدة معالجة والأخرى ضابطة، وتشتركان في قيم المتغيرات المستقلة نفسها) تكون مختلفة. وبالإمكان حساب الحصيلة المتوقعة لكل من الحصيلة المحتملة للمجتمعين الإحصائيين المعالَج وغير المعالَج. تسمح الفروق، بين القيم المتوقعة للحصائل المحتملة بالنسبة للمشاركين وغير المشاركين في وجود المعالجة وغيابها، بتخمين كل من متوسط تأثير المعالجة على المعالَج (ATT) ومتوسط تأثير المعالجة (ATE) ومتوسط تأثير المعالجة على غير المُعالَج (ATU). شهد هذا النموذج توسعًا ليتجاوز مجموعات المجتمع الإحصائي هذه لتوقع تأثيرات المعالجة الهامشية للمجتمعات الإحصائية الفرعية المحددة خصيصًا (Moffitt 2008).

٤ الانحدار التبديلي هو انحدار يتم من خلاله تخمين معادلتي انحدار منفصلتين، ويُطلق على كل معادلة اسم "نظام". يتم تعيين كل ملاحظة إلى أحد النظامين.

ما المطلوب لإتمام الانحدارات التبديلية داخلية النشأة؟

بصفة عامة، المتطلبات مماثلة لتلك الخاصة بنموذج الاختيار من هيكمان. ومع ذلك، قد يكون من الضروري أن تكون لمعادلة المشاركة قدرة اكبر على تفسير الانحدار المقرر إيجاد حل له، علمًا بأنه يلزم تقدير معامِلات أكثر. بمساعدة الحزمة "movestay" في برنامج STATA، يمكن تطبيق التقنية آليًا (Lokshin and Sajaia 2004)، كما هو موضح في الملحق الأول، القسم ٦-٤.

مزايا وعيوب الانحدارات التبديلية داخلية النشأة

تتشارك الانحدارات التبديلية داخلية النشأة في العديد من المزايا والعيوب مع نماذج المعالجة داخلية النشأة. ومع ذلك، الميزة الإضافية الكبرى هي أنه يمكن تقدير قيم فريدة لكل من متوسط تأثير المعالجة على المعالَج (ATT) ومتوسط تأثير المعالجة (ATE) ومتوسط تأثير المعالجة على غير المُعالَج (ATU). أما العيب فهو الصعوبة المحتملة في تحقيق التقارب عند تقدير المعادلة، لا سيما إذا كان نموذج الاختيار ركيكًا، بالإضافة إلى أن وحدات المساعدة الرديئة في معادلة الاختيار يمكن أن تؤدي إلى تقديرات متحيزة. تعتمد هذه المقاربة أيضًا على الافتراضات القوية نفسها حيث إن المعالجة داخلية النشأة تؤثر على الانحدارات.

٥-١٠ الملخص

يمكن استخدام مجموعة متنوعة من الطرق غير التجريبية عندما يتعذر إجراء تجارب عشوائية. إلا أن كل الطرق تتطلب بيانات بشأن المجتمعين الإحصائيين المعالَج وغير المعالَج، بدلاً من بيانات التدخل القبلية والبعدية (بصرف النظر عن السلسلة الزمنية المتقطعة). حتى إذا كان الواقع المضاد ضمنيًا، كما هي الحال في المقاربات الانحدارية، مثل المتغيرات المساعدة، فإن تقييم الأثر يتطلب بيانات مستقاة من مجموعة مقارنة. كل الطرق بإستثناء الاختلاف في الاختلافات (DiD) والتأثيرات الثابتة والوحدات الضابطة الاصطناعية، تتطلب توصيفًا للنموذج يتضمن المتغيرات المحددة للاختيار المؤدي إلى المشاركة في البرنامج. يعطي نموذج الاختيار أيضًا رؤى إضافية حول محددات المشاركة، التي قد تحمل فائدة في حد ذاتها.

اختيار الطريقة الملائمة لدراسة معينة يكون مشروطًا بالسياق والبيانات (الجدول ٥-٥). ويتضمن ذلك طبيعة المعالجة، وعدد الملاحظات المتوفرة، وأهمية الاختيار من الوحدات غير الملحوظة، وتأثير المعالجة محل الاهتمام.

الجدول ٥-٥: موجز طرق تقييم الأثر غير التجريبية

الطريقة	نوع المعالجة المقيَّمة	الحد الأدنى للبيانات المطلوبة	إمكانية التصحيح فيما يتعلق باختيار وحدات قابلة للملاحظة	إمكانية التصحيح فيما يتعلق باختيار وحدات غير قابلة للملاحظة	تأثير المعالجة المتوقع
الاختلاف فى الاختلافات (DiD)/ التأثيرات الثابتة	ثنائية التفرع / ثنائية التفرع، مستمرة (فترتان أو أكثر)	طولية (فترتان أو أكثر)	نعم	نعم، وحدات غير ملحوظة وغير متغيرة بالنسبة إلى الزمن	ATT
الوحدات الضابطة الاصطناعية	ثنائية التفرع	طولية مع عدة دورات من الملاحظات القبلية للمعالجة	نعم	نعم، تصحيح تقريبي	ATT
مطابقة درجة الميل	ثنائية التفرع	مقطعية	نعم	لا	ATT، ATU، ATE
الانحدار القوي المزدوج	ثنائية التفرع	مقطعية	نعم	لا	ATE
تصميم انقطاع الانحدار	ثنائية التفرع	مقطعية، بمحاداة عتبة الأهلية	نعم	نعم	LATE
المتغيرات المساعدة	ثنائية التفرع، مستمرة	مقطعية، بما في ذلك وحدة مساعدة صالحة	نعم	نعم	LATE
انحدار تأثيرات المعالجة داخلية النشأة	ثنائية التفرع	مقطعية، بما في ذلك وحدة مساعدة صالحة	نعم	نعم، بموجب هيكل حدود الخطأ للمتغيرات-المتغيرات المشتركة المفترضة المحددة	ATE
الانحدار التبديلي داخلي النشأة	ثنائية التفرع	مقطعية، بما في ذلك وحدة مساعدة صالحة	نعم	نعم، بموجب هيكل حدود الخطأ للمتغيرات-المتغيرات المشتركة المفترضة المحددة	ATT، ATU، ATE

ATE = متوسط تأثير المعالجة، ATT = متوسط تأثير المعالجة على المعالَج، ATU= متوسط تأثير المعالجة على غير المعالَج، LATE = متوسط تأثير المعالجة المحلي.
المصدر: المؤلفان

- **طبيعة المعالجة:** إذا كان متغير المعالجة ثنائي التفرع (مثل تحديد إتمام المشاركة في البرنامج من عدمها) أو مستمرًا (مثل انخفاض وقت السفر بفضل البنية التحتية الجديدة)، فإن هذا يؤثر على النماذج التي يمكن استخدامها معه. إن النماذج التي تستخدم معادلة مشاركة (المعالجة داخلية النشأة، الانحدارات التبديلية داخلية النشأة، مطابقة درجة الميل) لا تنطبق إلا على المعالجات ثنائية التفرع، كما هي الحال مع تصاميم الاختلاف في الاختلافات (DiD) والوحدات الضابطة الاصطناعية. تنطبق انحدارات التأثيرات الثابتة والمتغيرات المساعدة عندما تكون المعالجة مستمرة.

- **توافر بيانات خط الأساس:** لا يمكن تطبيق تصاميم الاختلاف في الاختلافات (DiD) ونماذج التأثيرات الثابتة والوحدات الضابطة الاصطناعية إلا عندما تتوافر بيانات طولية مُجمعة، مثل عندما تتوافر ملاحظات قبلية وبعدية للتدخل، وكذلك مع وبدون. تتطلب الوحدات الضابطة الاصطناعية بالإضافة إلى ذلك ملاحظات متعددة لما قبل التدخل. تتعزز طرق أخرى إذا تم تنفيذها اعتمادًا على البيانات الطولية المُجمعة، ولكن يمكن تطبيقها على البيانات المقطعية العرضية إذا أمكن توافر فترة ملاحظة واحدة.

- **وجود الوحدة المساعدة:** تعتمد المتغيرات المساعدة ونماذج تأثيرات المعالجة داخلية النشأة والانحدارات التبادلية داخلية النشأة على وجود متغير واحد على الأقل ملائم كوحدة مساعدة. يتعين أن تفي الوحدة المساعدة بقيد الاستثناء، وأن تكون ذات صلة. كما أن متغير التعيين الخاص بتصميم انقطاع الانحدار يُعد أيضًا وحدة مساعدة في الأساس.

- **الاختيار من الوحدات غير الملحوظة:** إذا كان من المتوقع أن يحدث الاختيار من وحدات غير ملحوظة، فإذًا يتعذر استخدام التقنيات القائمة على درجة الميل بصورة مقطعية عرضية (أي بدون بيانات خط أساس). علاوة على ذلك، إذا كان الوحدات غير الملحوظة هذه أو تأثيراتها تتباين بمرور الوقت، فإذًا ستؤدي التأثيرات الثابتة والاختلاف في الاختلافات (DiD) أيضًا إلى تقديرات متحيزة.

- **قياس الأثر محل الاهتمام:** تعطي الطرق المختلفة قياسات أثر مختلفة. بواسطة نماذج الاختلاف في الاختلافات (DiD)/التأثيرات الثابتة، والوحدات الضابطة الاصطناعية، ومطابقة درجة الميل، والانحدارات التبديلية داخلية النشأة، يمكن تقدير متوسطات تأثير المعالجة على المعالَج (ATT) أو التأثير على أولئك الذين اختاروا المشاركة في التدخل. ومن خلال الانحدارات القوية المزدوجة وانحدارات المعالجة داخلية النشأة، يمكن تقدير متوسطات تأثير المعالجة (ATEs) أو التأثير على متوسط المجتمع الإحصائي (المشاركين وغير المشاركين). أما المتغير المساعد وتصميم انقطاع الانحدار، فيسمحان بتقدير متوسطات تأثير المعالجة المحلي (LATEs)، أو التأثيرات على أولئك الذين يقعون عند هامش المشاركة فيما يتعلق بالوحدة المساعدة. قد تكون لمتوسطات تأثير المعالجة على المعالَج (ATT) أهمية أكبر في تفسير التأثيرات الممتدة حتى اللحظة الراهنة، بينما قد تكون لمتوسطات تأثير المعالجة (ATEs) أهمية أكبر في تعزيز القرارات المستنيرة حول توسيع البرامج.

المراجع*

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*. 105 (490). pp. 493–505.

Asian Development Bank. 2014. *Food Stamps and Medicard: Impact Evaluation Report*. Manila.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2010. IVREG2: STATA Module for Extended Instrumental Variables/2SLS and GMM estimation. Statistical Software Components, S425401. Revised 9 February 2016. Boston College Department of Economics. http://ideas.repec.org/c/boc/bocode/ s425401.html.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica. 82. pp. 2295–2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*. 14 (4). pp. 909–946. https://sites.google.com/site/rdpackages/rdrobust/Calonico- Cattaneo-Titiunik_2014_Stata.pdf.

Grover, D. 2013. Sampling Recommendations for Second Wave Impact Evaluation of the Pantawid Pamilya Program Applying Regression Discontinuity Design. Project Document.

Heckman, J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement. 5: 475–492.

Imai, K. and M. Ratkovic. 2014. Covariate Balancing Propensity Score. *Journal of the Royal Statistical Society*: Series B. 76 (1). pp. 243–263.

Imbens, G. and K. Kalyanaraman. 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*. 79. pp. 933–959.

Jalan, J. and M. Ravallion. 2001. Does Piped Water Reduce Diarrhea for Children in Rural India? World Bank Policy Research Working Paper. WPS 2664. Washington, DC: World Bank.

Khandker, S. R., D. F. Barnes, and H. A. Samad. 2012. The Welfare Impacts of Rural Electrification in Bangladesh. *The Energy Journal*. 33 (1). pp. 187–206.

King, G. and R. Nielsen. 2016. Why Propensity Scores Should Not Be Used for Matching. Working Paper. http://j.mp/2ovYGsW.

* يشير بنك التنمية الآسيوي (ADB) إلى "Vietnam" (فيتنام) على أنها "Viet Nam" (فيت نام).

Lokshin, M. and Z. Sajaia. 2004. Maximum likelihood estimation of endogenous switching regression models. *The STATA Journal*. 4 (3). pp. 282–289.

Lunceford, J. K. and M. Davidian. 2004. Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study. *Statistics in Medicine*. 23. pp. 2937–2960.

Maddala, G. S. and F. D. Nelson. 1975. Switching Regression Models with Exogenous and Endogenous Switching. Proceedings of the Business and Economics Statistics Section. American Statistical Association. pp. 423–426.

Moffitt, R. 2008. Estimating Marginal Treatment Effects in Heterogeneous Populations. *Annales D'Économie Et De Statistique*. (91/92). pp. 239–261.

Mu, R. and D. Van de Walle. 2011. Rural Roads and Local Market Development in Vietnam. *Journal of Development Studies*. 47 (5). pp. 709–734.

Orbeta, A., A. Abdon, M. del Mundo, M. Tutor, M. Theresia Valera, and D. Yarcia. 2014. Keeping Children Healthy and in School: Evaluating the Pantawid Pamilya Using Regression Discontinuity Design Second Wave Impact Evaluation Results. Unpublished draft.

Quistorff, B. and S. Galiani. 2017. The synth_runner package: Utilities to Automate Synthetic Control Estimation Using Synth. Version 1.6.0. August 2017. https://github.com/bquistorff/synth_runner.

Reiersol, O. 1945. Confluence Analysis by Means of Instrumental Sets of Variables. *Arkiv for Matematik, Astronomi och Fysik*. Band 32A. pp. 1–119.

Rosenbaum, P. R. and D. B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 70. pp. 41–55.

Sween, J. A. and D. Campbell. 1965. The Interrupted Time Series as Quasi- Experiment: Three Tests of Significance. Northwestern University.

Thistlewaite, D. L. and D. T. Campbell. 1960. Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. *Journal of Educational Psychology*. 51. pp. 309–317.

مراجع إضافية للإطلاع

Angrist, J. D. and S. Pischke. 2009. *Mostly Harmless Econometrics*: An *Empiricists' Companion*. Princeton, New Jersey: Princeton University Press.

Caliendo, M. and S. Kopeinig. 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. IZA Discussion Paper. No. 1588. http://ftp.iza.org/dp1588.pdf.

Funk, M. J., D. Westreich, C. Wiesen, T. Stürmer, M. A. Brookhart, and M. Davidian. 2011. Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology*. 173 (7). pp. 761–767. https://doi.org/10.1093/ aje/kwq439.

Imbens, G. and T. Lemieux. 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*. 142. pp. 615–635. https:// doi.org/10.1016/j.jeconom.2007.05.001.

O'Neill, S., N. Kreif, R. Grieve, M. Sutton, and J. S. Sekhon. 2016. Estimating Causal Effects: Considering Three Alternatives to Difference- in-Differences Estimation. *Health Services & Outcomes Research Methodology*. 16. pp. 1–21. http://doi.org/10.1007/s10742-016-0146-8.

Pokropek, A. 2016. Introduction to Instrumental Variables and Their Application to Large-Scale Assessment Data. *Large-Scale Assessments in Education*. 4 (4). https://link.springer.com/content/ pdf/10.1186%2Fs40536-016-0018-2.pdf.

Ryan, A. M., J. F. Burgess, Jr., and J. B. Dimick. 2015. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. Health Services Research. 50 (4). pp. 1211–1235. https://www.ncbi. nlm.nih.gov/pmc/articles/PMC4545355/pdf/hesr0050-1211.pdf.

Wooldridge, J. M. 2012. *Introductory Econometrics: A Modern Approach*. Fifth Edition. Mason, Ohio: South-Western, Cengage Learning.

الفصل ٦
تحديد ما يجب قياسه وكيفية قياسه: جمع البيانات لتقييم الأثر

رسائل أساسية

- البيانات ضرورية من أجل تقييم الأثر. لذا، فإنه من المهم أن تتوافر أعداد كافية من الملاحظات بشأن المعالجة، ومشاركة البرنامج، والتدخلات الأخرى، والمتغيرات الشرطية للمشاركة، وحصائل المتغيرات المتنبأ بها، والحصائل مع وجود التدخلات وفي غيابها.
- قد تأتي البيانات من عدة مصادر، ولكن في أغلب الأحيان يتطلب تقييم الأثر إجراء مسوحات جديدة، وكثيرًا ما تشمل بدورها عدة مستويات، مثل الأسر المعيشية والشركات والمرافق والمجتمعات والوكالات.
- تصميم أداة المسح الموثوق بها يتطلب عناية كبيرة واختبارًا ميدانيًا جوهريًا.
- يجب أن يعكس أخذ العينات تأثيرات تداعيات التدخل، والاختيار بالطريقة العنقودية، وتحليل المجموعة الفرعية المستهدفة. عادةً ما يتعين إجراء عملية أخذ عينات على عدة مراحل.
- يجب أن تخضع عملية جمع البيانات لإشراف متيقظ مع التحري عن صحة البيانات، والتحقق لضمان الاتساق، والتوثيق الكافي.

٦-١ أهمية البيانات عالية الجودة

لا يُعد تقييم الأثر صالحًا إلا إذا صلحت البيانات التي يستند إليها. فالبيانات رديئة الجودة أو الملاحظات غير السديدة أو البيانات التي تفتقر بشكل كبير إلى المعلومات الكافية يمكن أن تجعل من المستحيل التمكن من تقدير الأثر أو يمكن أن تؤدي إلى نتائج زائفة. لذلك، ضمان ملاءمة البيانات التي يتم جمعها لا يقل أهمية عن ضمان خلو تصميم تقييم الأثر من أي تحيز محتمل. إذا كان في التصميم ما يعيبه، أو تم إجراء تقييم الأثر بدون الاقتداء بالخطة الموضوعة، فإنه في بعض الأحيان تكون هناك إمكانية لتصحيح الأوضاع عن طريق الاستخدام الملائم للكميات المناسبة من البيانات عالية الجودة والغنية بالمعلومات بما يكفي، بينما القصور في جودة البيانات يعني غالبًا عدم إمكانية تصحيح الدراسة على الإطلاق وإن كانت جيدة التصميم.

توضع إستراتيجية تصميم عملية جمع البيانات بالشكل الذي يسمح بالإجابة عن أسئلة التقييم، التي نشتقها من نظرية التغيير (كما ناقشنا في الفصل ٢). ينبغي للإستراتيجية أن تحدد نوع البيانات المطلوبة، بما يتضمن كلاً من العوامل الكمية والنوعية، للإجابة عن أسئلة التقييم (يوفر المربع ٦-١ نظرة عامة إلى مصطلحات المسح). إن وُجدت خطة التقييم، فهي تحدد بوضوح البيانات المقرر استخدامها للإجابة عن كل سؤال تقييم.

لا يقتصر الأمر على جمع البيانات الصحيحة، بل هناك اعتبار رئيسي يتخطى ذلك ليشمل ضمان جمع البيانات بالطرق الصحيحة. وهذا يعني أنه ينبغي تقليص نسبة التحيز والخطأ إلى أدنى حد. والخطأ نوعان: خطأ أخذ العينة، وخطأ غير مرتبط بأخذ العينة (Banda 2003). يشير خطأ أخذ العينة إلى جمع البيانات عن وحدات غير ممثلة للمجتمع الإحصائي محل اهتمام الدراسة. وقد يرجع ذلك إلى أن بروتوكول أخذ العينات قد أحدث تحيزًا، أو أن أُطر العينة لم تكن ممثِّلة للمجتمعات الإحصائية المستهدفة، أو عدم إجابة المجيبين بشكل انتقائي. الخطأ غير المرتبط بأخذ العينات يمكن أن ينشأ بسبب التصميم الرديء للمسح، أو عدم دقة المجيبين، أو خطأ من مندوب الإحصاء، أو مشكلات في إدخال البيانات. الحد من كلا نوعي الخطأ يمثل تحديًا أساسيًا.

لتجنب كثرة البيانات بدون داعٍ، يجب على فريق التقييم الإمعان في تقديره لماهية البيانات المتوفرة لديه بالفعل من المصادر الإدارية ومستودعات بيانات أخرى. أحيانًا لا تكون هناك حاجة إلى جمع البيانات الأولية التي يمكن الاستغناء عنها باستخدام البيانات الحالية، على الرغم من ندرة مثل هذه الموارد في الدول ذات الدخل المنخفض والمتوسط. من الممكن أيضًا "الارتكاز" على مسح حالي بدلاً من بدء مسح جديد خصيصًا لتقييم الأثر (كما ناقشنا في القسم ٦-٣).

يقدم هذا الفصل فكرة عامة عن مصادر البيانات، وتصميم المسح وإدارته، وأخذ العينات. يبدأ الفصل بنظرة عامة موجزة حول تحديد ماهية البيانات المطلوبة (القسم ٦-٢)، قبل مناقشة مصادر البيانات (القسم ٦-٣)، وتصميم المسح (القسم ٦-٤)، وأخذ العينات (القسم ٦-٥)، وإدارة جمع البيانات (القسم ٦-٦). يقدم الملحق الثاني مزيدًا من التفاصيل لوصف هذه الموضوعات.

٦-٢ تحديد ماهية البيانات المطلوبة

يعتمد المجتمع الإحصائي المستهدف الذي يمثل محور المسح والأسئلة المقرر طرحها على تصميم التدخل والأهداف ونظرية التغيير فيما يتعلق بكيفية تحقيق هذه الأهداف. تؤثر آلية تقديم التدخل على تصميم أخذ العينات، بما أن وحدة التعيين ستكون دائمًا المستوى الأول لأخذ العينات. ويتم تحديد وحدات أخذ العينات إستنادًا إلى التأثيرات المُراد تقييمها. مثلاً، إذا كان التأثير المرجو هو الحد من الفقر، فيلزم توافر بيانات عن الأسر المعيشية، أما إذا كان التأثير المرجو هو طلب العمالة، فيلزم توافر بيانات على مستوى الشركات.

تعد أسئلة التقييم، المستقاة من نظرية التغيير، الخطوة الأولى في تحديد ماهية البيانات المقرر جمعها. وينبغي لها أن تحدد (١) الحصائل الأولية والثانوية المعنية، وكذلك أي حصائل وسطية، و(٢) المتغيرات والعوامل السياقية الضرورية للعملية التي قد تسهم في إحداث التدخل لتأثيرات، و(٣) قياس مشاركة المستفيدين المستهدفين في البرنامج، وقد يكون هذا القياس مستمرًا أو "متدرجًا" (مثل التدريج بناءً على حجم القروض أو بُعد المسافة عن الطريق) أو ثنائي التفرع (وصول الكهرباء أو العيش في قرية المشروع)، و(٤) مؤشرات نشاط البرنامج، مثل الإنشاء. بالإضافة إلى ذلك، تحتاج التصاميم غير التجريبية إلى متغيرات لتمثيل المشاركة في البرنامج، التي يمكن أن تمثل أيضًا متغيرات ضابطة في المقاربات القائمة على الانحدار. إذا تم استخدام مقاربة المتغير المساعد، فإن تضمين أكثر من وحدة مساعدة محتملة يساعد في الحد من مخاطر عدم تمتع الوحدة المساعدة بالخصائص الضرورية.

ما إن نحدد ماهية البيانات، يلزم تحديد مصادر البيانات كخطوة تالية. في بعض الحالات، يمكن لمصادر إدارية حالية أو مسوحات حالية أن توفر البيانات الضرورية. إذا لم يتوفر أي منها، فيجب تحديد تفاصيل مسح جديد. لضمان أقصى درجات الدقة، يجب تحديد المجتمعات الإحصائية ذات الصلة المطلوب تضمينها في المسح، بأقصى قدر ممكن من التفاصيل. فعلى سبيل المثال، في مسح الشركات، يجب تحديد ما إذا كان المالك أو الرئيس التنفيذي أو المدير أو الموظف يشكل مصدرًا مناسبًا للمعلومات. قد يكون المالك الشخص الأكثر إحاطة بتاريخ الشركة واتجاهها الإستراتيجي، والرئيس التنفيذي سيكون الأكثر إحاطة بمستوى أداء نموذج الأعمال الحالي وظروف السوق الجارية، والموظفون سيكونون الأكثر إحاطة بمعلومات دقيقة أكثر عن المخرجات والأرباح وما إلى ذلك. في مسح الأسر المعيشية، غالبًا ما يتم توجيه نماذج مختلفة إلى مجيبين مختلفين. كمثال على مسح الأسر المعيشية، تُطرح أسئلة بخصوص صحة الأطفال واستهلاك الأسرة من الغذاء على أحد أفراد الأسرة من الإناث. بالنسبة لمسح المجتمع، غالبًا ما تكون الطريقة الأفضل لتقديم المسح في مكان جماعي يعج بقادة المجتمع للحصول على رد جامع مع مراعاة اتباع إجراءات أخلاقية بخصوص سرية البيانات.

٣-٦ مصادر البيانات

قد تأتي البيانات من المصادر التالية:

- *التعداد:* البيانات المجمعة عن المجتمع الإحصائي بالكامل. عادةً ما تُجرى التعدادات السكانية لجميع البالغين المقيمين في دولة ما مرة كل ١٠ سنوات، إلا أنها أحيانًا تكون أكثر تواترًا. وقد يُجرى التعداد الصناعي لكل الشركات بشكل أكثر تواترًا، وعادة ما يكون محصورًا على الشركات المسجلة. تُجرى أيضًا التعدادات الزراعية في بعض البلدان، وتكون مسوحات العينات أكثر تواترًا، وتزداد بشكل أكبر عبر بلدان آسيا والمحيط الهادئ. عادة ما تكون بيانات التعداد مفيدة في توفير إطار أخذ العينات أو البيانات التي يمكن استخدامها للحصول على مجموعات مقارنة مطابقة. ومع ذلك، يجب الانتباه جيدًا إذا كان المجموعات المستهدفة المهمة لا يشملها التعداد، مثل سكان البدو وأولئك الذين ليس لهم محل إقامة منتظم.

- *المسوحات:* البيانات التي يتم جمعها من عينة من مجتمع إحصائي معني. يُجرى المسح باستخدام أداة مسح واحدة أو أكثر. بالنسبة لتقييم الأثر، عادة ما يكون المجتمع الإحصائي المعني هو المجتمع الإحصائي المستفيد المستهدف. يجب وضع إستراتيجية لأخذ العينات لضمان تمثيل العينة لهذا المجتمع الإحصائي. على مجموعة المقارنة أن تمثل أيضًا هذا المجتمع الإحصائي وليس السكان عامةً.

- *نُظم المعلومات الجغرافية* والاستشعار عن بُعد هما بمثابة وسيلتان تمكنان من استخراج بيانات ذات صلة بالخصائص الجغرافية لمنطقة ما أو الخصائص التي يمكن تحديدها استنادًا إلى موقع جغرافي، مثل قياسات بُعد المسافة عن المرافق أو هطول الأمطار أو التضاريس. من خلال الاستشعار عن بُعد، تستطيع الحصول على معلومات تساعد على مراقبة الأحداث التي يمكن أن تؤثر في الحصائل (مثل الفيضانات) أو المعالجة (مثل تمديد الطرق).

- *البيانات الإدارية* تتضمن معلومات يتم جمعها بشكل دوري كجزء من نُظم إحصائية أو إدارية. تتضمن الأمثلة بيانات الفواتير من المرافق، والبيانات المدخلة إلى نُظم معلومات إدارة التعليم من خلال تعداد المدارس السنوي.

- *يمكن جمع مصادر أخرى للبيانات* في الوقت الفعلي من أجهزة مختلفة تسجل بيانات حركة المرور ومستويات التلوث وما إلى ذلك. ومن الممكن أيضًا استخدام البيانات المبلغ عنها ذاتيًا من المزارعين الذين يستخدمون تطبيقات الهاتف المحمول، مثلاً، حيث تستطيع التطبيقات نفسها إمداد المزارعين بالنصائح المخصصة.

الارتكاز على مسح حالي

في بعض الأحيان يكون من الممكن الارتكاز على مسح آخر، مثل مسح دخل الأسرة والإنفاق على المستوى الوطني أو مسح الصحة والتقاعد الوطني، بدلاً من إجراء مسح منفصل خصيصًا بهدف تقييم الأثر. مثل هذه المقاربة يمكن أن تساعد على تحسين الكفاءة وتمكين إدخال عينات كبيرة إذا كان المسؤولون عن تنفيذ المسح الآخر راغبين في إضافة وحدات كافية لجمع البيانات من أجل تقييم الأثر. على الأرجح أيضًا أن تكون البيانات الواردة عن المسوحات الوطنية المنفذة من الوكالة الإحصائية الوطنية، ذات جودة عالية.

ومع ذلك، لتحقيق فائدة من هذه البيانات، تحتاج فرق المشروع في الغالب إلى تخصيص الوقت والجهد في التنسيق والتمويل الإضافي. سيكون من الضروري عادةً أن تطلب وتمول نوعين من التعديلات على المسح المقرر الارتكاز عليه. أولاً، قد يكون من الضرورة زيادة حجم العينة في مجالات المشروع (ما يُسمى بـ "العينة المعززة") من أجل الوصول إلى أحجام عينة كافية للمجتمعات الإحصائية المعالَجة. ثانيًا، يجوز إضافة وحدة محددة حسب المشروع، لا سيما لمعالجة الجوانب العملية للمشروع. ويشترط أن تكون هذه الوحدة الإضافية ذات صلة بمجالات المشروع.

المربع ٦-١: مصطلحات المسح

أداة المسح (Survey instrument): شكل مسبق للتصميم (استبيان) يُستخدم لجمع البيانات الأولية في أثناء مسح ما. عادة ما تكون للمسح أكثر من أداة مسح واحدة، مثل مسوحات الأسر والمرافق.

الوحدة (Module): قسم من أداة المسح له مجال تركيز معين، مثل استخدام الطاقة أو التعليم أو التغذية. قد تتطلب الوحدات المختلفة مجيبًا مختلفًا في الدراسة ضمن الأسرة أو الشركة المعنية. بصفة عامة، ليس من الممارسات الجيدة أن يكون هناك مجيبون مختلفون ضمن الوحدة، رغم أن هذا قد يكون مطلوبًا ضمن بعض الوحدات.

المجيب (Respondent): الفرد الذي يجيب عن السؤال. جرت العادة على أن يكون لكل سؤال مجيب واحد فحسب.

مندوب الإحصاء (Enumerator): الشخص الذي يجري المسح. لا بد من أن يكون مندوبو الإحصاء على درجة عالية من التدريب حتى يتمكنوا من جمع بيانات عالية الجودة.

خط الأساس (Baseline): البيانات التي تُجمع قبل أن يبدأ التدخل. عمليًا، تنشأ بيانات خط الأساس، في بعض الأحيان، بعد بدء التدخل. ما يهم هو أن تُجمع البيانات قبل تأثر المستفيدين المستهدفين بالتدخل بأي شكل من الأشكال. يساهم توافر بيانات خط الأساس في تعزيز تصميم تقييم الأثر المحتمل.

خط الوسط (Midline): البيانات التي تُجمع في منتصف سريان التدخل، وهي تركز على المتغيرات والعمليات المتوسطة بدلاً من الحصائل النهائية.

خط النهاية (Endline): البيانات التي تُجمع عند نهاية التدخل.

ما بعد خط النهاية (Post-endline): البيانات التي تُجمع بعد انتهاء سريان التدخل ببضعة أعوام. تسمح بيانات ما بعد خط النهاية بإجراء تحليلات إذا ما استمرت فوائد التدخل.

البيانات الطولية المُجمعة (Panel data): البيانات التي تُجمع من الوحدات نفسها في أشواط متتالية من جمع البيانات. عادة يكون جمع البيانات الطولية المُجمعة ضروريًا لتصاميم تقييم الأثر الأكثر صرامة، تلك التي تُطبق فيها نماذج الاختلاف في الاختلافات (DiD) أو الآثار الثابتة. تعتمد هذه المقاربة على القدرة على إعادة تحديد موقع المجيبين.

المصدر: المؤلفان

غالبًا ما يكون الارتكاز على المسوحات الحالية أرخص من إجراء مسح جديد. ومع ذلك، قد لا تكون هذه الطريقة أو المقاربة أسرع، مع الأخذ في الاعتبار الوقت المطلوب للتفاوض والتنسيق مع المنظمة صاحبة المسح. لا تُعد هذه المقاربة خيارًا إلا إذا كان توقيت المسح الوطني مناسبًا لفترات بيانات خط الأساس و/أو خط النهاية لتقييم الأثر.

٤-٦ تصميم المسح

تصميم المسح هو مهمة أساسية في تصميم تقييم الأثر، ويتطلب استثمارًا كبيرًا من الوقت. تستدعي هذه المهمة العمل الجماعي بين أعضاء الفريق المشتمل على خبراء في تنفيذ المسح، وخبراء في قطاع التدخل، ومقيِّمين للأثر. يعالج هذا القسم القضايا المختلفة المتعلقة بتصميم المسح.

ربط تصميم المسح بتصميم التقييم

يجب أن يتضمن المسح أسئلة تستوعب المتغيرات الآتية:

- المشاركة في أنشطة التدخل.

- الحصائل التي تُقاس عليها التأثيرات، بما في ذلك أي حصائل قد لا تكون متوقعة أو متعمدة في أثناء تصميم المشروع. لا تُطرح أسئلة مباشرة بالنسبة لبعض الحصائل، بل تُشتق الإجابة من مجموعة متنوعة من الأسئلة. في بعض الحالات، مثل إنفاق الأسر، قد يتطلب حساب حصيلة واحدة الإجابة على العديد من الأسئلة.

- خصائص الأسرة أو الشركة إما لإجراء مطابقة أو التحقق من الموازنة في غياب المطابقة.

- العوامل الشرطية التي قد تؤثر على الأثر، مثل الموقع أو النوع الاجتماعي.

- المتغيرات المساعِدة أو المتغيرات الشرطية للأهلية في حالة المتغير المساعد أو تصاميم انقطاع الانحدار، على التوالي.

- الحصائل المتوسطة ومتغيرات العملية إلى جانب السلسلة السببية، مثل الوعي بالأنشطة الداعمة للتدخل؛ إمكانية الوصول، والتكلفة، وكمية هذه الأنشطة، وتوفر المنتجات أو الخدمات التكميلية. تسمح هذه المتغيرات بإجراء تحليلات وقائعية ووقائعية مضادة إلى جانب السلسلة السببية. حيثما يخضع تقدير الأثر لنمذجة هيكلية، تزداد الحاجة إلى عدد أكبر من المتغيرات عن تلك المطلوبة في الطرق المختزلة.

- التفاصيل التي تسمح بإعادة تحديد موقع المجيب أو التواصل معه في مراحل التحقق من صلاحية المسح أو متابعته.

سيتم تحديد كل مجموعة من المتغيرات من خلال تصميم التقييم، الذي بدوره يستند إلى نظرية التغيير التي تتساءل عن الكيفية التي ستتحقق بها الحصائل المرجوة بفضل التدخل (كما ناقشنا في الفصل ٢).

يجب أن تقدم نظرية التغيير أيضًا خططًا لتحليل المجموعة الفرعية. فمثلاً، هل سيؤثر التدخل بشكل مختلف حسب النوع الاجتماعي، أو سيكون أثره مختلفًا بين الفقراء وغير الفقراء أو وسط الأقليات؟ إذا كانت الإجابة نعم، فيجب إذًا جمع البيانات لتحديد هذه المجموعات الفرعية وتمثيلها.

أدوات المسح المتعددة

أداة المسح الأكثر استخدامًا في تقييمات الأثر هي مسح الأسر المعيشية، حيث إن معظم الحصائل المعنية ترتبط برفاهية الأسرة أو سلوكها. إلا أن عمق التحليل وجودته يتحسنان عادةً عند الاستعانة بأدوات مسح إضافية. نسرد فيما يلي أكثر أدوات المسح شيوعًا:

- *مسح الأسر المعيشية (Household survey):* تُجمع البيانات عن طريق زيارة منازل الأسرة المعيشية. يكون المجيب المستهدف عادةً هو رب الأسرة، إلا أن الوحدات المختلفة قد تكون موجهة إلى مجيبين آخرين بخلافه. وتتمثل الوحدات الشائعة في قائمة أفراد الأسرة المتضمنة للمعلومات التالية: الاسم والسن والنوع الاجتماعي والمستوى التعليمي والعلاقة برب الأسرة بالنسبة لجميع أفراد الأسرة الآخرين، وتَرد وحدات أخرى تعتمد على مجال تركيز الدراسة، مثل التعليم والصحة والأنثروبومترية (القياسات البشرية) والتوظيف والدخل والإنفاق/الاستهلاك.

- *مسح المشاريع (Enterprise survey):* تُجمع البيانات على مستوى المشروع بحيث يكون مالك المشروع أو المدير هو المجيب المستهدف. تتضمن الوحدات التوظيف والمبيعات والنفقات والائتمان. يمكن استخدام وحدة السلوك لقياس متغيرات مثل السلوك تجاه المخاطر ومعدل الخصم الذاتي.

- *مسح المرافق (Facility survey):* تُجمع البيانات على مستوى المرفق، مثل مدرسة أو عيادة صحية، بحيث يكون مدير المرفق (مدير المدرسة أو رئيس الأطباء) هو المجيب المستهدف. ستتناول الوحدات الخدمات كمًا ونوعًا. ويجوز أيضًا إدخال وحدات فرعية تستهدف المعلمين والأطباء.

- *مسح المجتمع (Community survey):* تُجمع البيانات على مستوى المجتمع (قد تكون قرية أو منطقة إدارية في مناطق حضرية مثل البرنجيه في الفلبين). يكون المجيبون المستهدفون من قادة المجتمع، ومن الممكن تضمين أشخاص مثل العمدة (المسؤول عن إدارة شؤون القرية) وسكرتير القرية في الهند. في بعض الأوقات، يجب استجواب القرويين العاديين للتحقق من المعلومات التي يقدمها زعماء القرية أو للتعرف على آراء القرويين. عادة ما يُنفذ المسح على مجموعات محددة بدلاً من أشخاص فرادي. تغطي الأسئلة النموذجية مواضيع تتعلق بمرافق القرية (المدارس، مكاتب البريد، إلخ)، والبُعد عن المرافق الأخرى التي يفتقر المجتمع إليها، ومشاريع التنمية، وسبل العيش الأساسية. من المهم بشكل خاص الإحاطة بمشاريع التنمية الأخرى في المناطق التي تنتمي إليها مجموعتي المعالجة والمقارنة.

- *مسح الوكالات (Agency survey):* تُجمع البيانات على مستوى الوكالة، مثل المنظمة غير الحكومية أو مكتب المقاطعة الحكومي المنفذ للتدخل بحيث يكون رئيس الوكالة في منطقة التدخل المجيب المستهدف. قد تغطي الوحدات الموارد والموظفين (كمًا ونوعًا)، والإجراءات المتبعة من الوكالة.

- *مسح العاملين (Worker survey):* بيانات تُجمع من الأفراد العاملين، مثل المعلمين أو العاملين في مجال الصحة، الذين يمثلون المجيبين المستهدفين. قد تتضمن الوحدات المؤهلات وبيئة العمل والرضا عن الوظيفة والمهارات. إذا تلقى العاملون تدريبًا كجزء من التدخل، يجوز للمسح أن يختبر المعرفة المكتسبة والتغيرات المستهدفة عمليًا.

يمكن استخدام مصادر بيانات متعددة لدراسة واحدة سواء عن طريق تطبيق مسوحات متعددة للدراسة (مثل المجتمع والأسرة المعيشية) أو عن طريق ربط بيانات المسح الخاصة بالتقييم بمصدر بيانات قائم، مثل بيانات هطول الأمطار من دراسة زراعية. للقيام بذلك، يكون من المفيد غالبًا جمع إحداثيات نظام الموقع العالمي (GPS) للقرويين أو المجيبين الأفراد.

عند إجراء عدة مسوحات كجزء من الدراسة، من الضروري استخدام رموز تعريفية تسمح بربط المسوحات معًا. مثلاً، الرمز التعريفي في مسح المعلمين قد يكون ccc.ss.ww، حيث ccc هو رمز تعريف المجتمع المكون من ثلاثة أرقام، وss هو رمز تعريف المدرسة المكون من رقمين (يتم تمييز المدرسة بالصيغة ccc.ss)، وww هو رمز المعلم المكون من رقمين.

تصميم أسئلة المسح

أمر أساسي أن تكون صياغة أسئلة المسح جيدة، حتى يتسنى للمسح أن يكون فعالًا. يتعين أن تحمل الصياغة الصفات الأربع التالية: موجزة وهادفة وبسيطة ومحددة (Iarossi 2006).

تكون معظم الأسئلة في مسح تقييم الأثر إما كمية أو مسبقة الترميز، أي يوجد نطاق محدود من الإجابات. من الأهمية بمكان الاحتفاظ بكل المعلومات الكمية المهمة على أنها مستمرة، حتى وإن بدا من الأبسط اختزال المتغيرات إلى فئات. في الأوقات التي تُستخدم الفئات فيها، من المتوقع أن تكون الإجابات قائمة بذاتها ومفهومة بالكامل. إذا لم ينطبق ذلك، فيجب إذًا وضع تعليمات تنص على "حدد كل ما ينطبق".

يحتوي المسح عادة على نمط التخطي، الذي يوجه مندوب الإحصاء إلى السؤال التالي الذي يجب طرحه. على سبيل المثال، إذا أجاب صاحب المشروع الصغير أنه لم يقترض في العام الماضي، فيجب تخطي الأسئلة المتعلقة بمصدر القروض التي حصل عليها العام الماضي واستخداماتها. من بين الأسباب الرئيسية التي تشجع على اختبار المسح بشكل كامل هو أن أنماط التخطي قد يصيبها الخطأ في أثناء مراجعات المسح.

كل سؤال يجب أن يرتبط بإجابة واحدة، وينبغي للأسئلة ألا تتضمن افتراضات ضمنية. ليس من الضروري البدء من الصفر عند تصميم مسح. هناك العديد من المسوحات المتوفرة، لا سيما من برامج المسح العالمية الكبرى، التي استنفدت موارد طائلة في تطوير أدوات المسح وفقًا لأفضل الممارسات. تتضمن الأمثلة الجديرة بالذكر مسوحات قياس مستويات المعيشة من البنك الدولي والمسوحات الديموغرافية والصحية. يمكن نسخ الوحدات وتكييفها من هذه المسوحات، بعد الحصول على الموافقة حيثما تطلب الأمر. تتضمن عملية التكييف إجراء تعديلات، من قبيل ضمان تضمين المدخلات المحلية والممارسات الإدارية في المسح الزراعي، باستخدام وحدات قياس محلية (التي يجب إعداد جداول مكافئة لها) وربط الأسئلة المتعلقة بالتعليم بنظام التعليم المحلي. لأغراض المقارنة بين الدراسات والجمع بين نتائجها، قد يكون من المفيد أن تُجمع البيانات باستخدام مقاربة معيارية. بالنسبة لمعظم الموضوعات المتخصصة، قد تكون نقطة البداية هي تحديد المسوحات الكبرى ذات الصلة التي أُجريت مؤخرًا والحصول على نسخ من أدوات المسح الخاصة بها.

إن تكييف الوحدات انطلاقًا من أداة مثبتة الفعالية سيجنبك الأخطاء التي يرتكبها مصممو المسوحات الذين يعوزون إلى الخبرة. قد تتضمن هذه الأخطاء ما يلي: (١) الإفراط في طرح الأسئلة غير الضرورية في المسح، (٢) طرح أسئلة على المجيب يتعذر الإجابة عنها بشكل معقول، ومثال على ذلك طرح أسئلة عن قيم المجاميع، مثل الدخل الزراعي بدلاً من تبني مقاربة مبنية على بيانات مقسمة ومفصلة أكثر دقة عن طريق السؤال عن تكاليف المدخلات وكميات الإنتاج لمحصول معين أو ماشية وما إلى ذلك، (٣) عدم الوضوح أو الغموض في الأسئلة، (٤) جمع خصائص مستمرة، مثل السن، كفئات لا تفسر الشيء الكثير وذات أهمية قليلة للتحليل، و(٥) تبني نمط التخطي غير الملائم والتصميم غير المنطقي فيما يتعلق بتسلسل المسح (يوفر القسم ٤ في الملحق الثاني مزيدًا من التفاصيل حول تصميم المسح).

الاختبار ثم الاختبار ثم الاختبار مرة أخرى

يرى كل من بريسر وبلير أنه من الأهمية بمكان أن تخضع أدوات المسح لعمليات اختبار قبلي قبل تنفيذ المسح (Presser and Blair 1994). يجب أن يتم الاختبار أولاً داخل الفريق، ومن ثم مع مندوبي الإحصاء، وأخيرًا اختبار تجريبي تحت ظروف ميدانية من قِبل مندوبي الإحصاء. غالبًا ما تكون هناك حاجة إلى أن يتم الاختبار الميداني على مدى عدة أسابيع حتى تخضع أدوات المسح لعدة مراحل من المراجعة. على الرغم من أن العملية قد تكون مستهلكة للوقت، فهي أساسية للحصول على تقييم أثر ذي مغزى.

إذا لم يخضع المسح رديء التصميم للاختبار كما ينبغي، فستكون النتائج عبارة عن بيانات رديئة الجودة (غالبًا تكون لا قيمة لها) ونتائج غير حاسمة أو زائفة.

الوثائق الداعمة

كل أداة مسح يجب أن يصاحبها دليل إرشادي يصوغ الدافع العام لإجراء المسح، والغرض من كل وحدة، والإجابات عن أي أسئلة قد يتم طرحها على مندوبي الإحصاء. يجب أن تتعامل هذه الأدلة الإرشادية أيضًا مع *قضايا مثل الإقدام على المجتمع* (كيفية إقدام فريق المسح على التواصل مع قادة المجتمع) وعملية اختيار المجيبين والاستبدال إذا تعذر تحديد موقع المجيبين.

الترجمة

في معظم الحالات، لا تكون الإنجليزية هي اللغة الأم للمجتمع الإحصائي الذي ينتمي إليه المجيب، وحينها يلزم ترجمة المسح. ولكن أحيانًا إذا كانت مهارة مندوب الإحصاء جيدة في اللغة الإنجليزية وكان على فهم جيد للأسئلة، فيمكن وقتئذ الاستبقاء على أدوات المسح باللغة الإنجليزية. في تلك الحالات، يجب أن يتضمن تدريب مندوب الإحصاء مناقشة المصطلحات اللغوية المحلية للعناصر غير الشائعة أو العناصر التي ليست لها ترجمة واضحة مباشرة. في حالات حيث تكون الترجمة ضرورية، يتعين إتمام الترجمة قبل الاختبار الميداني بحيث يمكن تحديد أي قصور في الترجمة.

أدوات المسح الإلكترونية

يزداد استخدام الحواسيب المحمولة الصغيرة (النت بوك) والحواسيب اللوحية (التابلت) بشكل شائع بالنسبة لجمع البيانات إلكترونيًا. على الرغم من أن إعداد الاستبيانات الإلكترونية يستهلك وقتًا أطول مبدئيًا مقارنة بالاستبيانات الورقية، فهناك العديد من المزايا لهذه المقاربة: (١) يمكن في كثير من الأحيان رصد عملية جمع البيانات للتأكد من أن مندوبي الإحصاء يقضون وقتًا كافيًا مع المجيبين في الميدان، (٢) عادة ما يكون من الأسهل إجراء تعديلات على المسح، حتى ولو مرة واحدة في الميدان إذا لزم الأمر، على الرغم من أن المقاربة الأفضل تنطوي على اتصال مندوبي الإحصاء بالإنترنت لتحديث المسح، (٣) يتم تضمين نمط التخطي والتحقق من الاتساق وحدود النطاق في برنامج الجهاز، كما هي الحال في العبارات الموسومة، مما يسمح بتجنب أخطاء مندوبي الإحصاء، (٤) إمكانية استبعاد أي إجابات غير صالحة، (٥) إمكانية إرسال البيانات إلى مقر الدراسة في الوقت الفعلي لفحصها، بحيث يمكن توسيم المجيبين الذين يحتاجون إلى إعادة النظر أو اعتبار إجاباتهم خطأ بسبب أخطاء في إجراء الإحصاء، (٦) ليست هناك حاجة لإدخال البيانات بشكل منفصل كما هي الحال في الطريقة اليدوية التي تستغرق وقتًا طويلاً وتؤدي إلى أخطاء محتملة، على الرغم من أن القدرة على تحرير البيانات قد تكون محدودة في الأنظمة الأساسية المدعومة مثل Survey Monkey، وأخيرًا (٧) في ذلك توفير في الورق وحماية للبيئة! يوضح المربع ٦-٢ حزم البرامج الرائدة لجمع البيانات إلكترونيًا.

المربع ٦-٢: منصات لإجراء المقابلات الشخصية بمساعدة الحاسوب

تتوافر العديد من المنصات البرمجية المجانية التي تمكّن من جمع البيانات إلكترونيًا (من دون استعمال الورق)، مما يجعل المسوحات أكثر دقة وأسرع وأقل كلفة. بالإضافة إلى البرامج المجانية المذكورة أدناه، تتوفر أيضًا العديد من المنصات نظير مقابل مادي.

CSPro هي أقدم وأكبر منصة لإجراء المقابلات الشخصية بمساعدة الحاسوب (CAPI) وطورها مكتب تعداد الولايات المتحدة. وهي عبارة عن حل مغلق المصدر يمكن تشغيله على نظامي Windows وAndroid. تكون الاستبيانات مبرمجة بلغة فريدة لهذه المنصة. ومن الممكن مزامنة البيانات في الوقت الفعلي عن طريق استخدام المنصة جنبًا إلى جنب مع Dropbox أو مشاركة الملفات نظير إلى نظير FTP. وتتوفر إمكانية لتصدير البيانات المجمعة مباشرة إلى البرامج الإحصائية الشائعة. https://www.census.gov/ international/software/cspro/index.html /population

Open Data Kit (ODK) تم تطويرها بالتعاون مع جامعة واشنطن وشركة غوغل (Google) لتكون منصة مفتوحة المصدر تتيح جمع البيانات عبر أجهزة التابلت (الحواسيب اللوحية) والهواتف المحمولة التي تعمل بنظام أندرويد (Android). وتقدم هذه المنصة عدة أدوات للمساعدة على إنشاء مسوحات بتنسيق xml معياري. لكن، يلزم تثبيت خادم Aggregate المخصص لتجميع البيانات من أجل استخدام هذه المنصة. وتوجد مجموعة متنوعة من البيانات الاستشعارية، بما في ذلك الإحداثيات من نُظم تحديد المواقع العالمية (GPS) والكاميرات. ويمكن مراقبة إدخال البيانات في الوقت الفعلي عبر Google Maps أو أي جهة استعلامات أخرى. https:// opendatakit.org/

Survey Solutions هي منصة لإجراء المقابلات الشخصية بمساعدة الحاسوب الخاصة بالبنك الدولي. وهي عبارة عن منصة سحابية مغلقة المصدر يستضيفها البنك الدولي لإدارة الاستبيانات والبيانات، وتعمل على أجهزة التابلت المشغلة بنظام Android. على الرغم من أن البرنامج مجاني، فإن استخدام منصة البيانات السحابية يكون مدفوعًا. وهي تستخدم تراكيب C# كلغة برمجة الاستبيانات، التي يمكن أيضًا إتمامها من خلال أدوات تصميم عبر الإنترنت. وعلى غرار ODK، يمكن أن تسمح بتضمين مختلف البيانات الاستشعارية مباشرة، ويمكن مراقبة إدخال البيانات في الوقت الفعلي. وتسمح هذه المنصة أيضًا بتصدير البيانات مباشرة إلى أهم البرامج الإحصائية. http://support.mysurvey.solutions/

المصدر: المؤلفان

يتمثل العيب الأساسي في أن العديد من حزم برامج جمع البيانات إلكترونياً - ولكن ليس كلها - تطبق معايير بها شيء من الصرامة، حيث إنها قد تقيد ترتيب تخطي الأسئلة، أو تنسيق التسجيل، أو قد تجعل من الصعب تدوين ملاحظات أو إجراء حسابات يدوية. يمكن أن تؤدي هذه الصرامة إلى حدوث خطأ إذا لم يتم اختبار الجهاز مسبقًا بشكل كافٍ، وبالتالي يمكن زيادة العبء على الاختبار المسبق. قد يواجه الجهاز أيضًا مشكلات، ويلزم الوصول المنتظم إلى مصادر الطاقة، وربما أيضًا شبكة Wi-Fi أو إشارة هاتف محمول قوية.

دور البيانات النوعية

تتعزز تصاميم تقييم الأثر باستخدام الطرق المختلطة التي تلعب فيها البيانات الكمية والنوعية أدوارًا تكاملية. يمكن أن تلعب البيانات النوعية الأدوار التالية:

- يمكن استخدام البيانات النوعية في المرحلة التكوينية لدعم تصميم التقييم والمسح بالمعلومات الوافية.
- يمكن للبيانات النوعية أن تلتقط مشكلات حساسة أو غير مستوعبة بشكل كامل متعلقة بتنفيذ التدخل، مثل عوائق المشاركة ومشكلات التنفيذ... إلخ.
- بإمكان البيانات النوعية أن تساعد على تفسير استنتاجات الدراسة وتعميم الخلاصات.

يمكن استخدام مجموعة من التقنيات النوعية، مثل المقابلات المفتوحة مع موظفي المشروع وقادة المجتمع، وتواريخ الحياة الشفوية، ومجموعات التركيز، ورسم الخرائط، والمقاطع العرضية لجمع هذه البيانات. من المفيد جدًا لأعضاء فريق الدراسة الأساسي قضاء بعض الوقت في الميدان في أثناء أداء بعض هذه الممارسات. يجب إخضاع الأدوات المقرر استخدامها لجمع البيانات النوعية إلى اختبار مسبق على غرار أدوات المسح المنظمة. في بعض الحالات، قد يكون جمع البيانات النوعية جزءًا من مرحلة تشخيص أولية قبل تصميم تقييم الأثر الكامل.

بدائل بيانات خط الأساس

تزداد فعالية تصاميم تقييم الأثر في ظل وجود بيانات خط الأساس. إذا جُمعت بيانات خط الأساس في وقت مبكر بما فيه الكفاية، يمكن للمقيِّمين استخدامها أيضًا للتحقق من الموازنة (بين مجموعة المعالجة والمجموعة الضابطة) على مستوى بيانات خط الأساس، وللمطابقة مع المؤشرات التي لا تتأثر بالضرورة بالتدخل، ولحساب مقدِّر أثر التباين المزدوج، ومن شأن ذلك أن يقلل من متطلبات حجم العينة.

إلا أن بيانات خط الأساس لا تتوفر على الدوام. في بعض الحالات، من الممكن استخدام بدائل لخط الأساس. ولكن ذلك الحل أيضًا لا يكون متيسرًا أو مصدرًا جيدًا للبيانات في العديد من الحالات. إلا أن بعض تقييمات الأثر عالية الجودة قد أُجريت مع غياب بيانات خط الأساس الرسمية، وذلك في ظل توافر مصادر أخرى كالبيانات الإدارية. يمكن استخدام الطرق التالية للاستعاضة عن بيانات خط الأساس:

- مسح حالي يتطرق إلى نفس مجال التدخل المعني، بشرط أن يكون قد أُجري في وقت قريب من بدء التدخل، الذي تم فيه جمع بيانات حول حصائل محل اهتمام وخصائص مطلوبة لعقد المطابقة. يجب وضع بيانات التعداد في الاعتبار أيضًا لأن لها أفضلية بارزة في التغطية العامة. ومع ذلك، قد لا تكون بيانات التعداد هذه قريبة بما يكفي من وقت بداية التدخل ويكون مدى شمول البيانات محدودًا.

- البيانات الإدارية، مثل نُظم معلومات إدارة التعليم، لها ميزة التغطية الوطنية، إلا أنها غالبًا تتعلق بالمرافق العامة (المدارس الخاصة قد تكون مستثناة، مثلاً).
- توفر بيانات القمر الصناعي معلومات حول استخدام الأراضي الذي يمكن أن يستعمل لتحليل الحصائل، مثل أنماط زراعة المحاصيل والتصحر.
- يمكن الاعتماد على قدرة المجيبين على التذكر، ولكن يجب أن يكون المسح محصورًا على أشياء أساسية (مثل الأحداث الحياتية ومشتريات الأصول الكبيرة) على الأرجح أن يتذكرها المجيبون بشكل دقيق. إن توافرت عدة خصائص لم تتأثر بالتدخل، فيمكن جمعها بقيمها الوقائعية (الفعلية)، مثل تعليم أفراد الأسرة البالغين، والدين، وجنس رب الأسرة، وهكذا.

خطة التحليل

من شأن وضع خطة للتحليل أن يساعد على استهداف البيانات المُراد جمعها، عن طريق صياغة آليات إستراتيجية تحديد العناصر بشكل واضح. قد تتضمن الخطة توصيف جميع النماذج المقرر تقديرها والجداول التي يتعين تضمينها في التقرير. تكتمل خطة التحليل بعد تحديد كل المتغيرات بشكل كامل وربطها بالأسئلة الواردة في الاستبيان. يضمن وضع خطة تحليل كاملة عدم حدوث أي ثغرات لاحقًا تهم البيانات المطلوبة، وتجنب جمع أي بيانات غير ضرورية.

من الممكن استخدام البيانات التي تم جمعها في أثناء إجراء المسح التجريبي، أو إنشاء بيانات صورية للاستبيان من أجل اختبار التحاليل المعنية عمليًا. تساعد هذه المقاربة على تحديد أي أخطاء غير متوقعة في الترميز أو غيرها من المشكلات التي قد تتطلب مراجعة لأدوات المسح.

٦-٥ أخذ العينات

للإتيان بتقييم أثر صالح، يجب تصميم إستراتيجية لأخذ العينات من أجل التأكد أن العينة التي تم جمعها (١) ممثلة للمجتمع الإحصائي المعالج (المستفيد)، و(٢) تسمح بتحديد مجموعة ضابطة أو مقارنة صالحة. عندما يكون من المتوقع نشوء تداعيات، فسيكون لذلك أيضًا انعكاسات على تصميم العينة.

سوف يستند أخذ العينات إلى الاختيار العشوائي لوحدات المسح من إطار أخذ العينات يحدد بدوره كل الوحدات المؤهلة. عادة ما تكون هناك ضرورة لاستخدام تصميم العينة العنقودية. على سبيل المثال، مسح الشركات لن يختار بشكل عشوائي كل الشركات في الدولة، ولكن سيأخذ أولًا عينة من عدد من المقاطعات أو المقاطعات الفرعية، ومن ثم عينة من الشركات داخل المقاطعات أو المقاطعات الفرعية المختارة. عادة ما يلجأ المقيّمون إلى التصاميم العنقودية، ويظهر ذلك جليًا حيث إن التدخلات غالبًا ما تُقدم على مستوى عنقودي، مثل المجتمع أو المنطقة. ومع ذلك، تتطلب التصاميم العنقودية حجم عينة أكبر من العينة العشوائية لاكتساب قوة إحصائية مكافئة. يجب أن تسمح حسابات القوة الإحصائية بتنفيذ التصميم العنقودي (كما ناقشنا في الفصل ٧ والملحق الثاني).

أحيانًا لا يتوفر إطار العينة سوى للمرحلة الأولى من أخذ العينة العنقودية. على سبيل المثال، أن تتوفر قائمة بالقرى ولكن ليس كل الأسر المعيشية في تلك القرى. في هذه الحالة، يجب إعداد قائمة إحصائية بلا شك. يتم اختيار القرى من إطار أخذ العينات، وبعد ذلك تزور الفرق القرى المختارة لإنشاء قائمة بالأسر المعيشية المؤهلة في تلك القرى (أي "قائمة إحصائية") لأخذ عينة عشوائية منها.

قد يتضمن تصميم العينة القيام بأخذ عينات طبقية تقسم أولاً المجتمع الإحصائي المؤهل إلى مجموعات (مثل ريفية وحضرية، وفقيرة وغير فقيرة، وكل الأسر في كل مقاطعة معنية بالمسح)، ثم تؤخذ عينة عشوائية من كل مجموعة. يمكن استخدام التقسيم الطبقي عندما يكون من المخطط إجراء تحليل لمجموعة فرعية، وذلك لضمان الحصول على حجم عينة كافٍ من كل مجموعة فرعية. يمكن إتمام تحليل المجموعة الفرعية من دون إتمام خطوة التقسيم الطبقي، ولكن يستدعي ذلك العمل على عينة أكبر عامةً. يوفر القسم ٢ في الملحق الثاني مزيدًا من التفاصيل حول تصميم أخذ العينات.

يتعين عدم الخلط بين أخذ العينات العشوائي، الذي يُعد ضروريًا لضمان الحصول على عينة ممثلة، والتعيين العشوائي للتدخل. استخدام طريقة أخذ العينات العشوائي لا يجعل الدراسة تقييمًا عشوائيًا، لأنه يتطلب تعيينًا عشوائيًا.

اختيار مجموعة المقارنة

كيف تختار مجموعة مقارنة صالحة هو موضوع محوري في هذا الكتاب الذي بين يدي القارئ. ينبغي لتصميم التقييم تحديد إستراتيجية أخذ العينات، وليس العكس. لذا، إذا كان سيتم استخدام التعيين العشوائي، فهذا التعيين العشوائي ما يحدد المجتمعين الإحصائيين المعالَج والمقارَن، وكذلك تصميم المسح. تتطلب التجربة العشوائية العنقودية المضبوطة تصميم العينة العنقودية. عند استخدام طرق غير تجريبية، يجب أن يهدف تصميم العينة إلى جمع البيانات من مجموعات المقارنة المرجحة لأن تكون مؤهلة بُغية تجنب الاضطرار إلى التخلص من كميات كبيرة من البيانات. بالنسبة للتصاميم العنقودية، يجوز تحديد المجتمعات المطابقة باستخدام مجموعة بيانات مختلفة، مثل بيانات التعداد أو البيانات الإدارية.

التداعيات

إذا كان من المتوقع أن تطرأ تداعيات مهمة للتدخل تتجاوز نطاق المستفيدين، فإن تحليل تأثيرات التداعيات يعتمد على غير المستفيدين الذين يتم تضمينهم في العينة. ويختلف هؤلاء غير المستفيدين عن مجموعة المقارنة التي يجب أن تتألف من غير مستفيدين لا يختبرون أي تداعيات.

تحليل المجموعة الفرعية

في بحث أجراه كل من ليندن وبانيرجي ودوفلو، خلصوا إلى أن تغايرية الأثر تعني أن الأثر قد يتباين بطرق مختلفة: الأثر تعني أن الأثر قد يتباين بطرق مختلفة: حسب النوع والفئة الاجتماعية والموقع والوقت من العام، إلخ. مثلاً، تبين أن التعلم بمساعدة الحاسوب في جمهورية الصين الشعبية يحسن من درجات الرياضيات بين الثلث السفلي والوسطي والأعلى من توزيع الاختبار القبلي حسب الانحرافات المعيارية ٠,٤٣ و٠,٣٥ و٠,٣٣ على التوالي (Linden, Banerjee, and Duflo 2003). والمُلَاحَظِ أن التأثير كان ملموسًا أكثر بالنسبة للطلاب الأضعف أداءً.

من الممكن اكتساب بعض الرؤى حول التغايرية المتوقعة، استنادًا إلى نظرية التغيير. لكن، من الأفضل إجراء تحديد مسبق لتحليل المجموعة الفرعية المستهدفة في بروتوكول تصميم الدراسة من أجل تجنب التنقيب عن البيانات. ومع ذلك، هذا لا يمنع الإبلاغ عن أي تغايرية مهمة قد يتم تسليط الضوء عليها في أثناء العمل الميداني أو التحليل، إذا كان يمكن دعمه بالأدلة والاستدلالات السليمة.

يجب أن يسمح أخذ العينات بتحليل المجموعة الفرعية المخطط له. قد يتبين أن التدخل ذو فعالية بشكل عام، ولكن قد لا يكون له أثر على النساء. رغم ذلك، قد يرجع ذلك القصور في تحديد الأثر إلى أن حجم العينة ليس كافيًا للكشف عن الأثر الواقع على عينة فرعية. قد لا تتوصل الدراسة إلى أي أثر يُذكر على الرجال أيضًا!

٦-٦ إدارة جمع البيانات

لضمان الاتساق مع التحاليل المستهدفة، يجب إدارة جمع البيانات من قِبل الفريق نفسه الذي سيجري تقييم الأثر. تَعَذر ذلك، يجب أن تضمن الآليات الأخرى إذًا التنسيق المناسب بين الفريقين، بحيث تفي البيانات بمتطلبات تقييم الأثر.

إذا كان فريق تقييم الأثر مسؤولاً عن جمع البيانات، فيجوز له استئجار مندوبي الإحصاء وتدريبهم، ويجوز له استئجار شركة مسح محلية. يوفر القسم ٥ في الملحق الثاني مزيدًا من المعلومات حول كيفية إدارة هذه العملية. على سبيل المثال، تقييم أثر "برنامج الطوابع الغذائية في منغوليا" المدعوم من بنك التنمية الآسيوي تم إجراؤه من قِبل شركة Oxford Policy Management التي تعاقدت من الباطن مع مركز التدريب والأبحاث السكانية التابع لكلية الاقتصاد في جامعة منغوليا الوطنية، من أجل إدارة جمع البيانات. أما بالنسبة لمنظمات تقييم الأثر التي لديها مكاتب محلية، مثل Innovations for Poverty Action التي لها مقرات في العديد من الدول بما فيها الهند والفلبين، فعادةً ما تشكل فرقها الخاصة.

في كلتا الحالتين، يتطلب جمع البيانات الإدارة الحريصة من قِبل أعضاء فريق تقييم الأثر. يجب أن يشاركوا في الاختبار المسبق للتأكد من تمثيل المتغيرات الأساسية بشكل مناسب، وتدريب مندوبي الإحصاء بما يضمن الفهم الصحيح للاستبيان، وتوفير الإشراف الميداني المستقل في أثناء تنفيذ المسح.

غالبًا ما يتطلب تدريب مندوبي الإحصاء فترة زمنية طويلة لضمان الفهم المناسب لتنفيذ المسح، بما في ذلك الاختبار الميداني. يحسن هذا التدريب من جودة المسح ويحد من الأخطاء غير المرتبطة بأخذ العينات. يشكل تقمص الأدوار جزءًا مهمًا من التدريب، حيث يجري مندوبو الإحصاء العملية المقررة مع بعضهم البعض باستخدام أدوات المسح. يستطيع المشاركون الآخرون ملاحظة العملية ومناقشتها. من خلال نشاط تقمص الأدوار، يمكن إحداث تغيرات إضافية على الأدوات لإزالة أي غموض أو أخطاء (مثلما يحدث في نموذج التخطي). ينبغي للتدريب أن يغطي أيضًا الجوانب الأخلاقية من جمع البيانات، مثل الموافقة المستنيرة.

في أثناء جمع البيانات، قد يساعد اضطلاع المدير الميداني للمسح والمشرف الميداني المستقل الخاص بفريق تقييم الأثر بعمليات فحص دورية، للكشف عن جودة البيانات، على تنفيذ المسح بشكل مناسب. من الشائع بشكل متزايد إدخال البيانات إلكترونيًا في الميدان، وفي هذه الحالة تساعد عمليات فحص تناسق البيانات المضمنة في البرنامج على ضمان جودة البيانات. يمكن أيضًا لجمع البيانات الإلكتروني أن يسمح بنقل البيانات ومراقبتها في الوقت الفعلي.

من بين طرق إجراء عمليات فحص الجودة هي التحقق من صلاحية عينة صغيرة من البيانات. يمكن إتمام ذلك من قِبل المشرف الميداني عن طريق إعادة مسح أسر معيشية مختارة. وبدلاً من ذلك، يمكن الاستعانة بشركة مستقلة أو خبير لإعادة المسح، ربما عن طريق الهاتف حيثما كان ذلك ممكنًا. مندوبو الإحصاء الذين يكونون على معرفة بطبيعة مجموعة من أجوبة المسح القياسية قد يجدون إغراءً في التقصير في واجباتهم أو حتى الغش في جمع البيانات ما لم يكونوا مدركين بوجود تدابير مطبقة للكشف عن مثل هذه الممارسات الخاطئة. لهذا السبب، يجب التحقق من صحة البيانات دائمًا من خلال المجيبين أنفسهم، على الأقل على أساس عينة فرعية عشوائية، وسيفيد أكثر أيضًا التأكد من تسجيل الإحداثيات الجغرافية من نظام تحديد المواقع العالمي (GPS)، وفي ذلك طمأنة من أن المواقع المستهدفة زارها مندوب الإحصاء بالفعل.

عادة ما تكون شركة المسح مسؤولة عن التنقية الأولية للبيانات، إلا أن عادةً ما يتطلب الأمر جهدًا إضافيًا لتنقية البيانات من قِبل الفريق المسؤول عن إجراء التحليل. في الغالب، تنتج عن العملية الأخيرة استفسارات محددة للتحقق من إجابات معينة. لهذا السبب، يجب الاحتفاظ بنماذج الاستبيان المكتملة الأصلية لحين الانتهاء من تحليل البيانات. ويتعين تضمين بروتوكول أو وصف لمقاربة تنقية البيانات فضلاً عن الاحتفاظ بنسخة من البيانات غير المنقاة. ذلك الشرط يجب تضمينه في العقد المبرم مع الأشخاص المسؤولين عن جمع البيانات.

الاعتبارات الأخلاقية

جزء كبير من أخلاقيات تقييم الأثر (كما ناقشنا في الفصل ٨) يتعلق بجمع البيانات.

تكمن المشكلات الأساسية المتعلقة بالحفاظ على أخلاقيات جمع البيانات في ما يلي:

١. **الحصول على الموافقات الأخلاقية الضرورية:** إذا كان فريق الأبحاث ينتمي إلى مؤسسة أكاديمية، فإذًا على الأرجح ستطلب المؤسسة موافقة من مجلس الإدارة المعني بالمراجعة (تحذير: يمكن أن يستغرق ذلك بعض الوقت!). كما قد يكون للدولة التي تُجرى فيها الدراسة متطلبات موافقة أخلاقية.

٢. **الموافقة المستنيرة من المجيبين:** بالنسبة للتجارب العشوائية المضبوطة العنقودية، قد يكون من المقبول الحصول على الموافقة على المستوى العنقودي، مثل قادة المجتمع بالنسبة للقرى، وكبار المعلمين بالنسبة للمدارس. يوفر المربع ٦-٣ مثالاً لنص الموافقة المستنيرة.

٣. **تعويض المجيبين:** بالنسبة للمسوحات التي تتضمن العديد من الأسئلة، قد يتطلب الأمر تقديم تعويضات لأصحاب المعلومات نظير الوقت الذي كرسوه للإجابة عن الأسئلة لضمان ارتفاع معدل إكمال المسوحات. يجب ألا يؤثر التعويض على الحصيلة، وهذه مسألة يحتاج فريق الدراسة إلى وضعها في الاعتبار بجدية. بالنسبة للأسر المعيشية، يكفي تقديم هدية عينية بسيطة، مثل قلم رصاص وقلم جاف ومفكرة. يجوز أيضًا تقديم تبرع إلى القرية، لا سيما في حالة مجتمعات المقارنة.

المربع ٦-٣: مثال لنص الموافقة المستنيرة

مرحبًا، اسمي _____، وأنا أجري مسح حول معدل استخدام الطريق. الجهة المسؤولة عن إجراء هذا البحث هي _____. ستساعد هذه المعلومات على تحديد ما إذا كانت تنمية الطريق نتج عنها انخفاض في الأسعار ومصالح للمستهلكين. عادة يستغرق المسح ما يصل إلى ٣٠ دقيقة لإكماله. نطمئنك بأن المعلومات التي تقدمها/تقدمينها ستظل سرية. يسعدنا كثيرًا مشاركتك.

المشاركة طوعية في هذا المسح، ويمكنك الامتناع عن الإجابة عن أي سؤال أو جميع الأسئلة حسب رغبتك. يمكنك التوقف عن استكمال المسح في أي لحظة. ولكننا نأمل أن تشارك/تشاركي في هذا المسح لأن رأيك يهمنا.

والآن، هل لديك أي سؤال بخصوص المسح؟ انتظر الرد.

هل توافق/توافقين على إجراء المسح؟ نعم أم لا.

المصدر: المؤلفان

التوقيت

هناك مشكلتان متعلقتان بتوقيت المسح: (١) متى تجمع البيانات، و(٢) كم تستغرق عملية المسح. تعتمد المشكلة الأخيرة على حجم العينة وتوزيعها الجغرافي وأوقات التنقل وطول الاستبيان وكم عدد الزيارات الضرورية لاستكمال الاستبيان. عادةً يكون من الضروري القيام بأكثر من زيارة واحدة بسبب عدم إتاحة المجيب. قد تساعد الزيارات المتعددة أيضًا على التخفيف من تعب المجيب في حالة كانت المسوحات طويلة.

ثمة أيضًا سؤال يطرح حول عدد جولات جمع البيانات المطلوب إجرائها. عندما يكون المسح مطلوبًا، يكون من الضروري الحصول على بيانات خط النهاية بالطبع. كما أكدنا في جميع أجزاء هذا الكتاب، ننصح بشدة بالحصول على بيانات خط الأساس. بالنسبة للتدخلات التي تستغرق مدة أطول، ننصح أيضًا بمسح في منتصف المدة، وذلك للتركيز على جوانب العملية أو الحصائل المتوسطة غالبًا. وأخيرًا، سيكون من المفيد للغاية إجراء مسح ما بعد خط النهاية بعد انتهاء التدخل لتحديد ما إذا كانت الفوائد ما تزال مستمرة أو امتدت إلى مجتمع إحصائي أوسع.

يتقيد توقيت المسوحات بعاملين. أولاً، توقيت جمع بيانات خط الأساس وخط النهاية غالبًا ما يرتبط بعوامل محددة حسب المشروع لبداية الأنشطة الجاري تقييمها ونهايتها وتوفر القدرة المالية لتمويل جمع البيانات. ثانيًا، بالنسبة لأي مسح، توجد اعتبارات للتوقيت محددة حسب السياق، مثل عدم القدرة على التنقل بسهولة في موسم الأمطار، والاحتفالات الكبرى، والهجرة الموسمية، ومواسم الذروة التي تؤثر على إتاحة المجيبين، والتقويم المدرسي بالنسبة للمسوحات المتعلقة بالتعليم.

الكثير من المتغيرات لها طابع موسمي، لذلك من الأحرى إجراء كل جولة من جولات المسح في الوقت نفسه من العام. بالطبع البيانات التي يتم الحصول عليها من مناطق المعالجة والمقارنة بحاجة إلى الجمع في الوقت نفسه.

من الضروري إتاحة الوقت الوفير لإتمام جمع البيانات حسب أعلى المقاييس. إن الوقت المطلوب من البدء في تصميم المسح إلى تنفيذه غالبًا ما يمتد لبضعة أشهر أو ما يزيد على ذلك. ويتم خلال هذه الفترة وضع تصميم المسح واختباره وتعزيزه بالوثائق، وتدريب مندوبي الإحصاء، وتحديد العينة، وترتيب اللوجستيات المطلوبة لجمع البيانات.

غالبًا ما يستغرق تنفيذ المسح من أسبوعين إلى ٦ أسابيع حسب حجم العينة والتغطية الجغرافية، ولكن قد يتطلب الأمر وقتًا أطول بالنسبة للمسوحات الكبيرة أو أدوات المسح الطويلة. يلزم شهرٌ إضافي أو أكثر لإدخال البيانات وتنقية البيانات الأولية، وذلك بالنسبة للمسوحات الورقية.

الأذونات

في العديد من الدول، يحتاج فريق المسح إلى إذن رسمي لإجراء المسح. وقد يزداد الأمر صعوبة عند إجراء المسوحات في مناطق المقارنة، حيث لا يكون هناك وجود للمشروع. في مثل هذه الحالات، يلزم بشكل خاص الحصول على أذونات صريحة من السلطة المعنية على المستوى المحلي. قد يحتاج مدير المشروع إلى العمل على تسهيل الحصول على هذه الأذونات.

أرشفة البيانات

لتسهيل استخدام البيانات التي تم جمعها، يُنصح بإتباع بعض الممارسات الجيدة كالقيام بأرشفة البيانات المجموعة والوثائق الداعمة لها، مثل الاستبيانات، مع إتاحتها للجمهور. وفي نفس السياق، من الملاحظ زيادة الإقبال على أرشفة الملفات المستخدمة في إجراء التحليل أيضًا. عادة ما تصبح البيانات متاحة للعامة بمجرد توافرها بعد اكتمال التحليل بعام أو اثنين.

يجب إزالة الهوية من البيانات قبل السماح للجمهور بالوصول إليها. وهذا يعني أن أي معرفات يمكن أن تسمح بتحديد هوية الأسرة المعيشية أو الأفراد أو الشركات، يجب إزالتها من الإصدارات العامة (ولكن مع الحفاظ عليها في الأرشيف غير العام بحيث يمكن إعادة زيارة المجيبين لاحقًا). قد يلزم تضمين الأرشفة في العقد بالنسبة لفريق تقييم الأثر أو الجهة المكلفة بإجراء المسح، بما أنها تتطلب تخصيص بعض الجهود والموارد الإضافية. يجب أيضًا وضع علامات معيارية للملفات وتشفير كلمة مرور التخزين للحفاظ على أمن البيانات. يوفر القسم ٦ في الملحق الثاني مزيدًا من التفاصيل حول إدارة البيانات.

المراجع

Banda, J. P. 2003. Nonsampling Errors in Surveys. Paper prepared for Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys. United Nations Secretariat, New York. 3–5 December 2003.

Iarossi, G. 2006. The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents. Washington, DC: World Bank.

Linden L., A. Banerjee, and E. Duflo. 2003. Computer-Assisted Learning: Evidence from a Randomized Experiment. Poverty Action Lab Paper 5.

Presser, S. and J. Blair. 1994. Survey Pretesting: Do Different Methods Produce Different Results? *Sociological Methodology*. 24. pp. 73–104.

مراجع إضافية للإطلاع

Grosh, M. and P. Glewwe. 2000. Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Volumes 1, 2, and 3. Washington, DC: The World Bank. http://go.worldbank.org/NTQLJEEXQ0.

Pollock, H. D., E. Chuang, and S. Wykstra. 2015. *Reproducible Research: Best Practices for Data and Code Management*. New Haven, CT: Innovations for Poverty Action. http://www.poverty-action.org/ sites/ default/files/publications/IPA%27s%20Best%20Practices%20 for%20Data%20and%20Code%20 Management_Nov2015.pdf.

Statistical Services Centre, University of Reading. 2009. International Household Survey Network: Survey Quality Assessment Framework. http://www.ihsn.org/sites/default/files/resources/SQAF_ Draft.pdf.

Stecklov, G. and A. Weinreb. 2010. Improving the Quality of Data and Impact-Evaluation Studies in Developing Countries. Impact Evaluation Guidelines Technical Notes No. 1. Washington, DC: Inter-American Development Bank. http://www.iadb.org/document. cfm?id=39376032.

United Nations Department of Economic and Social Affairs, Statistics Division. 2005. Household Sample Surveys in Developing and Transition Countries. Studies in Methods Series F No. 96. New York. http://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf.

الفصل ٧
تحديد حجم العينة لجمع البيانات

رسائل أساسية

- القوة الإحصائية هي احتمال توصل الدراسة إلى أثر ملحوظ، في حال وجوده في الأساس.
- يلزم إجراء حسابات القوة الإحصائية من أجل فهم حجم العينة المطلوبة لاكتشاف أثر بالاحتمال المنشود. ويتعين إجراء هذه الحسابات بشكل منفصل لكل متغير ناتج.
- من الشائع جدًا وجود دراسات ضعيفة وغير فعالة، نظرًا أسباب. وهذه مخاطرة يجدر التقليل منها؛ حيث إن الدراسة الضعيفة قد تخلص بالخطأ إلى أن التدخل ليس له تأثير قابل للاكتشاف.
- تتطلب معظم تقييمات أثر التدخلات الإنمائية أخذ عينات عنقودية. ما يهم أكثر بالنسبة للقوة الإحصائية هو عدد المجموعات العنقودية، وليس عدد وحدات العينة الممثلة في كل مجموعة عنقودية.

٧-١ حسابات القوة الإحصائية: مقدمة

ترتكز تقييمات الأثر على عينة من مناطق التدخل والمناطق الضابطة/مناطق المقارنة لتقدير التأثيرات على الحصائل. وكلما زاد حجم العينة، زاد الاحتمال بأن تكون ممثلة للمجتمع الإحصائي الذي أُخذت منه العينة.

مثلاً، يهدف برنامج للتوسيع الزراعي إلى تعميم الفائدة على ١٠ آلاف مزارع. سيكون الأمر مكلفًا للغاية إذا كان المسح سيغطي جميع الأسر المعيشية التي ينتمي إليها المزارعون، وكذلك بالنسبة للمجتمع الإحصائي المقارن المكافئ. كم عدد المزارعين الذين يجب أن يشملهم المسح؟ للإجابة عن هذا السؤال، لا بد من إجراء حساب للقوة الإحصائية.

يجب إجراء حسابات القوة لتحديد حجم العينة الكافية لتكون الدراسة مؤهلة لاكتشاف تأثيرات للتدخل ذات دلالة إحصائية.

إذا كان حجم العينة صغيرًا للغاية، فإذًا الدراسة تكون "ضعيفة القوة الإحصائية" وتنشأ مخاطرة بعدم اكتشاف الدراسة لأثر ملحوظ حتى إن كان موجودًا بالفعل. أما إن كانت العينة أكبر مما ينبغي، فستشكل عبئًا ثقيلًا على ميزانية الدراسة. يوضح هذا القسم الغرض من وراء حسابات القوة الإحصائية والمعلومات المطلوبة لإجرائها. كما سنوضح اعتبارات إضافية في الملحق الثاني.

أهم المبادئ التي يجب أن يسترشد بها مديرو تقييم الأثر نسردها فيما يلي: (١) يجب إجراء حسابات القوة الإحصائية كجزء من تصميم الدراسة، (٢) يجب التحقق من صحة حسابات القوة الإحصائية بشكل مستقل من قِبل شخص يتمتع بالمهارات الإحصائية الضرورية، و(٣) يلزم الحصول على حجم عينة كافٍ لتجنب مخاطرة الاستثمار في دراسة لا تثمر أي نتائج حاسمة.

قد يبدو أن العينات الصغيرة باستطاعتها توفير الوقت والموارد المالية، ولكن ذلك سيترتب عنه انخفاض احتمال اكتشاف تأثيرات ملحوظة للتدخل. عندما يخلص التقييم منخفض القوة الإحصائية إلى غياب أي أثر ملحوظ، يكون من المستحيل تحديد ما إذا كان غياب الأثر ناتجًا عن حقيقة أن التدخل لم يفلح فعلاً أم إذا كانت الدراسة منخفضة القوة الإحصائية. في مثل هذه الحالات، قد يوفر تقييم الأثر النذر اليسير من المعلومات المفيدة.

٢-٧ ماذا نعني بالقوة الإحصائية؟

القوة الإحصائية للدراسة هي مدى احتمال أن تحدد الدراسة بشكل صحيح أثرًا حقيقيًا وجوهريًا لتدخل ما. هناك أربعة سيناريوهات فيما يتعلق بالأثر الحقيقي واستنتاجات الدراسة الجوهرية (الجدول ٧-١). في اثنين من هذه السيناريوهات الأربعة، تتوصل الدراسة إلى الخلاصة الصحيحة: التدخل فعال وتم اكتشاف أثر ملحوظ، أو أن التدخل غير فعال ولم يتم اكتشاف أثر ملحوظ. في الحالتين الأخريين، تكون الخلاصات مغلوطة. عندما لا يكون التدخل فعالاً ولكن تخلص الدراسة إلى أنه فعال، فهذا يُسمى خطأ من النوع الأول (Type I). وعندما يكون التدخل فعالاً ولكن لا تكتشف الدراسة أي أثر ملحوظ، فهذا يُسمى خطأ من النوع الثاني (Type II). نحصل على قيمة القوة الإحصائية للدراسة عند طرح ١٠٠ من احتمال خطأ النوع الثاني. لتفسير هذا القول، يلزم اكتساب بعض المعارف الإحصائية.

الجدول ٧-١: الأخطاء المحتمل ارتكابها عند تقدير الأثر

	لم يتم اكتشاف أثر ملحوظ	تم اكتشاف أثر ملحوظ
التدخل غير فعال	لا يوجد خطأ (الخلاصة صحيحة)	خطأ من النوع الأول (إيجابية كاذبة)
التدخل فعال	خطأ من النوع الثاني (سلبية كاذبة)	لا يوجد خطأ (الخلاصة صحيحة)

المصدر: المؤلفان

حسب المصطلحات الإحصائية، الفرضية الصفرية (H_0) هي ألا يكون للتدخل أي أثر. نعني بالأثر هنا الفرق في متوسط الحصائل (أو التغير في الحصائل بالنسبة لتصميم التباين المزدوج) في مجموعة التدخل (المعالجة) (T) (μ) ومجموعة المقارنة (μC). الفرضية الصفرية مفادها تكافؤ مجموعتي المعالجة والمقارنة، بعبارة أخرى H_0: μT = μC؛ أو بشكل مكافئ، H_0: μT - μC = 0. لنفترض أن المتغير الناتج الرئيسي لتقييم أثر التوسع الزراعي هو دخل المزرعة، إذًا الفرضية الصفرية تنص على أن الدخل هو نفسه في منطقتي التدخل والمقارنة. ويُفترض هنا أن المَعْلم/المقياس محل الدراسة هو المتوسط، وينطبق ذلك على أغلب الحالات. ولكن في أحيان أخرى، قد يكون معلمًا آخر، مثل التباين: يشجع التمويل الأصغر تقليل الاستهلاك (تباين أقل في النفقات) أو أن الري يحد من مدى تغير المحصول. رغم أن الصيغ تختلف، فإن المبادئ الأساسية نفسها لحسابات القوة الإحصائية تنطبق كما ناقشنا في هذا الفصل.

يُحسب متوسط الحصيلة لكل مجموعة على أساس عينة من تلك المجموعة. ويُستخدم متوسط العينة، $\bar{X}$ كتقدير لمتوسط المجتمع الإحصائي الذي يرمز له بالحرف μ. عند استخدام عينة ما، يكون من المستبعد أن يكون متوسط العينة ($\bar{X}$) مساويًا تمامًا للمتوسط الحقيقي (متوسط المجتمع الإحصائي) (μ).

عند خط الأساس، أي قبل التدخل، يجب أن يكون متوسط الخصائص هو نفسه في منطقتي المعالجة والمقارنة إن أمكن (أي، $\mu_T = \mu_C$). ويُطلق على ذلك اختبار موازنة خط الأساس، ويجب إجراؤه على الدوام حيثما توافرت بيانات خط الأساس.

إذا افترضنا أن التدخل ليس له تأثير، فهذا يعني أن متوسطا الحصيلة الحقيقيين لكل من مجتمعي المعالجة والمقارنة الإحصائيين سيكونان متطابقين ($\mu_T = \mu_C$). عند خط النهاية. ومع ذلك، بما أن متوسطي العينة لكل من مجموعتي المعالجة والمقارنة من المستبعد أن يتطابقا تمامًا من حيث قيم المجتمع الإحصائي لكل منهما، فهما لن يتساويا مع بعضهما البعض. حتى إذا كان التدخل عديم الأثر، فإن متوسط الحصيلة في مجموعتي المعالجة والمقارنة من المرجح أن يختلف إلى حد ما بسبب خطأ أخذ العينة. تُختبر الدلالة الإحصائية للفرق $\bar{X}_T - \bar{X}_C$ لاحتساب ذلك. إذا لم يتم اكتشاف أي دلالة إحصائية، فإن الدراسة تخلص إذًا إلى أن الفرق الملحوظ قد يكون ناتجًا عن خطأ أخذ العينة، ومن ثم تكون الفرضية الصفرية مقبولة، تلك التي تنص على أن التدخل عديم الأثر. أما إذا كان الفرق ذا دلالة إحصائية، فهذا يعني أنه يوجد فرق كبير يتجاوز تفسيره بأنه خطأ أخذ العينة، وتكون الفرضية البديلة مقبولة، تلك التي تنص على أن التدخل له أثر، عوضًا عن الفرضية الصفرية.

تجدر الإشارة إلى أن الفرضية الصفرية سواء كانت مقبولة أو مرفوضة، فمن غير الممكن أن نتيقن بنسبة ١٠٠٪ من صحة حكمنا عليها. توجد مخاطرة تتمثل في حدوث خطأ: سواء رفض الفرضية الصفرية بشكل غير صحيح، أي تكون إيجابية كاذبة (خطأ من النوع الأول) أو قبولها بشكل غير صحيح، أي تكون سلبية كاذبة (خطأ من النوع الثاني) (الجدول ٧-١). من الضروري فهم هذه الأخطاء لاستيعاب مفاهيم القوة الإحصائية بشكل أكبر.

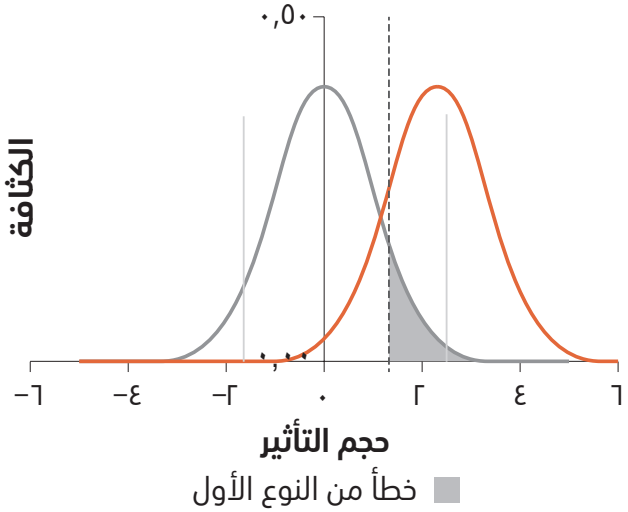
يُظهر الشكل ٧-١ التوزيع الاحتمالي لمتوسطي عينتين مأخوذتين من مجتمع إحصائي له متوسط مجتمع إحصائي حقيقي يساوي صفرًا وتوزيع طبيعي. المساحة أسفل المنحنى بين الخطين العموديين المرسومين عند $-x_1$ وx_1 تمثل نسبة ٩٥٪. هذا يعني أنه إذا أُخذت عينة من هنا، يكون هناك احتمال بنسبة ٩٥٪ أن يقع متوسط العينة في المدى $-x_1$ إلى x_1. على نحو مماثل، توجد فرصة نسبتها ٥٪ أن يكون متوسط العينة إما أقل من $-x_1$ أو أكبر من x_1.

لاختبار الفرضية الصفرية بأن متوسط المجتمع الإحصائي يساوي صفرًا، يجب توافر مدى مقبول من القيم التي يجب أن يقع متوسط العينة فيها من أجل الخلوص إلى أن الفرق بين متوسط العينة ومتوسط المجتمع الإحصائي المفترض المساوي لصفر يعكس خطأ في في أخذ العينة. يتحدد هذا المدى على أساس مستوى الدلالة الإحصائية المحدد، ويُطلق عليه α. من الشائع اختيار مستوى دلالة إحصائية ٥٪، وهذا المستوى هو خطأ من النوع الأول: احتمال رفض الفرضية الصفرية بشكل غير صحيح.

بمجرد تحديد مستوى الدلالة، يتم تحديد مدى متوسطات العينة الذي ضمنه تكون الفرضية الصفرية مقبولة، بالصيغة $\pm t_{\alpha/2} s_x$، حيث t_α هو النسبة الإحصائية t (t-statistic) وs_x هو الخطأ المعياري للمتغير x. يُستخدم الحد t لأن النسبة ٥٪ منقسمة بالتساوي إلى ٢,٥٪ في كل طرف. تعتمد t-statistic أيضًا على حجم العينة، ولكن مع الدلالة البالغة ٥٪، تكون قيمة t حوالي ٢. لذا إذا كان متوسط العينة يقع ضمن خطأين معياريين للصفر، فإن الفرضية الصفرية بأن متوسط المجتمع الإحصائي يساوي صفرًا، تكون مقبولة. إذا كان متوسط المجتمع الإحصائي الحقيقي يساوي صفرًا، فهذا يشير ضمنيًا إلى أن خلال ٥٪ من الوقت سيكون متوسط العينة خارج هذا المدى وسيتم رفض الفرضية الصفرية بشكل خاطئ. أي أن هناك احتمال بنسبة ٥٪ لحدوث خطأ من النوع الأول.

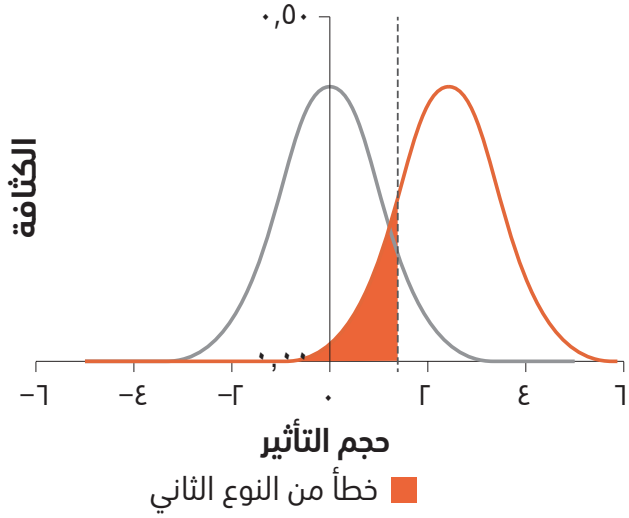
يوضح الشكل ٧-٢ احتمال حدوث خطأ من النوع الثاني. في هذه الحالة، يكون متوسط المجتمع الإحصائي الحقيقي، μ_2، أكبر من صفر. يوضح الخط الأحمر توزيع متوسطات العينة، علمًا أن هذا هو متوسط المجتمع الإحصائي الحقيقي. أما الخط الرمادي في الشكل، فهو يوضح التوزيع من الشكل ٧-١ الذي عندما تكون الفرضية الصفرية ومدى متوسطات العينة ضمنه، تكون الفرضية الصفرية مقبولة. يمكن أن نرى بسهولة أن نسبة كبيرة من متوسطات العينة المحتملة الواردة في التوزيع (حيث متوسط المجتمع الإحصائي الحقيقي يساوي ٢) تقع ضمن مدى مقبولية الفرضية الصفرية بأن المتوسط الحقيقي يساوي الصفر. لا شك في أن قرابة ٤٠٪ من التوزيع يقع ضمن هذا المدى. وهذا يعني أن في ٤٠٪ من الوقت، الاختبار الإحصائي سيقبل الفرضية الصفرية بأن المتوسط الحقيقي يساوي الصفر حتى إن كان في الحقيقة أكبر من الصفر. هذا هو الخطأ من النوع الثاني. إذا كانت هذه التوزيعات تمثل تأثيرات المعالجة، فإن في نصف الوقت تقريبًا سيخلص التحليل إلى أن التدخل عديم الأثر على عكس الواقع.

الشكل ٧-١: خطأ من النوع الأول



المصدر: المؤلفان

الشكل ٧-٢: رسم توضيحي لخطأ من النوع الثاني

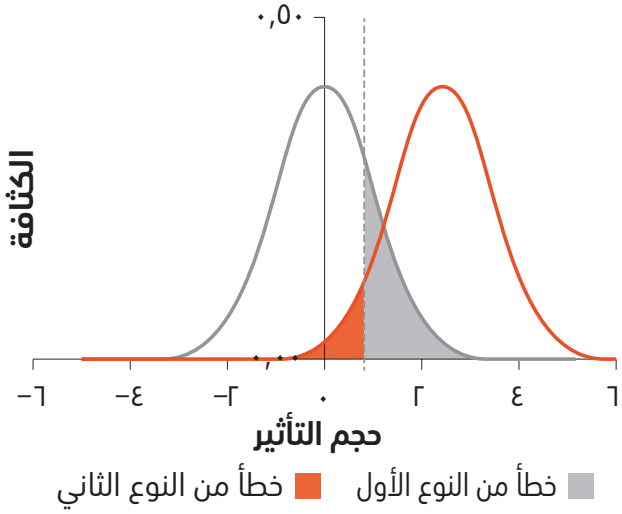


المصدر: المؤلفان

هناك طريقة للحد من حدوث خطأ من النوع الثاني تتمثل في تقليل فاصل الثقة (CI). وفي هذا الصدد، يعرض الشكل ٧-٣ مثالاً على هذه الحالة. عن طريق تحديد مستوى دلالة ١٠٪ بدلاً من ٥٪، فالخطأ من النوع الثاني يقل من ٤٠٪ إلى قرابة ٢٥٪. وهذا الانخفاض في الخطأ من النوع الثاني يكون على حساب زيادة خطأ النوع الأول إلى ١٠٪. يتضح من هنا أنه توجد علاقة معاوضة بين الخطأ من النوع الأول والخطأ من النوع الثاني. في العلوم الاجتماعية، يتم تعيين الخطأ من النوع الأول إلى ٥٪ بشكل قياسي.
أما العلوم الطبيعية، فقد تتطلب مستويات أعلى من الثقة.

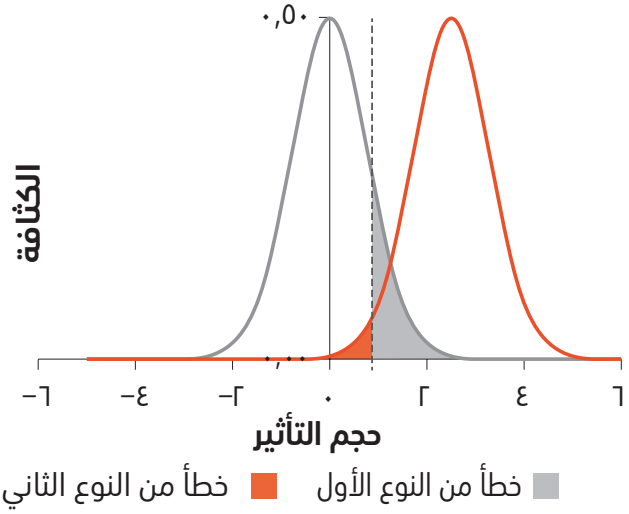
ينخفض الخطأ من النوع الثاني أيضًا مع زيادة حجم العينة. عندما يكون حجم العينة أكبر، يكون التوزيع أطول وأنحف. فمع كبر العينة، على الأرجح أن يكون متوسط العينة أقرب إلى متوسط المجتمع الإحصائي الحقيقي. كما هو موضح في الشكل ٧-٤، ينخفض الخطأ من النوع الثاني كلما ارتفع حجم العينة. هذا هو جوهر حسابات القوة الإحصائية: ما حجم العينة المطلوبة لتحصل الدراسة على مستوى مقبول من خطأ النوع الثاني؟ عادة ما يكون مستوى القبول عند ٢٠٪. بما أن القوة الإحصائية تساوي ١٠٠ ناقص الخطأ من النوع الثاني، فإن احتمال الخطأ من النوع الثاني البالغ ٢٠٪ يمكن صياغته بطريقة أخرى وهي بقول إن القوة الإحصائية تبلغ ٨٠٪. بالنسبة للقوة الإحصائية التي تكون أعلى من ذلك، يلزم استخدام عينة أكبر.

الشكل ٧-٣: الحد من خطأ من النوع الثاني عن طريق تقليل مستوى الدلالة



المصدر: المؤلفان

الشكل ٧-٤: الحد من خطأ من النوع الثاني عن طريق زيادة حجم العينة

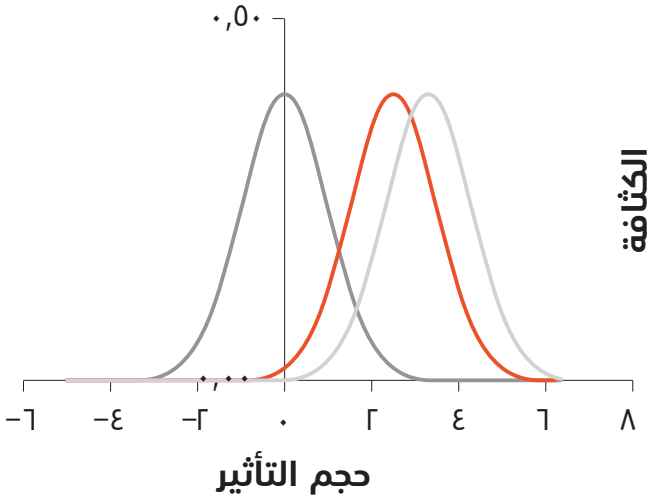


المصدر: المؤلفان

٣-٧ حسابات القوة الإحصائية لتصاميم الدراسة البسيطة

كلما كان تأثير متوسط المجتمع الإحصائي الحقيقي أكبر، كان من الأسهل اكتشاف مدى اختلافه عن الصفر كفرق دلالة إحصائية. يوضح الشكل ٧-٥. ذلك الأمر. إذا كان متوسط المجتمع الإحصائي أكبر $\mu 2 < \mu 3$، فسينخفض الخطأ من النوع الثاني من ٤٠٪ إلى ٢٠٪.

الشكل ٧-٥: كلما كبر التأثير، كان اكتشافه أسهل



المصدر: المؤلفان

بالطبع، لا يعرف المقيِّمون تأثير متوسط المجتمع الإحصائي (فهذا ما تسعى الدراسة إلى تقديره)، وبالتالي لا يستطيعون اختياره. ولكن المعلوم هنا هو الحد الأدنى لحجم التأثير (MES) أو الحد الأدنى للتأثير القابل للاكتشاف (MDE)، أي ما مدى كبر أو صغر التأثير الذي تستطيع الدراسة اكتشافه. يجب أن يستند الحد الأدنى لحجم التأثير (MES) إلى التجارب السابقة للتدخلات المماثلة وما ستؤول إليه المشاورات مع صانعي السياسات. وقد يرتبط ذلك بأهداف السياسة للتدخل. افترض أن برنامج تدريب مهني يهدف إلى خفض بطالة الشباب بنسبة ١٠٪ وأي إنجاز أقل من ذلك سيعد فشلاً. إذًا، يمكن تسمية النسبة ١٠٪ الحد الأدنى لحجم التأثير. لذا، إحدى طرق تحديد الحد الأدنى لحجم العينة هي معرفة حجم الأثر الذي يجب أن يحدثه البرنامج لكي يعتبر صانعو السياسات البرنامج فعالاً.

كما هو موضح أدناه، يعتمد الحد الأدنى لحجم التأثير على قيم t–statistic لمستوى الدلالة (α) ومستوى القوة الإحصائية المختار ($1-\beta$)، وأيضًا الخطأ المعياري للمتغير الناتج (σ_y)، ونسبة العينة في مجموعة المعالجة (P)، وحجم العينة (n):

$$MES = (t_{\alpha/2} + t_{1-\beta})\sigma_y \sqrt{\frac{1}{P(1-P)n}} \quad (٧-١)$$

يمكن إعادة ترتيب الصيغة للحصول على حجم العينة المطلوبة:

$$n = \frac{\left(t_{\alpha/2} + t_{1-\beta}\right)^2 \sigma_y^2}{MES^2 P(1-P)} \qquad (٢-٧)$$

توضح هذه الصيغة ما يلي:

- كلما كبرت العينة، صغر الحد الأدنى لحجم التأثير (MES) (المعادلة ٧-١). بعبارة أخرى، كلما كبر الحد الأدنى لحجم التأثير المحدد، كانت العينة المطلوبة أصغر (المعادلة ٧-٢). ومع ذلك، ستكون النتيجة هي ضعف احتمال اكتشاف التأثير إن كان صغيرًا في الحجم. ولهذا السبب، سينتج عن رفع الحد الأدنى لحجم التأثير (MES) دراسة منخفضة القوة الإحصائية إذا كان التدخل ليس له أثر كبير كما هو متوقع.

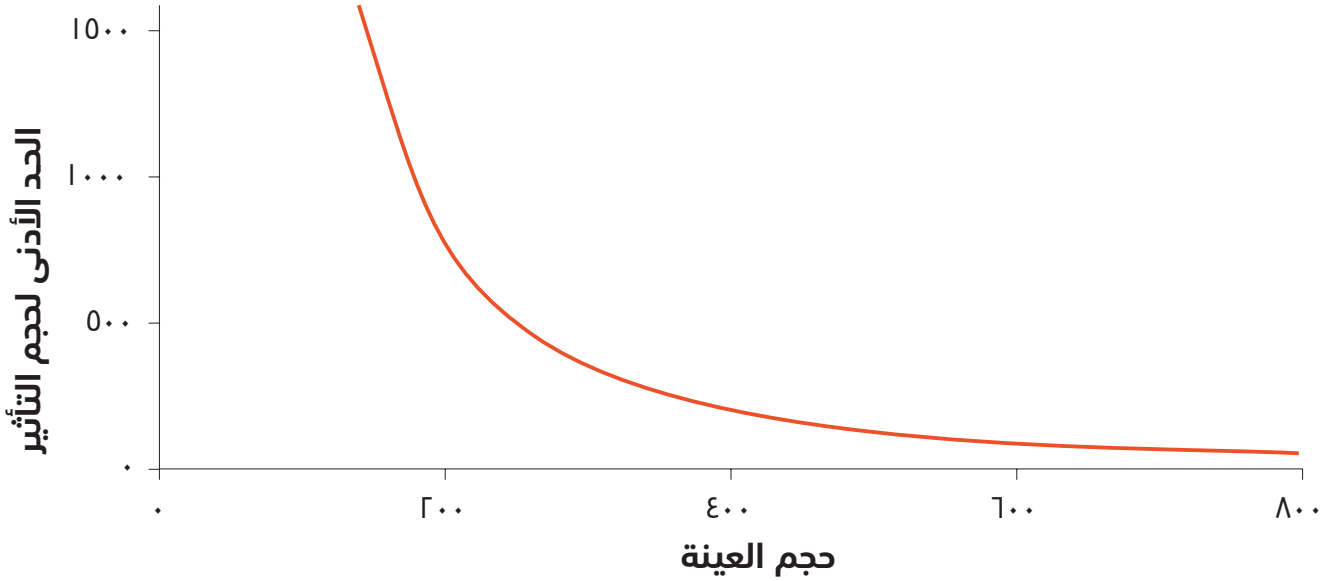
- يتقلص الحد الأدنى لحجم التأثير (MES) مع "العينة المتوازنة" (P=0.5)، وهذا يعني أنه عادة ما يكون من الأفضل الحصول على العدد نفسه من الملاحظات في مجموعتي المعالجة والمقارنة.

عند تطبيق هذه الصيغة، تكون قيمة σ_y مجهولة، بما أن حسابات القوة الإحصائية تتم قبل جمع أي بيانات. ويجب أن يؤخذ تقدير σ_y من مصدر بيانات آخر، ويُفضل من نفس الدولة أو السياق. وبطبيعة الحال، تختلف قيمة y حسب الحصيلة. ويتعين إجراء حسابات القوة الإحصائية بشكل منفصل لكل متغير ناتج. يجب اختيار العينة الأكبر حجمًا من بين أحجام العينة المطلوبة. تتوفر برامج لحساب القوة الإحصائية ولإنتاج المعلومات المطلوبة استنادًا إلى معاملات معلومة.

لنعد إلى مثال مشروع التوسيع الزراعي الذي طرحناه في بداية الفصل، ولنفترض أن متوسط دخل المزارع في الشهر كان يبلغ ١٥ ألف روبية هندية قبل بداية المشروع. وتجدر الإشارة إلى أن الهدف المسطر لهذا المشروع هو زيادة الدخل بمعدل ١٠٪. لقد أظهرت البيانات حول ولاية أندرا براديش الواردة عن مسح حديث حول الدخل والنفقات انحرافًا معياريًا للدخل يبلغ ١٢ ألف روبية. الزيادة البالغة ١٠٪ في الدخل هي عبارة عن التغير البالغ ١٥٠٠ روبية هندية، وهو يمثل الحد الأدنى لحجم التأثير (MES). مع α =٥٪ وقوة إحصائية ٨٠٪، فإن حجم العينة المطلوبة هو ٢٠٠٠ (الحد الأدنى لحجم التأثير هو ١٥٠٤).

يظهر بفحص المعادلة ٧-١ أنها ليست خطية في n. بعبارة أخرى، للكشف عن نصف الحجم، يجب أن يكون حجم العينة أكثر من الضعف. على وجه التحديد، لاكتشاف تأثير الزيادة بمعدل ٤٪ في الدخل، وهي زيادة مطلقة تساوي ٧٥٠ روبية هندية، يجب أن تتضمن العينة ٨٠٠٠ فرد (وهذا معناه أن الحد الأدنى لحجم التأثير هو ٧٥٢). وبهذا نجد أن الحد الأدنى لحجم التأثير يساوي ٤ مرات حجم العينة المطلوبة. ويوضح الشكل ٧-٦ هذه العلاقة.

الشكل ٧-٦: زيادة حجم العينة تؤثر عكسيًا في الحد الأدنى لحجم التأثير: العلاقة بين الحد الأدنى لحجم التأثير وحجم العينة



المصدر: المؤلفان

هذا الكتاب، تستخدم معظم تقييمات الأثر تصميمًا عنقوديًا. تتطلب هذه التصاميم تعديلاً على صيغة القوة الإحصائية.

٤-٧ حسابات القوة الإحصائية للتصاميم العنقودية

تتطلب معظم تقييمات الأثر للتدخلات الإنمائية تصاميم عنقودية تكون وحدة التعيين فيها مشتملة على وحدات متعددة تُجمع عنها البيانات. على سبيل المثال، من الممكن تعيين برنامج التوسيع الزراعي حسب المقاطعة أو القرية أو فئة المزارعين، ولكن الحصائل تُقاس على مستوى الأسر المعيشية. أما بالنسبة للتدخلات ذات الصلة بالتعليم، فيمكن تعيينها إلى المدارس، ولكن الحصائل تُقاس للطلاب فرادى. لهذا الأمر إنعكاسات كبيرة بالنسبة لأحجام العينات، التي غالبًا لا تحظى بالاهتمام الكافي في دراسات تقييم الأثر (Song and Herman 2010).

معامل الارتباط البيني العنقودي (ICC) أو ρ، هو مقياس لمدى تماثل الوحدات داخل كل مجموعة عنقودية. القوة الإحصائية تكون أعلى كلما زاد عدم تجانس (تغاير) الوحدات داخل المجموعة العنقودية الواحدة، كما ينعكس في ρ المنخفض. يُحسب الارتباط البيني العنقودي (ICC) باستخدام الصيغة التالية:

$$\rho = \frac{S_b^2}{S_b^2 + S_w^2} \qquad (٣-٧)$$

حيث S_b^2 هو تباين المتغير الناتج بين المجموعات العنقودية، و S_w^2 هوتباين المتغير الناتج داخل المجموعات العنقودية. لذلك، يمثل الارتباط البيني العنقودي (ICC) جزءًا من إجمالي التباين بين المجموعات العنقودية. حيثما لا يكون هناك اعتماد متبادل بين الأفراد داخل المجموعة العنقودية، يكون الارتباط البيني العنقودي (ICC) بالقيمة ٠. ولكن تكون قيمته ١ إذا كان هناك اعتماد متبادل مثالي بين الأفراد في المجموعة العنقودية الواحدة. كثيرًا ما تأتي هذه المعاملات بقيم تتراوح بين ٠,٢ و٠,٣ تقريبًا.

أفضل مصدر للارتباط البيني العنقودي (ICC) يمكن استخدامه في حسابات القوة الإحصائية هو مجموعة بيانات مماثلة لتلك التي سيتم استخدامها في التقييم. مجموعة البيانات المثالية ستكون تلك التي لها المتغير الناتج نفسه، ونوع المجموعة العنقودية نفسه، والمجتمع الإحصائي المستهدف نفسه. هناك مصدر آخر يتمثل في المقالات البحثية والتقارير المنشورة أو المقترحات البحثية المتاحة للجمهور، ووثائق التسجيل (مثل تلك المطلوبة لإجراء تجارب سريرية)، والخطط القبلية للتحليل.

بمجرد حساب حجم العينة (n) مع تجاهل الاعتمادية الإحصائية داخل المجموعات العنقودية، يجب ضرب حجم العينة في تأثير التصميم (DE) للحصول على مجمل العينات المطلوبة. يُحسب *تأثير التصميم* (DE) باستخدام الصيغة التالية:

$$DE = 1 + (m - 1)\rho \qquad (٤-٧)$$

حيث m هو عدد الأفراد في كل مجموعة عنقودية و ρ هو معامل الارتباط البيني العنقودي (ICC). وبُناءً على ذلك، نحصل على حجم العينة الحقيقي، مع احتساب الارتباط البيني للمجموعات العنقودية، باستخدام الصيغة أدناه.

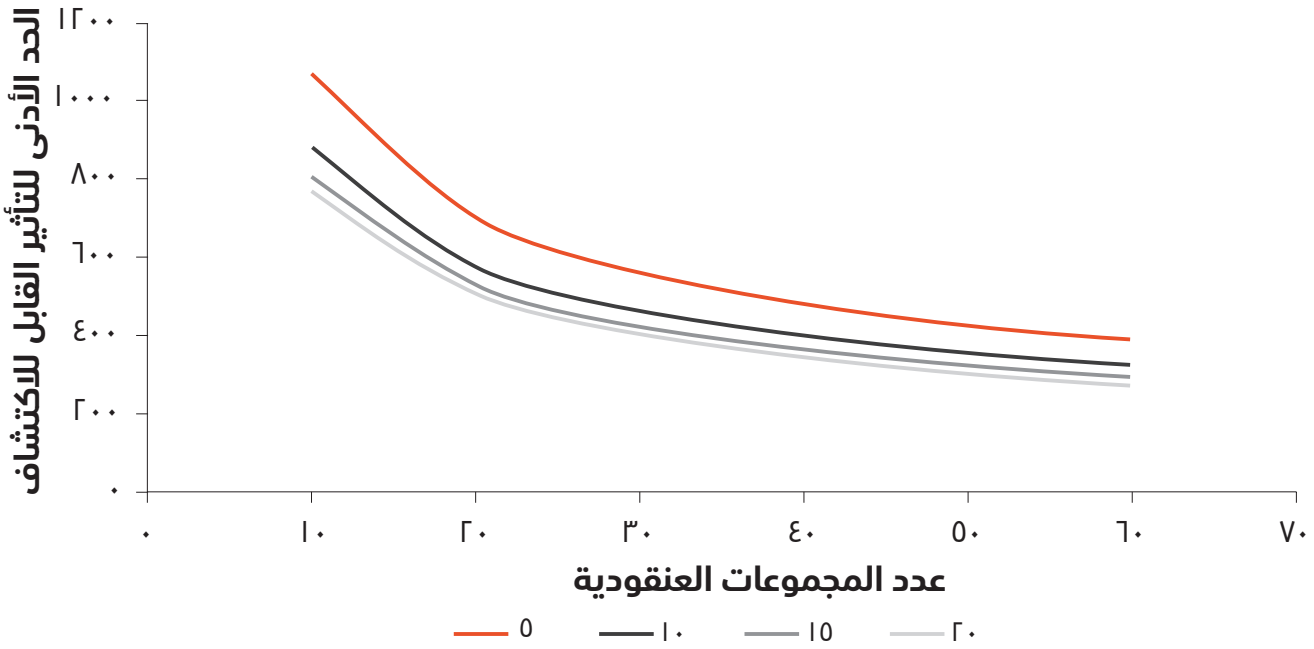
$$n = \frac{\left(t_{\alpha/2} + t_{1-\beta}\right)^2 \sigma_y^2}{MES^2 P(1-P)} \ (1 + (m-1)\rho) \qquad (٥-٧)$$

النقطة الأولى الجديرة بالذكر هي أن التصميم العنقودي يتطلب ملاحظات أكثر مقارنة بالتصميم البسيط. لنسترجع مرة أخرى المثال المتعلق بولاية أندرا براديش، الذي تم فيه الخلوص إلى أنه مع التصميم البسيط، يلزم أن يكون حجم العينة ٢٠٠٠ لاكتشاف زيادة ١٠٪ في الدخل. أما بالنسبة للتصميم العنقودي، فإذا كانت هناك ٤٠ ملاحظة عن ٥٠ مجموعة عنقودية، أي بعدد إجمالي للملاحظات يبلغ ٢٠٠٠ ملاحظة، فإن الحد الأدنى لحجم التأثير (MES) إذًا يبلغ ٧٠٠٠ روبية هندية تقريبًا، وهو بعيد بشكل مفرط عن القيمة المستهدفة البالغة ١٥٠٠ روبية هندية (ρ=0.2 مستخدم في هذا المثال). مع التصميم العنقودي، عندما تكون هناك ٤٠ ملاحظة لكل مجموعة عنقودية وحد أدنى لحجم التأثير (MES) يبلغ ١٥٠٠ روبية هندية، فيجب توافر ٤٤٥ مجموعة عنقودية أو حجم عينة إجمالي يبلغ ١٧٨٠٠. هذا أكبر قرابة ٩ أمثال المطلوب للتصميم البسيط! من الواضح، كما يُظهر هذا المثال، أن الفشل في احتساب المجموعات العنقودية عند إجراء حسابات القوة الإحصائية يمكن أن تنتج عنه دراسة ذات قوة إحصائية شديدة الانخفاض.

النقطة المهمة الثانية المستقاة من المعادلة (٧-٥) هي أن عدد المجموعات هو العامل الرئيسي المحدد للقوة الإحصائية للدراسة بالنسبة للتدخل العنقودي، بدلاً من عدد الملاحظات في كل مجموعة عنقودية. تتضح هذه الحقيقة وضوحًا كبيرًا مع مثال عددي.

في الشكل ٧-٧، يُظهر المحور الأفقي عدد المجموعات العنقودية (J)، ويتقابل كل خط مع عدد الملاحظات في كل مجموعة عنقودية (n). لنفترض أن لديك عينتين لهما ١٠٠ ملاحظة: (١) J = 10، n = 10 و(٢) J = 20، n = 5. أولى هذه التركيبات لها حد أدنى لحجم تأثير يبلغ ٨٢٠ تقريبًا، بينما ثاني تركيبة لها حد أدنى لحجم تأثير يبلغ ٧٠٠ تقريبًا. العينتان المتماثلتان في الحجم لا تشتركان في القوة الإحصائية نفسها لأن عدد المجموعات العنقودية يختلف. يمكن التصميم البسيط، الذي ينطوي على عدد أكبر من المجموعات العنقودية، من اكتشاف تأثير أصغر.

الشكل ٧-٧: زيادة عدد المجموعات العنقودية تأثيرها أكبر على القوة الإحصائية مقارنةً بزيادة عدد الملاحظات في كل مجموعة عنقودية



المصدر: المؤلفان

تُقدم تلك الصيغ مع افتراض الامتثال الكامل مع التجربة. إلا أن انخفاض معدل الامتثال يقلل من جدية الهدف المتمثل في معالجة التأثير. ولهذا، خيار الحد الأدنى لحجم التأثير (MES) يجب أن يأخذ في الاعتبار معدلات الامتثال. إن الإخفاق في القيام بذلك سيؤدي إلى دراسة ضعيفة القوة الإحصائية.

في واقع الأمر، قد تكون التصاميم الفعلية أكثر تعقيدًا. ويرجع ذلك إلى أنها قد تتضمن أخذ عينات طبقية وعدة أذرع علاجية (يناقش القسم ٣ في الملحق الثاني هذه القضايا). هذه المقاربة مختلفة أيضًا بالنسبة للمقاربات القائمة على الانحدار، مثل تصميم انقطاع الانحدار. ولا يتسع هذا الكتاب لتضمين حسابات القوة الإحصائية لمثل هذه الحالات بين طياته. الأمر الضروري الذي نريد لفت الأنظار إليه هو أن القوة الإحصائية مهمة، ويجب إجراء حسابات القوة ومراجعتها من قِبل شخص ما يتمتع بالخبرات الضرورية.

٥-٧ خطر الدراسات ضعيفة القوة الإحصائية

القوة الإحصائية هي احتمال الخلوص إلى أن التدخل له تأثير ملموس على نحو يطابق الواقع. عادة ما يُحسب حجم العينة بحيث يكون هذا الاحتمال ٨٠٪. إذا كانت الدراسة ضعيفة القوة الإحصائية، أي أن العينة صغيرة للغاية، فهي إذًا لها احتمال منخفض لاكتشاف أثر موجود بالفعل. ومن الشائع جدًا أن تتصف الدراسات بضعف قوتها الإحصائية، حيث تتراوح القوة بين ٥٠٪ إلى ٦٠٪ فحسب. ما سبب حدوث ذلك؟

لقد خلُص العديد من الباحثين من بينهم أنفين ونوردويج وفارلي وفازير إلى أنه عادة ما تكون الدراسات ضعيفة القوة الإحصائية للأسباب التالية (An-Wen et al. 2008 و Noordzij et al. 2010 و Fraley and Vazire 2014 و Rohrer 2010):

- اختيار مديري المشروع أو فرق الدراسة لما يعتقدون أنه حجم عينة ملائم، من دون دون إجراء حسابات القوة الإحصائية.

- عدم الأخذ بعين الاعتبار التجمع العنقودي للتدخلات، وحساب القوة الإحصائية للعينة مع تجاهل معامل الارتباط البيني العنقودي (ICC). نتيجةً لذلك، يتم أخذ عينة من بضع مجموعات عنقودية فحسب.

- افتراض أن يكون معامل الارتباط البيني العنقودي (ICC) أدنى مما هو عليه بالفعل. على الرغم من احتمالية الرجوع إلى مصادر البيانات القائمة، فهناك تباين ملحوظ في معاملات الارتباط البيني العنقودي (ICCs) للحصيلة نفسها في المنطقة نفسها. إن اعتبار درجة التجانس (معامل الارتباط البيني العنقودي المرتفع) أو عدم التجانس (معامل الارتباط البيني العنقودي المنخفض) للمجتمع الإحصائي ضمن كل مجموعة عنقودية يمكن أن يساعد على الوصول إلى افتراضات أكثر واقعية فيما يخص معامل معامل الارتباط البيني العنقودي (ICC).

- تعيين الحد الأدنى لحجم التأثير (MES) إلى قيمة مرتفعة للغاية. غالبًا ما تكون أهداف المشاريع غير واقعية. وعند الأخذ بهذه الأهداف كأساس للحد الأدنى لحجم التأثير (MES)، يقل احتمال أن تتمكن الدراسة من اكتشاف التأثيرات الأصغر، رغم أهميتها. وقد يؤدي ذلك أيضًا إلى تقييم أثر غير مكتمل الجوانب. سيبلغ التأثير المستوى المتوقع، ولكنه لن يتم تسجيله من قبل المقيّمين.

- وضع افتراضات شديدة التفاؤل فيما يخص الامتثال/المشاركة.

- من المحتمل افتراض أن المتغير الناتج على تباين أقل مما هو عليه فعليًا. وقد تكون هذه الحالة إذا كانت القيمة مأخوذة من مجتمع إحصائي غير قابل للمقارنة.

- إجراء حسابات القوة الإحصائية لمتغير ناتج واحد، ومن شأن ذلك عدم استخدام حجم عينة كافٍ لاستيعاب المتغيرات الناتجة الأخرى. يجب إجراء حسابات القوة الإحصائية لجميع الحصائل ذات الصلة واستخدام أكبر عينة مطلوبة.

- قد تكون الدراسة ذات قوة إحصائية كافية لتقدير متوسط تأثير المعالجة، ولكن ليس لإجراء تحليل مجموعة فرعية. لذا، قد تفيد الدراسة بعدم تجانس الأثر بين، مثلاً، الرجال والنساء أو المناطق الحضرية والريفية. ولكن في الحقيقة، لم تكن هناك قوة إحصائية كافية لإجراء مثل هذا التحليل. يجب أن تسمح حسابات القوة الإحصائية لإجراء أي تحليل مخطط له لمجموعة فرعية.

- هناك تناقص في تصميم الدراسة، مثل جمع البيانات فعليًا من عينة أصغر مما كان مخططًا له.

٦-٧ وسائل مساعدة لاعتبار حسابات القوة الإحصائية وإجرائها

- بالإضافة إلى الأخطاء المحتملة المذكورة أعلاه، من المفيد أن تكون على دراية ببعض القواعد العامة التي يمكن أن تكون بمثابة دليل أولي لتقدير متطلبات العينة اللازمة لتقييم الأثر:

- هناك قاعدة جوهرية شائعة تنص على أن حجم العينة المطلوبة لكل مجموعة هو $n = 16/mes^2$ حيث الحد الأدنى لحجم التأثير (MES) معبر عنه في الانحرافات المعيارية. إذا كان mes = 0.5، فإن n = 64 لكل مجموعة أو إجمالي حجم العينة يبلغ ١٢٨ في حالة وجود ذراع (مجموعة) معالجة واحد وذراع (مجموعة) مقارنة واحد مع غياب أي مجموعات عنقودية. بالنسبة للتدخلات الاجتماعية والاقتصادية، غالبًا ما يكون الحد الأدنى لحجم التأثير (MES) أصغر بكثير، لنقل قرابة ٠,١، وهذا يعطي إجمالي حجم عينة مطلوبة يبلغ ١٦٠٠ لكل ذراع. هذه الصيغة تناسب كل ذراع على حدة. إذًا بالنسبة لتصميم عاملي له أربعة أذرع (A، وB، وA+B، ومقارنة)، حجم العينة المطلوب هو: $4 \times 16 / mes^2$، مثل، ٢٥٦ مع mes (حد أدنى لحجم التأثير) يبلغ ٠,٥، و٤٨٠٠ مع mes (حد أدنى لحجم التأثير) يبلغ ٠,١.

- تتصف صيغ القواعد الأساسية للتصاميم العنقودية بأنها أكثر تعقيدًا، حيث تعتمد على عدة سمات للتدخل وتصميم الدراسة. هناك قاعدة أساسية مجردة تنص على أن ٦٠ مجموعة عنقودية (٣٠ معالجة و٣٠ مقارنة) يمكن أن تكون المتطلب الأدنى البحث بصفة عامة. في بعض الحالات المختارة حيث يكون معامل الارتباط البيني العنقودي (ICC) منخفضًا، والحد الأدنى لحجم التأثير (MES) مرتفعًا، قد تكون ٣٠ مجموعة عنقودية كافية (McNeish and Stapleton 2016). في مجموعة فرعية لهذه الحالات، من الممكن أن تنخفض القيمة دون ٣٠ مجموعة عنقودية مع مطابقة المجموعات العنقودية (مثل التطابق العشوائي للأزواج).

يرى رودنبوش وآخرون أنه ثمة بضعة خيارات من البرامج التي بإمكانها إجراء حسابات حجم العينة. أفضل برنامج معروف خصيصًا لهذا الغرض هو (Optimal Design) Raudenbush et al. 2011. يمكن أيضًا استخدام الحزم الإحصائية مثل STATA وSAS وSPSS وR لإجراء حسابات القوة الإحصائية. وتوجد أدوات متخصصة من المستبعد أن يتسنى لمديري المشروع الوقت لاستخدامها. هناك خيار يتميز بسهولة الاستخدام، وهو 3ie Excel Power Calculator والدليل الإرشادي ذو الصلة (Djimeu and Houndolo 2016).

نستعرض فيما يلي اعتبارات مهمة عند التخطيط لأحجام العينات:

- يجب إجراء حسابات القوة الإحصائية لجميع الحصائل مع استخدام أكبر حجم عينة مطلوب.
- يجب إضافة ١٠٪ أخرى على الأقل للسماح باستبدال المجيبين والمسوحات المكتملة جزئيًا.
- إذا كان من المقرر تنفيذ تحليل مجموعة فرعية، فإن حجم العينة ينطبق على المجموعة الفرعية.
- إذا كانت هناك مقارنة بين العلاجات (تصاميم أ/ب)، فإن الحد الأدنى لحجم التأثير (MES) ذا الصلة هو الفرق في التأثيرات وليس التأثير المطلق. ولذلك، على الأرجح أن تحتاج تصاميم أ/ب إلى عينات أكبر حجمًا.
- يؤثر الامتثال على حجم العينة المطلوبة. إذا كان الامتثال للبرنامج منخفضًا، سيكون حجم العينة المطلوبة أكبر.

كما ذكرنا آنفًا، حساب القوة الإحصائية يحتاج أيضًا إلى افتراضات تستدعي معامل الارتباط البيني العنقودي (ICC) وتباين المتغير الناتج. ومن المصادر الشائعة لهذه الافتراضات ما يلي: (١) تقييمات الأثر الحالية التي تدور حول الموضوع نفسه (قد يلزم إرسال طلب إلى مؤلفي الدراسة للحصول على البيانات)، (٢) دراسات كمية أخرى تنتمي إلى الدولة أو المنطقة نفسها، (٣) تحليل مجموعات البيانات العامة (مثل مسوحات الدخل والنفقات)، (٤) المراجعات المنهجية، و (٥) الكثير من الأوراق البحثية التي تتضمن مثل هذه البيانات لأغراض حساب القوة الإحصائية.

المراجع

An-Wen, C., A. Hróbjartsson, K. J. Jørgensen, P. C. Gøtzsche, and D. G. Altman. 2008. Discrepancies in Sample Size Calculations and Data Analyses Reported in Randomised Trials: Comparison of Publications with Protocols. *BMJ*. 337. a2299.

Djimeu, E. W. and D. Houndolo. 2016. Power Calculation for Causal Inference in Social Science: Sample Size and Minimum Detectable Effect Determination. 3ie Impact Evaluation Manual. Working Paper 26. Delhi: 3IE. www.3ieimpact.org/media/filer_public/2016/07/08/ wp26-power-calculation.pdf and http://www.3ieimpact.org/media/ filer_public/2016/03/22/3ie-sample-size-minimum-detectable-effect- calculator.xlsx.

Fraley, R. C. and S. Vazire. 2014. The N-Pact Factor: Evaluating the Quality of Empirical Journals with Respect to Sample Size and Statistical Power. *PLoS ONE*. 9 (10). doi:10.1371/journal.pone.0109019

McNeish, D. M. and L. M. Stapleton. 2016. The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. L. M. Educational *Psychology Review*. 28 (2). pp. 295–314. https://doi.org/10.1007/s10648- 014-9287-x.

Noordzij, M., G. Tripepi, F. W. Dekker, C. Zoccali, M. W. Tanck, and K. J. Jager. 2010. Sample Size Calculations: Basic Principles and Common Pitfalls. *Nephrology Dialysis Transplantation*. 25 (5). pp. 1388–1393.

Raudenbush, S. W., H. Bloom, J. Spybrook, and A. Martinez. 2011. Optimal Design Software for Multi-Level and Longitudinal Research (Version 3.01) [Software]. http://hlmsoft.net/od/.

Rohrer, J. E. 2010. Nonrandomized Impact Evaluation Studies: Errors and Tips. *Journal of Primary Care & Community Health*. 1 (2). pp. 70–72.

Song, M. and R. Herman. 2010. Critical Issues and Common Pitfalls in Designing and Conducting Impact Studies in Education: Lessons Learned From the What Works Clearinghouse (Phase I). *Educational Evaluation* and Policy *Analysis*. 32 (3). pp. 351–371. http://www.jstor.org/stable/40963082.

مراجع إضافية للإطلاع

Batistatou, E., C. Roberts, and S. Roberts. 2014. Sample Size and Power Calculations for Trials and Quasi-Experimental Studies with Clustering. *The Stata Journal*. 14 (1). pp. 159–175. www.stata-journal. com/article.html?article=st0329.

Bell, B. A., G. B. Morgan, J. A. Schoeneberger, J. D. Kromrey, and J. M. Ferron. 2014. How Low Can You Go? An Investigation of the Influence of Sample Size and Model Complexity on Point and Interval Estimates in Two-Level Linear Models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. 10. pp. 1–11. http://econtent. hogrefe.com/doi/abs/10.1027/1614-2241/a000062.

Reich, N. G., J. A. Myers, D. Obeng, A. M. Milstone, and T. M. Perl. 2012. Empirical Power and Sample Size Calculations for Cluster- Randomized and Cluster-Randomized Crossover Studies. PLoS ONE. 7 (4). e35564. http://doi.org/10.1371/journal.pone.0035564.

Schulz, K. F. and D. A. Grimes. 2005. Sample Size Calculations in Randomised Trials: Mandatory and Mystical. The Lancet. 365 (9467). pp. 1348– 1353. https://www.ncbi.nlm.nih.gov/ pubmed/15823387.

الفصل ٨
إدارة عملية تقييم الأثر

رسائل أساسية

- هناك أشياء يجب وضعها في الاعتبار عند التخطيط لتقييم الأثر، وهي المساهمات التي يمكن تقديمها بما يتخطى مجرد سرد الأدلة الحالية، والمطالبات بالأدلة، وقابلية تقييم التدخل.
- مشاركة الأطراف المعنية ضرورة لا تقتصر على ضمان الملاءمة، إنما لضمان نزاهة التصميم وسلامته أيضًا.
- تستغرق بعض تقييمات الأثر بضع سنوات لإجرائها، ومن ثم يجب إنشاء آليات مناسبة لضمان الإشراف على عملية التقييم خلال هذه الفترة، وهذا أمر قد لا يتناسب بشكل مثالي مع دورة المشروع.
- المراجعة المتأنية والمشاركة الفعلية من قِبل العاملين على المشروع ومقيِّمي الأثر الخبراء أمران ضروريان لضمان دقة تصميم المشروع، والتنفيذ، وجمع البيانات، وصلاحية التحليل.

٨-١ مقدمة إلى إدارة تقييمات الأثر

يعتمد تقييم الأثر المجدي، بشكل أساسي، على تسيير العمليات وفقًا لمنهجية صارمة دون أي خلل. يعالج هذا الفصل مسائل قد يتعرض لها الأشخاص المسؤولون عن إدارة تقييمات الأثر.

وتتضمن هذه المسائل ما يلي: (١) التخطيط لتقييمات الأثر، (٢) اختيار تصميم لتقييم الأثر، (٣) إعداد الميزانية وإدارة تقييم الأثر، (٤) تعبئة الموارد لتقييم الأثر، و (٥) تفسير استنتاجات تقييم الأثر.

٨-٢ التخطيط لتقييمات الأثر

ما التدخلات التي يناسب إجراء تقييم أثر لها؟

لا تكون لتقييمات الأثر أي قيمة إلا باستخدام الأدلة الناتجة. ويجب توجيه التقييمات لإنتاج أدلة قادرة على دعم تحسين إعداد البرنامج الإنمائي، من حيث "إثباتات حول صحة المفاهيم" التي بإمكانها دعم استمرارية التدخل أو تكراره أو ترقيته، أو لاستنباط رؤى محددة حول كيفية تعديل التدخل لتعزيز فعاليته. للقيام بذلك، يجب إعداد تصور مفاهيمي لكيف يسهم تقييم الأثر لتدخل معين في مجموعة الأدلة الناتجة بالفعل. وهذا يعني أن أولويات تقييم الأثر ينبغي لها الاعتماد على الفجوات بين الاحتياجات المعلوماتية للجماهير الأساسية والأدلة القائمة.

قد تتمثل نقطة البداية في مناقشات مع صانعي السياسات الإنمائية (من جانب الحكومات، والهيئات المانحة

والأكاديمية، وغير ذلك من المؤسسات) في قطاع و/أو منطقة جغرافية معينة. ويكون الغرض من هذه المناقشات فهم الاحتياجات المعلوماتية المتعلقة بالتأثيرات الناجمة عن التدخل والفرص المتاحة لإحداث تأثير. وفي بعض الأحيان، قد تتضمن تلك الاحتياجات الكثير من المتطلبات التي تقع خارج اختصاص تقييم الأثر، ويظل من الضروري فهم كيف يمكن لتقييم الأثر أن يلبي مثل هذه الاحتياجات من أجل تعزيز الاستخدام (IEG 2012). لتحديد ما إذا كان يمكن تلبية الطلبات المعلوماتية عن طريق تقييم الأثر، قد يتعين دراستها لمطابقتها مع الأسئلة التالية: (١) هل الاحتياج المعلوماتي مرتبط بنظرية التغيير للتدخل القائم أو المخطط له أو القابل للتجريب؟، (٢) هل هو متعلق بالحصائل/التأثيرات المشروطة بالاستجابات السلوكية (البشرية) للتدخل؟، و(٣)هل من الممكن إجراء تطبيق ميداني للتدخل المعني في الإطار الزمني لهذا الاحتياج المعلوماتي؟

الخطوة الثانية يمكن أن تنطوي على مقارنة الطلبات مع المعلومات المتاحة حاليًا. تتنامى مجموعة من المراجعات المنهجية بوتيرة سريعة، وهي مستندة إلى مخزون من المؤلفات الشاملة التي تتناول موضوعات وتحاليل وصفية محددة، من أجل تحديد الارتباطات بين سمات التدخل وتقديرات التأثير. وتمثل هذه المراجعات نقطة بداية جيدة لفهم الأدلة المتوفرة. للاستزادة، يجوز الاطلاع على مستودع تقييم الأثر وقواعد البيانات الأكاديمية للمبادرة الدولية لتقييم الأثر (3ie) للاجابة على الاستفسارات المبدئية، ويمكن إجراء عمليات بحث عبر الويب عن الدراسات الجارية. قد تُمنح الأولوية للدراسات التي تقيِّم التدخلات ذات الأدلة الضئيلة، أو الحصائل التي لم يتم التحقيق فيها، أو النظريات الأوسع التي لم يتم اختبارها.

تقييم الأثر يمكن أن يتم أيضًا تحت الطلب، بُناءً على طلب ورغبة من منفذي المشروع. وفي هذه الحالة، من الجيد إخضاع الاحتمالات ((قد لا نكون بصدد إجراء تقييم أثر للمشروع بكامله) لتمحيص كامل قبل تحديد ما إذا كان تقييم الأثر ملائمًا ومتى وحسب أي مقياس يكون كذلك. سينطوي ذلك على مناقشة تفصيلية لفهم الطلبات المعلوماتية اللازم تلبيتها، ونظرية التغيير، وما إذا كان تقييم الأثر مقبولاً، أو بعبارة أخرى، ما إذا كان التدخل ذا قيمة (Peersman et al. 2015). تعتمد قابلية التقييم على (١) وضوح نظرية التغيير والفرضيات ذات الصلة المطلوب اختبارها، و(٢) القدرة على إنتاج ملاحظات كافية تعزز القوة الإحصائية حول الحصائل المعنية، و(٣) توافر تصميم تقييم أثر يجعل "افتراض قيمة معالجة الوحدة المستقرة" ملائمًا للتدخل.

متى يجب إعداد تصميم تقييم الأثر؟

على الرغم من أن الأثر يتم قياسه بصفة عامة بعد بدء التدخل بسنوات، فمن الأفضل العمل على تصميم الدراسة الخاص به قبل التطبيق الميداني للتدخل، عبر تصميم تقييم أثر استباقي. تكون تصاميم تقييم الأثر الاستباقي، بشكل شبه دائم، أقوى من تلك التصاميم المعدة بمجرد تطبيق التدخل ميدانيًا. ونسرد فيما يلي المزايا الرئيسية للتصاميم الاستباقية: (١) يمكن حشد بيانات خط الأساس قبل أن تسري آثار التدخل ميدانيًا، (٢) يمكن الأخذ بعين الاعتبار احتمال التعيين العشوائي، (٣) يمكن الرجوع إلى الأطراف المعنية لمناقشة أهمية أسئلة التقييم وإقناعها في مرحلة مبكرة من العملية. كما تسمح بيانات خط الأساس بفحص الموازنة (إذا كانت مجموعتا المعالجة والمقارنة تشتركان في متوسط الخصائص)، وبإنتاج تقديرات أثر أكثر دقة.

التحدي الرئيسي الذي يواجه المقيمين عند الشروع في تقييم الأثر قبل التطبيق الميداني هو أن طرح التدخل قد لا يكون معلومًا بعد، وقد يكون عرضة للتغييرات بعد تجاوز نقطة أو مرحلة خط الأساس، مثل تبيُّن عدم استقرار المجتمعين الإحصائيين: الضابط والمعالج. وهذا يعني وهذا يعني أنه من الأفضل البحث في تقييم الأثر بمجرد فهم خطط تنفيذ المشروع المتوقع بشكل كافٍ. يجب التشجيع على المشاركة في مرحلة مبكرة للغاية لما لها من فائدة على احتمالات استخدام تقييم الأثر للتدخلات على أساس تجريبي، وذلك لتضمين فرص التوزيع العشوائي.

حيثما يكون هناك العديد من الاستثمارات في القطاعات نفسها أو ذات الصلة في المناطق نفسها، يكون من الممكن استخدام خط أساس واحد عبر عدة مشاريع مختلفة. قد ترفع مقاربة كهذه من تكاليف التنسيق الأولية، لكنها توفر تكلفة كبيرة في وقت لاحق.

ضمان اقتناع الأطراف المعنية

يوجد طيف واسع وعريض من الأطراف المعنية التي يمكن إشراكها في المراحل الأولية من التخطيط لتقييم الأثر. في الوكالة المنفذة، من الضروري فهم المقاربة ودعمها ليس فقط من قِبل قيادة الوكالة بل أيضًا من قِبل الفريق الميداني المنفذ للمشروع أو البرنامج. ويجب إشراك فريق الوزارة المباشر المسؤول، والحصول على دعم المناصب العليا في الوزارة للحد من مخاطر التداخل السياسي الذي يقوض نزاهة التصميم (في الأغلب عن طريق العمل بالتدخل الخاضع للتقييم في مناطق المقارنة). وتستدعي الضرورة أيضًا إشراك المجتمع ذي الصلة في مرحلة مبكرة، بما في ذلك تحديد الوكالات البحثية المحتملة التي تستطيع إجراء الدراسة.

تتزايد القدرة المحلية على تنفيذ تقييمات أثر عالية الجودة في معظم الدول، وتوجد وكالات بحثية إقليمية أيضًا يمكن إشراكها في عملية التقييم.

الاتفاق والتعاون بين الحكومة وفريق المشروع شيئان مهمان لفهم طرح المشروع وتحديد المشاركين وغير المشاركين وتجنب التداخل. يكون الإقناع القوي أهم ركن أساسي للتصاميم التجريبية عادة لضمان صلاحيتها، بحيث يعالج المشروع مجموعات المعالجة ولا يعالج المجموعات الضابطة.

تعالج تقييمات الأثر مسألة الإسناد. إلا أن البيانات التي يتم جمعها للدراسة قد تعالج نطاقًا عريضًا من الأسئلة المتعلقة بالتقييم، مثل الاستهداف وأسباب عدم المشاركة وقضايا التنفيذ. لضمان أقصى درجة من الملاءمة، يجب استشارة الأطراف المعنية حول أسئلة التقييم الإضافية التي قد تكون ذات أهمية بالنسبة لها، وقد يتم ذلك من خلال ورشة عمل تدور حول نظرية التغيير. تساعد ورش العمل هذه فريق الدراسة على فهم المشروع وتحديد أسئلة التقييم ذات الصلة بالتعاون مع الشركاء. وهي تمثل أيضًا فرصة لفريق الدراسة لتقديم تصميم تقييم الأثر المقترح.

٣-٨ اختيار تصميم تقييم الأثر

تقع إستراتيجية التحديد في قلب تقييم الأثر، وينبغي لجميع تقييمات الأثر تحديد الإستراتيجية المقصودة قبل البدء. في العديد من الحالات، يكون من المفيد الأخذ في الاعتبار أي إستراتيجيات تحديد قد تكون ملائمة قبل تضمين فرق الدراسة الخارجية. يدرج دليل القرار في الجدول ٨-١ الأسئلة الرئيسية لتوجيه اختيار التصميم، وكيف يمكنها أن تساعد على تحديد مقاربة مناسبة.

التأكد من الالتزام بمقاربة دقيقة

للحد من مخاطرة الفشل في اكتشاف التأثيرات الملموسة، يكون في الغالب من المفيد أن تتضمن المقترحات عدة إستراتيجيات تحديد ممكنة. في التصاميم غير التجريبية، قد تكون هناك ارتباطات غير متوقعة تؤثر في صلاحية مقاربات معينة، مثل المتغيرات المساعدة أو عدم توقع طرح البرنامج أو وجود مشكلات فنية في تنفيذ سلسلة انحدارات معينة. أما في التصاميم التجريبية، قد يكون هناك فشل في الالتزام ببروتوكول التوزيع العشوائي، أو إدارة غير سليمة للمعالجة، أو تداخل، أو انخفاض في معدلات المشاركة. في هذه الحالات، قد يكون من الضروري الرجوع إلى تصميم الدراسة، على سبيل المثال، لاستبدال تجربة عشوائية مضبوطة (RCT) بمتغيرات مساعدة. وتحسبًا لمثل هذه الحالات، يجب إرفاق تصميم التقييم بإستراتيجية تحديد احتياطية.

الجدول ٨-١: اختيار تصميم تقييم الأثر: المقاربة المتبعة لاتخاذ القرار

س١	هل هو تصميم استباقي؟	نعم>>س١-١ لا >> س٢
س١-١	هل التعيين العشوائي ممكن؟	نعم>>س١-٢ لا >> س٢
س١-٢	هل وحدة التعيين هي نفسها وحدة المعالجة والتحليل؟	نعم>>س١-٣ لا: **تجربة RCT عنقودية**
س١-٣	هل من المرجح أن ترد تأثيرات تداعيات مهمة؟	نعم: **تجربة RCT عنقودية** لا: **تجربة RCT بسيطة**
س٢	هل التجربة الطبيعية ممكنة؟	نعم: **تجربة طبيعية** لا >> س٣
س٣	هل هو تدخل متاح بشكل عام لم يتم اعتماده بشكل عام من قبل؟	نعم >> س٣-١ لا >> س٤
س٣-١	هل يمكن تحديد تشجيع صالح؟	نعم: **تصميم تشجيعي** لا >> س٤
س٤	هل يوجد الكثير من وحدات المعالجة للتعيين؟	نعم: س٥ لا: س٤-١
س٤-١	هل يوجد الكثير من الفترات المكرسة للملاحظات السابقة للتدخل؟	نعم: س٤-٢ لا: **فكِّر في بديل لتقييم الأثر**
س٤-٢	هل تتضمن الملاحظات وحدات غير معالجة يمكن أن تمثل وحدات مقارنة؟	نعم: **وحدات ضابطة اصطناعية**
س٥	هل توجد قاعدة لعتبة الأهلية (بما في ذلك الحد الزمني لوقت تقديم البرنامج)؟	نعم >> س٥-١ لا >> س٦
س٥-١	هل تم تطبيق القاعدة بدقة؟	نعم: **تصميم انقطاع الانحدار (RDD) (بما في ذلك ITS)** لا: **RDD مشوش**
س٦	هل من المرجح أن تؤثر الوحدات غير الملحوظة على الاختيار؟	نعم >> س٧ لا >> س٦-١
س٦-١	هل تتوفر بيانات خط أساس؟	نعم: **الاختلاف فى الاختلافات (DiD) أو تأثيرات ثابتة** لا: **مقاربات قائمة على درجة الميل**
س٧	هل الوحدات غير الملحوظة ستكون ثابتة على مر الزمن على الأرجح؟	نعم >> س٧-١ لا >> س٨
س٧-١	هل تتوفر بيانات خط أساس؟	نعم: **الاختلاف فى الاختلافات (DiD) أو تأثيرات ثابتة** لا >> س٨
س٨	هل يمكن تحديد قيد مُحدِّد أو وحدة مساعدة صالحة؟	نعم: **متغيرات مساعدة أو معالجة داخلية النشأة أو انحدار تبديلي** لا: فكِّر في بديل لتقييم الأثر

ITS = السلاسل الزمنية المتقطعة، RCT = التجربة العشوائية المضبوطة.

المصدر: المؤلفان

تتعزز جودة تصميم الدراسة بالمشاركة الوثيقة للأطراف المعنية، وبالأخص في المراحل المبكرة. ويمكن أن تتم هذه المشاركة من خلال آليات مختلفة، مثل ورش عمل تدور حول نظرية التغيير أو مجموعات استشارية أو المشاركة في مراجعة أقران رسمية.

تشكل العديد من الوكالات مجموعة استشارية للإشراف على تقييمات الأثر (وأشكال أخرى من التقييم أيضًا) (VeLure Roholt and Baizerman 2012). يتعين أن تتألف المجموعة الاستشارية من ٣ إلى ٤ أفراد كحد أدنى، بمن فيهم فرد واحد على الأقل لديه خبرة في تقييم الأثر وآخر لديه معرفة بالقطاع. من الأحرى عادة أن تتضمن المجموعة الاستشارية تمثيلاً لمجموعات تفكير محلية أو غيرها من المؤسسات الأكاديمية. قد تأخذ هذه المجموعة شكل مجموعة افتراضية مع إرسال التعليقات عبر البريد الإلكتروني أو بمقدور الأعضاء المشاركة في ورش عمل يتم فيها تقديم التصميم المقترح واستنتاجات الدراسة الأولية.

القضايا الأخلاقية

ترتبط القضايا الأخلاقية الرئيسية، التي تنشأ عند إجراء تقييم أثر، بما يلي:

- تأسيس مجموعة ضابطة (حيث تنشأ القضية غالبًا بالنسبة للتجارب العشوائية المضبوطة، كما ناقشنا في الفصل ٤)
- القضايا المتعلقة بالعوامل البشرية المتدخلة في جمع البيانات

يهدف عدد قليل للغاية من التدخلات إلى تغطية عامة فورية، وعلى الرغم من ذلك، تحقق بضع منها فقط مرادها. لذلك، غالبًا ما يكون هناك مجتمع إحصائي غير معالج بأي حال يتم استخراج منه مجموعة ضابطة أو مقارنة. تقييم الأثر لا يخلق المجتمع الإحصائي غير المعالج، فهو إما يستخدم تباينًا قائمًا في المعالجة، أو يجعل تعيين المعالجة أكثر منهجيًا. في العديد من الحالات، تغير التجارب العشوائية المضبوطة توقيت التعيين أو توفر محفزًا إضافيًا للمشاركة.

وفي قليل من الحالات يكون هناك أساس واضح للاعتراضات الأخلاقية على وجود مجموعة مقارنة أو ضابطة. هذا لا يعني أن نقول إن أعضاء مجموعة المقارنة قد لا يشعرون بمعالجة غير عادلة، لا سيما أنهم معرضون لجمع البيانات لتدخل لا يثمر أي فوائد مباشرة لهم.

قد تظهر إلى النور مجموعة أكثر تعقيدًا من القضايا الأخلاقية في مسعى تقييمات الأثر التي تتضمن معالجة وهمية للمجموعة الضابطة، بحيث يتم التخلص من التحيز الذي يتأتى من معرفة الأفراد بخضوعهم للمعالجة. تطرح القضية الأخلاقية مخاوف من أن يتعرض المجيبون لتدخل غير مكتمل الخصائص، وحينها قد تكون الموافقة المستنيرة أكثر صعوبة. بالنسبة لهذه الحالات، المعايير المطبقة يجب أن تكون مماثلة لتلك الخاصة بالأبحاث الطبية، ويتعين بذل الغالي والنفيس لتقديم التدخلات بكل صراحة بأسلوب غير مضلل.

تتعلق القضايا البشرية في الغالب بجمع البيانات، كما ناقشنا في الفصل ٦. وكما ذكرنا في هذا الفصل، قد يتطلب تصميم الدراسة وجمع البيانات موافقة أخلاقية من الوكالة البحثية أو في الدولة نفسها. ويتعين الالتزام بالممارسات الجيدة للحصول على الموافقة المستنيرة. ومن الممكن التفكير في تقديم تعويضات رمزية للمجيبين من مجموعة المقارنة أو مجتمعاتهم. ويجب وصف معالجة القضايا الأخلاقية في تقرير تقييم الأثر (يوفر المربع ٨-١ مثالاً لذلك).

المربع ٨-١: إعداد تقارير بشأن معالجة القضايا الأخلاقية

أُجريت دراسة لتشجيع غسل اليدين في مدينة كراتشي الباكستانية، وخَلُصت إلى المعلومات التالية فيما يتعلق بالقضايا الأخلاقية:

"كانت تحصل الأسر المعيشية الضابطة بشكل منتظم على كتب للأطفال أو مفكرات أو أقلام رصاص أو أقلام جاف ولكن لا يتلقون أي نصائح بشأن غسل اليدين أو معالجة المياه. وكان يوفر كل شخص بالغ في كل أسرة معيشية موافقة مستنيرة بالنيابة عن أسرته. وتم اعتماد هذا البروتوكول من قِبل مجالس المراجعة المؤسسية في مراكز السيطرة على الأمراض والوقاية منها (CDC) وHOPE ".

المصدر: Bowen et al. (2012).

٨-٤ التوقيت وإعداد الميزانية لتقييم الأثر

الجدول الزمني

خلال عملية متابعة المشروع التقليدية، يتم إجراء مسح خط الأساس في السنة الأولى أو الثانية من المشروع، ومسح خط النهاية في السنة الأخيرة. وكل من هذين التوقيتين قد يكونان ذا دلالة أو لا يحملان أي دلالة على الإطلاق بالنسبة لتقييم الأثر. يشكل مسح خط الأساس تحديًا في حد ذاته، حيث إنه معرض للتجاهل بسبب الانشغال الذي يطبع بداية المشروع. ويشكل مسح خط النهاية أيضًا تحديًا، حيث إن أفضل توقيت لإجرائه هو بمجرد إغلاق المشروع، وبالتالي لن تُعد هناك مشاركة من الجهات المانحة الرسمية.

في بعض الحالات، يتم إجراء مسح كجزء من إعداد المشروع، وقد يبدو أن هذا المسح يمكن استخدامه لتمثيل خط الأساس. عادة لا يكون ذلك مُيسرًا بسبب (١) الصغر الشديد لحجم العينات، فلا يتسنى لها أن تتحلى بقوة عينة كافية (كما ناقشنا في الفصل ٧)، و(٢) عدم كفاية المعلومات حول المشروع لتحديد عينة المعالجة أو العينة الضابطة أو تصميم أداة المسح، و(٣) افتقار أغلب هذه المسوحات إلى التفاصيل الكافية لحشد الحصائل المعنية أو العوامل الشرطية بدقة.

يجب إجراء مسح خط الأساس قبل البدء في تقديم الخدمات إلى المجتمع الإحصائي المستفيد، ولكن لا يزال بالإمكان إجراؤه بعد بدء المشروع. على سبيل المثال، المشروع الذي ينشئ بنية تحتية كبيرة على الأرجح أن يبدأ في إثمار الفوائد بعد مرور ٣ إلى ٤ سنوات على اتمام المشروع. ولهذا، يجب ألا يتم إجراء مسح خط الأساس في السنة الأولى، ولكن يمكن إجراؤه قبل فتح البنية التحتية للاستخدام بوقت قصير. من ناحية أخرى، إذا كان من المقرر إتاحة الخدمات بعد بدء المشروع بوقت قصير، كما هي الحال مع الائتمان الأصغر، يجب أن يسبق مسح خط الأساس تنفيذ المشروع.

يتحدد الموعد الذي يتعين إجراء خط النهاية فيه على أساس المدة الزمنية المطلوبة لملاحظة الأثر المرجو. فيجب أن يشير البعد الزمني لنظرية التغيير إلى طول المدة الزمنية لحين ظهور الأثر (الفصل ٢). يكون للتدخل "مسار للأثر" يسلكه الأثر الملحوظ ويعتمد قياسه على النقطة التي وصل إليها في وقت القياس.

لا ينبغي إجراء مسح خط النهاية إلا بمجرد توقع وجود أثر بقدر جد كافٍ بما يسمح باكتشافه. فرق التوقيت بين خط الأساس وخط النهاية قد يكون أطول أو أقصر من مدة المشروع أو مكون المشروع الخاضع للتقييم. في حالة مشروع بنية تحتية كبير، يجب أن يكون موعد مسح خط النهاية غالبًا عند اكتمال المشروع، أو حتى بعد ذلك ببضع سنوات. إلا أنه في حالة الائتمان الأصغر، على سبيل المثال، قد يكون من الممكن إجراء مسح خط النهاية بينما لا يزال تنفيذ المشروع جاريًا.

يجب تخصيص وقت كافٍ لتصميم المسح. غالبًا ما يستغرق تصميم المسح والتجريب وتدريب مندوب الإحصاء من ٣ إلى ٦ أشهر. يجب إخضاع أدوات المسح للاختبار الشامل. ومن شأن التعجيل بهذه العملية أن يؤثر سلبًا على جودة البيانات، ما يحط من ثمرة الدراسة (يوفر الفصل ٦ مزيدًا من المناقشة). بمجرد جمع البيانات، تلزم فترة إضافية من ٣ إلى ٦ أشهر لإدخال البيانات وتنقيتها وإجراء تحليل أولي لها. ويجب تخصيص مثل هذه الفترة من أجل إجراء تحليل لبيانات خط النهاية.

غالبًا ما يتعين جمع البيانات في جولتين كحد أدنى، وذلك في إطار تصاميم تقييم الأثر الدقيقة، ولكن تتعزز قوة التحليل أكثر إن تلا ذلك عملية جمع بيانات إضافية. مثلاً، إذا كان الفاصل الزمني بين خط الأساس وخط النهاية ٣ سنوات أو أكثر، فإذًا يجوز التفكير في إجراء مسح خط وسط (الجدول ٨-٢). قد يتطلب الأمر تنفيذ مسح موسمي مع بعض التدخلات التي تقترن بتأثيرات تتغير موسميًا أو خلال السنة. في التصاميم التجريبية، قد تدعو الحاجة إلى جمع البيانات بشكل أكثر تواترًا لمراقبة التعيين والمعالجة والتداخلات المحتملة. وربما ينطوي ذلك على إجراء مسوحات قصيرة عبر الهاتف أو شخصيًا مع عينة فرعية من أولئك الذين شملهم المسح.

الجدول ٨-٢: جدول زمني توضيحي لتقييم الأثر

	قبل سنة من المشروع	٣ إلى ٦ أشهر قبل بدء المشروع	السنة ٠	السنة ٣	السنة ٥
الجدول الزمني للمشروع	بدء تصميم المشروع	الموافقة على المشروع	إعداد وبدء المشروع	المراجعة النصفية	نهاية المشروع
الجدول الزمني لتقييم الأثر	بدء تصميم الدراسة	الانتهاء من تصميم الدراسة. تصميم واختبار أدوات المسح المساعدة	جمع بيانات خط الأساس	جمع بيانات خط الوسط: عينة أصغر موجهة بالعمليات	مسح وتحليل خط النهاية. إصدار التقرير بعد ٦ أشهر من إجراء المسح

المصدر: المؤلفان

إدارة التأخيرات والتغييرات في تصميم المشروع

قد يحدث أي تأخير في كل من تدخلات المشروع وفي تقييم الأثر. أحيانا قد يشكل تأخر التنفيذ، في بداية المشروع، فرصة سانحة لكسب مزيد من الوقت لإجراء مسح خط الأساس. من ناحية أخرى، قد يسبب التأخير في إجراء مثل هذا المسح إشكالية إذا تبين أن هذا المسح سيتم إجراؤه بعد طرح المشروع بوقت طويل. ويتعين الموافقة مقدمًا على جدول زمني للدراسة يتسم بالوضوح والمعقولية، ويأخذ في الاعتبار القضايا الموسمية المرتبطة بتوقيت جمع البيانات (كما ناقشنا في الفصل ٦).

كما يطرح التأخر في تنفيذ التدخل قضية أخرى، إذا تبين أن المشروع لا يتمتع بوقت أو بطرح كافٍ لتقديم الأثر المتوقع. قرار إجراء مسح خط النهاية من عدمه، وتوقيته، يجب أن يتسما بالمرونة بالنسبة للمشاريع التي يتم تنفيذها بوتيرة أبطء مما كان متوقعًا في البداية.

تخضع العديد من المشاريع لتغيرات من حيث التصميم في أثناء التنفيذ. ويحتاج فريق الدراسة إلى الوعي بهذه التغييرات في تصميم المشروع لما لها من إنعكاسات على الدراسة. على سبيل المثال، قد يستمر طرح المشروع في المناطق الضابطة، أو قد لا تعود مناطق المعالجة هدفًا للمشروع. ومن هنا تنشأ الحاجة إلى الاحتفاظ بقناة اتصال مستمرة مع فريق المشروع بغية تجنب أي مشكلات تداخل.

الإشراف على تقييم الأثر

يمكن أن يشكل طول المهلة الزمنية بين جولات المسح تحديًا حاسمًا داخل المؤسسات التي غالبًا تدير تقييمات الأثر (مثل الجهات المانحة والوكالات المنفذة المسؤولة عن تنفيذ المشروع وتمويله). علاوة على ذلك، يمكن أن يشكل معدل دوران الموظفين تحديات بالغة على ضمان استمرار التركيز على تقييم الأثر. على اعتبار أن أنشطة تقييم الأثر لا تشكل جزءًا هيكليًا من كل مشروع، يصبح من السهل فقدان التركيز على النقاط الأهم حين تطول فترة عدم النشاط بين جولات جمع البيانات. لضمان الاستمرارية، يجب وضع خطة طويلة الأجل للإشراف المستمر على تقييم الأثر حتى إذا تغير فريق عمل المشروع.

الميزانية

متوسط تكلفة جولة مسح واحدة لمختلف تقييمات الأثر المدعومة من بنك التنمية الآسيوي يبلغ ٢٠٠ ألف دولار (الجدول ٨-٣)، أو حوالي ٤٠٠ ألف دولار للدراسة بأكملها. ويبلغ متوسط الميزانية للدراسات المدعومة من 3ie، وهي منظمة شهيرة تجري تقييمات للأثر، ٤٥٠ ألف دولار. وهناك دراسات تقييم أثر تدعمها وكالات كبيرة، مثل الوكالة الأمريكية للتنمية الدولية والبنك الدولي، وعادة ما تتجاوز تكلفتها المليون دولار. يمكن تقليل هذه التكاليف عن طريق استخدام موارد تحليلية من جهات مانحة وهيئات حكومية، بدلاً من الرجوع إلى استشاريين دوليين باهظي التكلفة.

أهم عنصرين يؤثران على ميزانية تقييم الأثر هما تكاليف الخبراء وتكاليف المسح. تميل تقييمات الأثر في آسيا إلى كونها أقل في التكلفة نسبيًا عن المتوسط العالمي بسبب انخفاض تكاليف تنفيذ عمليات المسح إلى حد ما.

تعتمد تكاليف المسح على (١) عدد جولات المسح، و(٢) حجم العينة، و(٣) الموقع الجغرافي لجمع البيانات. من الضروري إجراء حسابات القوة الإحصائية لتحديد حجم العينة، والميزانية بالتبعية. عندما تكون التكلفة مقيدة، يمكن تقييد مناطق الدراسة في بعض الحالات للحد من ميزانية المسح. على سبيل المثال، بالنسبة لمشروع في ١٠ مقاطعات، يمكن إجراء المسح في ست منها فقط بدلاً من ١٠، وهذا بالطبع على حساب الصلاحية الخارجية.

الجدول ٨-٣: ميزانيات مختارة لتقييمات أثر مدعومة من بنك التنمية الآسيوي، (بالدولار)

العنصر	خلق فرص العمل (الفلبين)	تمديد خط المترو (جورجيا)	Medicard وطوابع الطعام (منغوليا)	تغير المناخ والمرأة (فيتنام)	العمالة القائمة على أعمال الطرق (المحيط الهادئ)	صغار ومتوسطو المزارعين (نيبال)
التصميم	تجربة RCT بسيطة	DiD	RDD	DiD	DiD	DiD
الفريق الدولي	١٣٠٠٠٠	١٢٠٠٠٠	١٢١٠٠٠	١٠٤٠٠٠	٨٠٠٠٠	١٠٢٣٠٤
الفريق الوطني	٩٦٠٠٠	٧٠٠٠٠	–	–	–	–
المسح	١٦٠٠٠٠	٤٤٠٠٠	٨٧٠٠٠	٧١٠٠٠	٦٠٠٠٠	١١٠٧٠٠
ورش العمل والسفر	١٠٠٠٠	٤٤٠٠٠	–	–	٣٥٠٠٠	١٩٠٠٠
غير ذلك	–	٢٢٠٠٠	١٩٧٠٠	–	–	١٨٥٦٠
الإجمالي	٣٩٦٠٠٠	٣٠٠٠٠٠	٢٢٧٧٠٠	١٧٥٠٠٠	١٧٥٠٠٠	٢٥٠٥٦٤

– = لا ينطبق، ADB = بنك التنمية الآسيوي، DiD = الاختلاف في الاختلافات (DiD)، تجربة عشوائية مضبوطة (RCT)، RDD = تصميم انقطاع الانحدار.

المصدر: المؤلفان

مصدر التمويل

مما لا شك فيه أن المشاريع غالبًا ما تحتاج إلى الانتظار حتى حلول موعد الإغلاق المالي قبل إجراء مسح خط النهاية والتمكن من استكمال تحليل تقييم الأثر اللاحق، ما قد يحول دون تمويل المشروع الخاضع للتقييم. وهذه مشكلة أساسية يجب على المنظمات وفرق المشاريع معالجتها عند التخطيط لجهود تقييم الأثر. بالنسبة للمؤسسات المالية الدولية، هناك طرائق متعددة الشرائح يمكنها التكيف بشكل أفضل مع هذه المشكلة. ولكن طرائق المشروع ربما قد لا تكون مصممة، أبدًا، لتلبية احتياجات تقييم الأثر. بدلاً من ذلك، تحتاج الحلول غالبًا إلى النصح على المستوى المؤسسي لحشد الموارد عبر المشاريع.

التعاقد على تقييمات الأثر

للحصول على تقييم الأثر عالي الجودة، لا بد من إشراك فريق من ذوي المهارات العالية يعكف على إجراء الدراسة. ومع ذلك، حتى مع توافر مثل هذا الفريق، يمكن أن تساعد عملية مراجعة الأقران الصارمة على التأكد من ملاءمة التصميم وجدواه في الميدان.

يتم إجراء معظم تقييمات الأثر على يد باحثين أكاديميين على الرغم من أن هناك شركات خاصة توفر أيضًا خدمات أكثر متعلقة بتقييم الأثر. على أي حال، فإن انتقاء فريق الدراسة سيكون تفاضليًا. من ناحية، من الأرجح أن الباحثون الأكاديميون سيقدمون تصاميم تقييم أثر أقوى ولكنهم سيبدون رغبة أقل في تقييم مشاريع معينة أو التركيز على أسئلة محددة لسياق بعينه. ولهذا، تنشأ مخاطرة "الاستيلاء البحثي" (يخضع للمناقشة لاحقًا في هذا الفصل). ومِن ناحية أخرى، فعلى الأرجح أن الشركات الاستشارية الخاصة ستكون أكثر "التزامًا ببنود العقد" ولكن سيقل احتمال تقديمها لتصاميم صارمة أو مبتكرة. وهناك مجال للاستثناء بالنسبة للمجموعة الناشئة من الشركات الخاصة التي تستثمر في إمكانات تقييم الأثر لديها (الملحق ٨-١). هناك جانبان فريدان ومحددان عند التعاقد على تقييم أثر:

(١) مجموعة المهارات اللازمة لإجراء تقييمات الأثر مختلفة عن تلك المطلوبة لتقييمات العمليات، التي تُجرى بشكل أكثر تواترًا. يدرج الملحق ٨-١ بعض الوكالات الرائدة التي تجري تقييمات الأثر. ربما تصادف أيضًا خبراء فرديين من خلال البحث في قوائم الخبراء التي توفرها 3ie.

(٢) يجب أن يتضمن الفريق مَن لديه خبرة سابقة في تصميم تقييمات الأثر وتنفيذها، ويُفضل في القطاع والمنطقة محل اهتمام الدراسة المقترحة. إذا كان جمع البيانات مطلوبًا لهذه الدراسة، فيجب أن تتضمن الخبرة عملاً ميدانيًا في مكان ما في دولة نامية. إذا كان جمع البيانات لا لزوم له، فإذًا يجب أن يتضمن الفريق من يمتلك خبرة في القطاع. يمكن استخدام تقارير تقييم الأثر السابقة التي أعدها الباحثون للمساعدة على فهم مهارات الفرق المقترحة.

الإشراف على إجراء تقييم الأثر

يزيد الإشراك الوثيق والمتواصل، بين فريق الدراسة والأطراف المعنية، من ملاءمة التقييم. يجب أن يبدأ هذا الإشراك والتفاعل في مرحلة تصميم الدراسة، وذلك من خلال تناول نظرية التغيير وأسئلة التقييم. ينبغي لفريق الدراسة أن يقدم تقريرًا أوليًا يضم تصميم الدراسة الكامل وإستراتيجية نموذجية وأدوات المسح. بعد ذلك، يجب الموافقة على العينة الفعلية (على المستوى العنقودي على الأقل) أيضًا بمجرد تحديدها. من الضروري أن يحدد العقد التقرير المراد تسليمه بعد كل جولة مسح ومحتويات هذا التقرير، مثل اختبارات الموازنة لبيانات خط الأساس.

بالإضافة إلى ذلك، يجب أن تكون متطلبات توثيق البيانات وأرشفتها واضحة من البداية. يجب مشاركة البيانات فورًا بعد تنقيتها، مع المكلفين بإدارة تقييم الأثر. على فرق الدراسة أن تقدم موافقة واضحة على الحد الأدنى لفحوص البيانات المقرر إجراؤها، بما يتضمن ذلك التحقق المستقل من صحة البيانات. تتضمن فحوص البيانات أيضًا فحص موازنة بيانات خط الأساس.

هناك صعوبة بشكل خاص بالنسبة لتقييمات الأثر وهي أنها غالبًا ما يتم إجراؤها على مدار فترة طويلة من الزمن، تتخللها فترة ركود وخمول في المنتصف. يجب توثيق بيانات خط الأساس وأرشفتها على النحو السليم. في حالة البيانات الطولية المُجمعة، يلزم العمل بآلية لإعادة تحديد موقع وحدات أخذ العينات الأولية نفسها، وكذلك يجب وضع بروتوكول يوضح ما يتعين فعله عندما يتعذر إعادة تحديد موقعها. لا بد أن تكون هناك قابلية للمقارنة بين المسوحات المقدمة عبر الجولات المختلفة، وفي الغالب يتم استخدام أدوات المسح نفسها لكل جولة. قد تطرأ مشكلات في الاتساق والمحفزات إذا طبقت كيانات مختلفة جولات مسح ذات نسق مختلف.

رقابة جودة المنتج النهائي

سيكون من البديهي والمفيد إبراز خطط النشر والتعميم بشكل واضح، ولا ننسى تحديد الخطوات المتبعة لإجراء رقابة الجودة المتوقعة. قد يخضع تقرير الدراسة الرئيسي للمراجعة من قِبل الأطراف المعنية ومراجعة أقران خارجية. وقد يتولى هذه المراجعة المجموعة الاستشارية المذكورة أعلاه، أو خبراء في تقييم الأثر يتم التعاقد معهم بشكل منفصل. ينبغي للمراجعين من الأقران أن يتضمنوا على الأقل خبيرًا في القطاع وباحث تقييم أثر مؤهلاً.

الاستيلاء البحثي

أغلب تقييمات الأثر يجريها الباحثون الأكاديميون وليس الشركات الاستشارية. من الأرجح أن يمتلك الباحثون الأكاديميون مجموعة المهارات المطلوبة لتصميم تقييمات الأثر وتنفيذها، إلا أن الشركات الخاصة بدأت في الاتجاه إلى الاستثمار لاكتساب هذه المهارات. ومع ذلك، يكون لدى الأكاديميين حافز يتمثل في نشر مقالات

مراجعة الأقران. يحدث "الاستيلاء البحثي" عندما يضطلع فريق الدراسة بتصميم الدراسة، أو إجراء تحليل لإنتاج أوراق بحثية يمكن نشرها في المجلات الأكاديمية، ولكنها لا تتناول بشكل مباشر أسئلة التقييم المعنية لجمهور التقييم. كما يكون للباحثين الأكاديميين مخزون من الأفكار حول التجارب وجمع البيانات لا تتوافق مع الحقائق الميدانية، وبالتالي يمكن أن تأخذ الدراسات منحنيات غير عملية أو غير متصلة بالموضوع (Barrett and Carter 2010).

يمكن التخفيف من مخاطرة "الاستيلاء البحثي" بعدة طرق. أولاً، يجب تحديد نطاق العمل بصورة واضحة قبل التعاقد مع وضع التدخل في نصب الأعين. ثانيًا، يجب ذكر أسئلة التقييم بوضوح ودون ترك أي مجال للريبة من حيث الإشارة والإسناد. ثالثًا، يتعين استخدام مجموعة استشارية تتضمن خبير تقييم أثر باستطاعته توفير فحص دقيق لفنيات المقترح. ويستطيع هذا الخبير أن يساعد في التعامل مع المناقشات الجارية مع الباحثين بحيث يوضحون المقومات التي تؤهلهم لمعالجة أسئلة التقييم المعنية بالمشروع أو اعتراضهم على التطرق إلى هذه الأسئلة بالتحديد. رابعًا، يجب مراقبة تنفيذ تقييم الأثر عن كثب لضمان معالجة أسئلة التقييم الأصلية بشكل كافٍ مع اختتام التصميم والبدء في جمع البيانات. خامسًا، يجب صياغة العقد وعملية المراجعة بحيث تسمح بعملية مراجعة نهائية يتم التأكد من خلالها من أن التقرير النهائي يجيب فعليًا عن أسئلة التقييم المتفق عليها.

٨-٥ مصادر الدعم لإجراء تقييم الأثر

المشورة

ربما يستفيد الممارسون الإنمائيون من غير الاقتصاديين من إجراء مناقشات، مع مقيِّمي الأثر وغيرهم من الاقتصاديين، للحصول على ملاحظات حول ما إذا كان تقييم الأثر لمشروعهم ذا ذا جدوى. على سبيل المثال، يتوفر لدى بنك التنمية الآسيوي فريق في "قسم البحث الاقتصادي والتعاون الإقليمي" يمكن استشارته. من الممكن أيضًا التواصل مع الأكاديميين ذوي الصلة أو استشارة المنظمات غير الحكومية المهتمة بموضوع تقييم الأثر.

العثور على خبير

تتطلب تقييمات الأثر مجموعة مهارات محددة، ويزداد الأمر صعوبة عند البحث عن خبراء في مجال معين، نظرًا لمحدودية المجالات التي يغطيها كل خبير وممارس. كما قد لا ينجذب الممارسون (الأشخاص أو المنظمات) للعطاءات التنافسية على تعيينات تقييم الأثر ما لم يعلموا بالفرصة المتاحة لتقديم عروضهم.

توجد أربع طرق للعثور على الخبراء:

١. يمكن الرجوع إلى قائمة أو لائحة الخبراء التالية:

- تحتفظ 3ie بلائحة تضم خبراء تقييم الأثر، يمكن البحث فيها حسب الدولة والقطاع (http://www.3ieimpact.org/en/evaluation/expert-roster/). لقد وافق جميع الخبراء الذين تشملهم اللائحة على إدراج أسمائهم.

- يدرج الموقع الإلكتروني لمعمل عبد اللطيف جميل لمكافحة الفقر (Jamal Latif Poverty Action Lab) قائمة بجميع المؤسسات التابعة له مع توضيح مجال اهتمام كل منها (https://www.povertyactionlab.org/affiliated-researchers).

- تدرج مؤسسة Innovations for Poverty Action في موقعها الإلكتروني الكيانات البحثية التابعة لها وشبكتها البحثية الأوسع (https://www.poverty- action.org/research).

- تدرج مؤسسة Bureau for Research and Economic Analysis of Development الباحثين الاقتصاديين الرائدين، الموجودين في الولايات المتحدة بشكل رئيسي، ولكن لا يعمل جميعهم في مجال تقييم الأثر (http://ibread.org/ bread/people).

٢. المؤلفون الذين أجروا تقييمات الأثر في تدخلات مماثلة للتدخل المقرر إجراء تقييمه أو تندرج تحت القطاع نفسه في المنطقة المحددة. يمكن العثور على تقييمات الأثر الحالية في المصادر التالية:

- 3ie Evidence Database، وهي عبارة عن قائمة شاملة لتقييمات الأثر في الدول النامية (http://www.3ieimpact.org/en/ evidence/). وهي مرتبطة بلائحة خبراء 3ie، ومن الممكن البحث فيها حسب القطاع والدولة والمنطقة.

- تتضمن قاعدة بيانات الأبحاث الاقتصادية، IDEAS، مجموعة كبيرة من الأبحاث المنشورة وغير المنشورة ضمن قاعدة بيانات اقتصادية (https://ideas.repec.org/). لا تنحصر هذه المراجع على موضوع تقييم الأثر، ولكن عند تضمين مصطلح البحث impact evaluation "تقييم الأثر"، ستظهر الكثير من الأبحاث ذات الصلة.

- يتضمن Google Scholar أكبر مجموعة شاملة من الأبحاث الأكاديمية، ولكن من الأرجح أيضًا أن أي عملية بحث يمكن أن تثمر عددًا كبيرًا من النتائج غير المتصلة بالموضوع المعني (https://scholar.google.com).

٣. يمكن سؤال ذوي الخبرة في إدارة تقييمات الأثر، عن توصياتهم بناءً على خبرتهم.

٤. يمكن الإعلان عن دعوات عامة للمقترحات. إن تقييمات الأثر عبارة عن دراسات كبيرة، وعادة ما تتطلب من المؤسسة أن تجري عملية جمع بيانات على نطاق واسع. يتزايد عدد الشركات الاستشارية من القطاع الخاص التي تحسِّن من إمكاناتها لإعداد تقييمات الأثر، إلا أن بعضها لا يزال يفتقر إلى مجموعة المهارات المطلوبة. يتضمن الملحق ٨-١ قائمة بالوكالات ذائعة الصيت التي تقوم بإعداد تقييمات أثر.

إذا كان من المقرر طرح الدراسة في عطاء تنافسي، فمن الضروري التأكد من إطلاع الشركات المؤهلة بشكل مناسب على هذه الفرصة. يمكن التعاقد مع بعض الوكالات، مثل 3ie، لتقديم إعلان على مواقعها الإلكترونية.

٦-٨ تفسير استنتاجات تقييم الأثر وتقديمها

التهديدات التي تعترض الصلاحية

يتعين أخذ التهديدات التي تواجه صلاحية التحليل في عين الاعتبار عند تفسير استنتاجات تقييم الأثر. يجب أن يكون الاعتبار الأول للصلاحية الداخلية، وما إذا كانت التأثيرات محددة بشكل ملائم بالنسبة للعينة الخاضعة للتحليل. بصفة عامة، توجد تهديدات محتملة أكثر على الصلاحية الداخلية بالنسبة للدراسة شبه التجريبية مقارنة بتجربة عشوائية مضبوطة منفذة على الوجه الصحيح.

بالنسبة للتجارب العشوائة المضبوطة، لا مناص من التأكيد على التنفيذ المتقن لبروتوكول التوزيع العشوائي المقصود، وأنه لم يكن هناك مجال يسمح بتقصير الموظفين الميدانيين في واجباتهم. يجب أيضًا تأكيد الحرص على إجراء عملية جمع البيانات بشكل متماثل بالنسبة للمجتمعين الإحصائيين: المعالَج وغير المعالَج. بخلاف ذلك، قد تعني "تأثيرات هوثورن" أن المستويات المتمايزة للملاحظة تؤثر على السلوك، أو على التقارير على الأقل. من الضروري أيضًا التيقن مما إذا كانت المعرفة المتمايزة للمعالجة، بين المجموعة المعالَجة والمجموعة الضابطة، قد تُحدث تحيزًا في الإجابات. أحد أشكال التحيز قد يكون "تأثير جون هنري"، وفيه تغير المجموعة الضابطة سلوكها في الإجابات بناء على علمها بأن المعالجة لا تشملها. وعلى النقيض، قد تحفز تأثيرات المعالجة الوهمية على إحداث تغييرات لدى المجموعة المعالَجة استنادًا إلى توقعات المعالجة.

بالنسبة للدراسات شبه التجريبية، في الغالب يكون هناك شاغل رئيسي يتمثل في التحيز المتغير المحذوف، بما أن ليس كل المتغيرات المحتملة يمكن ضبطها في المقاربات القائمة على الانحدار أو المطابقة. في هذه الحالات، يتعين إخضاع النماذج المطبقة لعلميات تمحيص صارمة. بالنسبة للمتغيرات المساعدة، يجب الانتباه جيدًا إلى وحدة المساعدة المختارة وما إذا كانت متداخلة مع الحصائل. أما بالنسبة لتصاميم انقطاع الانحدار، فيجب مراجعة الموازنة حول نقطة انقطاع الأهلية لمتغير التعيين.

بالنسبة لجميع الدراسات، من الضروري التيقن من أن المجتمعين الإحصائيين (المعالج والضابط) لا يشهدان أي برامج أخرى، بخلاف التدخل الخاضع للتقييم، تجنبًا للوقوع في خطر التداخل. من الأهمية بمكان أيضًا، بالنسبة لكل الدراسات، التأكد من أن أي مجموعة عنقودية يشملها التدخل قد تم اعتبارها بشكل ملائم في شكل نسبة من الأخطاء المعيارية في إطار تأثيرات المعالجة. وبعد نشر الدراسة، يجب إتاحة البيانات المستخدمة للجمهور، إن أمكن، لتمكين المزيد من التحليل.

أهم شيء يجب الانتباه إليه هو الصلاحية الخارجية، أو تعميم الاستنتاجات التي تتمتع بصلاحية داخلية. ينص تقييم الأثر على أثر تصميم معين في سياق محدد. سواء كانت هذه الاستنتاجات مترجمة إلى سياقات أخرى، بما يجعلها تمثل صلاحية خارجية للتصميم. افترض أن الدراسة وجدت أن تدخل التعلم بمساعدة الحاسوب بين طلاب المدرسة الابتدائية التي يتاح فيها للطالب استخدام الحاسوب لساعة واحدة في الأسبوع، يسبب زيادة في حصائل التعلم بنحو ١٠٪. هذا لا يعني أن التحليل سيؤول إلى النتيجة نفسها مع طلاب المدرسة الثانوية. ولا يشير ذلك ضمنيًا أيضًا إلى أن مضاعفة الوقت إلى ساعتين أسبوعيًا سيرفع من حصائل التعلم بمعدل ٢٠٪ بالضرورة.

ما الذي تعنيه النتائج؟

تتألف النتائج الرئيسية لدراسات تقييم الأثر من كم معين من تأثير المعالجة ودلالتها الإحصائية. عند تفسير هذه الاستنتاجات، من الضروري عدم "الانبهار بالتقييمات العالية". في أغلب الأحيان، تركز الدراسات التي تشمل عينات كبيرة العدد، على الدلالة الإحصائية للتأثيرات، إلا أن حجم التأثير نفسه لا يقل عن الدلالة الإحصائية أهمية، وهو ذو أهمية أكبر بالنسبة للكثير من صانعي السياسات. ففي العينة الكبيرة، تلك التأثيرات ذات الأحجام الضئيلة، التي تكون صغيرة جدًا بشكل يجعلها عديمة الصلة بالسياسة، قد تكون ذات دلالة إحصائية. لذلك، من الضروري جدًا العمل بمقياس لتفسير حجم التأثير.

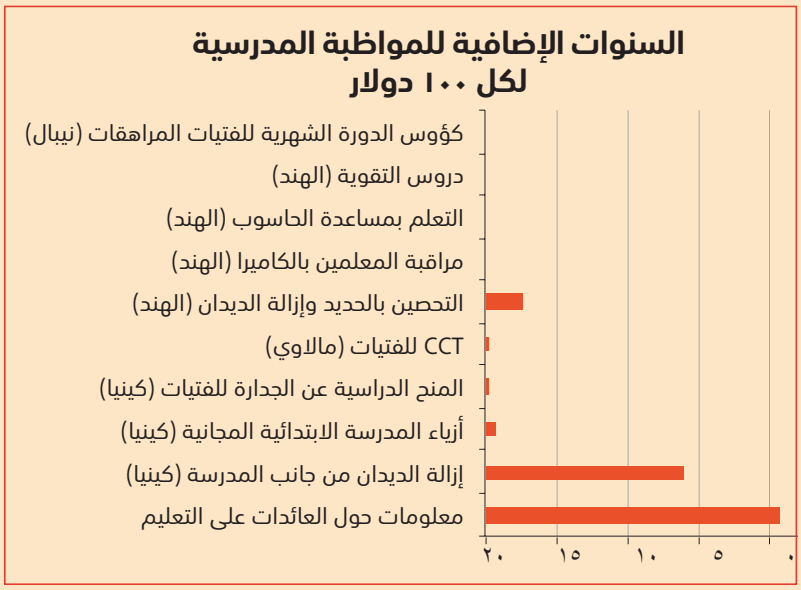
يمكن ذكر حجم التأثير بدلالة وحدات يستطيع صانعو السياسات فهمها بسهولة. على سبيل المثال، التحسينات في حصائل التعلم تصدر بها تقارير بشكل شائع كتحسن في درجات الامتحان بنسب انحرافات معيارية. من المعروف في أوساط خبراء القطاع أن الانحراف المعياري ٠,٢ يعبر عن "تحسن جيد". يتقدم الأطفال عادة بانحرافات معيارية تتراوح بين ٠,٢٠ و٠,٢٥ في العام. لذلك، التدخل الذي يحقق تحسنًا بنحو ٠,٢ يعادل التعلم المتحصل من التعليم المدرسي لسنة إضافية. في حالات حيث تكون مستويات المتغير الناتج مفهومة للغاية، مثل عائدات المحاصيل لكل هكتار، قد يكون من المثمر اعتبار التأثيرات على جميع المستويات. بالنسبة للعديد من التدخلات، يسهل فهم التحسينات المئوية التي يشهدها متغير الحصيلة من حالة عدم المعالجة إلى حالة المعالجة.

يمكن استخدام تحليل فعالية التكلفة لتقدير تكلفة التحسن المعياري في الحصائل المعنية (يوضح المربع ٨-٢ مثالاً لذلك). يكون تحليل فعالية التكلفة ملائمًا عندما لا تكون هناك حصيلة واحدة معنية. في الحالات الأخرى، يمكن استخدام تقديرات تأثير المعالجة كمدخلات في نماذج الفائض الاقتصادي التي تدعم تحليل فعالية التكلفة.

المربع ٨-٢: تحليل فعالية التكلفة لتدخلات زيادة المواظبة المدرسية

يتطلب تحليل فعالية التكلفة مقياسًا شائعًا للبسط. في الحالة الموضحة هنا، تقدّر كل الدراسات أثر التدخلات المتتالية على مدار سنوات للمواظبة المدرسية.

التدخلات الأربعة الأولى الموضحة لم يكن لها أثر بالغ.



توضح الأمثلة الأخرى أهمية اعتبار التكلفة. يُعد توفير المعلومات التي تثبت العائدات على التعليم للآباء هو التدخل الأكثر فعالية على الإطلاق، ليس لأن ذلك له أثر ملحوظ بشكل أكبر، بل لأنه أقل في التكلفة.

CCT = التحويل النقدي المشروط.
المصدر: استنادًا إلى Dhaliwal et al. (2012).

٨-٧ الخلاصات: التوجيهات التنموية القائمة على الأدلة

ستطبق نتائج تقييم الأثر على عدة أصعدة. فعلى سبيل المثال، يمكن أن تزود نتائج التقييم صناع القرارات، على جميع المستويات، بمعلومات ثرية. يمكن أن تصير عملية تقييم الأثر - في حد ذاتها ومن خلالها - نقطة ارتكاز للتفاعل بين الباحثين التطبيقيين والممارسين الإنمائيين. فهي تُعرف الممارسين الإنمائيين بنظريات ومفاهيم سلوكية ربما لم يضعوها بعين الاعتبار من قبل، وكذلك تعرضهم لتجارب جديدة من خلال التعرف على تدخلات مماثلة قد لا يكونوا على علم بها. على نحو مماثل، تقرِّب العملية الباحثين من كثب إلى الحقائق الميدانية، بحيث يفهمون كيف تتم ترجمة الاعتبارات النظرية على أرض الواقع. ما يعزز العمليات هو الصرامة التي تقترن بتنفيذ التدخل بموجب تقييم الأثر، حيث يمكن أن تساعد في فرض مفهوم أن البروتوكولات الميدانية يتم تفسيرها باستمرار ومتابعتها بدقة.

إن الإتيان بأدلة قوية على صحة النظرية، والفهم، والافتراضات الداعمة للبرامج الإنمائية يمكن أن يكون له استخدامات على مستويات عدة. فمن شأن ذلك أن يحسّن من التصور المفاهيمي طويل الأجل للتنمية، إضافةً إلى أن الدراسات المصممة التي يمكن نشرها في المجلات العلمية الأكثر شهرة سيكون لها وقع كبير وستجذب انتباه الأوساط الأكاديمية والجمهور بشكل أوسع. بل قد تمتد تأثيراتها لتشمل في مجموعة من التحسينات السياساتية والبرامجية في جميع أنحاء العالم. ومع ذلك، لا يمثل ذلك إلا النذر القليل مما يمكنها التأثير فبه يمكن أيضًا الاستناد إلى استنتاجات تقييم الأثر لتوفير المزيد من الأدلة المباشرة للدفاع عن التدخلات المجدية وتوسيعها، بل وتكرارها في مكان آخر. ويمكنها أن تُظهر كيف يمكن إنجاز أنواع معينة من المشاريع بشكل أفضل وتوفر دليلاً مفاهيميًا للابتكارات والتعزيزات. يمكن أن يجلب تقييم الأثر الانتباه إلى أمور ذات جدوى في إطار التنمية ويساعد على إعادة تنظيم المحفزات نحو المؤسسات والفرق التي توفر أفكارًا جديدة ونتائج إنمائية. بعبارة أخرى، يستطيع تقييم الأثر أن يسمح للممارسين الإنمائيين باتباع عمليات اختبار المنتج والتعلم والتحسين المستمر التي يقوم عليها نجاح الشركات المبتكرة في القطاع الخاص.

المراجع

Barrett, C. and M. Carter. 2010. The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections. *Applied Economic Perspectives and Policy*. 32 (4). pp. 515–548.

Bowen A., M. Agboatwalla, S. Luby, T. Tobery, T. Ayers, and R. M. Hoekstra. 2012. Association Between Intensive Handwashing Promotion and Child Development in Karachi, Pakistan: A Cluster Randomized Controlled Trial. *Archives Pediatric Adolescent Medicine*. 166 (11). pp. 1037–1044.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch. 2012. Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. Abdul Latif Jameel Poverty Action Lab (J-PAL). Boston: MIT.

Independent Evaluation Group (IEG). 2012. *World Bank Group Impact Evaluations: Relevance and Effectiveness*. Washington, DC: World Bank. DOI: 10.1596/978- 0-82139717-6.

Peersman, G., I. Guijt, and T. Pasanen. 2015. Evaluability Assessment for Impact Evaluation. A Methods Lab Publication. London: Overseas Development Institute.

VeLure Roholt, R. and M. L. Baizerman. 2012. *Evaluation Advisory Groups*: New *Directions for Evaluation, Number 136. Hoboken*, New Jersey: Wiley.

مراجع إضافية للاطلاع

Bamberger, M. 2006. Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints. *Doing Impact Evaluation* No. 2. Washington, DC: World Bank. http://lnweb18.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/757a5cc0bae22558852571770059d89c/$FILE/conduct_qual_impact.pdf.

Bedi, T., S. Bhatti, X. Gine, E. Galasso, M. Goldstein, and A. Legovini. 2006. Impact Evaluation and the Project Cycle. *Doing Impact Evaluation* No. 1. Washington, DC: World Bank. http://siteresources.worldbank.org/ INTISPMA/Resources/383704-1146752240884/doing_ie_series_01.pdf.

Pritchett, L. and J. Sandefur. 2013. Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix – Working Paper 336. Washington, DC: Center for Global Development. https://www. cgdev.org/sites/default/files/1425010_file_Szekely_Results_Based_ Social_Policy_FINAL.pdf.

Székely, M. 2011.Toward Results-Based SocialPolicy Design and Implementation – Working Paper 249. Washington, DC: Center for Global Development. https://www.cgdev.org/sites/default/files/context-matters-for-size_1. pdf.

الملحق ٨-١: قائمة المنظمات المشاركة في تقييمات الأثر

دوليًا

Abt Associates: شركة أبحاث مقرها في الولايات المتحدة تطبق برنامجًا محليًا على نطاق واسع، وترفع من مستوى اهتمامها بالتنمية الدولية (http:// www.abtassociates.com/).

American Institutes for Research (AIR): المعاهد الأمريكية للأبحاث هي وكالة أبحاث خاصة غير ربحية وتتخذ واشنطن العاصمة مقرًا لها. وتمتلك الوكالة موارد بحثية آخذة في النمو بشأن التنمية الدولية وقد أجرت تقييمات أثر في العديد من الدول. (http://www.air.org/).

Center for Effective Global Action (CEGA): مركز العمل العالمي الفعال (AGEC) عبارة عن مجموعة يقع مقرها في جامعة كاليفورنيا في بيركلي. (http://cega.berkeley.edu/).

Center for Learning on Evaluation and Results (CLEAR): مركز التعليم بشأن التقييم والنتائج (CLEAR) عبارة عن شراكة عالمية مدعومة من وحدة "مجموعة التقييم المستقلة" التابعة للبنك الدولي التي تدعم بناء القدرات لتقييم الأثر وكذلك مقاربات التقييم الأخرى. وتتفرع منها مراكز إقليمية في كل من جنوب وشرق آسيا (https://www.theclearinitiative.org).

تقييم الأثر الإنمائي (DIME)، البنك الدولي: إن DIME عبارة عن وحدة معنية بإدارة تقييمات الأثر ضمن "مجموعة اقتصاديات التنمية" للبنك الدولي (http://www.worldbank.org/en/research/dime).

Centre for the Evaluation of Development Policies (EDePo): مجموعة تقييم أثر مقرها في كلية لندن الجامعية (https://www.ifs.org.uk/ centres/EDePo/).

Catalyst Management Services (CMS): مجموعة تنتمي إلى القطاع الخاص تتخذ مدينة بنغالور الهندية مقرًا لها، وتتمتع بخبرة في تقييم الأثر. (http://cms.org.in/).

FHI360: مؤسسة استشارية مقرها في الولايات المتحدة تولي تركيزًا كبيرًا على الأبحاث والتقييم، ولها مكاتب إقليمية في جميع أنحاء آسيا (https://www.fhi360.org/).

IDInsight: تساعد IDInsight الوكالات على وضع تصاميم المشاريع والتقييم الصارم الدقيق لهذه المشاريع. يقع مقر IDInsight في الولايات المتحدة ولها مكتب في نيودلهي (http://idinsight.org/).

Innovations for Poverty Action (IPA): تجري مؤسسة IPA (ابتكارات من أجل التحرك لمكافحة الفقر) تجارب عشوائية مضبوطة (RCT) فقط. ويركز عملها على التمويل الأصغر، ولكنها عملت أيضًا في قطاعات أخرى. تمتلك مؤسسة IPA بضعة مكاتب إقليمية ولا تجري الدراسات إلا في الأماكن التي يكون لها فيها مقر. وفي المنطقة، لدى IPA مكتب في مانيلا. لمؤسسة IPA جهات تابعة وأغلبها تنتمي إلى الجامعات الأمريكية. وتتولى هذه الجهات التابعة مسؤولية تصميم تقييمات الأثر وتحليلها. أما جمع البيانات، فيتولاه المكتب الميداني. توفر

مؤسسة IPA دورات تدريبية وورش عمل بصفة دورية وعادةً ما تكون مركزة على موضوع بعينه. وقد سبق أن نظمت المؤسسة فعاليات مع بنك التنمية الآسيوي (ADB). تتعاون IPA وJ-PAL بشكل وثيق مع بعضهما البعض (https://www.poverty-action.org/).

Institute of Development Studies (IDS): معهد دراسات عليا معنية بالتنمية يتضمن "مركز تقييم الأثر" الذي يكرس جهوده في استكشاف طيف واسع من طرق تقييم الأثر بما يتجاوز التصاميم الكمية كبيرة الحجم. يُعد هذا المركز مشروعًا مشتركًا مع **ITAD**، وهي شركة أبحاث خاصة مقرها في المملكة المتحدة. (http://www.ids.ac.uk/).

International Food Policy Research Institute (IFPRI): يركز المركز الدولي لبحوث السياسات الغذائية (IFPRI) على الغذاء والزراعة والتنمية الريفية، بما يشمل ذلك موضوعات مثل تنمية الأسواق والبنية التحتية والتغذية. ويجري IFPRI تقييمات أثر في جميع هذه المجالات. يمتلك IFPRI مكاتب إقليمية في عدة دول حول العالم، بما في ذلك بنجلاديش، وجمهورية الصين الشعبية، والهند، وباكستان (http://www.ifpri.org/).

International Initiative for Impact Evaluation (3ie): المبادرة الدولية لتقييم الأثر (3ie) هي منظمة غير حكومية عالمية تعمل في مجال تقييمات الأثر، ولها مكاتب في نيودلهي وواشنطن ولندن. وهي تمول تقييمات الأثر من خلال أشكال متعددة من المنح، وتجري تقييمات أثر مختارة. تساعد 3ie في إجراء تقييمات الأثر لوكالات أخرى، بما في ذلك مراجعة الأقران الخارجية للمقترحات والمنجزات. وتركز المنظمة على دعم تصاميم الدراسة التي تجيب عن أسئلة ذات صلة بالسياسات، باستخدام طرق مختلطة تتضمن إما التصاميم التجريبية (التجارب العشوائية المضبوطة) أو غير التجريبية. وتوفر 3ie ورش عمل بصفة دورية. وقد سبق أن نظمت فعاليات مع بنك التنمية الآسيوي (ADB)، بما في ذلك "المؤتمر حول الأهمية العظيمة لإجراء تقييم الأثر" الذي انعقد في مقرات البنك الرسمية في سبتمبر عام ٢٠١٤. (http://3ieimpact.org/).

معمل عبد اللطيف جميل لمكافحة الفقر (J-PAL): معمل J-PAL هو أول وأشهر وكالة تكرس جهودها في برامج الأثر الإنمائي، وقد اشتركت - حتى تاريخه - في قرابة ٣٠٠ تجربة عشوائية مضبوطة.ويتمركز المعمل في معهد ماساتشوستس للتكنولوجيا (MIT)، ويمتاز بشبكة إقليمية مع تمثيل آسيوي جنوبي في نيودلهي. يتميز معمل J-PAL بشبكة من الجهات التابعة البحثية، وأغلبها من الجامعات الأمريكية والأوروبية، التي من بينها العديد من الكيانات الرائدة عالميًا العاملة في مجال تقييم الأثر. وتكون هذه الجهات التابعة مسؤولة عن تصميم الدراسة، الذي ينفذه مديرو البرنامج، وعادةً ما يقدمون دراسات عليا بشأن جمع البيانات أو ينخرطون في دور استشاري على السياق المحلي. لا يجري معمل J-PAL إلا التجارب العشوائية المضبوطة. وتوفر الوكالة دورة تدريبية قصيرة بعنوان J-PAL Executive Training وهي تدور حول تصميم التجارب العشوائية وتنفيذها. يتعاون J-PAL وIPA بشكل وثيق مع بعضهما البعض (https://www.povertyactionlab.org/).

Mathematica: تشارك Mathematica، وهي وكالة أبحاث خاصة مقرها في الولايات المتحدة، في العدد المتنامي لتقييمات الأثر في الدول النامية (https://cipre.mathematica-mpr.com).

NORC: مركز أبحاث الرأي الوطني NORC هي شركة أبحاث تستقر في الولايات المتحدة وتتميز باتصالات أكاديمية قوية وتجري تقييمات أثر في كل من الولايات المتحدة وبالخارج (http://www. norc.org).

Oxford Policy Management (OPM): أكسفورد لإدارة السياسات (OPM) هي شركة تابعة للقطاع الخاص، وتتخذ الولايات المتحدة مقرًا لها، وتعمل على زيادة إمكاناتها لإجراء تقييمات الأثر. لدى OPM مكتب إقليمي في نيودلهي (http://www.opml.co.uk).

RTI International: شركة أبحاث تستقر في الولايات المتحدة وتركز خبراتها في تقييم الأثر. (https://www.rti.org/).

Social Impact: شركة أبحاث صغيرة تتمركز في الولايات المتحدة وتولي اهتمامها الأساسي لتقييم الأثر. (https://socialimpact.com/).

إقليميًا ووطنيًا

Philippine Institute for Development Studies (PIDS): المعهد الفلبيني لدراسات التنمية - Philippine Institute for Development Studies هي وكالة أبحاث مدعومة من الحكومة آخذة في التركيز بشدة على تقييم الأثر. (https://www.pids.gov.ph).

Sambodhi: مجموعة من القطاع الخاص مقرها في نيودلهي، وتتمتع بخبرة في تقييم الأثر. (http://sambodhi.co.in/).

الملحق الأول
تطبيق طرق التقدير على تقييم الأثر

يقدم هذا الملحق مقدمة أكثر تفصيلاً عن كيفية تطبيق طرق تقييم الأثر. ويبدأ الملحق بتعريف أنواع الترميز المستخدمة في التحليل الاعتيادي والاعتبارات المتبعة عند اختيار الطرق. ثم يقدم الملحق نظرة عامة على تقنيات STATA والخطوات الروتينية المتبعة عند تخمين آثار التدخلات. وتستند تطبيقات العينة إلى مجموعات بيانات نموذجية يحتفظ بها برنامج STATA، أو إجراءات روتينية كتبها خبراء في تقنيات STATA وصارت متاحة للجمهور والمستخدمين، بحيث يمكن تكرارها عبر سلسلة من الأوامر البسيطة.

فيما يلي مُخطط مختصر لأقسام هذا الملحق:

القسم الأول: إطار الحصائل المحتملة
القسم الثاني: التجارب العشوائية المضبوطة
القسم الثالث: نماذج الاختلاف فى الاختلافات (DiD) والتأثيرات الثابتة
القسم الرابع: الوحدات الضابطة الاصطناعية
القسم الخامس: المقاربات المستندة إلى درجة الميل (المطابقة، والوزن، وعوامل التقدير القوية المزدوجة)
القسم السادس: المقاربات المستندة إلى المتغيرات المساعدة (مربعات صغرى ثنائية المرحلة، وانحدارات المعالجة داخلية النشأة، والانحدارات التبديلية داخلية النشأة)
القسم السابع: تصميم انقطاع الانحدار

القسم الأول: إطار الحصائل المحتملة

١-١ إطار Rubin/Neyman السببي ونظام الترميز المرتبط به

يرتبط تقييم ما إذا كان للبرنامج تأثير بترميزات السببية (أي السبب والتأثير) التي يمكن تمثيلها بالاستعانة بالترميز القياسي المأخوذ من إطار الحصائل المحتملة لإطار روبين نيمان للنتائج المحتملة (Rubin 1974). سنفترض أولاً أن هناك بيانات خاصة بالوحدات المُعَالجة "n" (والتي قد تكون أشخاصًا، أو منازل، أو مزارع، أو مؤسسات، أو شركات أعمال، وما إلى ذلك)، وأن النية هي تقييم أثر أحد التدخلات على تلك الوحدات.[1] ثانيًا، ومن أجل تبسيط الأمور، سنفترض أنه يمكن تحديد ما إذا كانت الوحدة تشارك في التدخل أو لا باستخدام متغير ثنائي يأخذ قيمة واحدة من اثنتين (١ إذا كان الشخص يشارك في التدخل، و٠ إذا لم يكن الشخص يشارك). ستُسمى الوحدات المشاركة

في التدخل باسم الوحدات المُعَالجة والوحدات غير المشاركة في التدخل باسم الوحدات غير المُعَالجة أو المقارنة.[٢] W_i يمكن أن يكون متغيرًا ثنائيًا (أو صوريًا) يبين ما إذا كان الفرد i تمت معالجته أم لا:

$W_i = 0$: الوحدة i غير مُعَالجة
$W_i = 1$: الوحدة i مُعَالجة

يمكن أن يُمثل المتغير Y "الحصيلة" التي يُفترض أن يسببها التدخل (المعالجة). الأمر المهم، أنه توجد في الحقيقة قيمتان من Y بالنسبة إلى أي وحدة معينة: (أ) القيمة التي يمكن أن تحدث إذا كانت الوحدة غير مُعَالجة، والتي يمكن الإشارة إليها بالرمز Y_0؛ و(ب) القيمة التي يمكن أن تحدث إذا كانت الوحدة مُعَالجة، والتي يمكن الإشارة إليها بالرمز Y_1:

Y_{0i} = قيمة Y بالنسبة للوحدة i إذا كانت غير مُعَالجة
Y_{1i} = قيمة Y بالنسبة للوحدة i إذا كانت مُعَالجة

من المهم أن نلاحظ هنا أن كلاً من Y_{0i} وY_{1i} مُعرّفة لجميع الوحدات، بغض النظر عما إذا كانت في حالة المُعَالجة أو كانت وحدة ضابطة. على سبيل المثال، ربما تكون هناك ملاحظات بأن Y_1 وحدة "مُعَالجة"، ولكن *إذا لم تكن الوحدة مُعَالجة فستكون الملاحظات هي* Y_0.

وبموجب هذا الترميز، سيكون الهدف من تقييم الأثر هو تقدير $Y_{1i} - Y_{0i}$. وهذا هو *تأثير المعالجة* (أثر التدخل) بالنسبة إلى الوحدة. والهدف هو تقديم تأثير المعالجة على كل وحدة في بعض الفئات محل الدراسة (أو على الأقل معرفة متوسط $Y_{1i} - Y_{0i}$ بالنسبة للوحدات في الفئة محل الدراسة). بيد أن المشكلة الأساسية في تقييم الأثر هي أنه بالنسبة لكل وحدة يمكن أن تكون هناك ملاحظات بـ Y_1 أو Y_0 عند كل نقطة زمنية، ولكن ليس لكليهما.

الجزء الأخير في الترميز الأساسي هو القيمة المُلاحظة لـ Y في كل وحدة، والتي يمكن الإشارة إليها بـ Y_i بالنسبة إلى الوحدة i. والقيمة المُلاحظة لـ Y هي قيمة الوحدة "الفعلية" من Y استنادًا إلى كون الوحدة مُعَالجة فعليًا، أو هي وحدة ضابطة.

[١] يمكن أن تكون أيضًا n من المنازل، أو المؤسسات، أو أي وحدة للملاحظة أخرى، ولكن حرصًا على تجنب اللغة المجردة، سيشير هذا النص عمومًا إلى وحدات n بأنهم أشخاص.

[٢] لاحظ أنه في هذه الحالة، ستكون هناك حالة معالجة واحدة فقط، وسيحصل الأشخاص على هذه المعالجة أو لا. يمكن توسعة الترميز بسهولة ليشمل حالات وحدات معالجة (وحدات ضابطة) متعددة.

هناك علاقة بسيطة جدًا بين Y_i، وY_{0i}، وY_{1i}، وW_i:

$$Y_i = Y_{0i} \times (1 - W_i) + Y_{1i} \times W_i$$

بمعنى آخر، إذا كانت $W_i = 0$ بالنسبة إلى الوحدة i، فإن Y_i المُلاحظة لهذه الوحدة هي Y_{0i}.
إذا كانت $W_i = 1$ فإن Y_i المُلاحظة لهذه الوحدة هي Y_{1i}:

$$Y_{0i} = Y_i \text{ إذا كانت } W_i = 0$$
$$Y_{1i} = \text{ إذا كانت } W_i = 1$$

٢-١ ثلاثة سيناريوهات لكيفية تحديد المتغيرات "W_i"

في ظل الترميز العام والإعداد في القسم ١-١، فإن القدرة على تقدير أثر التدخل تعتمد على عملية تحديد W_i في كل وحدة i. ومن منظور التقدير، هناك ثلاثة احتمالات:

الاحتمال الأول (التحديد العشوائي): تم تحديد W_i عشوائيًا كجزء من "تجربة عشوائية".

$$W_i \perp\!\!\!\perp (Y_{0i}, Y_{1i})$$

نكرر أن الفكرة الأساسية وراء التعيين العشوائي للأشخاص في المجموعة المعالجة والمجموعة الضابطة أنه يُتوقع أن تكون نفسها من حيث جميع السمات الملحوظة وغير الملحوظة. وإذا تم تعيين المشاركين عشوائيًا، فإنه يمكن تقدير أثر التدخل (تأثير المعالجة) من خلال مقارنة الحصائل المتوسطة بين مجموعة المعالجة والمجموعة الضابطة.

الاحتمال الثاني (التعيين غير المدمج): لم يتم تعيين W_i عشوائيًا، بل كان شرطيًا على بعض المتغيرات المُلاحظة (X_i)، ولا تعتمد على (مستقلة عن) Y_{0i} وY_{1i}. يُسمى هذا الاحتمال (أو الافتراض) بأسماء تُستخدم بالتبادل وهي "التعيين غير المدمج" أو "الاختيار حسب الملاحظات" (حيث إن المتغيرات X_i هي الأمور التي يمكن ملاحظتها)، أو افتراض "الاستقلال الشرطي". بصورة رسمية، يتم التعبير عن هذا الافتراض كما يلي.

$$W_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}) \mid X_i$$

الفكرة الأساسية وراء افتراض التعيين غير المدمج أنه طالما كان هناك عنصر مقارنة (شرط) خاص بـ X_i، فلا توجد علاقة بين W_i وأي من Y_{0i} أو Y_{1i}. مما يعني في الواقع أنه بالنسبة لقيمة X_i محددة، فإن W_i يتم تحديده عشوائيًا.

تجدر ملاحظة أن افتراض الاحتمال الثاني ليس بقوة الاحتمال الأول (التعيين العشوائي) على أي حال. ينبغي في التعيين العشوائي أن يكون للمشاركين في المجموعة المعالجة والمجموعة الضابطة التوقعات نفسها من حيث السمات الملحوظة وغير الملحوظة. أما في افتراض التعيين غير المدمج، يجب أن يفترض المُحلل أنه إذا كانت X_i على اليمين منضبطة، فإن المشاركين في المجموعة المعالجة والمجموعة الضابطة يشتركون في التوقعات من حيث جميع السمات الملحوظة وغير الملحوظة الأخرى.

أما في المواقف حيث يعتقد المرء أن افتراض التعيين غير المدمج صحيح، فإن طريقة تقييم الأثر الأساسية المُستخدمة في تقدير تأثيرات المعالجة هي طريقة المطابقة (الأكثر استخدامًا هي مطابقة درجة الميل)، أو تحليل الانحدار الذي يتحكم في متغيرات الخلط. سيتم في القسم الخامس مناقشة المطابقة وعلاقتها بتحليل الانحدار مع العوامل الضابطة.

الاحتمال الثالث (التعيين المدمج): W_i يعتمد على Y_{0i} وY_{1i}، حتى بعد اشتراط متغيرات X_i الملحوظة. وهذا أصعب ما يمكن معالجته، ولكنه يكون أيضًا في كثير من الأحيان أكثر الاحتمالات المعقولة، لا سيما إذا كانت الوحدة i هي التي تتخذ القرار بخصوص المشاركة في البرنامج أم لا (أي اختيار $W_i = 0$ أو 1)، وطالما كانت الوحدة i تراعي Y_i، فيحتمل أن W_i تعتمد على Y_{0i} وY_{1i}.

ماذا يمكن أن يفعل المُقيم إذا كان يحتمل أن يكون تعيين W_i مدمجًا؟ تشمل الطرق المُستخدمة في هذا الموقف تقدير الاختلاف فى الاختلافات (DiD) (القسم ٣)، والوحدات الضابطة الاصطناعية (القسم ٤)، والمقاربات المستندة إلى درجة الميل (القسم ٥). بالإضافة إلى ذلك، يمكن أن يبحث المُحلل عن المجموعات الفرعية في الفئات أو المواقف حيث يوجد التعيين الشرطي الملحوظ، ويقوم بتطبيق المتغيرات المساعدة (إذا لم يكن هناك نقطة انقطاع واضحة في الأهلية) (القسم ٦) أو تصميم انقطاع الانحدار (إذا كان المتغير لديه مشاركة شرطية ذات نقطة انقطاع حادة) (القسم ٧).

١-٣ تعريف تأثير المعالجة والتحيز في الاختيار

من أهم المعاملات التي يحاول المقيمون تقديرها هي *متوسط تأثير المعالجة* (ATE)، والذي يُعرّف بأنه

$$ATE = E[Y_{1i} - Y_{0i}]$$

بمعنى آخر، هذا متوسط تأثير المعالجة عبر جميع وحدات العينة. وقد يكون للوحدات المختلفة قيم مختلفة من $Y_{1i} - Y_{0i}$، ولكن يهتم المُقيمون عادةً بمتوسط الفئة بأكملها. بشكل بدهي، يجيب متوسط تأثير المعالجة عن السؤال التالي: *كيف يمكن أن تتغير Y إذا تمت معالجة العينة بالكامل بالنسبة إلى الموقف حيث لم تتم معالجة أي فرد؟*

من أهم المُعامِلات التي يحاول المقيمون تقديرها هي متوسط تأثير المعالجة على المعالَج (ATT)، والذي يُعرّف بأنه

$$ATT = E[Y_{1i} - Y_{0i} \mid W_i = 1]$$

بمعنى آخر، هذا هو متوسط تأثير المعالجة (متوسط $Y_{1i} - Y_{0i}$) *بالنسبة لأعضاء العينة الذين حصلوا على المعالجة*. بشكل بدهي، يجيب متوسط تأثير المعالجة على المعالَج (ATT) عن السؤال التالي: *كيف تتغير Y بالنسبة إلى الوحدات التي تمت معالجتها فعليًا؟*

عندما يريد المُقيّمون أن يفهموا ما إذا كان ينبغي توسعة البرنامج أو لا، فقد يهتمون أيضًا بمتوسط تأثير المعالجة في الوحدات غير المُعالجة (ATU)، أو بمتوسط التأثير الذي يمكن أن يُسببه التدخل على من لم يشاركوا بعد.

$$ATU = E[Y_{1i} - Y_{0i} \mid W_i = 0]$$

تحيز الاختيار هو التأثير الذي يريد المُقيمون التخلص منه باستخدام تقنيات تقييم الأثر. ويمكن تمثيل هذا بأنه الفرق بين متوسط الأثر للمشاركين في مجموعة المُعالجة، والمشاركين في المجموعة الضابطة قبل تطبيق أي علاج على أي من المجموعتين. ويتم تعريف هذا باستخدام المعادلة التالية:

$$E[Y_{0i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0]$$

هناك مفهومان إضافيان أساسيان في الاقتصاد القياسي ينبغي إدراكهما، واللذان ينطبقان في سياق طرق الانحدار المستندة إلى مَعلَمات (حيث يمكن نمذجة العلاقات بين المتغيرات والحصائل بطريقة واضحة في صورة دالة). المفهوم الأول هو *اختلاف التباين* والذي يعني أساسًا أن للبيانات قابلية للتغيير غير متساوية بالنسبة لجميع قيم المتغيرات. ولتكون مطمئنًا تجاه هذا الاحتمال، ينبغي أن تستخدم جميع المقاربات المستندة إلى مَعلَمات الأخطاء المعيارية القوية تجاه اختلاف التباين، والتي تتجمع في البيانات إذا كانت المعالجة مترابطة داخل المجموعات أو الأماكن المستهدفة. المفهوم الثاني هو *الارتباط الخطي المتعدد*، حيث قد لا تكون معاملات الانحدار دقيقة إذا تمت إضافة معاملات مستقلة مترابطة. ويمكن التحقق من هذا الأخير من خلال اختبار عوامل تضخم التباين، ومعاملات الارتباط على المستوى الثنائي.

القسم الثاني: التجارب العشوائية المضبوطة

التجارب العشوائية المضبوطة تُسمى غالبًا "المعيار الذهبي" في تقييم الأثر (Athey and Imbens 2017). ولفهم السبب، من المفيد أن نبحث في الكلمتين "عشوائية" و"تجربة" على حدة. إن السمة الأساسية المشتركة بين جميع التجارب أن المُقيم يعمد إلى *التعامل* مع السبب بهدف *كشف* نتائجه. وهذا ما يميز التجارب عن الدراسات الرصدية التي تركز أولاً على التأثير، ثم تحاول اكتشاف الأسباب، وهذه مهمة أصعب كثيرًا.

تضمن إجراءات التعيين العشوائي أن تكون المجموعة المعالجة والمجموعة الضابطة متشابهتين (حسب التوقعات) في جميع السمات المتوسطة الملحوظة وغير الملحوظة، حتى وإن تمت إضافة عدد كبير جدًا من الوحدات. بمعنى آخر، تميل كل صفة، بخلاف التدخل (المعالجة)، إلى أن تكون هي نفسها في جميع الوحدات في المجموعة المعالجة والمجموعة الضابطة، وذلك بسبب التعيين العشوائي. بالنسبة للأشخاص، يشمل ذلك الصفات الملحوظة (X_i) مثل العمر، والنوع، والثروة، وكذلك الصفات غير الملحوظة، مثل القدرات والسلوكيات.

التجربة العشوائية المضبوطة هي تجربة يغلب أن تكون فيها المجموعة المعالجة والمجموعة الضابطة متساويتين في جميع الصفات الملحوظة وغير الملحوظة، وهذا باستثناء التعرض للتدخل أو الخضوع للمعالجة. ومن أجل هذا تعزو الاختلافات في نتيجة Y_i بين المجموعة المعالجة والمجموعة الضابطة إلى المعالجة نفسها فقط، ولا تكون ناتجة عن أي سبب آخر. ويُمكّن هذا المُقيم من الحصول على تقدير غير متحيز لمتوسط تأثير المعالجة.

يتحقق المُقيمون من وقت لآخر من أن عملية التعيين العشوائي تفي بما هو متوقع منها، بحيث تنتج مجموعة معالجة ومجموعة ضابطة تبدوان متشابهتين (أو متوازنتين). ويفعل المُقيم ذلك بإجراء اختبارات إحصائية لتحديد ما إذا كان المشاركون في المجموعة المعالجة والمجموعة الضابطة متشابهين إحصائيًا (في المتوسط) من حيث الصفات الملحوظة أو لا. وإذا كانت المجموعة المعالجة والمجموعة الضابطة متشابهتين في هذه الصفات (على سبيل المثال، وفقًا لاختبارات t الإحصائية للاختلافات من حيث كل عدد من الصفات الملحوظة المهمة)، حينئذٍ يمكن أن يثق المُقيم في أن تقديرات تأثير المعالجة الناتجة عن التجربة العشوائية المضبوطة ستكون غير متحيزة.

عندما تكون جميع الصفات متوازنة بين مجموعة المعالجة والمجموعة الضابطة، فستكون كافية لتقدير متوسطات العينة للأشخاص كل على حدة في مجموعة المعالجة والمجموعة الضابطة. وبالتالي، ستمكن من حساب الفروق بين متوسطات العينة للتمكن من تقدير متوسط تأثير المعالجة. ولرفع مستوى الدقة، يتم عادةً التطبيق في إطار الاختلاف فى الاختلافات (DiD)، حيث تكون الفروق بين استبيانات خط الأساس وخط النهاية متوسطة بالنسبة لمجموعة المعالجة والمجموعة الضابطة، ثم المقارنة بينهما. يمكن تطبيق الطرق الإحصائية القياسية (مثل اختبار t للمجموعتين) لاختبار الفرضية اللاغية لعدم تأثير المعالجة على المجتمع الإحصائي ذي الصلة.

يمكن أن تفيد مقاربة الانحدار متعدد المتغيرات، حتى في سياق التجربة العشوائية. فأولاً، يمكن أن يساعد الانحدار متعدد المتغيرات في تصحيح أي حالة عدم توازن قد تحدث في مجموعة المعالجة والمجموعة الضابطة. ثانيًا، يمكن أن يساعد في تحسين مستوى دقة التقديرات، ويحد من الأخطاء القياسية. ثالثًا، يمكن أن يساعد في تقديم دليل على كيفية اختلاف تأثيرات المعالجة حسب صفات المجموعة المستفيدة. وسينتج الانحدار متعدد المتغيرات عن نماذج التأثيرات الثابتة.

القسم الثالث: نماذج الاختلاف فى الاختلافات (DiD) والتأثيرات الثابتة

١-٣ مقدمة إلى الاختلاف في الاختلافات

لا يوجد خارج التجربة العشوائية مجتمع إحصائي "ضابط"، حيث إن المعالجة ليست الصفة الوحيدة المختلفة بين المجتمعات المُعَالجة وغير المُعَالجة. ومن ثم فإن المصطلحات تتغير إلى "مجموعة المقارنة" أو "غير المُعَالجة". تعتمد طريقة الاختلاف فى الاختلافات (DiD) على البيانات الخاصة بالمجموعتين المُعَالجة وغير المُعَالجة عند نقطتين زمنيتين أو أكثر: كلتاهما قبل حدوث المعالجة وبعدها. لا سيما وأن البيانات ينبغي أن تتضمن قيم نتيجة المتغير محل الاهتمام والتي تمت قياسها قبل وبعد المعالجة، بالنسبة لكل من المشاركين المُعَالجين وغير المُعَالجين.

تقوم طريقة الاختلاف فى الاختلافات (DiD) بتخمين تأثير المعالجة عن طريق المقارنة بين التغيير قبل وبعد في متوسط النتيجة في المجموعة المُعَالجة، وبين التغيير الذي يشهده متوسط النتيجة في المجموعة غير المُعَالجة قبل التدخل وبعده. ويمكن أن يكون تقدير تأثير المعالجة الناتج غير متحيز إذا كان "افتراض الاتجاه المتوازي" ثابتًا. ويقترح هذا الافتراض أن التغيير أو الاختلاف الذي تشهده مجموعة المُعَالجة بين ما قبل التدخل وبعده ينبغي أن يتبع الاتجاه (الانحدار) نفسه مثل التغيير أو الاختلاف الذي تشهده المجموعة غير المُعَالجة بين ما قبل التدخل وبعده *إذا لم تتم معالجة المجموعة المُعَالجة*. بمعنى آخر، يقترح الافتراض أن التغيير المتوسط في المجموعة غير المُعَالجة يمثل التغيير المتوسط في الواقع المضاد في مجموعة المُعَالجة حال غياب المعالجة. ويكون هذا الافتراض باطلاً إذا كانت مجموعة واحدة فقط (وليس الأخرى) تُغير اتجاهها أثناء فترة المعالجة *نتيجة لعوامل غير مرتبطة بالمعالجة*.[٣] ولكن للأسف لا يمكن اختبار هذا الافتراض عادةً، مثله في ذلك مثل الافتراضات القائمة على المطابقة وتحليل المتغيرات المساعدة (Ryan et al. 2015).[٤]

[٣] إلى جانب الانتهاكات في افتراض الاتجاهات المتوازية، هناك تهديدات محتملة أخرى تؤثر على صلاحية مقاربة الاختلاف في الاختلافات (DiD). ومن بين هذه التهديدات، التغييرات التفاضلية في تكوين المجموعتين المُعَالجة وغير المُعَالجة (في أثناء فترة المعالجة). وتضم أيضًا *انحدار أشينفلتر* حيث يواجه المشاركون في البرنامج انخفاضًا مفاجئًا في متغير حصائل قبل المعالجة قبل انضمامهم إلى البرنامج مباشرةً، ثم يرتدّون إلى الحالة الطبيعية بعد البرنامج (ليس بالضرورة أن ينتج ذلك عن البرنامج). أخيرًا، يجب أيضًا أن يتم اختيار مواصفات الانحدار المستخدم في نمذجة وتقدير تأثير المعالجة بعناية.

[٤] يتعذر على المُقيم أن يختبر معقولية افتراض الاتجاهات المتوازية باستخدام بيانات فترتين زمنيتين مختلفتين، ولكن يمكن في بعض الحالات فعل ذلك باستخدام بيانات أكثر من فترتين. فإذا توفرت لدى المُقيم بيانات فترتين أو أكثر من فترات ما قبل المعالجة، فسيمكنه على سبيل المثال اختبار ما إذا كانت اتجاهات ما قبل المعالجة للمجموعتين المُعَالجة وغير المُعَالجة متوازية أو لا.

٢-٣ الاختلاف فى الاختلافات (DiD)

يمكن تقدير مُقدِّر الاختلاف فى الاختلافات (DiD) في متوسط الأثير المعالجة على المعالَج (ATT)، باستخدام مواصفات الانحدار التالية:

$$Y_{it} = \alpha + \beta_1 W_{it} + \beta_2 T_t + \beta_3 (W_{it} \times T_t) + \varepsilon_{it}$$

ففي المعادلة يمثل Y_{it} متغير النتيجة (قبل وبعد المعالجة)، بينما يكون W_{it} متغيرًا ثنائيًا يبين ما إذا كان الشخص قد حصل على المعالجة أو لا، بينما يكون T_t متغيرًا ثنائيًا يبين الفترات قبل وبعد المعالجة، في شكل تقاطع، وفي حد الخطأ. تأثير المعالجة هو المعامل β_3، والذي يساوي $(Y_{11} - Y_{10}) - (Y_{01} - Y_{00})$.

يتم في الواقع تقدير مُقدِّر الاختلاف فى الاختلافات (DiD) باستخدام معادلة الانحدار التالية، والتي تتحكم أيضًا في الملحوظات الأخرى، وX_i ومعاملاتها γ، حيث يكون تأثير المعالجة هو المعامل β_3.

$$Y_{it} = \alpha + \beta_1 W_{it} + \beta_2 T_t + \beta_3 (W_{it} \times T_t) + \gamma X_{it} + \varepsilon_{it}$$

تطبيق مُقدِّر الاختلاف فى الاختلافات (DiD) في STATA

يمكن تقدير مُقدِّر الاختلاف فى الاختلافات (DiD) بسهولة عن طريق إنشاء متغير صوري للزمن خاص بملحوظات ما قبل المعالجة، ومتغير صوري للمعالجة. بعدئذٍ تتم تسمية الانحدار الخطي باستخدام الأمر "regress"، متبوعًا بمتغير تابع، ومتغير صوري للزمن، ومتغير صوري للمعالجة، وحد التفاعل بين المتغيرات الصورية للزمن والمعالجة. وينتج معامل التفاعل تأثير المعالجة.

يستخدم المثال التالي المسح الطولي الوطني في الولايات المتحدة بهدف تقدير أثر عضوية الاتحاد على السجل الطبيعي للأجور الحقيقية. ويعالج المسح عام ١٩٨٠ باعتباره فترة ما قبل التدخل، وعام ١٩٨٨ باعتباره فترة ما بعد التدخل، مع إدخال الاتحاد مقام "المعالجة". ولعضوية الاتحاد تأثير مهم جدًا، مع معامل شديد لمدة التفاعل time#union.

الأوامر:

```
webuse nlswork.dta

*يتم إنشاء لوحة تضم فترتي الملاحظات
keep if year==88 | year==80
gen time=0
replace time=1 if year==88

reg ln_wage time##union, robust
```

النتائج:

Linear regression

Number of obs	=	3574
F(3, 3570)	=	73.67
Prob > F	=	0.0000
R-squared	=	0.0534
Root MSE	=	.47822

ln_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.time	.1377238	.0187794	7.33	0.000	.1009045	.1745432
1.union	.1404369	.0232693	6.04	0.000	.0948145	.1860593
time#union 11	.0885873	.0348885	2.54	0.011	.0201839	.1569906
_cons	1.698597	.0125059	135.82	0.000	1.674078	1.723117

٣-٣ نماذج التأثيرات الثابتة

نموذج التأثيرات الثابتة ثنائي الاتجاهات هو مقاربة تناظرية لمُقدِّر الاختلاف فى الاختلافات (DiD)، حيث يسمح بالتطبيق على أكثر من فترتي ملاحظة. يعمل نموذج التأثيرات الثابتة بكفاءة على تمركز الملاحظات لكل وحدة على مدار الزمن (يحذف القيمة المتوسطة)، بحيث يعكس الانحدار الفرق على مدار الوقت لكل وحدة. وهناك مقاربة عملية مقابل لتلك المقاربة التناظرية وهي تضمين المتغيرات الصورية لكل وحدة في الانحدار، وكذلك في الزمن.

$$Y_{it} = \alpha + \beta_1 W_{it} + \gamma X_{it} + \beta_2 T_t + \alpha_i + \varepsilon_{it}$$

ففي هذه المقاربة يضيف انحدار الحصيلة متجهًا للصفات الملحوظة X، والمعاملات γ، ومتغير الزمن T مع المعاملات المُقدرة، وإضافة α_i إلى الثابت α الذي يختلف باختلاف الشخص. يمتص متغير الأشخاص الصفات غير المتغيرة بمرور الزمن، بينما يمتص متغير الزمن تأثير الاتجاهات الكلية، بحيث ينتج المعامل في W متوسط تأثير المعالجة على المُعالَج (ATT) نفسه مثل المعامل في حد التفاعل WxT في الاختلاف فى الاختلافات (DiD).

تطبيق نماذج التأثيرات الثابتة في برنامج STATA

يمكن تطبيق نماذج التأثيرات الثابتة في برنامج STATA بسهولة باستخدام الأمر "xtreg"، متبوعًا بمتغير الحصيلة، والمتغيرات المستقلة، والمدة الزمنية، والمتغيرات الصورية، والخيار "fe".

يستخدم المثال التالي المسح الطولي الوطني في الولايات المتحدة الخاص بتقدير تأثير عضوية الاتحاد على السجل الطبيعي للأجور الحقيقية على مدار الفترة ١٩٧١–١٩٨٨، مع ضبط أسابيع العمل، وما إذا كان يوجد في المدينة عدد كبير من السكان أم لا، وعدد سنوات الخبرة العملية. صفات الأشخاص، مثل التعليم، غير متغيرة بالنسبة إلى الزمن، ولهذا لم تُضاف إلى النموذج. تأثير عضوية الاتحاد كبير للغاية، مع معامل مهم.

الأوامر:

```
webuse nlswork.dta
xtset idcode year
xtreg ln_wage union wks_work ttl_exp not_smsa i.year, fe robust
```

النتائج:

```
Fixed-effects (within) regression               Number of obs =     18855
Group variable: idcode                          Number of groups =   4128

R-sq:  within  = 0.1601                         Obs per group: min =      1
       between = 0.3129                                        avg =    4.6
       overall = 0.2418                                        max =     12

                                                F(15,4127)    =     92.63
corr(u_i, Xb)  = 0.1634                         Prob > F      =    0.0000

                                               (Std. Err. adjusted for 4128 clusters in idcode)
------------------------------------------------------------------------------
             |               Robust Std.
     ln_wage |      Coef.   Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       union |   .0964293   .0094099    10.25   0.000     .0779808    .1148778
    wks_work |   .0016111   .0001468    10.98   0.000     .0013233    .0018989
     ttl_exp |   .0394834   .0024047    16.42   0.000     .0347689    .0441978
    not_smsa |  -.0954969   .0185183     5.16   0.000    -.1318028   -.0591911
             |
        year |
          71 |   .0120917   0107149     1.13   0.259    -.0089153    .0330988
          72 |  -.0086902   .0125379     0.69   0.488    -.0332712    .0158908
          73 |  -.0285989   .0138411    -2.07   0.039    -.0557348   -.0014629
          77 |  -.0488101   .0169072    -2.89   0.004    -.0819574   -.0156628
          78 |  -.0396196   .0182474    -2.17   0.030    -.0753944   -.0038449
          80 |  -.1612624   .0204025    -7.90   0.000    -.2012623   -.1212626
          82 |  -.1952173   .0228232    -8.55   0.000    -.2399632   -.1504715
          83 |  -.1404167   .024931     -5.63   0.000    -.1892949   -.0915384
          85 |  -.2274674   .0276147    -8.24   0.000     -.281607   -.1733277
          87 |  -.2650911   .0312928    -8.47   0.000    -.3264419   -.2037404
          88 |  -.2309831   .0343385    -6.73   0.000    -.2983051   -.1636611
             |
       _cons |   1.488808   .0146594   101.56   0.000     1.460068    1.517549
-------------+----------------------------------------------------------------
     sigma_u |  .3682758
     sigma_e |  .25143585
         rho |  .68206727   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

القسم الرابع: الوحدات الضابطة الاصطناعية

٤-١ مقدمة إلى الوحدات الضابطة الاصطناعية

تسمح طريقة الوحدات الضابطة الاصطناعية بتبسيط افتراض الاتجاهات المتوازية في طرق الاختلاف في الاختلافات (DiD). لعمل ذلك، يتم استبدال متغير السلسلة الزمنية بمتغير مقطعي عرضي في البيانات الطولية المُجمعة الواردة في اللوحة. تعمل هذه المقاربة عن طريق أخذ مجموعة من ملحوظات المقارنة المحتملة، وتحديد اللوغاريتمية الموزونة بين وحدات المقارنة التي تُعظم المطابقة بين الاتجاهات في المتغيرات المستقلة، ومتغير الحصيلة للوحدة الموزونة الاصطناعية (Abadie et al. 2010).

٤-٢ مُقدِّر الوحدات الضابطة الاصطناعية

تفترض الوحدات الضابطة الاصطناعية أن حصائل عدم التدخل ناتجة عن نموذج عامل يتكون من عوامل عامة مشتركة وغير ملحوظة (مثل زمن التأثير الثابت) مع المعاملات α والمتغيرات الملحوظة المستقلة X، والمتغيرات μ غير الملحوظة المستقلة مع المعاملات ϕ، وحدُّ الخطأ ε، كما يلي:

$$Y_{it} = \alpha T_t + \gamma X_{it} + \phi \mu_{it} + \varepsilon_{it}$$

ويمكن افتراض أن هناك مجموعة من الأوزان ω_j تُطبق على الملحوظات غير المُعالجة بما يسمح لكل من الحصيلة والمتغيرات الملحوظة المستقلة بتقليد الأنماط في الوحدة المُعالجة قبل المعالجة. بحيث يكون كل من

$$\sum_{j=2}^{j1+} \omega_j^* Y_{jt} = Y_{1t} \forall t \in \{1, \dots T_0\}$$

9

$$\sum_{j=2}^{j1+} \omega_j^* X_{jt} = X_{1t}$$

وبموجب هذه الحالات، يمكن افتراض أن الأوزان هي تمثيل شكل مختصر من $\phi\mu$. وينتج عن هذا أنه يمكن استعادة متوسط تأثير المعالجة على المُعالَج (ATT) عبر تطبيق الأوزان على الملحوظات غير المُعَالجة، والمجموع، والمقارنة مع الملحوظات المُعَالجة.

$$\Delta = Y_{1t} - \sum_{j=2}^{j1+} \omega_j^* Y_{jt}$$

يمكن توسيع تطبيق هذه المقاربة، ليشمل العديد من الوحدات المُعَالجة من خلال تقدير متجهات الوحدة المُعَالجة بمفردها، وحساب متوسط تأثيرات المعالجة في الوحدات. ولا يتم اختبار الأهمية باستخدام استنباط التقارب التقليدي. بل ينبغي إجراء اختبارات وهمية على البيانات، ويتم تقدير احتمالية مضاعفة التأثيرات عشوائيًا حسب أكبر تقدير.

تطبيق الوحدات الضابطة الاصطناعية في برنامج STATA

يمكن تطبيق مقاربات الوحدات الضابطة الاصطناعية في حزمة برنامج STATA باستخدام الأمر "synth_runner"، متبوعًا بمتغير الحصيلة، والمتغيرات المستقلة، ومواصفة متغير المعالجة (Quistorff and Galiani 2017). وتسمح هذه الحزمة بالتطبيق على وحدات مُعَالجة متعددة. ولكن يجب قبل تشغيل هذا الأمر أن تتم موازنة اللوحة موازنة تامة، مع تحديد مُعرّف اللوحة والفترات الزمنية عبر استخدام "tsset". ويجب تعريف المعالجة باستخدام متغير ثنائي، ويجب أن تبدأ الوحدات المُعَالجة في المعالجة بعد الفترة الأولية، وعدم انسحاب أي وحدة من المعالجة.

يستخدم المثال التالي المسح الطولي الوطني في الولايات المتحدة لتقدير تأثير الانضمام إلى الاتحاد على الأجور الحقيقية على مدار الفترة ١٩٨٧-١٩٨٨، باتباع نموذج "فترة التدريب" من ١٩٨٠ إلى ١٩٨٥ والتي تنتج الأوزان. يقدر النموذج الأجور كدالة عامة عن العرق، والدرجة الجامعية، والخبرة، وساعات العمل، والسكن في الجنوب، والسكن في مدينة مركزية. لجعل البيانات تعمل جيدًا في المثال، سيتم استخدام مجموعة فرعية من بيانات المسح فقط. نتيجة لذلك، سيكون تأثير المعالجة أصغر كثيرًا عن المُقدر في المثال الأخير، وليس مهمًا.

الأوامر:

```
ssc install synth, all
net install synth_runner, from(https://raw.github.com/bquistorff/synth_runner/master/) replace

webuse nlswork.dta
*إنشاء وإعلان لوحة متوازنة من فئة فرعية من البيانات
drop if year<80
gen wage=exp(ln_wage)
by idcode: egen minunion=min(union)
keep if wage!=. & race!=. & collgrad!=. & c_city!=. & south!=. ttl_exp & hours!=. & union!=. & minunion==0
drop if union==1 & year<86
by idcode: egen obs=count(year)
keep if obs==6
tsset idcode year

synth_runner wage race wks_work grade collgrad c_city south ttl_exp hours, d(union)
```

النتائج:
نتائج ما بعد المعالجة: التأثيرات، قيم p، وقيم p القياسية

	estimates	pvals	pvals_std
c1	.5443459	.512112	.182283

القسم الخامس: المقاربات المستندة إلى درجة الميل (المطابقة، والوزن، وعوامل التقدير القوية المزدوجة)

٥-١ مقدمة إلى درجة الميل

تحاول المقاربات المستندة إلى درجة الميل أن تضع نماذج لعملية الاختيار الخاصة بالتدخل، وذلك باستخدام الصفات الملحوظة للمشاركين وغير المشاركين في البرنامج. وعلى وجه التحديد، تقوم تلك المقاربات على افتراض أنه بعد ضبط الصفات قبل المعالجة X_i، لن تكون هناك علاقة بين W_i وأي من Y_{0i} أو Y_{1i}. مما يعني في الواقع أنه بعد ضبط قيمة X_i، فإن W_i يتم تحديده عشوائيًا. وإذا استمر هذا *التعيين غير المدمج* أو *افتراض الاستقلال الشرطي* فعليًا، فمن المناسب استخدام نموذج من المشاركة في البرنامج لتقدير تأثير المعالجة.

يرى كل من روزينباوم وروبين أنه يمكن تقدير درجة الميل باستخدام الانحدار اللوغاريتمي (أو وحدة الاحتمالية) الخاص بـ W_i على المتجه X_i (إعطاء β معاملات وحد الخطأ η) وتوقع احتمالية (p_i p-hat) لكل فرد (Rosenbaum and Rubin 1983). ويُقصد بـ p-hat درجة الميل التي يمكن استخدامها بطرق مختلفة لتقليل تحيز الاختيار.

$$logitP(W = 1|X) = \beta X_i + \eta_i, \text{ or } \hat{p}_i = e(X_i, \hat{\beta})$$

هناك طرق مختلفة يمكن من خلالها تطبيق درجة الميل. ويشمل هذا مطابقة الحالة الفردية، وموازنة درجة الميل العكسي، وتخمين الانحدار بالإضافة إلى استخدام درجة الميل.

٥-٢ مطابقة درجة الميل

من أشهر تطبيقات درجة الميل ما يتم عبر "مطابقة درجة الميل" (Rosenbaum and Rubin 1985). بصورة أساسية، بالنسبة لكل مشارك حصل على المعالجة في البيانات الملحوظة، تجد هذه التقنية حالة مطابقة واحدة أو أكثر لم تحصل على المعالجة. والمطابقة هي الشخص/الوحدة غير المُعَالجة التي لها القيم نفسها في X_is مثل المُشارك المُعَالج. وبعد "المطابقة" في مجموعة المعالجة والمجموعة الضابطة، يمكن مقارنة متوسط متغير الحصيلة بين الوحدات المُعَالجة وغير المُعَالجة لتقدير أثر متوسط المعالجة. ودرجة الميل هي احتمالية أن إحدى الوحدات تحصل على معالجة معينة بموجب متجه المعامل المشترك X_i.

بيد أن المشكلة الأساسية في استخدام المطابقة لإجراء تقييم الأثر هي أن *التعيين غير المدمج* هو افتراض قوي نسبيًا. حتى وإن كان المُقيم قادرًا على التحكم في عدد كبير من الصفات الملحوظة (X_is)، فقد تستمر الحالة ويكون هناك عوامل غير ملحوظة مهمة (في η_i) والتي قد تؤثر على كل من W_i وY_i. وحتى تتمكن المطابقة من إنتاج على الأقل تخمينات غير متحيزة نسبيًا، يجب أن يجد المُقيم كل عامل ملحوظ مهم، ويتحكم فيه، والذي يرتبط بالتزامن مع كل من W_i وY_i.

هناك طرق متعددة لمطابقة الأشخاص المُعَالجين وغير المُعَالجين. ومن بين الخيارات المختلفة التي ينبغي أن يحددها المُقيم عند إجراء المطابقة أن يختار عدد المطابقات التي ينبغي إجراؤها لكل شخص مُعَالج أو غير مُعَالج، وهل ستتم المطابقة مع أو بدون استبدال الأشخاص الذين تمت مطابقتهم بالفعل، وهل سيتم تطبيق مطابقة "تامة" على قيمة درجة الميل، أو سيُسمح بالمطابقة عبر نطاق من الدرجات، ونوع اللوغاريتمية التي تطابق جميع الأشخاص المُعَالجين مع جميع الأشخاص غير المُعَالجين.

ربما يميل محلل البيانات إلى الاعتقاد بأن المطابقة لا تختلف كثيرًا عن الانحدار الخطي متعدد المتغيرات التقليدي. فكلا الطريقتان تعتمدان على افتراض التعيين غير المدمج (تمت مناقشته في القسم ١-٢). لكن في الحقيقة، فإن الانحدار الذي يتحكم إحصائيًا في المتغيرات المشتركة هو أحد أشكال المطابقة الوهمية. بيد أن الفرق يكمن أنه بينما لا تعتمد المطابقة على النمذجة الإحصائية، أو افتراضات النماذج العملية، فإن الانحدار يعتمد عليهم. وسواء يمكن لتحليل الانحدار أن يقدر تأثير المُعالجة غير المتحيزة أو لا، فإن ذلك لا يعتمد فقط على افتراض التعيين غير المدمج، بل أيضًا على الافتراض بأن النموذج الضمني للانحدار تم اختياره اختيارًا صحيحًا. في الوقت نفسه، فإن عدد الأشخاص في مجموعة البيانات الملحوظة صغير، لهذا تكون المطابقة والطرق اللا معلمية الأخرى أقل كفاءة (ذات قوة إحصائية أقل) من الطرق المعلمية، مثل تحليل الانحدار. بالإضافة إلى ذلك، عندما يكون هناك ارتباط بين المتغيرات المشتركة والاهتمام بالمعالجة، فقد تتأثر تقنية الانحدار الخطي التقليدية بالارتباط الخطي المتعدد، بينما لا ترتبط به تقنيات المطابقة.

تطبيق مطابقة درجة الميل في برنامج STATA

يمكن تطبيق مطابقة درجة الميل في حزمة برنامج ستاتا (STATA 13 وSTATA 14) باستخدام الأمر "teffects psmatch"، متبوعًا بمتغير الحصيلة، ومتغير المعالجة، والمتغيرات المستقلة التي تحدد المجموعات المُعَالجة وغير المُعَالجة قبل أن تفرق المعالجة بينها.

يمكن تطبيق مثال بسيط على التقنية باستخدام إحدى مجموعات البيانات النموذجية من برنامج STATA. يقيم المثال التالي تأثير الرعاية قبل الولادة في مرحلة الشهور الثلاثة الأولى من الحمل على وزن الجنين بالغرام عند الولادة، مع المطابقة على عوامل الوزن السائدة الأخرى. ولقد وجدت تأثيرًا إيجابيًا ليس مهمًا من الناحية الإحصائية.

الأوامر:

```
webuse cattaneo2.dta
teffects psmatch (bweight) (prenatal1 mbsmoke mmarried mage fbaby medu alcohol)
```

النتائج:

Treatment-effects estimation		Number of obs	= 4642
Estimator	: propensity-score matching	Matches: requested	= 1
Outcome model	: matching	min	= 1
Treatment model	: logit	max	= 68

bweight	AI Robust Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
prenatal1						
(Yes vs No)	43.97727	30.37611	1.45	0.148	-15.5588	103.5133

٣-٥ ترجيح الاحتمال العكسي

المطابقة ليست هي التقنية الوحيدة التي تستفيد من درجة الميل. ويمكن أيضًا استخدام مقاربة "ترجيح الاحتمال العكسي". وبموجب هذه المقاربة، تتم موازنة كل ملاحظة في الانحدار بمعكوس احتمال المشاركة، وهذا لإنشاء فئة وهمية تكون الصفات فيها متشابهة بين الفئات المُعَالجة وغير المُعَالجة.

$$\hat{\Delta}_{IPW} = n^{-1}\sum_{i=1}^{n}\left[\frac{W_i Y_i}{\hat{p}_i}\right] - n^{-1}\sum_{i=1}^{n}\left[\frac{(1-W_i)Y_i}{1-\hat{p}_i}\right]$$

مُقدر ترجيح الاحتمال العكسي، $\hat{\Delta}_{IPW}$، يمثل متوسط تأثير المعالجة (ATE).

تطبيق ترجيح الاحتمال العكسي في برنامج STATA

يمكن تطبيق ترجيح الاحتمال العكسي في حزمة برنامج ستاتا (STATA 13 وSTATA 14) باستخدام الأمر "teffects ipw" متبوعًا بمتغير الحصيلة، ومتغير المعالجة، والمتغيرات المستقلة التي تُحدد أو تختلف بين المجموعة المُعَالجة وغير المُعَالجة قبل المعالجة.

هناك مثال مشابه لمطابقة درجة الميل التي يمكن تطبيقها على نفس مجموعات البيانات النموذجية في برنامج STATA. يقيم المثال التالي تأثير الرعاية قبل الولادة في مرحلة الشهور الثلاثة الأولى من الحمل على وزن الجنين بالغرام عند الولادة، مع المطابقة على عوامل الوزن السائدة الأخرى. بخلاف مطابقة درجة الميل، كان التأثير هنا مهمًا عند مستوى ٥٪.

الأوامر:

```
webuse cattaneo2.dta
teffects psmatch (bweight) (prenatal1 mbsmoke mmarried mage fbaby medu alcohol)
```

النتائج:

Treatment-effects estimation — Number of obs = 4642

Estimator : inverse-probability weights

Outcome model : weighted mean

Treatment model : logit

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE prenatal1 (Yes vs No)	53.51234	26.99469	1.98	0.047	.6037186	106.421
POmean prenatal1 No	3317.837	25.29863	131.15	0.000	3268.252	3367.421

٤-٥ عوامل التقدير القوية المزدوجة

يمكن أيضًا الاستفادة من عوامل التقدير القوية المزدوجة (DR)، حيث تجمع هذه التقنيات ما بين دقة المقاربات المعلمية وقدرة تقنيات درجة الميل على عزل المشاركة من العوامل المشتركة الأخرى. تم تقديم مفهوم عوامل التقدير القوية المزدوجة من قبل روبينز وآخرون (Robins et al. 1994)، واستفاض في الشرح كل من لونسيفورد (Lunceford) ودافيديان (Davidian ,2004). بيد أن السمة الفريدة في عوامل التقدير القوية المزدوجة أنها تحتاج

فقط إما إلى نموذج درجة الميل، أو نموذج الانحدار المُحدد تحديدًا صحيحًا لتقدير تأثير المشاركة على الحصيلة تقديراً صحيحًا.

الخطوة الثانية بعد تقدير درجة الميل، يتم تعديل النماذج بما يناسب الحصيلة على المتغيرات المشتركة بالنسبة إلى المجموعة المُعَالجة ($W_i = 1$) والمجموعة غير المُعَالجة ($W_i = 0$) كل على حدة. ثم يتم الحصول على قيمة Y المتوقعة باستخدام كلا مجموعتي المعاملات مع المتغيرات المشتركة المُلاحظة، المُشار إليها بـ $m_1(X_i,\hat{\beta}_1)$ و$m_0(X_i,\hat{\beta}_0)$.

$$\text{postulated } E(Y_i|\, W_i = 1, X_i, \hat{\beta}_i) = \hat{Y}_1 = m_1(X_i, \hat{\beta}_1) = \alpha_1 + \beta_1 X_1$$
$$\text{postulated } E(Y_i|\, W_i = 0, X_i, \hat{\beta}_0) = \hat{Y}_0 = m_0(X_i, \hat{\beta}_0) = \alpha_0 + \beta_0 X_0$$

يتحكم المُقدّر في الإدماج بأخذ معكوس $\hat{p}_i$ أو الاحتمال المتوقع بأن الوحدة تشارك في المُعالجة، أي $1/\hat{p}_i$ للوحدة المُعَالجة و للوحدة غير المُعَالجة لموازنة البيانات المُلاحظة، والقيم المتوقعة من الانحدار التبديلي. تنشأ عن هذه الموازنة حسب هذه الكمية فئة وهمية تكون توزيعات المُدمجات بين المجموعات المُعَالجة وغير المُعَالجة فيها نفسها مثل التوزيع الإجمالي لجميع المُدمجات في المجتمع الإحصائي الإجمالي الأصلي.

تُطبق النتائج الناتجة المتوقعة والحقيقية من Y_1، Y_0، وp بعدئذٍ في مُقدر القوة المزدوجة المُعطاة في المعادلة التالية:

$$\hat{\Delta}_{DR} = n^{-1}\sum_{i=1}^{n}\left[\frac{W_i Y_i}{\hat{p}_i} - \frac{\{W_i - \hat{p}_i\}}{\hat{p}_i} m_1(X_i, \beta_1)\right] - n^{-1}\sum_{i=1}^{n}\left[\frac{(1 - W_i)Y_i}{1 - \hat{p}_i} + \frac{\{W_i - \hat{p}_i\}}{1 - \hat{p}_i} m_0(X_i, \beta_0)\right]$$
$$= \hat{\mu}_1, DR - \hat{\mu}_0, DR$$

عامل التقدير القوي المزدوج، $\hat{\Delta}_{DR}$ ، يمثل متوسط تأثير المعالجة (ATE).

عند تطبيق هذه التقنية، يُضاف عادةً في معادلة الاختيار جميع المتغيرات التي تؤثر في الحصائل التي تختلف في المستويات قبل المعالجة (ولكن لا تتأثر في المستويات بموجب المعالجة). ويمكن إضافة مجموعة من المتغيرات المُؤثرة الإضافية لتحسين مستوى الدقة.

تطبيق عوامل التقدير القوية المزدوجة في STATA

يمكن تطبيق ترجيح الاحتمال العكسي المتصاعد في برنامج STATA باستخدام الأمر "teffects aipw"، متبوعًا بمتغير الحصيلة، والمتغيرات المستقلة المُهيئة للحصائل، ومتغير المعالجة، والمتغيرات المستقلة التي تحدد المعالجة.

هناك مثال مشابه لمطابقة درجة الميل وترجيح الاحتمال العكسي يمكن تطبيقه على نفس مجموعات البيانات النموذجية في برنامج STATA. يقيم المثال التالي تأثير الرعاية قبل الولادة في مرحلة الشهور الثلاثة الأولى من الحمل على وزن الجنين بالغرام عند الولادة، مع المطابقة على عوامل الوزن السائدة الأخرى. لاحظ أن التأثير هنا مهم عند مستوى ١٠٪.

الأوامر:

```
webuse cattaneo2.dta
teffects aipw (bweight mbsmoke mmarried mage fbaby alcohol) (prenatal1 mbsmoke mmarried mage fbaby alcohol medu), aequations vce(robust)
```

النتائج:

```
Iteration 0 :  EE criterion = 4.898e-20
Iteration 1 :  EE criterion = 5.930e-26

Treatment-effects estimation                    Number of obs    =       4642
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
prenatal1 (Yes vs No)	51.17366	26.54113	1.93	0.054	−.8460014	103.1933
POmean						
prenatal1 (No)	3321.505	24.7926	133.97	0.000	3272.912	3370.097
OME0						
mbsmoke	−169.8073	39.23131	−4.33	0.000	−246.6993	−92.91539
mmarried	156.5209	43.76734	3.58	0.000	70.73846	242.3033
mage	.6265175	4.466949	0.14	0.888	−8.128541	9.381576
fbaby	−61.49277	42.76311	−1.44	0.150	−145.3069	22.32138
alcohol	−143.5852	80.82422	−1.78	0.076	−301.9978	14.82732
_cons	3249.977	108.006	30.09	0.000	3038.289	3461.665
OME1						
mbsmoke	−242.8674	27.31882	−8.89	0.000	−296.4113	−189.3236
mmarried	150.3321	26.14595	5.75	0.000	99.08699	201.5772
mage	1.205585	2.052459	0.59	0.557	−2.817161	5.22833
fbaby	−49.61927	19.38907	−2.56	0.010	−87.62115	−11.61739
alcohol	−14.2429	55.9502	−0.25	0.799	−123.9033	95.41748
_cons	3303.01	57.82122	57.12	0.000	3189.683	3416.338

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
TME1		.0965382				
mbsmoke	−.1866845	.0956205	−1.93	0.053	−.3758958	.0025269
mmarried	1.099739	.0096489	11.50	0.000	.9123265	1.287152
mage	.0697882	.0901322	7.23	0.000	.0508767	.0886998
fbaby	.6101609	.1912115	6.77	0.000	.433505	.7868168
alcohol	−.6779378	.020514	−3.55	0.000	−1.052705	−.3031703
medu	.1493282	.2875023	7.28	0.000	.1091215	.189535
_cons	−3.043925		−10.59	0.000	−3.607419	−2.480431

القسم السادس: المقاربات المستندة إلى المتغيرات المساعدة (مربعات صغرى ثنائية المرحلة، وانحدارات المعالجة داخلية النشأة، والانحدارات التبديلية داخلية النشأة)

١-٦ المفاهيم والافتراضات الضمنية في تحليل المتغيرات المساعدة

هناك طريقة أخرى لتقدير آثار البرنامج عند عدم فعالية افتراض التعيين غير المدمج وهي استخدام تحليل المتغيرات المساعدة. يمكن توضيح تحليل المتغيرات المساعدة بأفضل طريقة ممكنة باستخدام إطار الانحدار الخطي. افترض أن ما يلي هي معادلة الانحدار الخطي:

$$Y_i = \alpha + \beta_1 W_i + \gamma X_i + \varepsilon_i$$

في هذه المعادلة يرتبط تعيين المعالجة W_i بصفات غير ملحوظة للباحث ε_i، حيث إن $\text{cov}(W,\varepsilon) \neq 0$. هناك طريقة أخرى لتوضيح ذلك تكمن في أن هناك شيئًا في حد الخطأ (ε_i) في المعادلة التي لا يؤثر فقط في Y_i بل يؤثر أيضًا في W_i. بالإضافة إلى ذلك، حيث إن ε_i غير ملحوظة، فلا يمكن التحكم فيها ببساطة من خلال تضمين متغير ومن ثم اختبار الأثر غير المتحيز لـ W_i على Y_i.[٥]

الهدف من وراء تحليل المتغيرات المساعدة هو تحديد مكان و"اجتزاء" جزء W_i الذي يرتبط بـ ε_i واستخدام هذا الجزء الخارجي فقط في تقدير تأثير المعالجة. ولتتمكن من فعل ذلك، ينبغي أن يكون هناك متغير من نوع خاص يُسمى *المتغير المساعد* (أو *المساعد* اختصارًا).

[٥] هذه صياغة أخرى لمشكلة "تحيز الاختيار" (تمت مناقشتها في القسم ١-١) أو "التداخلية"، أو "تحيز المتغير المحذوف".

يجب أن يستوفي المتغير المساعد العملي (Z_i) شرطين. أولاً، يجب أن يكون مرتبطًا جزئيًا وبصورة كافية بالمعالجة (W_i)، بعد التحكم في Z_i.[٦] ثانيًا، يجب ألا يتأثر المتغير المساعدة إلا بالحصيلة (Y_i) من خلال المعالجة (W_i) وليس من خلال أي وسيلة أخرى، كأن يكون حد الخطأ الخاص بالحصيلة غير مرتبط بالمتغير المساعد. بمعنى آخر، بعيدًا عن التعامل مع W_i، ينبغي ألا يكون لـ Z_i أثر مستقل على Y_i. وحيثما يجد المُقيم متغيرًا مساعدًا واحدًا فقط خاصًا بمتغير المعالجة، فسيكون افتراض أن المتغير المساعد لا يرتبط بحد الخطأ محض اعتقاد وليست هناك وسيلة لاختباره والتحقق من صحته باستخدام البيانات. وحيثما يتمكن المُقيم من العثور على متغيرين مساعدين أو أكثر لمتغير معالجة واحد، فربما يتمكن المُقيم من إجراء اختبارات إحصائية إضافية للتأكد مما إذا كانت المتغيرات المساعدة صالحة أو لا (تُسمى هذه الاختبارات "اختبارات قيود التمييز الزائد").

وعندما يجد المُقيم متغيرًا (متغيرات) مساعدًا يحتمل أن يكون صالحًا، يجب أن يتمكن من توفير الافتراضات المنطقية لسبب عدم ارتباط المتغير المساعد بحد الخطأ. وبصورة أساسية، يجب أن يدافع المُقيم عن تفسيره لسبب أن المتغير المساعد يؤثر في الحصيلة من خلال متغير المعالجة فقط، وليس من خلال أي وسيلة أخرى. من الضروري إجراء الاختبارات المرتبطة بتحليل المتغيرات المساعدة (يشمل ذلك على سبيل المثال، اختبارات ما إذا كان المتغير المساعد يرتبط ارتباطًا جزئيًا وكافيًا بالمعالجة أو لا) لدعم هذه الافتراضات أيضًا.

٦-٢ انحدار المربعات الصغرى ثنائية المرحلة

إن أشهر طريقة لتقدير الآثار باستخدام المتغيرات المساعدة هي إجراء انحدار *المربعات الصغرى ثنائية المرحلة* (2SLS). ويمكن تمثيلها في نماذج الانحدار الخطي التالية:

$$\hat{W}_i = \beta_0 + \beta_3 Z_i + \beta_2 X_i + \eta_i$$
$$Y_i = \alpha + \beta_1 \hat{W}_i + \gamma X_i + \varepsilon_i$$

مثلما يتضح من اسم انحدار المربعات الصغرى ثنائية المرحلة، يتم الانحدار على مرحلتين. في المرحلة الأولى، ينحدر المُقيم بمتغير المعالجة على المتغير المساعد Z_i مع التحكم في X_is. بعد إجراء هذا الانحدار، يحسب المُقيم القيمة المتوقعة لـ W_i (يُسمى أحيانًا W_i -hat). في المرحلة الثانية، يقوم المُقيم بإجراء انحدار متغير الحصيلة Y_i على W-hat، بينما لا يزال يتحكم في X_is. المعامل المرتبط بـ W-hat من مرحلة الانحدار الثانية هو تقدير المتغيرات المساعدة لأثر المعالجة W_i على Y_i.

[٦] يمكن أن يتحقق المُقيم من صحة هذا بإجراء انحدار W_i على المتغير (المتغيرات) المساعد(ة) وX_i، وتحديد ما إذا كانت إحصاءات F المقدرة والمرتبطة بالمتغير (المتغيرات) المساعد(ة) كبيرة أم لا (عادةً ما تكون أكبر من ١٠).

تجدر ملاحظة أن مُقدر المتغيرات المساعدة لا يمثل متوسط تأثير المعالجة، أو متوسط تأثير المعالجة على المُعالَج، بل يقدر ما يُسمى *متوسط تأثير المعالجة المحلي* (LATE). وهذا هو متوسط تأثير المعالجة على أولئك المشاركين *الناتج عن المتغير المساعد* لتحديد الحصول على المعالجة أم لا.

تطبيق انحدار المربعات الصغرى ثنائية المرحلة في برنامج STATA

يتوفر برنامج STATA على الأمر "ivregress"، مما يسمح بعدد من انحدارات المتغيرات المساعدة. وهناك أيضًا خطوات روتينية يتبعها المستخدم، والتي تقدم التشخيصات التلقائية لنتائج الانحدار، والتي يمكن أن تفيد في تأكيد صلاحية المتغير المساعد (Baum et al. 2010). وسيكون تركيب الأخير هو "ivreg2"، متبوعًا بمتغير الحصيلة، وحصائل تكييف المتغيرات المستقلة، والمتغيرات المساعدة.

يستخدم المثال التالي بيانات سمات توظيف المرأة، ويقيم أثر الاتحادات على الأجور في عام ١٩٧٢ في الولايات المتحدة. وفي هذا المثال، يُمثل المتغير المساعد لعضوية الاتحاد بما تتميز به المرأة من انتماء عرقي (عرق الأسود أو عرق آخر)، ومحل إقامة (عيش في الجنوب) وسن. وكان الاستنتاج هو وجود تأثير إيجابي مهم لعضوية الاتحاد على الأجور. توضح تشخيصات المتغيرات المساعدة أن هذه المتغيرات صالحة، حيث إن القيمة p لاختبارات كل من التمييز الزائد والمنخفض تبين الأهمية والدلالة، وأن إحصائية اختبار التمييز الضعيف مقبولة.

الأوامر:

```
ssc install ranktest
ssc install ivreg2
webuse union3
ivreg2 wage wks_work ttl_exp not_smsa grade (union= black south age), robust
```

النتائج:

```
IV (2SLS) estimation
----------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                Number of obs   =      1203
                                                F(5, 1197)      =     91.92
                                                Prob > F        =    0.0000
Total (centered) SS     = 6529.780859           Centered R2     =    0.2639
Total (uncentered) SS   = 45121.78109           Uncentered R2   =    0.8935
Residual SS             = 4806.687808           Root MSE        =     1.999
```

wage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
union	228736	.7924016	2.81	0.005	.675657	3.781814
wks_work	.017831	.0048998	3.64	0.000	.0082275	.0274344
ttl_exp	.222175	.0452564	4.91	0.000	.1334742	.3108759
not_smsa	−.8238825	.1372525	−6.00	0.000	−1.092892	−.5548726
grade	.499475	.0456322	10.95	0.000	.4100376	.5889125
_cons	−2.298186	.5495487	−4.18	0.000	−3.375282	−1.221091

Underidentification test (Kleibergen-Paap rk LM statistic):		34.401
	Chi-sq(3) P-val =	0.0000
Weak identification test (Cragg-Donald Wald F statistic):		12.826
	(Kleibergen-Paap rk Wald F statistic):	12.721
Stock-Yogo weak ID test critical values:	5% maximal IV relative bias	13.91
	10% maximal IV relative bias	9.08
	20% maximal IV relative bias	6.46
	30% maximal IV relative bias	5.39
	10% maximal IV size	22.30
	15% maximal IV size	12.83
	20% maximal IV size	9.54
	25% maximal IV size	7.80

المصدر: Stock-Yogo (2005). تمت إعادة النسخ بموجب تصريح.
لاحظ: القيم الحرجة خاصة بـ Cragg-Donald F الإحصائية، وأخطاء i.i.d.

Hansen J statistic (overidentification test of all instruments):		61.165
	Chi-sq(2) P-val =	0.0000
Instrumented :	union	
Included instruments :	wks_work ttl_exp not_smsa grade	
Excluded instruments :	black south age	

٦-٣ مقاربات وظيفة التحكم في الاختيار داخلي النشأة في المعالجة

أحد الأشكال المختلفة للمتغيرات المساعدة التي كان هيكمان (Heckman 1979) أول من قدمها هو نمذجة عملية الاختيار للمعالجة بوصفها تنطوي على مكونات توضحها المتغيرات الملحوظة، وعوامل مشتملة على مكونات عن الحد المتبقي والتي تتم معالجتها بوصفها متغيرًا إضافيًا. ويتم هذا في صورة انحدار لوغاريتمي يتوقع أن يُنتج حدودًا إضافية للفئات المُعَالجة وغير المُعَالجة، والتي تُستخدم كعناصر انحدار في انحدار المرحلة الثانية.

ويجب أن يتضمن الانحدار اللوغاريتمي بعض المتغيرات المستقلة التي هي متغيرات مساعدة ترتبط بالمعالجة وليس بالحصائل، ولكنها تضم أيضًا متغيرات الحصيلة.

$$probitP(W=1|Z)=\beta_1 Z_i + \eta_i = \hat{p}_i$$

يتم تصنيف النظم وفقًا لحالة المشاركة الملحوظة حسب ما يلي.

$$Y_i = \begin{cases} Y_{1i} \text{ if } W_i > 0 \\ Y_{0i} \text{ if } W_i = 0 \end{cases}$$

يتم وضع إطار نظري لنموذج الحصيلة العام حسب الموضح أدناه، قبل مراعاة تحيزات الاختيار.

$$Y_{1i} = \alpha_1 + \gamma_1 Xi + \varepsilon_{1i}$$
$$Y_{0i} = \alpha_0 + \gamma_0 Xi + \varepsilon_{2i}$$

ترصد مقاربة وظيفة التحكم تأثير تحيز الاختيار عبر حد الانحدار الإضافي. وللحصول على هذا الحد، يُفترض أن حد الخطأ في الانحدار اللوغاريتمي وحد الخطأ في معادلات الحصيلة يتم توزيعهما معًا توزيعًا طبيعيًا باستخدام متوسط صفري وهيكل المتغيرات (المتغيرات المشتركة) كما يلي:

$$\text{cov}(\varepsilon_{1i}, \varepsilon_{2i}, \eta_i) = \begin{bmatrix} \sigma^2_{\varepsilon 0} & . & \sigma_{\varepsilon 0 \eta} \\ . & \sigma^2_{\varepsilon 1} & \sigma_{\varepsilon 1 \eta} \\ . & . & \sigma^2_{\eta} \end{bmatrix}$$

هناك معاملان إضافيان يُسميان نسبة معكوس الإخفاق، ومعدل مخاطر الاختيار، ويتم تقديرهما استنادًا إلى الكثافة العادية القياسية $\phi(.)$ ، والتوزيع التراكمي الطبيعي القياسي $\Phi(.)$ للاستخدام مع الفئات المُعَالجة (λ_{1i}) وغير المُعَالجة (λ_{1i}):

$$\lambda_{1i} = \frac{\phi(\beta Z_i)}{\Phi(\beta Z_i)}$$
$$\lambda_{0i} = -\frac{\phi(\beta Z_i)}{1 - \Phi(\beta Z_i)}$$

يتم التطبيق الأمثل لهاتين النسبتين في تخمين تأثير المعالجة عبر نماذج تأثيرات المعالجة داخلية النشأة، والتي تستخدم المعاملات نفسها في نوعي الحصائل. وعندما يحدث هذا، يمكن أن يتوقع النموذج متوسط تأثير المعالجة باعتباره معامل W بعد امتصاص تحيز الاختيار في نسبة معكوس الاخفاق (للمعالجين)، ونسبة المخاطر (لغير المعالجين) في المعادلة.

$$Y_i = \gamma X_i + \beta_1 W_i + \beta_2 W_i \lambda_{1i} + \beta_3 (1 - W_i) \lambda_{0i} + \alpha$$

تطبيق انحدار تأثيرات المعالجة داخلية النشأة في برنامج STATA

يمكن تطبيق تقنية الانحدار هذه في برنامجي STATA 13 وSTATA 14 باستخدام الأمر "etregress"، متبوعًا بمتغير الحصيلة، وحصائل تكييف المتغيرات، و"المعالجة" (متغير المعالجة = معالجة تكييف المتغيرات المستقلة).

يستخدم المثال التالي البيانات نفسها عن سمات توظيف المرأة مثل مثال 2SLS ويُقيم أثر الاتحادات على الأجور. وفي هذا المثال، تم تكييف عضوية الاتحاد بما تتميز به المرأة من انتماء عرقي (عرق الأسود أو عرق آخر)، ومحل إقامة (عيش في الجنوب) وسن. وكان الاستنتاج هو وجود تأثير إيجابي مهم لعضوية الاتحاد على الأجور.

الأوامر:

```
webuse union3
etregress wage wks_work ttl_exp not_smsa grade, treat(union=black south age) vce(robust)
```

النتائج:

Linear regression with endogenous treatment — Number of obs = 1203
Estimator: maximum likelihood — Wald chi2(5) = 577.00
Log pseudolikelihood = –3093.8702 — Prob > chi2 = 0.0000

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage						
wks_work	.0189489	.0045733	4.14	0.000	.0099853	.0279125
ttl_exp	.2072293	.044969	4.61	0.000	.1190916	.295367
not_smsa	–.8011546	.1204989	–6.65	0.000	–1.037328	–.5649812
grade	.4607586	.0430331	10.71	0.000	.3764153	.5451018
union	3.330139	.5341358	6.23	0.000	2.283252	4.377025
_cons	–2.061778	.4773057	–4.32	0.000	–2.99728	–1.126276
union						
black	.2237051	.1046224	2.14	0.032	.0186489	.4287613
south	–.5459861	.0878814	–6.21	0.000	–.7182305	–.3737417
age	.045354	.0116589	3.89	0.000	.022503	.068205
_cons	–1.680473	.2687548	–6.25	0.000	–2.207223	–1.153723
/athrho	–.77179	.2231879	–3.46	0.001	–1.20923	–.3343497
/lnsigma	.7684901	.080225	9.58	0.000	.611252	.9257283
rho	–.6479691	.1294794			–.8364484	–.3224237
sigma	2.156508	.1730058			1.842737	2.523705
lambda	–1.39735	.369339			–2.121241	–.6734593

Wald test of indep. eqns. (rho = 0): chi2(1) = 25.47 Prob > chi2 = 0.0000

٤-٦ مقاربات الانحدار التبديلي داخلي النشأة

ينتج عن تبني مقاربة هيكمان Heckman الحصول على متوسط تأثير المعالجة γ فقط، حيث إن معاملات متغيرات الحصيلة في معادلة الحصيلة ثابتة في الفئات المُعَالجة وغير المُعَالجة. ولكن الانحدار التبديلي داخلي النشأة يخفف من حدة هذا الافتراض من خلال السماح بنظامين للحصول على معاملات متباينة في حصائل تكييف المتغيرات. وعندما تتوفر معاملات متباينة γ، يمكن أن يتباين تأثير المعالجة بين الفئات المُختارة للمعالجة، والفئات غير المُختارة (Maddala and Nelson 1975). ويعني هذا أن متوسط تأثير المعالجة على المُعالَج يمكن تمييزه بمعزل عن متوسط تأثير المعالجة على غير المُعالَج.

على وجه أكثر تحديدًا، كل من نسبة الاخفاق المعكوس ونسبة المخاطر المُقدرة مثلما هو الحال في انحدار تأثيرات المعالجة داخلية النشأة، جنبًا إلى جنب مع المعاملات من معادلات الحصيلة، وقيم المتغيرات المستقلة، تنتج توقعات الحصائل بالنسبة إلى (١) المُعالَجين والمشاركين في المعالجة؛ و(٢) غير المُعالَجين وغير المشاركين في المعالجة؛ و(٣) المشاركين في المعالجة إذا لم تتم معالجتهم؛ و(٤) غير المشاركين في المعالجة إذا تمت معالجتهم.

$$
\begin{aligned}
1&: E(Y_{1i}|W_i = 1) &&= \gamma_1 X_{1i} + \sigma_{\varepsilon 1\eta}\lambda_{1i} \\
2&: E(Y_{0i}|W_i = 0) &&= \gamma_0 X_{0i} + \sigma_{\varepsilon 0\eta}\lambda_{0i} \\
3&: E(Y_{0i}|W_i = 1) &&= \gamma_0 X_{1i} + \sigma_{\varepsilon 0\eta}\lambda_{1i} \\
4&: E(Y_{1i}|W_i = 0) &&= \gamma_1 X_{0i} + \sigma_{\varepsilon 1\eta}\lambda_{0i}
\end{aligned}
$$

الفروق بين التوقعات التكيفية هي ما تُحدد تأثير المعالجة. على سبيل المثال، متوسط الفرق بين ١ و٣ هو متوسط تأثير المعالجة على المُعالَج (ATT)، وبين ٢ و٤ هو متوسط تأثير المعالجة على غير المُعالَج (ATU). والمتوسط الموزون، لمتوسط تأثير المعالجة على غير المُعالَج (ATU) ومتوسط تأثير المعالجة على المُعالَج (ATT)، هو متوسط تأثير المعالجة (ATE).

تطبيق الانحدار التبديلي داخلي النشأة في برنامج STATA

يُحيل كل من لوكشين وسجايا إلى إمكانية تطبيق تقنية الانحدار هذه في برنامج STATA باستخدام الحزمة التي يكتبها المستخدم "movestay" (Lokshin and Sajaia 2004). تُقدَّم الصيغة كما يلي: "movestay" (متغير الحصيلة، المتغيرات المستقلة للحصيلة)، "select" (متغير المعالجة = المتغيرات المستقلة للمعالجة).

يستخدم المثال التالي البيانات نفسها عن سمات توظيف المرأة مثل تأثيرات المعالجة داخلية النشأة، ويُقيم أثر الاتحادات على الأجور باستخدام المواصفات نفسها. ولقد وجد أيضًا تأثيرًا إيجابيًا كبيرًا لعضوية الاتحاد على الأجور، مثل الفرق في متوسط التوقعات التكيفية بالنسبة للمشاركين في الاتحاد وغير المشاركين فيه.

الأوامر:

```
ssc install movestay
webuse union3
movestay (wage wks_work ttl_exp not_smsa grade), select(union=black south age) robust
mspredict yc0, yc0
mspredict yc1, yc1
egen ate=mean(yc1-yc0)
display ate
```

النتائج:

Endogenous switching regression model — Number of obs = 1203
Wald chi2(4) = 267.27
Log pseudolikelihood = –3044.2858 — Prob > chi2 = 0.0000

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage0						
wks_work	.0213619	.0046435	4.60	0.000	.0122609	.030463
ttl_exp	.152945	.0465026	3.29	0.001	.0618016	.2440885
not_smsa	.7722602	.1197559	–6.45	0.000	–1.006977	–.537543
grade	.4186368	.0421965	9.92	0.000	.3359331	.5013405
_cons	–1.711028	.5054607	–3.39	0.001	–2.701713	–.7203435
wage1						
wks_work	.0212183	.0111049	1.91	0.056	–.000547	.0429835
ttl_exp	.1937367	.0956872	2.02	0.043	.0061931	.3812802
not_smsa	–.7446781	.2691096	–2.77	0.006	–1.272123	–.217233
grade	.4200887	.0691625	6.07	0.000	.2845327	.5556448
_cons	1.022297	.8620225	1.19	0.236	–.6672363	2.71183
select						
black	.2165804	.0771031	2.81	0.005	.0654612	.3676996
south	–.4740664	.0761678	–6.22	0.000	–.6233525	–.3247803
age	.030676	.0116606	2.63	0.009	.0078217	.0535303
_cons	–1.333329	.269485	–4.95	0.000	–1.86151	–.8051476
/lns0	.8242716	.0690175	11.94	0.000	.6889998	.9595435
/lns1	.8213944	.1002633	8.19	0.000	.624882	1.017907
/r0	–1.659308	.1414636	–11.73	0.000	–1.936572	–1.382044
/r1	–.413224	.1111876	–3.72	0.000	–.6311477	–.1953003
sigma0	2.280219	.1573751			1.991722	2.610504
sigma1	2.273668	.2279654			1.868026	2.767396
rho0	–.9301239	.0190791			–.9592613	–.8814082
rho1	–.3912067	.0941712			–.558842	–.1928546

Wald test of indep. eqns. : chi2 (2) = 158.64 Prob > chi2 = 0.0000

```
. display ate
2.8555474
```

القسم السابع: تصميم انقطاع الانحدار

١-٧ المفاهيم الأساسية في تصميم انقطاع الانحدار

تصميم انقطاع الانحدار (RDD) هو في الحقيقة حالة خاصة من المتغيرات المساعدة، حيث يمكن تطبيقه عندما يوجد نوع دقيق جدًا من الأدوات المساعدة. تستفيد هذه الطريقة من حقيقة أنه في بعض الحالات، أحيانًا ما تتغير احتمالية الحصول على المعالجة تغيرًا مفاجئًا أو متقطعًا عند نقطة الانقطاع لمتغير مُعين، وأن هذا يرتبط غالبًا بأهلية البرنامج (يُسمى غالبًا *متغير الإجراء، أو متغير الفرض، أو متغير التعيين* . وتجدر الإشارة إلى أن النص يستخدم، بعدئذ، مصطلح *متغير التعيين تسهيلاً على القراء). وفي هذا الصدد، هناك حالات تنطبق عليها* قاعدة الانقطاع كحينما يكون هناك شخص كانت قيمة متغير التعيين الخاصة به أكبر من "X_0" (نقطة الانقطاع)، فستكون فرصته في المعالجة ($W_i = 1$) أكبر مما إذا كانت قيمته أقل من X_0. بالإضافة إلى وجود قاعدة الانقطاع، فعند استيفاء الشروط الثلاثة الأخرى التالية، فسيجعل تصميم انقطاع الانحدار تقدير تأثير المعالجة ممكنًا من خلال المقارنة بين حصائل الأفراد فوق نقطة الانقطاع مباشرة وأسفلها مباشرة:

الشرط الأول: يتوقع متغير التعيين المعالجة، مع حالة واضحة من الانقطاع عند عتبة الأهلية أو الانقطاع. بمعنى آخر، ينبغي أن تتغير احتمالية المعالجة تغيرًا مفاجئًا عند نقطة انقطاع متغير التعيين. ويمكن اختبار صحة هذا أو خطأه من خلال تقدير الاتجاه أو رسمه في العلاقة بين متغير الحصيلة ومتغير التعيين على جانبي نقطة الانقطاع، وملاحظة ما إذا كان هناك "قفزة" عند نقطة الانقطاع أم لا. وإذا لم يكن هناك انقطاع واضح مثل هذا، فلن يكون تصميم انقطاع الانحدار مناسبًا.

الشرط الثاني: من بين المتغيرات الخارجية، ينبغي أن يكون فقط للمعالجة محل الاهتمام انقطاع عند نقطة الانقطاع. ويُقصد بذلك أنه ينبغي ألا يكون هناك سبب آخر لوجود قفزة عند نقطة الانقطاع باستثناء الاختلافات في التعيين للمعالجة محل البحث. وينبغي أن يكون هذا واضحًا في عدم حدوث انقطاع في المتغيرات المشتركة قبل المعالجة (أو الحصائل المحتملة) عند نقطة الانقطاع. ويمكن اختبار صحة هذا أو خطأه باستخدام البيانات. وبصورة أساسية يمكن أن يقوم المُقيم بعمل انحدار في اتجاه العلاقة بين المتغيرات المستقلة الأخرى في مقابل متغير التعيين عند النقطة فوق أو أسفل نقطة الانقطاع مباشرةً لتحديد ما إذا كانت هناك "قفزات" عند نقطة الانقطاع أم لا. وإذا كانت هناك قفزات، فيعني هذا أن تصميم انقطاع الانحدار قد لا يكون مناسبًا. وقد يحدث هذا إذا كانت البرامج الأخرى تستخدم معايير الأهلية نفسها المُتبعة في المعالجة محل الاهتمام.

الشرط الثالث: لا يمكن أن يتلاعب الأشخاص بقيم متغير التعيين ليتم اختيارهم في المعالجة. قد يحدث أن يجد بعض الأشخاص طريقة للتلاعب بمتغير التعيين للحصول على المعالجة، أو تجنبها. ورغم أنه لا توجد طريقة مُثبتة لتحديد ما إذا كان هذا قد حدث أم لا، يمكن أن يقوم المُقيم بإجراء "اختبار McCrary" (McCrary 2008). وهذا اختبار إحصائي يقيم ما إذا كانت هناك قفزة أو انقطاع في كثافة الملاحظات عند نقطة الانقطاع. وإذا كانت هناك قفزة، فقد يعني هذا أن بعض الأشخاص قد وجدوا طريقة للتلاعب بموضع وجودهم بالنسبة إلى نقطة الانقطاع. في هذه الحالة، قد لا يكون تصميم انقطاع الانحدار مناسبًا.

٢-٧ تصميم انقطاع الانحدار الحاد

في تصميم انقطاع الانحدار الحاد، تتغير احتمالية الحصول على المعالجة بشكل متقطع من 0 (لا توجد فرصة للحصول على المعالجة) إلى 1 (فرصة الحصول على المعالجة ١٠٠٪) عند نقطة الانقطاع الخاصة بمتغير التعيين X_0. بمعنى آخر، هناك معيار أهلية صارم مطبق جيدًا في البرنامج. يتم تمثيل تصميم انقطاع الانحدار الحاد باستخدام المعادلات التالية:

$$Y_{0i} = \alpha + f(X_o + \delta_i) + \varepsilon_{0i}$$
$$Y_{1i} = \alpha + \tau_i + f(X_0 - \delta_i) + \varepsilon_{1i}$$

حيثما تكون Y_0 هي حصيلة الموجود فوق عتبة الأهلية، وتكون Y_1 هي حصيلة الموجود ضمن عتبة الأهلية، وتكون$f(X_0)$ هي دالة متصلة حول عتبة الأهلية، تخمين τ هي تأثير المعالجة، وδ هي المسافة من عتبة الأهلية.

قد يبين الاختلاف في الأنظمة تأثير المعالجة كما يلي:

$$Y_1 - Y_0 = \tau + f(X_0 - \delta) - f(X_o + \delta) + (\varepsilon_1 - \varepsilon_0)$$

بسبب وجود بعض المتغيرات العشوائية في متغير التعيين (الشرط الأول)، وحيث لا يمكن أن يتلاعب الأشخاص بموضع وجودهم فوق أو أسفل نقطة الانقطاع لأهلية البرنامج (الشرط الثاني)، فإن تصميم انقطاع الانحدار يشبه تجربة عشوائية حول نقطة الانقطاع. يشبه الأشخاص الموجودون فوق نقطة الانقطاع مباشرةً الأشخاص الموجودين أدناها مباشرةً، باستثناء عندما يكون الأشخاص على أحد جانبي نقطة الانقطاع قد حصلوا أيضًا على المعالجة، بينما لم يحصل عليها الأشخاص الموجودون ناحية اليسار. ومن ثم فإن أي اختلافات في متوسط الحصائل بين الموجودين فوق نقطة الانقطاع وأدناها مباشرةً يمكن عزوها إلى المعالجة فقط.

ويسهل نسبيًا تقدير تأثير المعالجة باستخدام تصميم انقطاع الانحدار الحاد. وحيثما توجد نقاط ملاحظة كافية حول نقطة الانقطاع، يمكن تقدير تأثير المعالجة بإيجاد الفرق بين متوسط الحصيلة فوق نقطة الانقطاع وأسفلها مباشرة. وفي هذه الحالة، ليس من الضروري أن يفترض المُقيم أي نماذج عملية. وإذا كانت البيانات حول نقطة الانقطاع قليلة، يمكن أيضًا أن يقوم المُقيم بعمل نموذجين منفصلين لكل من (أ) الاتجاه في Y بطول متغير التعيين من أسفل نقطة الانقطاع صعودًا إلى نقطة الانقطاع، و(ب) الاتجاه في Y بطول متغير التعيين من فوق نقطة الانقطاع حتى نقطة الانقطاع. ويمكن عندئذٍ أن يأخذ المُقيم الفرق بين تقاطعات الخطين عند نقطة الانقطاع، وهذا هو تأثير المعالجة. ويمكن في هذه الحالة تطبيق مقاربات الانحدار المَعلمي، وغير المَعلمي، وشبه المَعلمي في تقدير تأثير المعالجة.

تتفوق مقاربة تصميم الانحدار الحاد بميزة مهمة عن أنواع التصاميم شبه التجريبية الأخرى، مثل المطابقة والاختلاف في الاختلافات. ونظرًا لكون معايير الأهلية معلومة، لم يعد الاختيار بناءً على الملحوظات مشكلة، كما يمكن التحكم في الملحوظات من خلال اختيار الفئات حول عتبة الأهلية. وفي الوقت نفسه، فما يقدّره تصميم انقطاع الانحدار هو متوسط تأثير المعالجة المحلي للفئات حول عتبة الأهلية، وليس متوسط تأثير المعالجة للفئة بكاملها.

٣-٧ تصميم انقطاع الانحدار المشوش

لا تُطبق في العديد من الحالات معايير الأهلية بدرجة عالية من الدقة، أو لا يصل تطبيق البرنامج إلى ١٠٠٪ من الفئة المؤهلة. وحيثما تكون احتمالية الحصول على المعالجة تتغير تغيرًا غير مستمر عند نقطة الانقطاع بمعدل أقل من ١ (أقل من ١٠٠٪)، يمكن استخدام تصميم انقطاع الانحدار المشوش.

لتقدير تأثير المعالجة عند تبني مقاربة تصميم انقطاع الانحدار المشوش، يجب أن يضع المُقيم نماذج منفصلة (أ) اتجاه Y بطول متغير التعيين من أسفل نقطة الانقطاع صعودًا إلى نقطة الانقطاع؛ و(ب) اتجاه Y بطول متغير التعيين من فوق نقطة الانقطاع حتى نقطة الانقطاع؛ و(ج) اتجاه W بطول متغير التعيين من أسفل نقطة الانقطاع صعودًا إلى نقطة الانقطاع؛ و(د) اتجاه W بطول متغير التعيين من أعلى نقطة الانقطاع حتى نقطة الانقطاع. يمكن أن يأخذ المُقيم الفرق بين تقاطعات الاتجاهين (أ) و(ب) عند نقطة الانقطاع (البسط في أداة تقدير تصميم انقطاع الانحدار المشوش) ويقسمه على الفرق بين تقاطعات الاتجاهين (ج) و(د) عند نقطة التقاطع (المَقام في أداة تقدير تصميم انقطاع الانحدار المشوش). ويمكن في هذه الحالة تطبيق مقاربات الانحدار المَعلمي، وغير المَعلمي، وشبه المَعلمي في تقدير تأثير المعالجة. وعلى الرغم من قيامهما على فكرة مختلفة قليلاً، فإن أداة تقدير تصميم انقطاع الانحدار المشوش هي نفسها مثل أداة تقدير المربعات الصغرى ثنائية المرحلة في تحليل المتغيرات المساعدة.

تطبيق تصميم انقطاع الانحدار في برنامج STATA

يحيل كالونيكو وآخرون إلى أن القيام بإصدار الأمر "Rdrobust" يُمكّن من أتمتة تصميم انقطاع الانحدار في برنامج STATA، كما يمكننا من حساب الخطأ المعياري القوي في اختلاف التباين (Calonico et al. 2014). ويُتبع التركيب "rdrobust" بمتغير الحصيلة ومتغير التعيين، مع خيار "مشوش" (متغير المشاركة) عند استخدام تصميم انقطاع الانحدار المشوش.

يستخدم مثال تصميم انقطاع الانحدار الحاد التالي (مأخوذ من Calonico, Cattaneo, and Titiunik 2014) البيانات الموجودة في حزمة "tsubordr" لتوضيح التطبيق في برنامج STATA. والمعالجة هي ما إذا كان عضو الحزب الديمقراطي سيفوز في انتخابات مجلس الشيوخ الأمريكي عن إحدى الولايات، والحصيلة هي الهامش في الانتخابات التالية. وقيمة الانقطاع لمتغير التعيين هي 0. وتبين النتائج أن تولي المنصب له تأثير إيجابي مهم على هوامش الفوز بالانتخابات القادمة.

الأوامر:

```
net get st0366.pkg
use rdrobust_rdsenate.dta
rdrobust vote margin, all
```

النتائج:

تقديرات انقطاع الانحدار الحاد باستخدام الانحدار متعدد الحدود المحلي.

Cutoff c = 0	Left of c	Right of c
Number of obs	343	310
Order loc. poly. (p)	1	1
Order bias (q)	2	2
BW loc. poly. (h)	16.794	16.794
BW bias (b)	27.437	27.437
rho (h/b)	0.612	0.612

Number of obs	=	1297
NN matches	=	3
BW type	=	CCT
Kernel type	=	Triangular

Outcome: vote. Running variable: margin.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	7.4253	1.4954	4.9656	0.000	4.49446	10.3561
Robust	–	–	4.2675	0.000	4.06975	10.9833

All estimates.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	7.4253	1.4954	4.9656	0.000	4.49446	10.3561
Bias-corrected	7.5265	1.4954	5.0333	0.000	4.59569	10.4574
Robust	7.5265	1.7637	4.2675	0.000	4.06975	10.9833

المراجع

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*. 105 (1). pp.495050 1–21.

Athey, S. and G. W. Imbens. 2017. Chapter 3. The Econometrics of Randomized Experiments. In E. Duflo and A. Banerjee, eds. *Handbook of Field Experiments*. Volume 1, 1st Edition. Amsterdam: North Holland.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2010. IVREG2: STATA Module for Extended Instrumental Variables/2SLS and GMM estimation. Statistical Software Components, S425401. Revised 9 February 2016. Boston College Department of Economics. http://ideas.repec.org/c/boc/bocode/
s425401.html.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*. 14 (4). pp. 909–946. https://sites.google.com/site/rdpackages/rdrobust/Calonico- Cattaneo-Titiunik_2014_Stata.pdf.

Heckman, J. J. 1979. Selection Bias as a Specification Error. *Econometrica*. 47 13). pp.15321619.

Lokshin, M. and Z. Sajaia. 2004. Maximum Likelihood Estimation of Endogenous Switching Regression Models. *The STATA Journal*. 4 (3). pp. 282–289.

Lunceford, J. K. and M. Davidian. 2004. Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects:
A Comparative Study. *Statistics in Medicine*. 23. pp. 2937–2960.

Maddala, G. S. and F. D. Nelson. 1975. Switching Regression Models with Exogenous and Endogenous Switching. *Proceedings of the Business and Economics Statistics Section, American Statistical Association*. pp. 423–426.

McCrary, J. 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*. 142 (2).
pp. 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005.

Quistorff, B. and S. Galiani. 2017. The synth_runner package: Utilities to Automate Synthetic Control Estimation Using Synth. Version 1.6.0. August 2017.
https://github.com/bquistorff/synth_runner.

Robins, J. M., A. Rotnitzky, L. P. Zhao. 1994. Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association*. 89 pp. 846–866.

Rosenbaum, P. R. and D. B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 70. pp. 41–55.

Rosenbaum, P. R. and D. B. Rubin. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*. 39 (1).

Rubin, D. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*. 66 (5). pp.688-27019. doi:10.1037/h0037350.

Ryan, A. M., J. F. Burgess, and J. B. Dimick. 2015. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. *Health Services Research*. 50 (4). pp. 1211–1235. https://www.ncbi.nlm.nih.gov/

pmc/articles/PMC4545355/pdf/hesr0050-1211.pdf.

الملحق الثاني
تصميم المسح الميداني

يلزم وجود بيانات عالية الجودة لإنتاج تقييم موثوق للآثار، وينبغي غالبًا استخدام المسوحات الأصلية لجمع هذه البيانات. يقدم هذا الملحق نظرة عامة على كيفية تصميم أدوات المسح الميداني وتنفيذها.

ينقسم هذا الملحق إلى الأقسام التالية:

القسم الأول: أنواع البيانات اللازمة لتقييم الآثار
القسم الثاني: أخذ العينات والتمثيل
القسم الثالث: اعتبارات إضافية بخصوص حجم العينة وحسابات القوة الإحصائية
القسم الرابع: تصميم أداة المسح
القسم الخامس: تنفيذ المسح
القسم السادس: إدارة البيانات

القسم الأول: أنواع البيانات اللازمة لتقييم الآثار

تخضع أنواع البيانات اللازمة لتقييم معين *لنظرية التغيير* التي تم تطويرها لهذا التقييم. وعلى وجه الخصوص، يمكن استخدام نظرية التغيير الخاصة بالتقييم في تحديد ما يلي:

- الحصائل/التأثيرات الأولية للتدخل (أي المؤشرات التي ترتبط ارتباطًا وثيقًا بالهدف الأساسي من البرنامج أو السياسة)، وكذلك الحصائل الثانوية التي قد تتأثر بالتدخل (الحصائل التي ليست الهدف الأساسي لتدخل معين هي أيضًا محل اهتمام)؛

- الحصائل أو المؤشرات المتوسطة التي يمكن استخدامها في تقييم مدى نجاح التدخل (أو فشله وعدم تحقيقه ما كان متوقعًا منه)؛

- السمات المعتدلة (المؤشرات التي لا تتأثر بأحد البرامج، ولكنها قد تكبح تأثيره، مثل النوع).

تزداد قيمة تقييم الآثار على نحو كبير عندما تتوفر البيانات عن الحصائل الثانوية والمتوسطة، وكذلك السمات المعتدلة. ولا يسمح هذا فقط بتحديد مدى نجاح البرنامج أو فشله، بل يسمح أيضًا بتحديد أي الحصائل المهمة الأخرى قد تأثرت، ولم كان التدخل فعالاً، أو كيف كان التدخل فعالاً، ومع من كان جيدًا، ومع من لم يكن جيدًا.

عند تحديد الحصائل، ينبغي الاهتمام اهتمامًا خاصًا بما إذا كان التقييم صُمم ليعمل على تحديد أي تأثير على تلك المؤشرات. ولكن المؤشرات المتغيرة تغيرًا شديدًا نادرة، وتخضع للخطأ الإحصائي، أو قد يتعذر رصد المؤشرات التي يحتمل أن يكون تأثير البرنامج عليها ضعيفًا دون الاستعانة بعينة كبيرة الحجم.

بالإضافة إلى الحصائل الأولية، والحصائل المتوسطة، والسمات المعتدلة، يُفضل أيضًا جمع المعلومات عن العوامل الأخرى بخلاف التدخل الذي قد تؤثر على الحصائل محل الاهتمام. إن وضع هذه العوامل في الحسبان يمكن أن يزيد من القوة الإحصائية للتقييم. وعند استخدام الطرق شبه التجريبية قد تكون معلومات عن تلك العوامل الأخرى ضرورية للتحديد، أو يمكن استخدامها في تقدير قوة إستراتيجية تحديد معينة.

بصورة عامة، يُفضل جمع البيانات التي تغطي فترات زمنية متعددة قبل وبعد بدء البرنامج أو السياسة. وتكون بيانات خط الأساس ضرورية في العديد من طرق التقييم - حتى في الطرق حيث لا تكون بيانات خط الأساس ضرورية جدًا (مثل التجارب العشوائية المضبوطة، أو تصميم انقطاع الانحدار) - ويمكن أن تكون بيانات خط الأساس ضرورية في التحقق من صحة الافتراضات التي تقوم عليها تلك الطرق (يُقصد بذلك تحديدًا، أن تكون سمات خط الأساس متشابهة في مجموعة المعالجة والمجموعة الضابطة). ويمكن أن تزيد بيانات خط الأساس أيضًا من قوة الافتراضات (يتم غالبًا خفض متطلبات حجم العينة بما يكفي لمعادلة تكاليف جمع بيانات خط الأساس بشكل كبير)، وتسمح بتحليل تأثيرات التدخل غير المتجانسة.

يمكن أيضًا اللجوء إلى خوض مراحل أو جولات متعددة من جمع البيانات بعد تنفيذ التدخل لتقدير الآثار طويلة وقصيرة الأجل. ويمكن أن يسمح هذا للمقيمين بتحليل المدة التي تستغرقها الآثار للتطور، ومدة سريانها.

القسم الثاني: أخذ العينات والتمثيل

يُعَد التمثيل أحد العناصر المهمة التي ينبغي مراعاتها سواء عند استخدام البيانات الأساسية أو الثانوية. وغالبًا ما يكون من الضروري استنتاج الخلاصات بخصوص المجتمع الإحصائي الكبير الذي لا يجدي معه جمع بيانات عن كل شخص في ذلك المجتمع. وفي مثل هذه الحالات، يجب استخدام طريقة مناسبة لأخذ العينات قصد جمع البيانات التي تمثل المجتمع محل الاهتمام. وستكون طريقة أخذ العينات مناسبة إذا كانت (١) تتجنب التحيز في إجراء الاختيار، و(٢) تحقق أعلى مستوى ممكن من الدقة في ظل مقدار الموارد المُتاحة.

ما المقصود بالعينة الممثلة؟

من المهم عند إجراء تقييم الأثر أن تُختار عينة المشاركين أو المجيبين التي تمثل *المجتمع الإحصائي محل الاهتمام*. في سياق التقييم، يُقصد بـ "المجتمع الإحصائي محل الاهتمام" مجموعة الأشخاص، أو القرى، أو المدارس، أو الوحدات الأخرى التي سيتم تقييم الحصائل الخاصة بها. ورغم أن هذه هي الحالة المثالية، فنادرًا ما يمكن إجراء تعداد يغطي كافة أفراد المجتمع الإحصائي محل الاهتمام. لذا، فالعينة هي مجموعة فرعية من المجتمع الإحصائي يمكن جمع معلومات بشأنها.

تُعد العينة *ممثلة* عندما تعكس تمامًا سمات المجتمع الإحصائي بأكمله. ويعني هذا، ضمنًا، أن أي تاثيرات للتقييم تم رصدها في العينة ستنطبق على المجتمع الإحصائي بأكمله. بيد أن السبب الرئيسي وراء عدم اعتبار العينة ممثلة هو أن تتأثر العينة بأحد أشكال الاختيار الانتقائي من المجتمع الإحصائي. على سبيل المثال، لا يُعد الأشخاص الذين يجيبون عن أسئلة مسح يُجرى عبر البريد أو الإنترنت ممثلين للمجتمع الإحصائي بأكمله، حيث إن الأشخاص الذي يجيبون عن أسئلة هذه الأنواع من المسوحات يختلفون عمن لا يجيبون. يمكن أن يتأثر أيضًا تمثيل العينة للمجتمع الإحصائي، إذا كان إطار أخذ العينة (يقصد به قائمة الأشخاص في المجتمع الذي ستؤخذ منه العينة) لا يغطي المجتمع الإحصائي المستهدف بشكل مناسب.

اختيار العينة

في البداية، يتم تعريف المجتمع الإحصائي محل الاهتمام وتحديده جيدًا. وبعد ذلك يتم أخذ العينات التي تمثل المجتمع الإحصائي، وفقًا لإستراتيجية أخذ العينات المناسبة، من خلال *أخذ العينة الاحتمالية*.[١] ونوضح فيما يلي الخطوات الأساسية لأخذ العينة الاحتمالية:

١. تحديد المجتمع الإحصائي محل الاهتمام.
٢. اختيار إطار أخذ العينات.
٣. اختيار الوحدات من إطار أخذ العينات.

إطار *أخذ العينات* هو قائمة الوحدات الأشمل التي يمكن الحصول عليها في المجتمع الإحصائي محل الاهتمام. وتضم المصادر الشائعة لإطار أخذ العينات، تعداد السكان، وقوائم المدارس أو العيادات في منطقة معينة، والخرائط التي تبين القرى والمدن في منطقة معينة، وما شابه. وعند إنشاء إطار أخذ العينات، ينبغي الاهتمام بالحصول على قائمة تضم أقصى قدر ممكن من الوحدات في المجتمع الإحصائي محل الاهتمام. ويمكن أن يُنتج إطار أخذ العينات الذي يستثني الوحدات الموجودة في المجتمع الإحصائي محل الاهتمام (أو يشمل الوحدات غير الموجودة) تحيزًا في التقييم. على سبيل المثال، يمكن أن يُنتج إطار أخذ العينات السيئ شيئًا مثل قوائم الدليل القديمة التي لا تسرد جميع الأشخاص أو الشركات.

[١] لاحظ أن تقنيات أخذ العينة غير الاحتمالية (مثل أخذ العينة الملائمة، وأخذ عينات كرة الثلج، وأخذ العينات العمدية، وأخذ العينات بالحصص) لن ينتج، على العموم، عينات ممثلة. ينبغي استخدام أساليب أخذ العينة غير الاحتمالية، فقط، عندما تحول القيود المادية أو اللوجيستية من إمكانية تطبيق أسلوب أخذ العينة الاحتمالية.

حالما يتم تحديد إطار أخذ العينات، يتم اختيار وحدات العينة من الوحدات المذكورة في الإطار. وعند استخدام أخذ العينة الاحتمالية، يتم اختيار الوحدات حسب احتمالية معروفة. ويضمن هذا أن الوحدات المُختارة متشابهة إحصائيًا مع الوحدات غير المُختارة. ومن ثم سيكون التأثير المسجل على مستوى الوحدات المُختارة هو نفسه على مستوى الوحدات غير المُختارة في العينة.

على الرغم من أن أسلوب أخذ العينات العشوائية البسيطة هو الطريقة الأسهل لأخذ العينة، يحيد المقيمون غالبًا عن استخدام العينة العشوائية البسيطة لتحسين مستوى الدقة، أو لخفض خفض تكاليف جمع البيانات، أو أو لعدم توفر أطر أخذ العينات المناسبة التي تغطي المجتمع الإحصائي محل الاهتمام بأكمله. وفي مقابل ذلك، غالبًا ما يلجأ المقيّمون إلى أنماط وأساليب مختلفة، من بينها أخذ العينات الطبقية وأخذ العينات العنقودية والجمع بين هاتين المقاربتين أو الأسلوبين.

أسلوب أخذ العينات الطبقية

قد يهتم المُقيم أحيانًا عند إجراء التقييم بالحصائل بين بعض المجموعات الفرعية بعينها. فقد يهتم مثلاً بكيفية تأثير التدخل التعليمي على الفتيات مقابل الفتيان، أو كيفية تأثير سياسة الضرائب على الأغنياء مقابل الفقراء. وفي مثل هذه الحالات تكون العينة الطبقية هي الأسلوب أو الخيار المناسب. ففي أسلوب أخذ العينات الطبقية يتم أولاً تقسيم الوحدات إلى مجموعات (تُسمى طبقات)، ثم يتم أخذ عينة عشوائية بسيطة من داخل كل مجموعة من تلك المجموعات. ومن ثم تحصل كل وحدة داخل المجموعات على نفس الفرصة ليتم اختيارها. وإذا كانت العينة داخل كل مجموعة كبيرة بما يكفي، فسيسمح هذا بالاستدلال عن هذه المجموعات الفرعية، وليس فقط عن المجتمع الإحصائي بأكمله. ورغم أنه يمكن الاستدلال بخصوص هذه المجموعات الفرعية دون الاستعانة بالعينة الطبقية، فإن هذا الأسلوب يضمن اختيار عدد كافٍ من الأشخاص من كل مجموعة فرعية ليسهل إجراء التحليلات. وعلى العكس من ذلك، قد لا يتم في أسلوب أخذ العينة العشوائية البسيطة اختيار عدد كافٍ من الأشخاص من كل مجموعة فرعية (مثل مجموعات الأقليات) داخل العينة التي سيتم تحليلها.

الميزة الثانية لأسلوب أخذ العينات الطبقية أن العينة الطبقية تحقق مستوى دقة أعلى من أسلوب أخذ العينات العشوائية البسيطة عند تطبيقه على حجم العينة نفسه. ولتحقيق هذا المستوى الأعلى من الدقة، تحتاج العينة الطبقية إلى حجم عينة إجمالي أصغر، مما يؤدي إلى خفض تكلفة جمع البيانات.

أسلوب أخذ العينة العنقودية (متعدد المراحل)

أحد الأشكال المختلفة الأخرى لأخذ العينات بخلاف أسلوب أخذ العينات العشوائية البسيطة هو استخدام أسلوب *أخذ العينة العنقودية (تُعرف باسم آخر هو أسلوب أخذ العينة متعدد المراحل)*. ويتم غالبًا تقسيم الوحدات الفردية إلى كُتل، مثل القرى، أو المدارس، أو المستشفيات. ووفقًا لهذا الأسلوب، يتم أخذ عينة من الكُتل أولاً، ثم الأفراد داخل تلك الكُتل. ويُنشئ هذا عينة لم يتم فيها توزيع أفراد العينة عشوائيًا عبر مساحة معينة، بل يتم تقسيمهم تقسيمًا جغرافيًا.

ويتفوق أسلوب أخذ العينة العنقودية بميزتين أساسيتين عن أسلوب أخذ العينة العشوائية البسيطة. أولهما أن هذا الأسلوب فعال في التحكم في تكاليف عملية جمع البيانات. فتكاليف هذا الأسلوب أوفر بالنسبة لفريق المسح الذي سيسافر من قرية إلى قرية، أو ينتقل من مدرسة إلى مدرسة، ويقضي وقتًا أطول في كل منهم، مقارنة بالتنقل بين الأفراد المُختارين في العينة والذين ينتشرون على مسافات واسعة جغرافيًا. ثانيًا، نتيجة لغياب أُطر أخذ العينات المناسبة التي تغطي المجتمعات الإحصائية بأكملها فمن الضروري أن نحدد كُتل العينات أولاً، ثم ننشأ قوائم كاملة تضم الأفراد من داخل كل كُتلة تم اختيارها. ومن أجل هذه الأسباب، تستخدم معظم المسوحات في الدول النامية أسلوب أخذ العينة العنقودية.

تُعد العينة العنقودية هي الاختيار الطبيعي في تقييم الآثار عندما تكون التدخلات عبارة عن كُتل، أو عند تطبيق البرامج على وحدة أعلى من وحدة الملاحظة. على سبيل المثال، تُطبق العديد من البرامج في المجال التعليمي على المدارس كلٍ على حدة، مما يؤدي إلى تكوين كُتل من الطلاب. ولهذا يُفضل، عمومًا، استخدام أسلوب أخذ العينة العنقودية في تقييم الآثار.

بيد أن العيب الأساسي في أسلوب أخذ العينة العنقودية هو انخفاض مستوى الدقة الناتج عنها مُقارنة بأسلوب أخذ العينة العشوائية أو أخذ العينة الطبقية، حتى وإن كانت الأساليب تتشابه في استعمال نفس إجمالي حجم العينة. ولهذا فإن قرار استخدام أسلوب أخذ العينة العنقودية يعتمد على تصميم البرنامج المُراد تقييمه، وتوفير التكاليف المترتبة عن استعمال أسلوب أخذ العينة العنقودية، وتوفر أُطر أخذ العينات، وأوجه المقارنة بينها وبين التكاليف المرتبطة بالحاجة إلى تكوين عينة إجمالية أكبر. تم تناول القوة الإحصائية (الدقة) في العينة العنقودية بمزيد من التفصيل في الفصل السابع.

على الرغم من أوجه التشابه بين أسلوبي أخذ العينة الطبقية والعنقودية من حيث كونهما يقسمان أولاً إطار العينة إلى مجموعات فرعية غير متداخلة، فإنهما يختلفان من حيث إن الأفراد في جميع الطبقات ستتم إضافتهم في العينة النهائية، ولكن سيُضاف إلى العينة أشخاص من المجموعات الفرعية من الكُتل المُختارة فقط.

بيد أن الأمر الأهم في النهاية هو أن تمثل العينة المُختارة المجتمع الإحصائي محل الاهتمام. ويتم تحقيق هذا بالتأكد من أن إطار أخذ العينات يغطي المجتمع الإحصائي محل الاهتمام تغطية تامة ودقيقة، ومن تطبيق إستراتيجية أخذ العينات المناسبة (سواء تم اختيار أسلوب أخذ العينة البسيطة، أو الطبقية، أو العنقودية) في تحديد الوحدات أو الأفراد من هذا الإطار.

القسم الثالث: اعتبارات إضافية بخصوص حجم العينة وحسابات القوة الإحصائية

إذا وُجدت إستراتيجية أخذ العينات التي يتمخض عنها أخذ العينة المُمثلة للمجتمع الإحصائي محل الاهتمام، فإن الأمر الثاني المهم هو أن تكون العينة كبيرة بما يكفي لإنتاج قوة إحصائية مناسبة لتحديد أثر البرنامج. ويتحدد حجم العينة المطلوب لتقييم الأثر من خلال حسابات قوة العينة، مثلما تم توضيحه في الفصل السابع. ويتناول هذا القسم الاعتبارات الإضافية بخصوص حسابات القوة بخلاف تلك التي تم تناولها في النص الأساسي.

بيانات خط الأساس وموجات البيانات

إحدى ميزات التصميم التي تؤثر في متطلبات القوة الإحصائية وحجم العينة هي عدد موجات البيانات التي سيتم جمعها. وأبرز هذه الموجات تتمثل في توفر بيانات خط الأساس، حيث يمكن أن تؤدي إلى زيادة القوة الإحصائية زيادة كبيرة. ويُعزى ذلك إلى أن تضمين المتغيرات المشتركة لخط الأساس في تخمين تأثير المعالجة سيؤدي إلى خفض التشوش (التباين) في الحصيلة إلى حد كبير. بيد أن المكسب الأكبر للقوة ينتج عادة من تضمين بيانات خط الأساس في الحصائل قبل تطبيق البرنامج، وعلى وجه الخصوص إذا كانت الحصائل ترتبط ارتباطًا كبيرًا ومتزايدًا على مدار الوقت (مثل حصائل الصحة والتعليم). إن خفض حجم العينة المطلوب الناتج عن تضمين المعاملات المشتركة لخط الأساس يمكن أن يعادل تكاليف جمع البيانات الإضافية (قد يعادل هذه التكاليف، جزئيًا، على أقل تقدير).

بالنسبة للحصائل التي تصبح أقل ارتباطًا على مدار الزمن (مثل أرباح الأعمال)، فإن التحكم في قيم خط الأساس للحصيلة يحقق مكسبًا أقل في القوة الإحصائية. بالنسبة لهذه الحصائل، قد يفضي امتداد المعالجة على مدار فترة معينة إلى تجاهل جمع بيانات خط الأساس. وسيتم، بدلاً من ذلك، تخصيص الموارد لجمع موجات متعددة من بيانات المتابعة (McKenzie 2012).

الافتراضات المتعددة

عادة ما تستخدم التقييمات أكثر من حصيلة واحدة لتقييم أثر التدخل. كما تهدف، في أغلب الأحيان، إلى المقارنة بين أكثر من معالجتين. وبالتالي، تتمكن تلك التقييمات من اختبار عدة افتراضات معًا بالتزامن (عدد الحصائل مضروبًا في عدد المقارنات بين مجموعات المعالجة المختلفة). وتفترض هذه المقاربة القياسية في اختبار الافتراضات أن يتم إجراء اختبار واحد. وعندما يتم اختبار افتراضات متعددة، ينبغي تعديل حسابات القوة الإحصائية لمراعاة احتمالية أن ينتج أيٌ من اختبارات الافتراضات نتيجة مهمة على سبيل المصادفة العشوائية. بمعنى آخر، تكون احتمالية الحصول على تأثير مهم مصادفةً أعلى في الاختبارات المتعددة عنها في الاختبار المفرد، وذلك بسبب زيادة عدد المحاولات.

تنقسم طرق تعديل الافتراضات المتعددة إلى فئتين أساسيتين: الطرق التي تتحكم فيما يُسمى معدل الاكتشاف الخاطئ (FDR)، والطرق التي تتحكم فيما يُسمى معدل الخطأ الجماعي (FWER). وتقوم الإجراءات التي تتحكم في معدل الاكتشاف الخاطئ بإجراء التعديلات من خلال التحكم في النسبة المتوقعة من النتائج المهمة (رفض افتراضات العدم) التي هي اكتشافات خاطئة (الأخطاء من النوع الأول). تتحكم إجراءات التحكم في معدل الخطأ الجماعي في احتمالية خطأ من النوع الأول على *الأقل مرة واحدة*. تُستخدم عادة إجراءات التحكم في معدل الخطأ الجماعي بمعدل أكبر عند حساب القوة وحجم العينة، لأنها أكثر تحفظًا من إجراءات التحكم في معدل الاكتشاف الخاطئ.

تتوفر العديد من طرق التحكم في معدل الخطأ الجماعي الفردية، وأبسطها هي طريقة Bonferroni التي طورها Dunn عام ١٩٦١. يتم في طريقة Bonferroni الحصول بسهولة على مستوى الدلالة المُعدل (α) بقسمة مستوى الدلالة المطلوب على إجمالي عدد الافتراضات، واستخدام مستوى الدلالة الجديد هذا في حساب حجم العينة. وعلى سبيل المثال، إذا كان مستوى الدلالة المطلوب ٠,٠٥، وكان هناك افتراضان (كأن يتم اختبار نتيجتين)، فسيكون مستوى الدلالة ٠,٠٢٥ (٠,٠٥÷٢).

بيد أن عيب طريقة Bonferroni أنها تفترض أن الافتراضات مستقلة، ومن ثم تكون متحفظة جدًا. وعلى الرغم من تطوير هذه الطريقة لتكون أقل تحفظًا، (مثل طرق Holm/Hochberg)، فإن أفضل ممارسة هي استخدام مقاربات "التناقص" التي تراعي الارتباط بين الحصائل المُختَبرة (مثل تلك التي وضعها Westfall وYoung عام [١٩٩٣]، وRomano وWolf عام [٢٠٠٥]).

القسم الرابع: تصميم أداة المسح

بالإضافة إلى أخطاء أخذ العينات التي ناقشناها أعلاه، يمكن أيضًا *أن تؤدي الأخطاء غير المرتبطة بأخذ العينات* إلى خفض مستوى الدقة (القوة)، وتحيز نتائج تقييم الأثر. تضم الأخطاء غير المرتبطة بأخذ العينات الأخطاء الناتجة عن بعض المشكلات، مثل عدم الاستجابة (عدم وجود بيانات بعض الأشخاص، أو عدم اكتمالها)، والتناقص (انسحاب أو فقدان الأشخاص بين جولات المسح)، وأخطاء القياس (وجود اختلاف بين قيمة المؤشر المُسجلة وقيمته الحقيقية [Banda 2003]). وفي حين تظهر أخطاء أخذ العينة في أثناء مرحلة أخذ العينات، فإن الأخطاء غير المرتبطة بأخذ العينات تظهر في أثناء مرحلة جمع البيانات. وتتأثر تلك الأخطاء بكل من طريقة تصميم الاستبيان، وكيفية صياغة الأسئلة، وسلوك الفريق الميداني، وطريقة جمع البيانات والتحقق من صحتها.

عند تقييم الأثر تترتب عن الأخطاء غير المرتبطة بأخذ العينات تكلفة باهظة، حيث إن صحة التقييم تعتمد اعتمادًا تامًا على صحة البيانات. وعلى الرغم من تأثر جميع تحليلات البيانات بالأخطاء غير المرتبطة بأخذ العينات، فإنه يُزعم أنها تمثل خطرًا أكبر على تقييم الأثر، حيث إن الأخطاء التي لا تتم معادلتها في المجموعة المعالجة والمجموعة الضابطة تؤدي إلى خفض إمكانية تفسير نتائج التقييم خفضًا كبيرًا. لذا من المهم التخطيط لعملية جمع البيانات لخفض الأخطاء غير المرتبطة بأخذ العينات إلى أكبر حد ممكن، والتأكد من أن الأخطاء المتبقية تتم معادلتها بالتساوي ما بين مجموعة المعالجة والمجموعة الضابطة. والطريقة المُثلى لتحقيق ذلك هي وضع المعايير الموحدة وتضمين اختبارات جودة متكررة عند مختلف المراحل.

بالطبع تخضع عملية جميع البيانات دائما لبعض القيود المهمة، من بينها القيود المالية (الميزانية المتاحة لعملية جمع البيانات) والقيود التشغيلية (إمكانات من ينفذون المسح) والقيود المرتبطة بقدرة من تتم محاورتهم واستعدادهم لتقديم المعلومات. يتعين أخذ هته القيود في عين الاعتبار لتطوير إستراتيجية جمع البيانات التي ستوفر البيانات بنسبة خطأ منخفضة، وتغطية المؤشرات المهمة المطلوبة للتقييم. لهذا فإن عملية جمع البيانات تقوم على تحديد الأولويات، والموازنة بين تحقيق أهداف التقييم والأخذ بعين الاعتبار لقيوده.

وضع الاستبيان

أثناء إعداد الاستبيان، تبدأ عملية الموازنة بين الأهداف المرجوة والقيود التي تعترض جمع البيانات. فهناك توجه كبير، عند إعداد الاستبيانات، إلى جمع أكبر قدر ممكن من البيانات. لكن، من المؤكد أن عملية جمع البيانات غير الضرورية تكون مكلفة. ولا يؤدي وضع استبيانات طويلة غير ضرورية إلى زيادة التكاليف وحسب، بل يؤثر أيضًا على جودة جميع البيانات المُجمعة.

جرت العادة أن تشتمل عمليات تقييم الأثر على مستويات متعددة من المسوحات والاستبيانات، بهدف رصد أوجه التباين عند تطبيق البرنامج و/أو السياسة، وآثار تعديل المتغيرات. على سبيل المثال، ربما يكون هناك مسح منشأة يُطبق على المنشآت التي تقدم البرنامج (مثل المدارس في حالة التدخل التعليمي)، أو مسح للقرية أو المجتمع لرصد وجود برامج أخرى في الموقع، والسمات المشتركة للقرية، ومسح المنازل لتحديد سمات الفئة المستهدفة في البرنامج. ويمكن أن تتحسن كفاءة المسح عند إعادة صياغة الأسئلة لتصل إلى المستوى المطلوب.

يمكن أن تقتصر عملية إعداد الاستبيانات وأدوات المسح الأخرى على التقييم الحالي فقط. في الوقت نفسه تحاول معظم المسوحات أن ترصد المتغيرات التي كانت هدف جهود المسح السابقة. ومن أجل هذا، غالبًا ما يكون من المفيد مراجعة وحدات المسوحات السابقة قبل البدء في تصميم مسح جديد.

وفيما يلي أفضل المصادر لنماذج المسوحات:

- نماذج مسوحات قياس مستويات المعيشة (تشتمل على العديد من الوحدات المأخوذة من بلاد عدة): http://go.worldbank.org/ UK1ETMHBN0

- نماذج المسوحات من البرنامج الديموغرافي والصحي: http:// dhsprogram.com/What-We-Do/ Questionnaires.cfm

- المسوحات المُستخدمة في دراسات ديناميكيات القرية في جنوب آسيا:
- http:// vdsa.icrisat.ac.in/vdsa-questionaires.aspx

- نماذج المسح المذكورة في أدوات تقييم الأثر الخاصة بالبنك الدولي:
- http://siteresources.worldbank.org/EXTIMPEVALTK/Resourcs/
- 8811875-1346101602804/4.10_Household_Questionnaires.zip

يقدم هذا القسم نظرة عامة على مقاربة شائعة في إعداد الاستبيانات، ويقدم اقتراحات خاصة بكل مرحلة في تلك العملية. والمغزى الأساسي من هذه الاقتراحات هو الحد من الأخطاء غير المرتبطة بأخذ العينات، وللتأكد من اتساق الاستبيان.

الخطوات الأساسية المتبعة لوضع الاستبيان

يتم إعداد معظم الاستبيانات وتنظيمها في شكل وحدات، ويُقصد بذلك أن يتم تجميع أنواع الأسئلة المتشابهة في وحدات أو أقسام مختلفة وفق نموذج المسح. وتتوافق معظم الوحدات عادةً مع الموضوعات، ولكنها يمكن أن تتوافق أيضًا مع مختلف المجيبين مثلاً. ويُسهل استخدام الوحدات كل من مهمة إنشاء الاستبيان وتنظيمه، وينشأ فواصل بين الأسئلة، مما يساعد في الحد من شعور مندوب الإحصاء والمُحاوَر بالإرهاق.

عند استخدام مقاربة الوحدات، يمكن تقسيم عملية وضع الاستبيان إلى سبع خطوات أساسية:

الخطوة الأولى: تحديد الوحدات
الخطوة الثانية: تصميم الأسئلة الخاصة
الخطوة الثالثة: التوافق
الخطوة الرابعة: وضع الترتيب
الخطوة الخامسة: الاختبار المسبق للاستبيان
الخطوة السادسة: المراجعة
الخطوة السابعة: تكرار الخطوتين الخامسة والسادسة

تنطوي كل خطوة على اعتبارات مهمة قد تؤثر تأثيرًا كبيرًا على جودة بيانات خط النهاية.

الخطوة الأولى: تحديد الوحدات

يعتمد اختيار وحدات بعينها لإضافتها إلى الاستبيان على طبيعة التقييم. وينبغي اختيار الوحدات بحسب قائمة المؤشرات المُحددة وفقًا لنظرية التغيير التي تحدد مسار عملية التقييم. وينبغي أن تبدأ عملية تحديد الأولويات عند هذه المرحلة، على أن يتم فقط تضمين الوحدات وفقًا للموضوعات الأهم في التقييم.

الخطوة الثانية: تصميم الأسئلة الخاصة

إن تصميم الأسئلة الخاصة هو المهمة الأصعب في عملية وضع الاستبيان. وحيثما أمكن، يمكن الاستعانة بأسئلة طُرحت في استبيانات قائمة بالفعل وتم استخدامها والتحقق من صحتها في سياق معين. وحيثما أمكن، يمكن طرح الأسئلة في شكل كمي ومستمر، حتى وإن بدا أن التصنيف حسب الفئات أسهل بالنسبة لمندوبي الإحصاء. وتساعد المتغيرات المستمرة في إيجاد المزيد من الاحتمالات التحليلية في وقت أبعد من المتغيرات الفئوية.

تؤثر التفاصيل الصغيرة ظاهريًا المتعلقة بطريقة صياغة الأسئلة تأثيرًا كبيرًا على نوع الخطأ ومقداره في الإجابات المستخلصة خلال عملية المسح. لذا، ينبغي أيضًا الاهتمام بصياغة الأسئلة بطرق تحد من "التحيز للمقبول" الذي ينشأ عندما يخبر المجيبون مندوبي الإحصاء "بما يريدون سماعه". ويمكن أن يُشكل التحيز للمقبول خطرًا في تقييمات الأثر التي تنتج نوعًا من المنفعة: قد يشعر المجيبون عن أسئلة الاستبيان في مجموعة المعالجة بضرورة إعطاء ردود إيجابية إذا عرفوا أن المسح يرتبط ببرنامج معين، أو قد يعتقد المجيبون أن ردودهم قد تؤثر على مزايا لاحقة. وفي ظل كثرة المشكلات المرتبطة بتصميم الأسئلة، من الأفضل مراجعة بعضٍ من الكتب الكثيرة التي أُلفت في هذا الموضوع (مثل Krumpal 2013 وGrimm 2010) وإجراء اختبارات قبلية موسعة للأسئلة المستعصية (في الخطوة الخامسة).

من المهم أيضًا أن يتم تصميم الأسئلة بطريقة تُسهل الإجراءات إلى أقصى قدر ممكن (لكل من مندوب الإحصاء والمجيب)، وتحقق الاتساق في جميع المقابلات. فيما يلي بعض القواعد التي تحقق هذا:

- كتابة الأسئلة كاملة، بحيث يمكن أن يُجري المُحاور المقابلة بقراءة كل سؤال في الاستبيان.
- تضمين تعريفات دقيقة لجميع المفاهيم و/أو المصطلحات الأساسية المستخدمة في الاستبيان.
- جعل الأسئلة أقصر ما يمكن واستخدام المفردات اليومية الشائعة.
- طرح الأسئلة في شكل وحدات من الطبيعي أن يفكر المجيب فيها، والتي تحد من الحاجة إلى أي حسابات عقلية للوصول إلى الإجابة.
- تصميم الاستبيان، بحيث يكون لمعظم الأسئلة خيارات إجابة ذات رموز مُحددة مسبقًا.
- الاهتمام باللغات واللهجات المحلية (حتى وإن تحدث المجيبون بلغة وطنية فصحى، تسير اللقاءات بسلاسة أكبر وبأخطاء أقل عند إجرائها باللغة المحلية).
- إضافة مربعات اختيار في الاستبيان، بحيث يتم التحقق من صحة تسجيل الإجابات.

إذا كان المجيبون سيملئون النماذج بأنفسهم، فينبغي إيلاء اهتمام خاص بجعل النماذج سهلة القراءة والفهم. ويمكن أن تشتمل النماذج التي يملؤها المجيبون بأنفسهم على أخطاء قياس أكثر من الاستبيانات القائمة على طرح الأسئلة، ولكن تكون تلك النماذج التي يملؤها المجيبون بأنفسهم ضرورية في بعض الأحيان (على سبيل المثال، في الفصول الدراسية التي عدد طلابها كبير).

الخطوة الثالثة: ترتيب الوحدات

تتمثل الخطوة التالية في ترتيب الوحدات حسب ظهورها في الاستبيان. وينبغي أن يطابق هذا الترتيب ذلك الترتيب المتبع أثناء إجراء المقابلة.

يُفضل أن يتم تجميع الوحدات المتشابهة مفاهيميًا معًا، حيث سيسهل هذا عملية طرح الأسئلة، لأن التنقل الكثير بين الموضوعات سيكون مرهقًا بالنسبة لمن يتم إجراء المقابلات معهم. في حالة إجراء المقابلة مع أكثر من شخص في الاستبيان، فإن تجميع الأسئلة حسب المجيب يمكن أن يسمح لمندوب الإحصاء بإجراء المقابلات مع الأشخاص كل على حدة. ومن الأفضل بوجه عام أن يتم إعداد استبيانات منفصلة لمختلف المجيبين، ولكن يصعب أحيانًا أن تعرف مسبقًا الشخص الأكثر علمًا بموضوع معين (على سبيل المثال، عند إجراء المسح في مؤسسة ما). وحينئذٍ من الأفضل أن يتم تجميع أكبر عدد ممكن من الوحدات المتشابهة لتجنب اضطرار مندوبي الإحصاء إلى التنقل بين مختلف المجيبين والدخول في أخذ وردّ معهم.

فيما يلي بعض الاعتبارات الأخرى المهمة عند تنظيم الوحدات:

- ينبغي تنظيم الوحدات، بحيث تكون البداية مع الأسئلة التي يسهل الإجابة عليها.
- ينبغي طرح الوحدات الأكثر حساسية مع اقتراب نهاية الاستبيان.
- ينبغي وضع الوحدات الأهم في البداية، حيث يكون المجيبون أكثر حماسًا.

الخطوة الرابعة: التوافق

بعد الانتهاء من صياغة أسئلة كل وحدة وتعديلها، وترتيب الوحدات، ينبغي تحقيق التوافق في الاستبيان. والمقصود بذلك أن يتم تحرير الاستبيان لتحقيق الاتساق والسلاسة. يقتضي تحقيق التوافق اتخاذ الاجراءات التالية:

- التأكد من اتساق صياغة الأسئلة (على سبيل المثال، ينبغي أن تشير الأسئلة المتقاربة المتشابهة إلى الفترة نفسها حيثما أمكن)؛
- التحقق من اتساق رموز الأسئلة عبر مختلف الأسئلة في النموذج (هذا أمر مهم أيضًا لخفض الأخطاء المحتملة في مرحلة تحليل البيانات)؛
- تضمين "رموز التخطي"، والتي تشير إلى الأسئلة التي سيتم تجاوزها بُناءً على إجابات الأسئلة السابقة، والتحقق منها جيدًا في الاستبيان.

عند وضع الاستبيانات لمسح خط النهاية (بعد تطبيق البرنامج)، ينبغي مراجعة الأسئلة للتحقق من اتساقها مع الأسئلة المستخدمة في دورات المسح السابقة إن وجدت.

الخطوة الخامسة: الاختبار المسبق

عند الانتهاء من صياغة المسح، وإجراء عملية تحديد الأولويات المتكررة، من المهم إجراء الاختبار القبلي لأدوات المسح. ينطوي الاختبار القبلي على تجربة الاستبيان على عدد من الأشخاص (يُفضل ألا يكونوا من بين أعضاء مجموعتي المعالجة والمقارنة المشاركين في التقييم)، وذلك بهدف ملاحظة أداء الاستبيان. من الأفضل غالبًا أن يتم إجراء الاختبار المسبق على مدار عدة أسابيع، مع الاستعانة بمندوبي الإحصاء المتوقع مشاركتهم في المسح، وتحت متابعة دقيقة. ومن شأن هذا الاختبار القبلي أن يُبرز المشكلات التي قد تظهر خلال العمل الميداني، ويرفع مستوى كفاءة مندوب الإحصاء.

ينبغي أن يدقق مندوبو الإحصاء في العوامل التالية في أثناء إجراء الاختبار المسبق:

- التوقيت: ما المدة التي يستغرقها الاستبيان حتى ينتهي؟ ما المدة التي تستغرقها كل وحدة (مع مراعاة أن الاستبيان في مرحلة تجريبية قد يستغرق وقتًا أطول بسبب أخطاء الصياغة وعدم خبرة مندوبي الإحصاء؟ ويفيد هذا أيضًا في قياس الوقت الذي يستغرقه كل سؤال على حدة في أثناء كل اختبار مسبق، بحيث يمكن تنسيق الاستبيان بأفضل شكل ممكن.
- شعور مندوبي الإحصاء والمجيبين بالإرهاق.
- هل توجد أسئلة يصعب على مندوبي الإحصاء شرحها، أو يصعب على المجيب فهمها؟

- ما الأسئلة التي وجد المجيبون صعوبة في الإجابة عنها؟ هل يجدون صعوبة في استرجاع الأحداث؟ هل يشعرون بعدم الراحة؟

- الاتساق بين مختلف مندوبي الإحصاء. هل يمكن معالجة أي حالة من عدم الاتساق من خلال التدريب، أو ينبغي إعادة صياغة الأسئلة؟

ينبغي، إن أمكن، إجراء الاختبار المسبق في مواقع متعددة ومختلفة. وبالإضافة إلى نماذج المسح، يُنصح أيضًا بمحاكاة إجراءات المسح الأخرى في هذه المرحلة (تنظيم فريق المسح، إجراءات الإشراف، وإجراءات التحقق من الجودة، إلخ)، ويتم هذا عادة كجزء من جولات الاختبار المسبق.

الخطوة السادسة: المراجعة

ينبغي مراجعة الاستبيانات وفقًا لمدخلات الاختبار القبلي. ومن الممارسات الجيدة في هذا السياق أن يتم إشراك جميع أفراد فريق جمع البيانات في عمليتي الاختبار المسبق والمراجعة. فغالبًا ما يكون مندوبو الإحصاء أفضل معرفة بما كان جيدًا وما لم يكن جيدًا، وسوف يلاحظ مختلف الأشخاص المشاركين في العملية مشكلات مختلفة.

الخطوة السابعة: تكرار الخطوتين الخامسة والسادسة

ينبغي تكرار عمليتي الاختبار المسبق والمراجعة حتى يصل الاستبيان إلى شكل مرضٍ، ويُقصد بذلك أن يصل الاستبيان إلى أوضح وأسهل شكل ممكن، ويعكس حصيلة الموازنة المتقنة بين الأهداف المرجوة والقيود التي تعترض جمع البيانات. وفي نهاية هذه العملية المتكررة، ينبغي أن تخضع النسخة النهائية من الاستبيان لاختبار مسبق نهائي؛ للتأكد من حل جميع المشكلات.

القسم الخامس: تنفيذ المسح

تنسيق العمل الميداني

إن العملية الميدانية للمسح عملية معقدة، وتعتمد اعتمادًا كبيرًا على سياق المسح. ولكن بوجه عام، فإن أفضل وسيلة للتأكد من جودة البيانات هي (١) الحرص على تدريب مندوبي الإحصاء تدريبًا جيدًا، والإشراف عليهم إشرافًا جيدًا، و(٢) تطبيق نظام لمراقبة الجودة.

تنظيم العمل الميداني

يتم غالبًا تنظيم المسوحات بأفضل شكل ممكن، بحيث يصبح كل فريق من الفرق الميدانية مسؤولاً عن مجموعة فرعية مُحددة من المواقع. وتضم الفرق الميدانية عمومًا مشرفًا، ومندوبي إحصاء (يكون الفريق غالبًا مسؤولًا عن استبيانات فردية، أو عقد لقاءات مع أنواع معينة من المجيبين على المسح)، ويتم أحيانًا إضافة مختص بيانات إلى الفريق ليكون مسؤولاً عن إدخال البيانات والتحقق منها في أثناء العمل الميداني.

فيما يتعلق باختيار فرق المسح المناسبة للمواقع المختلفة، من الجيد تعيين فرق المسح عشوائيًا في المواقع، والتأكد من أن كل فريق مسؤول عن عدد متساوٍ من وحدات المعالجة والوحدات الضابطة. وفي الحالة المثالية، لا يتم إعلام فرق المسح بوحدات المعالجة والوحدات الضابطة، بيد أن هذا الأمر لا يكون ممكنًا على الدوام.

تدريب مندوبي الإحصاء

لا يجوز بأي حال من الأحوال التغاضي عن أهمية الحرص على تدريب مندوبي الإحصاء تدريبًا جيدًا. حيث يمكن أن تؤثر جودة التدريب تأثيرًا خطيرًا على جودة البيانات الناتجة.

ويؤكد تدريب مندوبي الإحصاء على ثلاثة جوانب:

١. الفهم العام لهيكل المسح وأهدافه

أولاً، ينبغي ألا يقتصر تدريب مندوبي الإحصاء على تعليمهم كيفية تنفيذ مسؤولياتهم الخاصة فقط. من الأجدر أن يُشرح لهم أيضًا، أثناء التدريب، الهيكل العام للمسح وأهدافه.

إن الحرص على استيعاب مندوبي الإحصاء لأهداف المسح العامة يمنحهم فهمًا أفضل لمدى أهمية الدور الذي يضطلعون به في السياق الأعم للعملية. بيد أن الأمر الأهم يتمثل في أن فهم الهيكل العام ومسؤوليات الآخرين يُمكن أن يُسهل التنسيق في أثناء العمل الميداني، لا سيما وقت وقوع الأحداث غير المتوقعة (الحتمية). إن أعمال الإحصاء شاقة ومرهقة في أغلب الأحيان، حتى وإن كان المسح في أبسط شكل ممكن. لذا، فإن الفهم العام لأهداف المسح يمكن أن يفيد أيضًا في الحفاظ على روح الحماس طوال فترة العمل.

٢. توحيد المعايير والمقاييس

ثانيًا، ينبغي إجراء التدريب بهدف توحيد المعايير والمقاييس ويشمل ذلك إقامة التدريب في موقع مركزي مما يُسهل التأكد من اتساق التدريب (من الناحيتين الكيفية والكمية) بين جميع الفرق الميدانية المشاركة في المسح.

٣. الممارسة العملية

أخيرًا، ينبغي أن ينطوي التدريب على قدر كبير من الممارسة العملية. فكلما زاد تدريب مندوبي الإحصاء العملي، زادت سرعتهم ودقتهم ومستوى الاتساق بينهم. يمكن أن يتم التدريب العملي في بيئة فصل دراسي حيث يعقد مندوبو الإحصاء لقاءات مع بعضهم بعضًا، ويتبادلون تمثيل دور المجيبين، وينبغي أيضًا عقد جزء من التدريبات الميدانية "في الموقع".

أدلة التدريب التوجيهية

إن الجزء المحوري والأساسي في التدريب هو مجموعة مفصلة وواضحة من أدلة التدريب التوجيهية. وينبغي إعداد الأدلة التوجيهية لكل عضو في فريق المسح، بحيث توضح ما يلي:

١. الغرض من المسح؛

٢. والهيكل العام للمسح (الأماكن التي ستذهب إليها الفرق، وجميع الاستبيانات التي سيتم طرحها في كل موقع)؛

٣. والمهام التي سيؤديها كل عضو في الفريق؛

٤. والإجراءات المتبعة في الحالات الطارئة؛

٥. والمبادئ العامة للتعامل مع المشكلات غير المتوقعة (المشكلات الحتمية)؛

٦. والتفاصيل المتعلقة بكيفية تعبئة الاستبيانات (تُضاف إلى الأدلة التوجيهية غالبًا نُسخٌ من الاستبيانات تشتمل على توضيحات؛ هذا على الرغم من ضرورة إضافة تعليمات تفصيلية في الاستبيانات نفسها على قدر الإمكان)؛

٧. وتعريف مصطلحات معينة، بحيث لا توجد اختلافات في تفسيرها.

اختبارات الجودة

ينبغي أن يُضاف إلى إطار العمل الميداني نظامٌ من اختبارات متعددة للجودة على مدار المراحل المختلفة للمسح، وأن يتم على أيدي أشخاص مختلفين. وينبغي أن تتحقق هذه الاختبارات من مستوى الدقة، والاتساق، والكمال. ينبغي أن يتحقق مندوبو الإحصاء من عدم وجود أخطاء بعد الانتهاء من كل مقابلة مباشرة، وينبغي أن يراجع المشرفون لاحقًا ما قام به المندوبون، ولكن يُفضل إجراء هذا قبل مغادرة الموقع. فإجراء تلك المراجعات المتكررة في الميدان يتيح رصد الأخطاء سريعًا، ومن ثم تكون تكلفة إصلاحها زهيدة. وينبغي أن يُجري المشرفون أيضًا مراجعات انتقائية عشوائية، وأن يعيدوا المقابلة مع المجيبين حول مجموعة فرعية صغيرة من الأسئلة، للتأكد من دقة تنفيذ المسح بينما لا يزال من الممكن إجراء التصحيحات.

تُفضي جميع طرق أخذ عينات المجيبين، تقريبًا، إلى اختيار مجيبين لا يمكن الوصول إليهم وقت إجراء المسح. لذا ينبغي تطبيق نظام لإعادة زيارة المجيبين الذين لم يُعثر عليهم، أو المجيبين غير الكاملة معلوماتهم، أو استبدالهم.

القسم السادس: إدارة البيانات

تهدف عملية جمع البيانات أساسًا إلى الحصول على البيانات ليتم استخدامها في التحليلات. الخطوة الأخيرة في هذه العملية، وليست الأقل في الأهمية، هي جمع البيانات وإعداد مجموعات البيانات الفعلية. من المهم أيضًا جمع البيانات بحرص وإدخالها وتصحيح الأخطاء وإعداد البيانات لجعل الأخطاء، غير المتعلقة بأخذ العينات، عند أدنى حد ممكن.

جمع البيانات وإدخالها

لا تزال العديد من المسوحات تقوم أساسًا على جمع البيانات باستخدام الأوراق والأقلام، وذلك بسبب التعود على هذا الأمر أو لضرورة القيام به، ولكن يزداد معدل استخدام أجهزة الحاسوب اللوحية أو المحمولة.لاستخدام أجهزة الحاسوب في جمع البيانات في الميدان فوائد ظاهرة، من بينها (١) القدرة على استخدام برمجيات مُبرمجة تشتمل على أكواد التخطي الخاصة بالاستبيان، ومراجعات الاتساق في الوقت الفعلي، و(٢) إلغاء الحاجة إلى إدخال البيانات في أجهزة الحاسوب في خطوة منفصلة وإضافية. بيد أن استخدام أجهزة الحاسوب في جمع البيانات ميدانيًا يمكن أن يزيد أيضًا من مخاطر خسارة البيانات، لذا ينبغي وضع نظام يُمكن الفرق من إجراء نسخ احتياطي للبيانات في الميدان. ربما تحتاج إدارة المسح الإلكتروني المزيد من الاختبارات القبلية، فقد يكون لنماذج الإدخال خاصية Skip Logic (تجاوز المنطق) ووحدات وصيغ لا ترقى إلى القدر نفسه من المرونة التي تتميز بها النماذج الورقية، حيث يمكن لمندوبي الإحصاء التعديل أو كتابة التعليقات.

عند جمع الردود المستخلصة من إجراء المسح باستخدام الأوراق والأقلام، ينبغي تطبيق إجراءات عملية للتأكد من جودة عملية إدخال البيانات. ويمكن استخدام برمجيات إدخال البيانات لتسهيل عملية إدخال البيانات وتحسين الجودة؛ بفضل عناصر اختبار الاتساق المدمجة. وفي حالة دمج عملية إدخال البيانات في عمليات المسح الميدانية، يمكن أيضًا رصد الأخطاء وتصحيحها سريعًا بالاستعانة بأجهزة الحاسوب في أثناء المقابلات. ويمكن أيضًا تحسين جودة عملية إدخال البيانات بالاستعانة بإجراء "الإدخال المزدوج"، حيث يقوم بإدخال البيانات شخصان مختلفان كلٌ على حدة. وينبغي أن تُضاف إلى بيانات المسح توقيت إجراء المقابلات، وأسماء مندوبي الإحصاء مع كل ملاحظة.

تنقية البيانات وإعدادها

ينبغي البدء في إعداد البيانات للتحليل حالما يتم إدخالها، بيد أن الجزء الأهم في هذه العملية هو "التحقق من البيانات" و"تنقيتها". ويُقصد بهذا عملية التحقق من البيانات الأولية لرصد الأخطاء، وإجراء التصحيحات. وينبغي في أثناء عملية تنقية البيانات أن يتم مجددًا مراجعة الإجابات للتأكد مما يلي:

- الاتساق (هل هناك أي إجابات متناقضة وردت خلال المسح؟)
- الصلاحية (هل هناك أي قيم "متطرفة" جدًا بحيث يصعب تصديقها؟)
- الاكتمال (هل تفتقد أي نماذج مقدارًا كبيرًا من الإجابات؟ هل هناك متغيرات معينة مفقودة بشكل لا يمكن التغاضي عنه؟)
- التوزيع (هل تتبع البيانات التوزيع المتوقع، وهل هناك أي قيم متطرفة؟)
- النزاهة (هل يحيط الشك بأي شكل من الأشكال بالبيانات الخاصة بأي فريق مسح معين، أو مندوب إحصاء معين؟)

ينبغي تطوير بروتوكول معياري يتناول إجراءات كشف أي من هذه المشكلات والتعامل معها. يشمل ذلك سلسلة من الإجراءات الصعبة والمهمة، بما في ذلك مقدار "التنقية" الواجب تطبيقها قبل التحليل (ما أنواع الأخطاء الواجب إصلاحها؟ وكيف يمكن تمييزها؟ وكيف يمكن تصحيحها؟) وينبغي أن يتم التصحيح عادة عبر ملف نصي موثق تمت مشاركته بين مسؤولي التقييم، بحيث يمكن نسخه ومراجعته إذا لزم الأمر.

عند الكشف عن ملحوظات إشكالية، ينبغي ترميزها باعتبارها قيم متطرفة، بدلاً من حذفها من مجموعة البيانات، بحيث يمكن إجراء المزيد من التحليلات وضمان اتسام عملية تصحيح البيانات بالشفافية. وينبغي التدقيق، قدر الإمكان، للكشف عن أسباب القيم المتطرفة. كما ينبغي دراسة العلاقات بين مندوبي الإحصاء والعوامل الأخرى لتحديد الأخطاء المحتملة في عملية تنفيذ المسح.

من الممارسات الجيدة أن يتم التحقق من كل عينة عشوائية من الملحوظات على حدة (بمعزل عن فريق المسح)، من خلال إعادة إجراء المسح مع المجيبين على أسئلة مُحددة. ويمكن إجراء هذا هاتفيًا، على الأقل في البداية، مقابلات مباشرة إذا لم تكن الأرقام الهاتفية لعدد كبير من المجيبين تعمل. تحقق هذه العملية الثقة في مصداقية الخطوات المتبعة في تنفيذ المسح، وينبغي إخطار مندوبي الإحصاء، ومديري المسح أنه سيتم إجراء المسح بصورة مستقلة، بحيث يحفز هذا على الالتزام بالممارسات الجيدة في تنفيذ المسح.

المراجع

Banda, J. P. 2003. Nonsampling Errors in Surveys. Paper prepared for Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys. United Nations Secretariat, New York. 3–5

December 2003.

Dunn, O. J. 1961. Multiple Comparisons among Means. *Journal of the American Statistical Association*. 56 (293). p52-64.

Grimm, P. 2010. Social Desirability Bias. In J. Sheth and N. Malhotra, eds. *Wiley International Encyclopedia of Marketing*. Part 2. Marketing Research.

Krumpal, I. 2013. Determinants of Social Desirability Bias in Sensitive Surveys:

A state-of-the-art review. *Quality and Quantity.* 47 (4). pp. 2025–2047.

https:// doi.org/10.1007/s11135-011-9640-9.

McKenzie, D. 2012. Beyond Baseline and Follow-Up: The Case for More T in Experiments. *Journal of Development Economics*. 99. pp. 210–221.

Romano, J. P. and M. Wolf. 2005. Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Statistical Association*. 100. pp. 94–108. doi: 10.1198/016214504000000539.

Westfall, P. H. and S. S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* John Wiley & Sons.

تقييم أثر التدخلات الإنمائية
دليل عملي

تقييم الأثر هي مقاربة تجريبية تهدف إلى تقدير التأثيرات السببية للتدخلات، من حيث كل من حجم التأثير والدلالة الإحصائية. بيد أن الاستخدام الموسع لتقنيات تقييم الأثر يمثل أهمية بالغة، نظرًا لتوفيره معلومات جيدة عن عمليات التنمية. إضافة إلى ذلك، فإنه يجعل الاستثمارات والسياسات التنموية أكثر اعتمادًا على الأدلة وأكثر فعالية. وللمساعدة في دعم زيادة تطبيق مقاربات تقييم الأثر، يقدم هذا الكتاب المفاهيم والطرق والاعتبارات الأساسية المُستَخدمة عند التخطيط لعملية تقييم الآثار، وتصميمها، وإدارتها، وتطبيقها، مع دعم ذلك بالأمثلة. ومن بين الموضوعات التي تم تناولها أهداف تقييم الآثار، والمبادئ الأساسية، والمنهجيات الخاصة، وبعض التوجيهات بخصوص التنفيذ الميداني. ويشتمل الكتاب على معلومات تناسب مختلف أنواع جماهير القراء، بدءًا من هؤلاء المهتمين بفهم الأدلة على "كل ما يصلح" في التنمية، ووصولاً إلى من سيساعدون في تطوير قاعدة الأدلة، مثل الباحثين التطبيقيين.

حول بنك التنمية الآسيوي

تتجسد رؤية بنك التنمية الآسيوي في القضاء على الفقر في منطقة آسيا والمحيط الهادئ. وتتمثل مهمة البنك في مساعدة البلدان النامية الأعضاء على الحد من الفقر، وتحسين الظروف المعيشية وجودة حياة شعوبها. وعلى الرغم من كثرة قصص النجاح في المنطقة، فلا تزال موطنًا لنسبة كبيرة من الفقراء على المستوى العالمي. يلتزم بنك التنمية الآسيوي بالعمل على الحد من الفقر من خلال التنمية الاقتصادية الشاملة، والتنمية المستدامة البيئية، والتكامل الإقليمي.

يقع مقر بنك التنمية الآسيوي في مانيلا، ويمتلكه ٦٧ عضوًا، من بينهم ٤٨ من المنطقة. وتتمثل أدواته الرئيسية لمساعدة أعضائه من البلدان النامية في عقد حوارات حول السياسات، وتقديم القروض والاستثمارات في رأس المال، وعرض الضمانات والمنح، فضلاً عن توفير المساعدة التقنية.

ASIAN DEVELOPMENT BANK
6 ADB Avenue, Mandaluyong City
1550 Metro Manila, Philippines
www.adb.org

"تقييم الأثر هو مقاربة تجريبية لتقدير الآثار الناتجة عن التدخلات، من حيث الحجم والأهمية الإحصائية. ويعد التوسع في استخدام تقنيات تقييم الأثر أمرا بالغ الأهمية لاستخلاص المعارف بدقة من العمليات الإنمائية لكي تصبح الاستثمارات والسياسات الإنمائية أكثر فعالية وقائمة على الأدلة.

وللمساعدة في دعم المزيد من استخدام مقاربات تقييم الأثر، يقدم هذا الكتاب مفاهيم وأساليب واعتبارات أساسية لتخطيط وتصميم وإدارة وتنفيذ تقييم الأثر، مع أمثلة توضيحية .

وتتراوح مواضيع الدليل بين أغراض تقييم الأثر والمبادئ الأساسية والمقاربات المحددة وإلإرشادات المتعلقة بالتنفيذ الميداني، كما يحتوي الدليل على مواد لمجموعة من المستفيدين، من مهتمين بفهم الأدلة حول "ما ينجح" في التطوير، إلى أولئك الذين يساهمون في توسيع قاعدة الأدلة كباحثين تطبيقيين.

محتويات الكتاب :

مقدمة: تقييم الأثر من أجل التنمية القائمة على الأدلة
استخدام نظريات التغيير لتحديد أسئلة تقييم التأثير
المفاهيم الأساسية لتقييم الأثر
التجارب العشوائية المضبوطة
التصاميم غير التجريبية
ماذا وكيفية القياس: جمع البيانات لتقييم التأثير
تحديد حجم العينة لجمع البيانات
إدارة عملية تقييم الأثر

التقييم

المتابعة والتقييم

قضايا عامة في المتابعة والتقييم

مركز لأثر
IMPACT CENTER
مستقبل أفضل
BETTER FUTURE

رقم الإيداع: ١٤٤٦/٩٥٦٧
ردمك: ٦-٣-٩٢٢٦٥-٦٠٣-٩٧٨