



REVIEW ARTICLE

Application of Randomized Controlled Trial in Evidence-Based Policymaking

Fatemeh Morovvti¹, Zeynab Armaniyan^{2*} , Mohammad Hassan Ebrahim Kani³

1. Master of Cognitive Sciences, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran

Email: f.morovvati@ut.ac.ir

2. Master of General Psychology, Faculty of Educational Sciences and Psychology, University of Shahid Beheshti, Tehran, Iran

* Corresponding Author's Email: Zeynab.armaniyan@gmail.com

3. PhD student of Cognitive Sciences, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran

Email: mh.ebrahimkani@ut.ac.ir



<https://doi.org/10.22059/jppolicy.2025.103607>

Received: 22 April 2025
Accepted: 25 August 2025

ABSTRACT

Every day, various policies are adopted and implemented to improve conditions and achieve goals; however, there is often little solid evidence regarding their effectiveness. In response to this issue, the approach of evidence-based policymaking—particularly through the establishment of behavioral insights teams—is expanding, where policymakers assess sufficient evidence on the effectiveness of policies before implementation and make decisions accordingly. One of the rigorous methods for this evaluation is Randomized Controlled Trials (RCTs). This study, conducted through a narrative review method, examines the capacity of RCTs, along with providing practical guidance for implementation, demonstrates that adhering to three key features—using large samples, considering a control group, and applying random assignment—enables causal inference through this method. Proper application and compliance with the requirements of this method can lead to benefits such as cost savings and greater policy impact, which is why RCTs have become the gold standard in this field.

Keywords: Behavioral Insights, Behavioral Policymaking, Evidence-Based Legislation, Evidence-Based Policymaking, Randomized Controlled Trial.

Citation: Morovvti, Fatemeh; Armaniyan, Zeynab; Ebrahim Kani, Mohammad Hassan (2025). Application of Randomized Controlled Trial in Evidence-Based Policymaking. *Iranian Journal of Public Policy*, 11 (3), 225-239.

DOI: <https://doi.org/10.22059/jppolicy.2025.103607>

Published by University of Tehran



This Work Is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)



مقاله مروری

کاربست روش کارآزمایی تصادفی کنترل شده در خطمشی گذاری شواهدمحور

فاطمه مروتی^۱، زینب آرمانیان^{۲*}، محمدحسن ابراهیم کنی^۳

۱. کارشناسی ارشد علوم شناختی، دانشکده روانشناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران
رایانامه: f.morovvati@ut.ac.ir

۲. کارشناسی ارشد روانشناسی عمومی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید بهشتی، تهران، ایران
* رایانامه نویسنده مسئول: Zeynab.armanivan@gmail.com

۳. دانشجوی دکتری علوم شناختی، دانشکده روانشناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران
رایانامه: mh.ebrahimkani@ut.ac.ir

 <https://doi.org/10.22059/jppolicy.2025.103607>

تاریخ دریافت: ۲ اردیبهشت ۱۴۰۴
تاریخ پذیرش: ۳ شهریور ۱۴۰۴

چکیده

هر روزه خطمشی‌های مختلفی به منظور بهبود وضعیت و دستیابی به اهداف تصویب و اجرا می‌شوند، اما عمدتاً شواهد دقیقی از اثربخشی آن‌ها وجود ندارد. در پاسخ به این مشکل، رویکرد خطمشی‌گذاری شواهدمحور، به‌ویژه در قالب مراکز بینش‌های رفتاری، در حال گسترش است که در آن خطمشی‌گذاران پیش از اجرا، شواهد کافی درباره اثربخشی خطمشی‌ها را بررسی و طبق آن تصمیم‌گیری می‌کنند. یکی از روش‌های دقیق این ارزیابی، کارآزمایی‌های تصادفی کنترل شده (RCTs) است. این پژوهش که با روش مرور روایی انجام شده، به بررسی ظرفیت RCTها پرداخته و در کنار ارائه راهنمایی عملی برای اجرای آن، نشان داده رعایت سه ویژگی استفاده از نمونه‌های بزرگ، در نظر گرفتن گروه شاهد و تخصیص تصادفی، سبب فراهم شدن استنباط علی از این روش می‌شود. استفاده اصولی و رعایت الزامات این روش می‌تواند منجر به مزایایی مانند صرفه‌جویی در هزینه‌ها و اثرگذاری بیشتر شود و به همین خاطر RCTها به استاندارد طلایی در این زمینه تبدیل شده‌اند.

واژگان کلیدی: بینش‌های رفتاری، خطمشی‌گذاری رفتاری، خطمشی‌گذاری شواهدمحور، قانون‌گذاری شواهدمحور، کارآزمایی تصادفی کنترل شده.

استناد: مروتی، فاطمه؛ آرمانیان، زینب؛ ابراهیم کنی، محمدحسن (۱۴۰۴). کاربرد روش کارآزمایی تصادفی کنترل شده در خطمشی‌گذاری شواهدمحور. فصلنامه سیاستگذاری عمومی، ۱۱ (۳)، ۲۲۵-۲۳۹.

DOI: <https://doi.org/10.22059/jppolicy.2025.103607>



ناشر: دانشگاه تهران

مقدمه

بسیاری از خط‌مشی‌ها و برنامه‌ها پیش از آنکه مورد ارزیابی قرار گیرند و شواهد کافی در مورد اثربخشی آن‌ها وجود داشته باشد، به‌طور گسترده و در سطح یک کشور و جهان اجرا می‌شوند؛ زیرا خط‌مشی‌گذاران عموماً طبق استدلال‌های منطقی و نظرات و برداشت‌های شهودی خود این‌طور جمع‌بندی می‌کنند که کارآمد و موثر خواهند بود. اما برخی از همین خط‌مشی‌ها پس از گذشت مدت زمانی از اجرایی شدن و صرف منابع و هزینه‌های زیاد، نتیجه مطلوب را حاصل نمی‌کنند و یا اثر کافی ندارند یا در حالت بدتر آسیب‌ها و مشکلاتی جدید ایجاد می‌کنند (Haynes, Goldacre & Torgerson, 2012; Wilson & Juarez, 2015). برای مثال، برنامه ارباب مستقیم^۱ برنامه‌ای دولتی با هدف کاهش سطح جرم و جنایت در میان نوجوانان آسیب‌پذیر در آمریکا بود. ایده‌ی طرح این بود که نوجوانان با واقعیت‌های زندگی مجرمین آشنا شوند تا با دیدن عواقب رفتارهای مجرمانه، از این رفتارها دوری کنند. اما پس از مدت‌ها، شواهدی از آزمایش‌های انجام‌شده به دست آمد که نشان می‌داد این مداخلات نه تنها بر روی پیشگیری از جرم موثر نبودند، بلکه سبب افزایش جرم در نوجوانان می‌شدند (Halpern, 2015). با نگاه به این مثال و مثال‌های مشابه، می‌توان اینگونه نتیجه گرفت که خط‌مشی‌گذاری شواهدمحور به‌علت تکیه بر شواهد پژوهشی از اثربخشی برنامه‌ها و خط‌مشی‌ها، بهترین روشی است که برای اجتناب از اشتباه در اتخاذ خط‌مشی‌ها در حوزه‌های مختلف وجود دارد و سبب می‌شود از اتلاف منابع جلوگیری و خط‌مشی‌های کارآمدی وضع شود؛ در واقع ممکن است در اثر اجرای خط‌مشی‌های بدون شواهد، میلیاردها تومان صرف برنامه‌هایی شود که هیچ تأثیری ندارند و یا صدمات و آسیب‌هایی از روی ناآگاهی به انسان‌ها و جامعه وارد شود (Wilson & Juarez, 2015). شواهد قابل استفاده جهت خط‌مشی‌گذاری موثر، انواع متنوعی دارند و کارآزمایی‌های تصادفی کنترل‌شده (RCTs) یکی از قابل استنادترین انواع شواهد را فراهم می‌کنند و بر این اساس در این پژوهش، صرفاً به RCTها پرداخته می‌شود. اجرای کارآزمایی‌های تصادفی کنترل‌شده پیش از تصمیم‌گیری پیرامون تصویب قوانین و خط‌مشی‌ها سبب می‌شود قوانین موثر شناسایی شوند و بودجه به سمت موثرترین قوانین و خط‌مشی‌ها که بازدهی بالاتری دارند، هدایت شود و همچنین مبتنی بر شواهد به دست آمده، قوانینی که اثربخش نیستند و نتایج مورد انتظار را ارائه نمی‌دهند، شناسایی و حذف شوند. مطالعه حاضر با هدف نشان دادن کاربرد کارآزمایی‌های تصادفی کنترل‌شده در خط‌مشی‌گذاری شواهدمحور تهیه شده‌است. یک اصل اساسی در خط‌مشی‌گذاری خوب این است که تصمیمات براساس شواهد قوی و آنچه از موثر بودن آن اطمینان وجود دارد، گرفته شود. تصمیم‌گیری مبتنی بر شواهد^۲ یا خط‌مشی‌گذاری مبتنی بر شواهد^۳ تأکید می‌کند که خط‌مشی‌ها باید مبتنی بر بهترین شواهد موجود از تحقیقات و آزمایش‌ها اتخاذ شوند تا پیش از تخصیص منابع این اطمینان حاصل شود که اولاً آن خط‌مشی در زمینه مدنظر امکان اجرایی شدن دارد، ثانیاً مورد پذیرش واقع می‌شود و نهایتاً مؤثر و کارآمد است (Haby & Sieber, 2021). در مقابل رویکرد شواهدمحور، رویکرد خط‌مشی‌گذاری شهودی^۴ وجود دارد که عموماً مبتنی بر حدس و گمان است و طبق آن تصمیم‌گیری در مورد انتخاب خط‌مشی‌ها و برنامه‌ها براساس شهود شخصی خط‌مشی‌گذار انجام می‌شود. با توجه به سوگیری شناختی «خودتاییدی»^۵ وقتی خط‌مشی یا برنامه‌ای اجرا شد، تمایل وجود دارد که فقط موارد موفقیت‌آمیز آن دیده شود و مواردی که اثرگذار نبوده‌اند و یا حتی تأثیر معکوس داشته‌اند استثنا و غیر پایدار تلقی شوند؛ زیرا طبق این سوگیری افراد تمایل دارند مواردی را بپذیرند که منطبق با باورهای قبلی آنها بوده است و مواردی که منطبق نیستند را رد کنند (Halpern, 2015). همچنین، رویکرد

1. Scared Straight
2. Randomized Controlled Trial
3. Evidence-Based Decision Making
4. Evidence-Based Policymaking
5. Intuition-Based Policy
6. Confirmation Bias

«خط‌مشی‌گذاری مبتنی بر شواهد» در نقطه مقابل رویکرد «شواهد مبتنی بر خط‌مشی یا شواهد خط‌مشی محور»^۱ نیز قرار دارد که در آن شواهد به صورت انتخابی و برای حمایت از یک خط‌مشی از پیش تعیین شده استفاده می‌شوند (Haby & Sieber, 2021).

معرفی کارآزمایی‌های تصادفی کنترل شده

کارآزمایی‌های تصادفی کنترل شده، آزمایش‌هایی به حساب می‌آیند که به محققان این امکان را می‌دهند تا مشخص کنند به چه میزان یک مداخله منجر به یک نتیجه مشخص می‌شود. با استفاده از این روش، افراد می‌توانند اطمینان پیدا کنند که یک چیز علت چیز دیگر است که به آن استنباط علی گفته می‌شود. این استنباط علی بر اساس سه ویژگی ساده ولی اساسی RCT حاصل می‌شود: (۱) استفاده از نمونه‌های بزرگ (۲) استفاده از گروه شاهد (گروهی که مداخله را دریافت نمی‌کند) و (۳) تخصیص تصادفی (به معنای اختصاص دادن تصادفی نمونه به گروه شاهد و گروه(های) مداخله). با اعمال درست این ویژگی‌ها، می‌توان انتظار داشت که دو (یا چند) گروه در ویژگی‌های مختلف آشکار و نهان، کاملاً شبیه به یکدیگر باشند و در این شرایط وقتی مداخله‌ای بر روی یکی از این گروه‌ها اجرا شود، می‌توان اینطور نتیجه‌گیری کرد که هر اثر مشاهده‌شده در این گروه، الزاماً برگرفته از مداخله‌ی اجرا شده است؛ نه هر علت دیگری (Hansen, 2019; Hallsworth & Kirkman, 2020).

پیشینه پژوهشی

ریشه استفاده از RCT در خط‌مشی‌های عمومی به دهه ۱۹۲۰ برمی‌گردد. پیشرفت‌های اولیه در استفاده از کارآزمایی‌های تصادفی کنترل شده در مطالعه خط‌مشی‌های عمومی همزمان با استفاده آن توسط روان‌شناسی برای مطالعه رفتار انسان بود. در ابتدای دهه ۱۹۲۰ سیاست‌گذاران شروع به استفاده از آزمایش کردند و دانش آن را منتشر ساختند؛ مانند مطالعاتی برای بررسی تاثیر ارائه شیر در مدارس بر سلامتی کودکان. اما این مطالعات در ابتدا با شکست‌هایی مواجه شد که ناشی از دانش ناکافی از اصول اجرای آزمایش و سختی‌های اجرای آن بود، با این حال این مطالعات تا چند دهه بعد و حدوداً تا سال ۱۹۳۹ ادامه پیدا کردند و بعد از آن دوباره بی‌استفاده ماندند (John, 2020). علاقه دوباره به استفاده از آزمایش‌های میدانی به دهه ۱۹۶۰ و مطالعات کمپل بر روی ویژگی‌های آماری آزمایش‌ها برمی‌گردد. گسترش خط‌مشی‌های رفاهی در دهه ۱۹۶۰ تقاضا برای ارزیابی محکم‌تر خط‌مشی‌ها را بیشتر کرد. در دهه‌های ۱۹۷۰ و ۱۹۸۰ مطالعات زیادی در حوزه‌هایی چون بهزیستی، مالیات، محیط کار، جرم و جنایت و اشتغال با تمرکز بر دولت آمریکا صورت گرفت. یکی از رویدادهای مهم در این تاریخچه، قانون فدرال اصلاحات علوم آموزشی در سال ۲۰۰۲ بود که در آن کنگره آمریکا بر استفاده از کارآزمایی‌های تصادفی‌سازی شده تاکید کرد و در نتیجه باعث رشد قابل توجه کارآزمایی‌هایی آموزشی در این کشور شد. علاقه فزاینده به RCT باعث شد تا این ایده پیشنهاد شود که آزمایش‌های سیاستی و اجتماعی، راهی موثر برای بهبود حکمرانی است (John, 2020). به کارگیری RCT‌ها در خط‌مشی‌گذاری‌های اجتماعی جهان در سال‌های اخیر رشد بی‌سابقه‌ای پیدا کرده است و در موضوعات مختلف اعم از جرم و عدالت، آموزش، رفاه اجتماعی، بهداشت، آموزش و غیره اجرا می‌شود. طبق بررسی‌های کمپل^۲ از RCT‌های باکیفیت انجام شده در جهان تا سال ۲۰۱۴، آمریکای شمالی در تولید RCT‌های سیاستی پیشتاز و دارای رتبه اول در جهان است و کانادا، انگلستان و استرالیا در رتبه‌های بعدی قرار دارند. قابل توجه اینکه در این لیست سه کشور در حال توسعه (هند، کنیا و مکزیک) نیز هستند که در خط‌مشی‌گذاری‌های خود از روش RCT استفاده می‌کنند (Ames & Wilson, 2016). همچنین از سال ۲۰۱۱ تا ۲۰۱۵ در انگلستان مجموعه‌ای از موسسات و مراکز دولتی و خصوصی برای ایجاد «اکوسیستم شواهد»^۳ و تقویت استفاده از شواهد توسط سیاست‌گذاران و شاغلین تاسیس شد.

1. Policy-Based Evidence
2. Causal Inference
3. Campbell
4. Evidence Ecosystem

هدف این مراکز، گردآوری و ایجاد شواهدی از برنامه‌ها و اقدامات کارآمد و سپس ارائه این شواهد به مسئولین و افراد است؛ این مراکز در جهان، مراکز «what works» نامیده می‌شوند. در مجموع، اکوسیستم شواهد، شواهد را تولید، گردآوری و طبق نیاز خط‌مشی‌گذاری ترجمه می‌کند و نهایتاً کمک می‌کند تا شواهد به بهترین نحو در خط‌مشی‌گذاری مورد استفاده قرار گیرند (Halpern, 2015). در سال‌های اخیر، تیم‌های بینش‌های رفتاری با تکیه بر رویکرد خط‌مشی‌گذاری رفتاری، پیش‌تاز تکیه بر خط‌مشی‌گذاری شواهدمحور و اجرای کارآزمایی‌های تصادفی کنترل‌شده هستند. به‌عنوان مثال تیم علوم اجتماعی و رفتاری آمریکا که در سال ۲۰۱۴ تأسیس شده‌است، مجری مداخلات خط‌مشی‌گذاری رفتاری در دولت آمریکا به‌شمار می‌آید و پیشرو در امر ارزیابی خط‌مشی و انجام RCT است (Haby & Sieber, 2021). همچنین شناخته‌شده‌ترین مرجع انجام RCT در دنیا، تیم بینش‌های رفتاری انگلستان است که در سال ۲۰۱۰ در دفتر نخست‌وزیری به‌عنوان اولین موسسه بینش‌های رفتاری در جهان شروع به کار کرد و طبق گزارش‌ها تا سال ۲۰۲۰ بیش از ۵۰۰ کارآزمایی تصادفی کنترل‌شده در حیطه‌های سیاستی مختلف طراحی و اجرا کرده‌است (Pew, 2015). از دیگر مراکز می‌توان به آزمایشگاه اقدام برای فقر عبداللطیف جمیل^۳ اشاره کرد که تا سال ۲۰۱۴، بیش از ۷۰۰ طرح RCT در حوزه مقابله با فقر و ارتقای رفاه طراحی و پیاده کرده‌است (Haby & Sieber, 2021). در جدول شماره ۱ برخی RCTها که در کشورهای پیشرو در این زمینه انجام شده‌اند به طور مختصر و به عنوان نمونه آورده شده‌است.

جدول ۱. نمونه‌هایی از کاربرست کارآزمایی تصادفی کنترل شده در جهان

شماره	سال	کشور	نام و هدف طرح	روش	نتیجه و اثر	منبع
۱	۲۰۰۳	انگلیس (وزارت کار و بازنشتی)	اثربخشی سه طرح متفاوت با هدف افزایش برگشت به کار بازنشستگان و نهایتاً انتخاب موثرترین طرح	اجرای RCT و مقایسه‌ی سه طرح بازگشت به کار	موثر نبودن هیچ یک از طرح‌ها به طور معنی‌دار و جلوگیری از هزینه‌های زیاد و اعطای مزایای بی‌اثر پیش از اجرا در سطح کلان و صرفه‌جویی در هزینه‌ها	Purdon, et al., 2006
۲	۲۰۰۷	آمریکا	بررسی اثرات و مزایای حضور و مشارکت پرستار خانواده در کنار مادر از زمان بارداری تا دوسالگی کودک	انجام RCTهای متعدد و مقایسه‌ی وضعیت حضور پرستاران با وضعیت عدم حضور آن‌ها	رشد عاطفی-اجتماعی، سلامت و پیشرفت بیشتر در کودکان با پرستار خانواده و ارتکاب جرم کمتر در بین آن‌ها در سنین بالاتر؛ و اجرا در سطح کلان و سرمایه‌گذاری بر روی این طرح	MacMillan, et al., 2009
۳	۲۰۱۳	هند	بررسی اثرات اعطای وام‌های خرد کارآفرینی به کارآفرینان	اجرای مجموعه RCTهایی بر روی وام‌گیرندگان خرد از بانک‌های خصوصی در محله‌های فقیرنشین	سرمایه‌گذاری بیشتر و خرید کالای بادوام توسط وام‌گیرندگان اما موثر نبودن وام‌ها بر تعداد کسب‌وکارها، سود شرکت‌ها و شاخص‌های مختلف توسعه. در نتیجه اعطای وام چنان‌که تصور می‌شد راه‌حل شگفت‌آوری در پیشرفت کارآفرینی نبود.	Pettersson, 2014
۴	حدوداً ۲۰۱۳	انگلیس (بنیاد وقف آموزش)	آزمایش ابتکارات نوآورانه برای بهبود پیشرفت دانش‌آموزان	RCT و مطالعات تحقیقاتی	استخراج مجموعه RCTها و نتایج آن‌ها و ارائه در سایت بنیاد جهت استفاده‌ی معلمان از طرح‌های موثر و علمی	Alexander & Letwin, 2013
۵		انگلیس (تیم بینش‌های رفتاری)	بررسی اثربخشی نامه‌ی طراحی‌شده با اصول رفتاری بر میزان پرداخت مالیات	انجام RCT	افزایش ۱۵ درصدی پرداخت مالیات و افزایش درآمد مالیاتی دولت به میزان ۳۰ میلیون پوند	Behavioural Insights Team, 2012
۶		انگلیس (تیم بینش‌های رفتاری)	بررسی اثربخشی پیامک‌های یادآور و طراحی‌شده با اصول رفتاری بر پرداخت به موقع عوارض	اجرای RCTهای مختلف و متعدد	موثر و سبب افزایش پرداخت به موقع عوارض و اضافه شدن ۸۶۰ هزار پوند در هفته به درآمد صندوق دولت	Haynes, Green, Gallagher, John & Torgerson, 2013

1. Social and Behavioral Sciences Team (SBST)
2. Behavioral Insights Team (BIT)
3. Abdul Latif Jameel Poverty Action Lab (J-PAL)

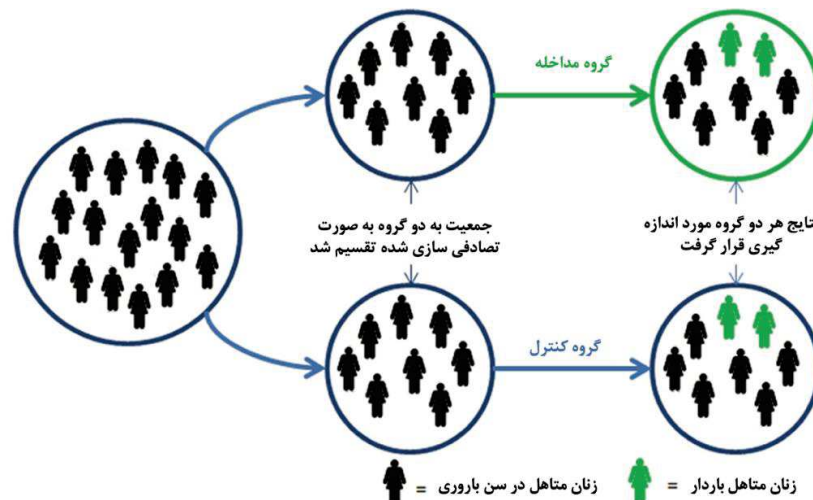
Ames & Wilson, 2016	افزایش ۳۹ درصدی ثبت نام کنندگان معادل ۱۰۰,۰۰۰ ثبت نام بیشتر در سال صرفاً به دلیل افزودن یک جمله به یک صفحه وب	اجرای RCT به صورت آنلاین	بررسی اثربخشی یک جمله‌ی طراحی شده با اصول رفتاری بر میزان ثبت نام اهدای عضو	انگلیس	۲۰۱۳	۷
Zakani, Ebrahim Kani, Morovvati, Givzad & Ghaderi, 2021	موثر بودن پیامک‌ها و افزایش ۳۷ درصدی تزریق دوز دوم واکسن و افزایش ۳۰ درصدی تزریق دوز سوم و سپس ارسال موثرترین پیامک به صورت انبوه با صرف هزینه‌ی بسیار کم	انجام دو RCT (برای هر دوز یک RCT)	بررسی اثربخشی پیامک‌های طراحی شده با اصول رفتاری بر میزان تزریق دوز دوم و سوم واکسن کرونا	ایران (مرکز بینش‌های رفتاری ایران و وزارت بهداشت)	۱۴۰۰	۸

باتوجه به پیشرفت‌های رویکرد خط‌مشی‌گذاری شواهدمحور در سالیان اخیر، چارچوب‌هایی برای معرفی و به‌کارگیری این رویکرد و طراحی و اجرای RCT توسط چند مورد از مراکز بینش‌های رفتاری ارائه شد؛ در این میان می‌توان به چارچوب «آزمایش، یادگیری و تطبیق» و چارچوب TESTS که به‌ترتیب در سال‌های ۲۰۱۲ و ۲۰۲۲ توسط تیم بینش‌های رفتاری ابداع شد (Haynes, Goldacre & Torgerson, 2012; Kettle & Persian, 2022) و همچنین چارچوب BASIC که توسط مرکز همکاری و توسعه اقتصادی مطرح شده، اشاره کرد (Hansen, 2019).

چستی و ضرورت بکارگیری کار آزمایی‌های تصادفی کنترل شده

در حال حاضر اغلب هنگام ارزیابی اثربخشی یک خط‌مشی (به عنوان یک مداخله)، آن سیاست اجرا و سپس با مشاهده نتایج کلی، درباره‌ی میزان اثربخشی آن قضاوت می‌شود. با این حال این رویکرد دارای مشکلات گسترده‌ای است. یکی از اشکالات این است که با این رویکرد نمی‌توان تشخیص داد آیا این سیاست مدنظر سبب ایجاد این نتایج شده است یا عامل خارجی کنترل نشده دیگری مسبب ایجاد اثرات بوده است (Haynes, Goldacre & Torgerson, 2012). برای مثال ممکن است طرح تشویق فرزندآوری در کشور اجرا شود، اما همزمان شرایط اقتصادی در کشور بهبود پیدا کند و سبب افزایش فرزندآوری شود؛ در اینجا به اشتباه رشد فرزندآوری به مداخله نسبت داده می‌شود. همچنین ممکن است عده‌ی خاصی تحت تاثیر این مداخله قرار بگیرند و همه‌ی مردم از این خط‌مشی متأثر نشوند (Haynes, Goldacre & Torgerson, 2012). در این شرایط معمولاً نتایج اثرگذاری به شکل کلی و غیردقیق و غیرعلی ارزیابی می‌شود. کار آزمایی‌های تصادفی کنترل شده برای هر یک از این چالش‌ها راهکاری ارائه داده که با رعایت آن‌ها رابطه‌ی علیت بین اجرای مداخله و دریافت نتایج برقرار می‌شود و به سیاست‌گذاران اطمینان می‌دهد که نتیجه‌ی حاصل شده، ناشی از مداخله است و نه موارد دیگر (White, Sabarwal & De Hoop, 2014). این راهکارها که بیانگر ویژگی‌های اصلی و تعریف مفهومی یک طرح کار آزمایی تصادفی کنترل شده هستند، عبارتند از: اولاً «انتخاب تصادفی» نمونه‌ای معرف جامعه هدف و تقسیم جمعیت به دو یا چند گروه بر اساس «تخصیص تصادفی»؛ ثانیاً ارائه یک مداخله به یک گروه (یا ارائه چند مداخله به چند گروه به طور مجزا) و عدم ارائه مداخله به گروهی دیگر یا همان گروه کنترل؛ و نهایتاً اندازه‌گیری پیامدهای ازپیش تعیین شده (متغیر وابسته) برای هر گروه و مقایسه آن‌ها (Haynes, Goldacre & Torgerson, 2012). اگر برای مثال یکی از خط‌مشی‌های مربوط به طرح فرزندآوری به‌صورت طرح RCT اجرا شود و فرضاً نتایج طبق شکل ۱ حاصل شود، بدین معناست که اجرای سیاست فرزندآوری مدنظر از عدم اجرای آن بهتر نیست. از این نتیجه فهم می‌شود که اگر هیچ گروه کنترلی وجود نداشت، ممکن بود از همان تعدادی که اقدام به فرزندآوری کرده‌اند به اشتباه اینگونه نتیجه‌گیری شود که طرح جدید موثر است و همین خطای ارزیابی سبب می‌شد طرحی پرهزینه و ناکارآمد برای مدت طولانی در کشور اجرایی و سبب اتلاف منابع گردد. در حالیکه در این مثال فرضی، گروهی که مداخله بر روی آن‌ها اجرا نشده نیز همین میزان فرزندآوری را داشته‌اند و

تفاوت معناداری بین دو گروه وجود ندارد. اما اگر فرضاً میزان فرزندآوری در گروه مداخله نسبت به گروه کنترل، به لحاظ آماری، دارای اختلاف معنی‌دار باشد، می‌توان گفت مداخله اثرگذار بوده و در نتیجه با اطمینان می‌توان این سیاست را در مقیاس بزرگ اجرا کرد.



شکل ۱ - کاربست کارآزمایی تصادفی کنترل شده (RCTs) در دو گروه کنترل و مداخله (عدم موفقیت مداخله)

به عنوان نمونه، یکی از حوزه‌های موضوعی که تاکنون RCT‌های زیادی برای اثربخشی سیاست‌ها در آن صورت گرفته است، حوزه مصرف انرژی است (Carroll, Lyons & Denny, 2013; Houde, Todd, Sudarshan & Armel, 2013). در یکی از این طرح‌ها که عموماً با هدف سنجش تاثیر سیاست‌های کاهش مصرف انرژی چون استفاده از کتورهای هوشمند، بهینه‌سازی وسایل سرمایشی و گرمایشی، و ارسال قبوض انجام می‌گیرند، شرکت خدمات انرژی آپاور^۱ با همکاری چندین شرکت برق محلی در ایالات متحده آمریکا و در سال ۲۰۱۱، طی مجموعه RCT‌هایی در پی آن بود تا اثرگذاری مداخلات و سیاست‌هایی برای کاهش مصرف انرژی را در چند ایالت چون کالیفرنیا، واشینگتن و تگزاس ارزیابی کند (Alcott, 2011). در این طرح بیش از ۶۰۰ هزار خانوار که به لحاظ متغیرهای مختلف چون درآمد، نژاد، سطح تحصیلات و منطقه جغرافیایی گستره وسیعی را پوشش می‌دادند، مشارکت داشتند و گروه مداخله گزارش‌های مصرف خانگی را دریافت می‌کرد که شامل ویژگی‌هایی چون مقایسه مصرف خانوار با همسایگان خود، پیام‌های انگیزشی و توصیه‌هایی رفتاری برای مدیریت مصرف می‌شد و به گروه کنترل صرفاً قبض برق معمولی را دریافت می‌کرد. این بازخوردها به صورت ماهانه یا دوماهه ارسال می‌شد و مداخله در بیشتر مناطق بین ۶ تا ۱۲ ماه ادامه یافت. نتایج مطالعه حاکی از آن بود که گروه مداخله به طور متوسط در ایالت‌های مختلف، حدود ۲٪ برق کمتری مصرف کرده است و در مناطقی که پیام‌ها به صورت مداوم و بلندمدت ارسال می‌شدند، این کاهش مصرف به ۳٪ رسیده بود. در برخی مناطق حتی پس از پایان طرح نیز کاهش مصرف ادامه داشت و تا ۱۲ ماه پس از انجام آخرین مرحله از مداخله مشاهده می‌شد. موفقیت‌های این شرکت به حدی بود که طبق گزارش Renewable Energy World (2010)، باراک اوباما، رئیس جمهور وقت ایالات متحده آمریکا، در سخنرانی خود در دفتر آپاور بر اهمیت بهینه‌سازی مصرف انرژی تأکید و اقدامات این شرکت را الگویی برای سایر شرکت‌ها معرفی کرد. در مطالعه‌ای دیگر در انگلستان، تاثیر نصب دوربین بر روی لباس افسران نیروهای پلیس مورد ارزیابی از طریق روش RCT قرار گرفت (Behavioural Insights Team, 2017). این دوربین‌ها یکی از نوآوری‌های جهانی جدید در فضای بهبود عملکرد نیروهای پلیس محسوب می‌شود؛ با این حال لازم است تا کارآمدی آنها مورد ارزیابی‌های سخت‌گیرانه قرار

1. Opower

گیرد. در این مطالعه که بر روی ۱۷۴۵ نیروی پلیس در انگلستان انجام می‌شد، تاثیر همراه داشتن این دوربین‌های پوشیدنی در دو موضوع غیبت نیروهای پلیس در محل انجام وظیفه و کارآمدی عملیات توقف و بازرسی متهم (در مواردی چون کشف مواد مخدر یا اشیاء ممنونه) مورد ارزیابی قرار گرفت. نتایج حاکی از آن بود که نصب این دوربین‌ها بهبودهای معناداری از نظر آماري هم در میزان غیبت افسران و هم در نتایج توقف و بازرسی ایجاد کرده‌است. به‌صورت مشخص در یک بازه شش ماهه، افسران پلیس مجهز به این دوربین‌ها، ۳٫۳ روز غیبت کمتری داشتند (کاهش بیش از ۲۰٪) و استفاده از دوربین‌ها باعث حدوداً ۱۷٪ بهبود در نتایج مثبت فرایند توقف و بازرسی شد. در یک طرح دیگر موسوم به Save More Tomorrow تلاش شد تا تاثیر طرحی که هدف آن افزایش تدریجی پس‌انداز بازنشستگی کارکنان در آمریکا بود، مورد ارزیابی قرار گیرد (Thaler & Benartzi, 2004). در این طرح که نسخه اول آن بر روی ۱۰۰۰ نفر از کارکنان شرکت اتصالات برق کالیفرنیا^۱ صورت گرفت، از یک گروه خواسته می‌شد از همین ماه، میزان پس‌انداز بازنشستگی خود را افزایش و سهم بیشتری از حقوق خود را به این موضوع اختصاص دهند؛ در مقابل از گروه مداخله درخواست می‌شد تا موافقت خود را برای اعمال این طرح بر روی حقوق سال آینده اعلام کنند و نیازی نبود تا در کوتاه‌مدت تغییری در حقوق‌شان رخ دهد. در نتیجه این طرح که با استفاده از سوگیری شناختی زمان حال^۲ صورت گرفته بود، نرخ پس‌انداز بازنشستگی گروه دوم در طول سالیان متمادی به شکل قابل توجه و پایداری نسبت به گروه اول افزایش یافت. یک RCT می‌تواند برای مقایسه اثرگذاری روش‌های مختلف اجرای بخش‌های کوچکتر یک خطمشی یا برنامه خاص نیز استفاده شود. در واقع آزمایش بخش‌های کوچکتر یک برنامه خاص نیز سیاستگذاران را قادر می‌سازد تا به طور مستمر خطمشی مدنظر را اصلاح کنند و بر جنبه‌های خاصی از خطمشی که بیشترین تاثیر را دارد، توجه کنند (Haynes, Goldacre & Torgerson, 2012). از این رو می‌توان گفت اگر چه اصول طراحی تمام RCTها یکی است، اما عموماً در خطمشی‌گذاری‌ها دو نوع RCT را می‌توان اجرا کرد (Ames & Wilson, 2016): ۱) RCTهای بهینه‌ساز^۳ که به دنبال بهبود اثربخشی خطمشی‌های موجود هستند. ۲) RCTهای سیاستی^۴ که به دنبال آزمایش اثربخشی انواع مختلف خطمشی‌ها، برنامه‌ها یا مداخلات جدید و موجود هستند.

راهنمای عملی استفاده از ابزار کارآزمایی‌های تصادفی کنترل شده

همانگونه که در بخش پیشینه اشاره شد، مدل‌ها و مفهوم‌پردازی‌های مختلفی، عموماً توسط مراکز بینش‌های رفتاری به‌عنوان پیشروان اجرای آزمایش‌ها و ارزیابی‌های دقیق و سخت‌گیرانه در خطمشی‌گذاری، برای اجرای یک طرح RCT مطرح شده که همه آنها با گذشت از فرعیات، چارچوب یکسان و واحدی دارند. در این بخش مراحل که برای اجرای هر RCT لازم است، طبق چارچوب «آزمایش، یادگیری و تطبیق» تیم بینش‌های رفتاری انگلستان که در سال ۲۰۱۲ منتشر شد (Haynes, Goldacre & Torgerson, 2012)، مطرح می‌شود.

آزمایش

مرحله ۱: تعیین دو یا چند مداخله سیاستی جهت مقایسه اثرگذاری آنها. RCTها معمولاً زمانی انجام می‌شوند که عدم اطمینان در مورد اثربخشی دو یا چند مداخله وجود داشته باشد. اغلب RCTها برای مقایسه برنامه جدید (مداخله) در برابر برنامه فعلی انجام می‌شود تا اثر آنها با هم مقایسه گردد.

1. California Electrical Utility
2. Present Bias
3. Optimisation RCTs
4. Policy RCTs

مرحله ۲: تعریف نتیجه مدنظر برای مداخله‌ها و نحوه سنجش آن. نتیجه‌ی عینی و قابل سنجش هر چه به هدف واقعی نزدیک‌تر باشد بهتر است. اما اگر هدف نهایی قابل سنجش نباشد، می‌توان از «نتیجه جایگزین» استفاده کرد. برای مثال در ارزیابی یک خط‌مشی با هدف کاهش ارتکاب جرم، شاید سنجش میزان واقعی ارتکاب جرم پس از اجرای مداخله ممکن نباشد و در عوض می‌توان میزان افراد دستگیرشده پس از ارتکاب جرم را به عنوان نتیجه (متغیر وابسته) جایگزین انتخاب کرد. پاسخ به این سوال که در هر RCT تعریف کدام نتیجه مناسب‌تر است، بهتر است از تعامل بین متخصصین RCT که می‌دانند چه نتیجه‌ای از نظر فنی و تخصصی مناسب‌تر است و نیز سیاست‌گذاران که می‌دانند چه نوع داده‌ای قابل دسترس‌تر است و هزینه جمع‌آوری آن کمتر است، حاصل شود.

مرحله ۳: تصمیم‌گیری درباره واحد تصادفی‌سازی. پس از بررسی اینکه چه نتیجه‌ای باید اندازه‌گیری شود، لازم است مشخص شود چه کسی یا چه چیزی تصادفی شود. آنچه تصادفی می‌شود به عنوان واحد تصادفی‌سازی شناخته می‌شود. واحد تصادفی‌سازی، اغلب، افراد هستند اما ممکن است موسسات یا مناطق جغرافیایی نیز مدنظر قرار بگیرد. این سوال که در هر RCT واحد تصادفی‌سازی بهتر است چه باشد، معمولاً بستگی به ملاحظات عملیاتی آن کارآزمایی دارد و مشاوره با یک متخصص می‌تواند در این زمینه کمک کننده باشد.

مرحله ۴: تعیین حجم نمونه. برای اعتبار بیشتر نتیجه‌ی حاصل‌شده از RCT، تشخیص اثرات و تفاوت‌های ظریف (اندازه اثر کوچک) و اطمینان از اینکه نتایج، حاصل شانس و احتمال نبوده، لازم است کارآزمایی با حجم نمونه کافی اجرا گردد و پیشنهاد می‌شود در صورت عدم افزایش زیاد هزینه، حجم نمونه بزرگتری انتخاب شود. برای این کار می‌توان از نرم‌افزارهای آماری کمک گرفت.

مرحله ۵: تخصیص هر واحد تصادفی‌سازی شده به یکی از مداخلات خط‌مشی. تخصیص «تصادفی» واحدها به گروه‌های مداخله و کنترل، RCT را از سایر ارزیابی‌های خط‌مشی متمایز می‌کند و به دلیل اینکه این گروه‌ها در تمام خصوصیت‌های آشکار و نهان (مانند وضعیت اقتصادی-اجتماعی، رخداد‌های اجتماعی یا شخصی و غیره) با یکدیگر یکسان هستند، امکان مقایسه موثر نتایج بین آنها فراهم می‌گردد.

مرحله ۶: اجرای مداخلات بر روی گروه‌های مختلف. پس از تخصیص تصادفی، مداخله به گروه‌های مدنظر ارائه می‌شود. در این مرحله نظارت بر اجرای دقیق مداخله برای اطمینان از ارائه مطلوب و استاندارد ضروری است.

یادگیری

مرحله ۷: اندازه‌گیری نتایج و تعیین تاثیر مداخله. نتایج باید در زمان و به روشی که از قبل تعیین شده بود (مرحله ۲)، اندازه‌گیری شوند. این زمان بستگی به سرعت تأثیر مداخله دارد؛ به طور مثال پیامکی که صبح روز انتخابات جهت تشویق به رای‌دهی ارسال می‌گردد، می‌تواند پس از ۲۴ ساعت اثر آن بررسی شود؛ اما بررسی اثر مداخله‌ای که جهت تشویق به فرزندآوری اعمال می‌شود به بازه‌ای در حدود یک سال نیازمند است. به جز بررسی نتیجه‌ی اصلی، جمع‌آوری داده‌های فرآیندی و جانبی نیز می‌تواند مفید باشد و بینش‌ها و فرضیات جدیدی را به دست دهد.

تطبیق

مرحله ۸: اجرای مداخله در مقیاس بزرگ. طبق نتایج حاصل شده، می توان مداخله را بر روی مقیاس بزرگ اجرا کرد. از نظر عموم افراد، کسب نتایج مثبت مطلوب تر است، اما نتایج منفی یا بی اثر نیز ارزشمندند. کارآزمایی که بی اثر باشد یا اثر منفی مداخله را نشان دهد، به همان اندازه ارزشمند است که اثربخش بودن مداخله را نشان می دهد؛ زیرا از اتلاف و هدررفت منابع مالی، انسانی و زمانی جلوگیری می کند و سبب استفاده از این منابع در راستای مداخلات اثرگذارتر می شود. انواع یافته های مثبت، منفی یا بی اثر حاصل از کارآزمایی باید ثبت و منتشر شوند تا سایر نهادهای سیاستی بتوانند از آن بهره گیرند.

مرحله ۹: بازگشت به مرحله اول و بهبود درک خود از آنچه موثر است. RCTها باید به عنوان بخشی از فرآیند مستمر نوآوری و بهبود خط مشی ها در نظر گرفته شوند. تکرار نتایج و استفاده از یافته های آزمایشی مستمر برای شناسایی راه های جدید برای بهبود دائمی نتایج مفید است. بهبود مستمر خط مشی ها مهمترین جنبه این روش است.

مزایای استفاده از کارآزمایی های تصادفی کنترل شده

کارآزمایی های تصادفی کنترل شده امروزه در جهان در حوزه های متنوعی اعم از خط مشی گذاری، توسعه بین المللی، تجارت و غیره مورد استفاده قرار می گیرند تا حدی که بعضاً انجام ندادن آنها کاری غیرمعمول و اقدامی بی پروا تلقی می شود (Haynes, Goldacre & Torgerson, 2012). در این بخش به برخی از مزیت ها و ویژگی های RCTها می پردازیم که به نظر می رسد سبب شده تا اجرا و به کارگیری آنها تا این اندازه گسترش یابد:

(۱) کم هزینه و مقرون به صرفه بودن: هزینه اجرای RCT به نحوی طراحی آن ها بستگی دارد، اما اغلب کم هزینه هستند و با برنامه ریزی اصولی می توان آنها را با هزینه پایین طراحی کرد. از جهت اینکه نتیجه این آزمایش ها می تواند نشان دهنده این باشد که یک خط مشی یا برنامه ارزش هزینه کردن دارد یا خیر. صرف هزینه ای بسیار کمتر از هزینه های کلان خط مشی اصلی برای ارزیابی تأثیرات آن بر روی نمونه کوچک تر، قطعاً منطقی و مقرون به صرفه خواهد بود؛ زیرا می تواند از صرف هزینه و بودجه در محل های ناکارآمد جلوگیری و به تخصیص آن به کارآمدترین برنامه، در بین چند برنامه موثر، کمک کند. به طور کلی بهتر است به جای در نظر گرفتن هزینه های اجرای یک RCT به هزینه های عدم اجرای RCT توجه کرد (Haynes, Goldacre & Torgerson, 2012).

(۲) سرعت قابل قبول اجرا: سرعت اجرای RCT وابسته به معیاری است که برای سنجش اثربخشی نتیجه یک مداخله انتخاب شده است. اگر معیار، بلندمدت باشد طبعاً زمان بیشتری برای بررسی نتایج لازم است، اما همیشه نیاز نیست نتایج طولانی مدت سنجیده شود، بلکه در مواردی که نتایج کوتاه مدت یا میان مدت ارتباط نزدیکی با نتایج بلندمدت دارد، می توان از آن ها استفاده کرد و پاسخ سریع تری از نتایج دریافت کرد (Buck & McGee, 2015). اغلب اوقات بررسی نتایج در زمان کوتاهی امکان پذیر است و از این جهت RCTها دارای سرعت اجرا و بررسی مطلوبی هستند و به دلیل قابلیتی که در بررسی نتایج در زمان های مختلف دارند به قدرت خط مشی گذاران در تشخیص برنامه ها و خط مشی های موثر کمک می کنند (Buck & McGee, 2015; Ames & Wilson, 2016).

(۳) سادگی و سهولت اجرا: برخلاف تصویری که ممکن است در ابتدا ایجاد شود، معمولاً طراحی و اجرای RCTها دشوار و پیچیده نیست و اگر در مرحله طراحی دقت لازم جهت رعایت اصول و اعمال نکات جزئی صرف شود، با پشتیبانی متخصصان و همکاری مسئولین اجرایی، می‌توان به سادگی و با سهولت آنها را عملیاتی کرد (Haynes, Goldacre & Torgerson, 2012).

(۴) قابلیت تعمیم نتایج: تعمیم‌پذیری یافته‌های پژوهشی به دلیل تفاوت‌های موقعیت و زمانی همواره یک چالش است. اما یک RCT به دلیل انتخاب تصادفی نمونه و تخصیص تصادفی آنها تا حد مطلوبی این چالش را رفع می‌کند و با افزایش حجم نمونه تعمیم‌پذیری با اطمینان بیشتری صورت می‌گیرد. همچنین طراحی این آزمایش‌ها به گونه‌ای هستند که معمولاً در شرایط مختلف قابل تکرار هستند و از این جهت نیز امکان تعمیم‌پذیری آنها قابل افزایش است (Buck & McGee, 2015).

(۵) ارائه شواهد برای دفاع از خطمشی‌ها: به کاربردن RCTها پیش از اجرای نهایی خطمشی‌ها می‌تواند این اطمینان را ایجاد کند که حاکمیت می‌تواند از تصمیمات و خطمشی‌های اتخاذشده خود دفاع کند، زیرا شواهد واضحی از کارآمدی و اثربخشی و مقرون‌به‌صرفه بودن آنها در اختیار دارد که همین امر منجر به افزایش اعتبار حاکمیت خواهد شد (Stéphane, 2020). همچنین این روش یک چشم انداز مطلوب برای پرداختن به موضوعات و رویکردهای مناقشه‌برانگیز، به کمک ارائه شواهد واقعی ایجاد می‌کند. پیشنهادات گوناگون می‌توانند از طریق اجرای RCTها مورد ارزیابی قرار گیرند و میزان کارآمدی هر یک مشخص شود (Ames & Wilson, 2016).

(۶) بستر اصلاح مستمر خطمشی‌ها: با توجه به اینکه کارکرد اصلی RCTها ارزیابی تأثیرات علی است، اجرای آنها می‌تواند بینش‌های خوبی به خطمشی‌گذاران پیرامون دلایل و فرآیندهای موفقیت یا شکست یک خطمشی بدهد و در بهینه‌کردن سیاست‌ها و برنامه‌ها در طول زمان کمک کند (Haynes, Goldacre & Torgerson, 2012).

(۷) قابلیت استفاده در شرایط پیچیده و در حال تغییر: دانش تخصصی طراحی و اجرای RCT این امکان را فراهم کرده است تا با طراحی مدل‌های پیشرفته‌تر مثل Step Wedge در شرایط پیچیده و یا در حال تغییر بتوان از RCTها برای ارزیابی اثربخشی برنامه‌ها استفاده کرد. تجربیات موفق جهانی نیز که در چنین حوزه‌ها و یا شرایطی RCTهای موفقیت آمیز و مفیدی اجرا کرده‌اند نیز می‌توانند تایید کننده این ادعا باشند (Buck & McGee, 2015).

نهایتاً می‌توان گفت نهادهای حاکمیتی در موقعیت بسیار مناسبی برای اجرای RCTها قرار دارند. این نهادها عمدتاً منابع قابل توجهی دارند و خطمشی‌ها و مداخلاتی را اجرا می‌کنند که بر جمعیت‌های بزرگی از مردم اثر می‌گذارد و می‌تواند تأثیرات پایداری بر کشور داشته باشد. با توجه به اینکه برای اجرای RCTهای خوب و قوی به جمعیت‌های بزرگ نیاز است و نهادهای حاکمیتی به این جمعیت‌ها دسترسی کافی دارند، می‌توانند مرجع بسیار مناسبی برای اجرای RCTها باشند. همچنین از آنجایی که نهادهای حاکمیتی داده‌های بسیار زیادی از شهروندان و رفتار آنها در اختیار دارند، می‌توانند با هزینه کم و با سهولت بیشتری نسبت به مجموعه‌های خصوصی، RCTها را اجرا کنند و اثربخشی تصمیمات و مداخلات را بسنجند (Ames & Wilson, 2016).

الزامات استفاده از ابزار کارآزمایی‌های تصادفی کنترل شده

جهت اجرای یک RCT موفق لازم است برخی شرایط و الزامات وجود داشته باشد که در ادامه به آن‌ها اشاره می‌گردد:

(۱) برای نتایج اثربخش، معیارهای تعریف شده و قابل مشاهده وجود داشته باشد: پیش از اجرای یک RCT باید معیاری که مؤثر بودن یک مداخله را نشان می‌دهد (متغیر وابسته) به شکل عینی و قابل ارزیابی تعریف شده و قابل مشاهده باشد (Ames & Wilson, 2016).

- (۲) جمعیت‌های مورد آزمایش (حجم نمونه) به اندازه کافی بزرگ باشد: RCT‌ها با نمونه‌های کوچکتر، عملی هستند، اما هر چه تعداد نمونه بیشتر باشد، تمایز بین گروه‌های مداخله و کنترل، قدرت آماری آزمون و احتمال استنباط نتایج معتبر افزایش می‌یابد (White, Sabarwal & De Hoop, 2014). انجمن ارزیابی آمریکایی‌نشان داده اکثر RCT‌های سیاستی و اجتماعی بین ۱۰۰۰ تا ۱۰۰۰۰ شرکت کننده دارند. در کنار آن، تیم بینش‌های رفتاری حجم نمونه ۱۰۰۰ تا ۵۰۰۰ نفر به ازای هر کدام از گروه‌ها را برای اجرای این آزمایش‌ها مناسب می‌داند (Kettle & Persian, 2022)، همچنین نمونه می‌تواند سطوح مختلفی مانند سطح فرد، خانواده، گروه‌ها و حتی مناطق جغرافیایی داشته باشد (Ames & Wilson, 2016).
- (۳) داده‌های مناسب در حوزه مورد ارزیابی وجود داشته باشد: برخی از حوزه‌های سیاستی، به لحاظ داده‌های دردسترس غنی هستند (مانند مالیات، بهداشت و آموزش)، در حالی که در برخی حوزه‌ها داده‌های ثبت‌شده کمتری وجود دارد (مثل بی‌خانمانی یا برخی آسیب‌های فرهنگی). RCT‌ها در شرایطی که به داده‌های غنی‌تری دسترسی وجود داشته باشد، با کیفیت و سهولت بیشتری انجام می‌شوند (Ames & Wilson, 2016).
- (۴) برای اینکه RCT‌ها شواهد معنی‌داری در مورد اثربخشی ارائه دهند، باید مداخله‌ی مدنظر برای تمام افراد به طور یکسان اجرا شود؛ در صورتی که مداخله استاندارد و کاملاً یکسان نباشد، نمی‌توان اطمینان حاصل کرد که نتایج به‌دست‌آمده ناشی از مداخله بوده‌است (Ames & Wilson, 2016).
- (۵) زیرساخت‌های مناسب و مسئولین آگاه وجود داشته باشند: برای اجرای RCT‌های موفق همزمان به زیرساخت مناسب و به‌روز و مسئولین و دست‌اندرکاران آگاه نیاز است. اگر در سازمانی دیوان‌سالاری فراوان، زیرساخت‌های ضعیف و قدیمی و مسئولین غیرآگاه و فاقد روحیه همکاری و همیاری وجود داشته باشد، احتمال اجرای RCT کاهش می‌یابد (Ames & Wilson, 2016).
- لازم به ذکر است در برخی موضوعات یا حوزه‌های سیاستی استفاده از RCT‌ها غیرممکن و نامناسب است که در این موارد باید از روش‌های ارزیابی دیگری استفاده کرد. دو عاملی که اغلب استفاده از RCT‌ها را منع می‌کنند عبارتند از عدم امکان تعریف گروه کنترل و حجم نمونه خیلی کوچک. در برخی موضوعات جداسازی گروه کنترل اصلاً امکان‌پذیر نیست و یا غیراخلاقی است. همچنین در برخی موارد جامعه هدف بسیار کوچک است و نمی‌توان از آن، نمونه‌ی کافی گرفت و آن را تخصیص تصادفی داد.

ضرورت ظرفیت‌سازی در راستای کاربست کارآزمایی‌های تصادفی کنترل‌شده

هنگامی که نهادهای حاکمیتی به استفاده از کارآزمایی‌های تصادفی کنترل‌شده روی می‌آورند، باید به ایجاد یک جریان مداوم از این آزمون‌ها و ایجاد یک اکوسیستم شواهد در کشور توجه کنند. شواهد باید به عنوان بخش جدایی‌ناپذیر از فرآیند خط‌مشی‌گذاری در نظر گرفته شوند و نتایج مثبت و منفی RCT‌ها مورد پذیرش قرار بگیرند. حاکمیت برای تغییر اولویت‌ها و تصمیم‌گیری‌ها، به مجموع این نتایج نیاز دارد (Ames & Wilson, 2016). فرصت‌های زیادی برای اجرای RCT وجود دارد، اما تبدیل آن به بخش ضروری از فرآیند خط‌مشی‌گذاری نیازمند ایجاد ظرفیت و فراهم کردن شرایط لازم است. سه ظرفیت حیاتی برای استفاده از این فرصت‌ها عبارتند از: فرهنگ و رهبری، تجربه و تخصص، و شفافیت.

فرهنگ سیاسی و رهبری: فرهنگ سیاسی و رهبری بزرگترین ظرفیت برای ایجاد خط‌مشی‌گذاری مبتنی بر شواهد است. سیاست‌گذاران باید باور داشته باشند که کارآمدی خط‌مشی‌ها، برنامه‌ها و مداخلات را نمی‌توان با شهود تعیین کرد. فرهنگ سیاسی باید به این باور برسد که می‌توان آزمایش کرد، یاد گرفت و تطبیق داد (Halpern, 2015). برای شروع خط‌مشی‌گذاری شواهد محور، اعتماد لازم است. می‌بایست به نهادهای اعتماد شود تا آنها بتوانند از آزمایش‌ها درس بگیرند. مسئولین نیز باید اعتماد کنند

که شکست سریع و کوچک بهتر از انکار شکست است. شهروندان نیز باید اعتماد کنند که حاکمیت از این آزمایش‌ها درس خواهد گرفت و خط‌مشی‌ها را بهبود خواهد بخشید. حمایت مسئولین حاکمیتی از خط‌مشی‌گذاری شواهد محور بسیار مهم است. تعیین دستورالعمل‌هایی که انجام RCT را بخش جدایی‌ناپذیر از فرآیند خط‌مشی‌گذاری می‌دانند، تخصیص بودجه برای آنها پیش از تصمیم‌گیری‌های نهایی، و دادن مجوزها و دسترسی‌های لازم برای انجام این آزمایش‌ها همگی بر عهده مسئولین است. مهمترین عامل ایجاد ظرفیت برای این رویکرد، شوق و تعهد مسئولین به RCT‌ها به عنوان ورودی‌های اصلی برای تصمیم‌گیری‌ها و خط‌مشی‌گذاری‌ها است (Ames & Wilson, 2016).

تجربه و تخصص: استفاده از دانش و مهارت افراد باتجربه و متخصص در روش‌های شواهدمحور و RCT‌ها یکی از ظرفیت‌های مهم برای ایجاد رویکرد خط‌مشی‌گذاری شواهد محور است. با توجه به نیاز RCT‌ها به تجربه و تخصص، پیشنهاد می‌شود که مراکزی به صورت متمرکز مسئولیت ارائه خدمات مرتبط با ارزیابی‌های خط‌مشی‌ها و RCT‌ها را بر عهده بگیرند. این مراکز استانداردهای ارزیابی را تعیین، خط‌مشی‌گذاری مبتنی بر شواهد را ترویج و منابع مورد نیاز را فراهم می‌کنند. همچنین سه کارکرد دیگر نیز بر عهده این مجموعه‌ها قرار دارد: تولید شواهد، ترجمه شواهد^۲ و تسهیل جذب شواهد؛ این مراکز به تولید و جمع‌آوری شواهد از سراسر جهان می‌پردازند، شواهد را طبق نیاز خط‌مشی‌گذاری ترجمه می‌کنند، و جذب شواهد توسط نهادهای خط‌مشی‌گذاری را تسهیل می‌کنند. مشارکت دانشگاهیان نیز برای ظرفیت‌سازی این رویکرد بسیار مهم است (Alexander & Letwin, 2013; Ames & Wilson, 2016).

شفافیت: عدم شفافیت، اعتبار ارزیابی‌ها را کاهش می‌دهد. ثبت نتایج کارآزمایی‌ها و ارزیابی خط‌مشی‌ها راه حلی برای این مشکل است. سیاستگذاران باید آمادگی داشته باشند تا تأثیرات کم یا عدم تأثیر برنامه‌ها را به صورت شفاف اعلام کنند. انتشار نتایج منفی و دانستن اینکه چه کاری نباید انجام شود به اندازه نتایج مثبت و دانستن اینکه چه کاری باید انجام گیرد، مفید است. همچنین ثبت نتایج سبب می‌شود بخش‌های مختلف حاکمیت به نتایج آزمایش‌ها دسترسی داشته باشند و از آنها استفاده کنند. هرچه مجموعه داده‌های موجود پیوستگی بیشتری داشته باشند و نهادهای مختلف به شبکه آن شواهد دسترسی بیشتری داشته باشند، ظرفیت گسترش رویکرد خط‌مشی‌گذاری شواهد محور افزایش پیدا می‌کند (Ames & Wilson, 2016).

توصیه‌های سیاستی جهت کاربست کارآزمایی‌های تصادفی کنترل شده در مجلس شورای اسلامی

در ادامه پیشنهادها و توصیه‌هایی جهت ترویج و بکارگیری روش کارآزمایی تصادفی کنترل‌شده در فرایندهای مجلس شورای اسلامی ارائه می‌شود:

۱. بهترین راه برای ترویج و بکارگیری RCT‌ها، شروع تلاش است. یادگیری از تجربه و کار بر روی RCT‌ها منجر به توسعه بلندمدت قابلیت RCT می‌شود. نتایج نظرسنجی از نمایندگان مجلس استرالیا نشان داد از میان نمایندگانی که در گذشته از نتایج RCT برای تصمیم‌گیری‌ها استفاده کرده بودند، ۶۷ درصد RCT‌ها را به عنوان یکی از سه اولویت خود در تصمیم‌گیری‌های سیاستی برای آینده انتخاب کرده بودند و در مقابل تنها ۱۲ درصد از نمایندگانی که قبلاً از RCT استفاده نکرده بودند، این روش را در آینده اولویت خود در تصمیم‌گیری می‌دانستند (Ames & Wilson, 2016). بنابراین صرف شروع استفاده از RCT‌ها می‌تواند به افزایش تقاضا برای بکارگیری آن‌ها در آینده کمک کند.

1. Generating Evidence Synthesis
2. Translating Evidence
3. Facilitating Evidence Absorption

۲. توصیه دوم این است که نمایندگان مجلس پیش از تصمیم‌گیری، شواهد RCT‌های مرتبط با قوانین، طرح‌ها و لوایح را از مراکز پژوهشی مطالبه نمایند و تصمیم‌گیری‌ها و تخصیص بودجه را مبتنی بر این شواهد انجام دهند. همچنین نمایندگان می‌توانند رویکرد «پرداخت برای موفقیت» را اتخاذ کنند (Result for Emerica, 2019)، که طبق آن بودجه صرفاً در طرح‌ها و برنامه‌هایی که پیش از اجرای اصلی و طبق نتایج RCT‌ها شواهدی از اثربخشی آن‌ها وجود دارد، هزینه گردد.

۳. پیشنهاد می‌شود درصدی از بودجه‌ی برنامه‌ها و طرح‌ها به ارزیابی پیشینی و کارآمدی طرح تخصیص پیدا کند. به‌عنوان مثال و طبق پیشنهادات جهانی می‌توان یک درصد از بودجه کل طرح را به این موضوع اختصاص داد (Result for Emerica, 2019).

۴- به دلیل تخصصی‌بودن RCT‌ها پیشنهاد می‌شود کارگروهی تخصصی به منظور طراحی و اجرای این روش مرتبط با قوانین، لوایح و طرح‌های مجلس ایجاد شود (ذیل مرکز پژوهش‌های مجلس و در یکی از دفاتر آن) و یا مرکزی با این تخصص و برای این هدف به استخدام مرکز پژوهش‌های مجلس شورای اسلامی درآید. این مرکز به صورت متمرکز به تامین نیازهای روز مجلس شورای اسلامی و ارائه نتایج RCT‌ها به نمایندگان پیردازد و دقیق‌ترین راهبردها را جهت بهبود برنامه‌ها و طرح‌ها ارائه دهد. همچنین منابع تخصصی و علمی، فنی و انسانی مورد نیاز را متمرکز نماید و از برون‌سپاری و پراکندگی بکاهد. در کنار آن، از ظرفیت متخصصین و جامعه دانشگاهی نیز استفاده نمایند و نهایتاً سیاستگذاری شواهدمحور و روش کارآزمایی تصادفی کنترل شده را ترویج و بر فرایند آن نظارت کند.

۵- ترویج و گسترش و اجرای RCT‌های مطلوب نیازمند افزایش انعطاف‌پذیری نهادهای بالادستی و حمایت مسئولین در ارائه دسترسی‌ها، مجوزها و اختیارات مورد نیاز به متخصصین دست‌اندرکار است تا بتوانند نتایج را ارزیابی و راهبردهای مبتکرانه ارائه نمایند. برای نمونه در آمریکا حمایت اوپاما در صدور دستورالعمل التزام دولت بر استفاده از شواهد و در انگلستان دسترسی و مجوز دفتر نخست وزیر برای ایجاد مرکز BIT بر گسترش استفاده از روش RCT و دستیابی به ارزش افزوده‌های آن برای حکمرانی بسیار موثر بود (Result for Emerica, 2019).

۶- آخرین توصیه حمایت از اشتراک‌گذاری داده‌ها و شواهد جهت استفاده است. نتایج کارآزمایی‌های تصادفی کنترل شده باید در قالبی کاربرپسند منتشر شود و در اختیار بخش‌های مختلف مجلس و نهادهای حکمرانی و مسئولین قرار گیرد (البته با رعایت ملاحظات محرمانگی). عدم شفافیت ارزیابی‌ها، اعتبار و فواید ارزیابی‌ها و تصمیمات مبتنی بر شواهد را کاهش می‌دهد، درحالی‌که با ایجاد دسترسی، پیوند داده‌ها بین نهادها رخ می‌دهد و از این ظرفیت می‌توان در جهت بهبود نتایج بهره برد.

نتیجه‌گیری

کارآزمایی‌های تصادفی کنترل شده در قلب جنبش شواهدمحور در دو دهه اخیر خط‌مشی‌گذاری عمومی جهان هستند. RCT‌ها به «استاندارد طلایی»^۲ مشهورند، چرا که با توجه به ارزیابی‌های علمی که انجام می‌دهند، بهترین روش علمی موجود برای بررسی موثر بودن یک مداخله به حساب می‌آیند. خط‌مشی‌گذاری شواهدمحور تاکید می‌کند که خط‌مشی‌ها باید مبتنی بر بهترین و قوی‌ترین شواهد موجود از تحقیقات و با توجه به مقرون به صرفه بودن، میزان اثربخشی، مقبولیت برای ذینفعان، و پذیرش افکار عمومی طراحی و اجرا شوند. این فرایند کمک می‌کند تا نسبت به کارآمدی قوانین و مقررات پیشنهادی اطمینان حاصل شود. با توجه به اهمیت کارآزمایی‌های تصادفی کنترل شده و مزایایی که در خط‌مشی‌گذاری عمومی دارند و همین‌طور تجارب جهانی و موفق این

1. Pay for Success
2. Gold Standard

رویکرد، ضروری است رویکرد خط‌مشی‌گذاری شواهدمحور نیز در حکمرانی ایران علی‌الخصوص در خط‌مشی‌گذاری‌ها و قانون‌گذاری‌های مجلس شورای اسلامی اتخاذ شود و از مزایای آن جهت بهبود کارآمدی خط‌مشی‌ها و قوانین بهره‌برده شود.

سیاست‌گذاری

این پژوهش با حمایت مادی و معنوی دفتر مطالعات حکمرانی مرکز پژوهش‌های مجلس شورای اسلامی و مرکز بینش‌های رفتاری ایران انجام شده‌است.

References

- Alexander, D., & Letwin, O. (2013). What Works: evidence centres for social policy. London: Cabinet Office.
- Allcott H. (2011), "Social Norms and Energy Conservation", *Journal of Public Economics*, (95), 1082-1095.
- Ames, P., & Wilson, J. (2016). Unleashing the potential of randomised controlled trials in Australian governments: Doctoral dissertation, Harvard University.
- Behavioural Insights Team. (2012). Applying behavioural insights to reduce fraud, error and debt. Cabinet Office, London, 185, 186.
- Behavioural Insights Team. (2017). The Behavioural Insights Team update report 2016-17. Behavioural Insights Team. <https://www.bi.team/publications/the-behavioural-insights-team-update-report-2016-17/>
- Buck, S., & McGee, J. (2015). Why government needs more randomized controlled trials: Refuting the myths. Houston, TX: Laura and John Arnold Foundation.
- Carroll, J., Lyons, S., & Denny, E. (2014). Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45, 234-243.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91-108.
- Haby, M., & Sieber, S. (2021). Evidence, policy, impact: WHO guide for evidence-informed decision-making. In *Evidence, policy, impact: WHO guide for evidence-informed decision-making* (pp. 92-92).
- Hallsworth, M., & Kirkman, E. (2020). Behavioral insights. MIT Press.
- Halpern, D. (2015). What works? The rise of experimental government. In *Inside the nudge unit: How small changes can make a big difference*. Random House.
- Hansen, P. G. (2019). Tools and ethics for applied behavioural insights: the BASIC toolkit. Organisation for Economic Cooperation and Development, OECD.
- Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomised controlled trials. Cabinet Office-Behavioural Insights Team.
- Haynes, L. C., Green, D. P., Gallagher, R., John, P., & Torgerson, D. J. (2013). Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management*, 32(4), 718-730.
- Houde, S., Todd, A., Sudarshan, A., Flora, J. A., & Armel, K. C. (2013). Real-time Feedback and Electricity Consumption: A Field Experiment Assessing the Potential for Savings and Persistence. *The Energy Journal*, 34(1), 87-102. <http://www.jstor.org/stable/41969212>
- John. P. (2020): Improving Causal Claims in Public Policy through Randomized Designs in the Field, *Journal of Comparative Policy Analysis: Research and Practice*.
- Kettle, S., & Persian, R. (2022). Target, explore, solution, trial, scale: An introduction to running simple behavioural insights projects. BIT Handbook.
- MacMillan, H. L., Wathen, C. N., Barlow, J., Fergusson, D. M., Leventhal, J. M., & Taussig, H. N. (2009). Interventions to prevent child maltreatment and associated impairment. *The lancet*, 373(9659), 250-266.
- Pettersson, J. (2014). Randomized controlled trials: strengths, weaknesses and policy relevance.
- Pew. (2015). Legislating evidence-based policymaking: A look at state laws that support data-driven decision-making. Issue Brief, 862-872. Retrieved from <https://www.pewtrusts.org/-/media/assets/2015/03/legislationresultsfirstbriefmarch2015.pdf>
- Purdon, S., Stratford, N., Taylor, R., Natarajan, L., Bell, S., & Wittenburg, D. (2006). Impacts of the job retention and rehabilitation pilot. *Renewable Energy World*. (2010, March 5). President Obama speaks at headquarters of smart grid firm Opower. *Renewable Energy World*. <https://www.renewableenergyworld.com/power-grid/smart-grids/president-obama-speaks-at-headquarters-of-smart-grid-firm-opower/>
- Result for Emerica. (2019). 9 Ways to Make Federal Legislation Evidence-Based: 2019 What Works Guide for Congress. Retrieved from <https://results4america.org/wp-content/uploads/2019/02/What-Works-Legislation-2019-9-Ways-to-Make-Legislation-More-Evidence-Based.pdf>
- Stéphane, J. (2020). How can governments leverage policy evaluation to improve evidence informed policy making?
- Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112(S1), S164-S187.
- White, H., Sabarwal, S., & de Hoop, T. (2014). Randomized Controlled Trials (RCTs): Methodological Briefs-Impact Evaluation No. 7 (No. innpub752).
- Wilson, T. D., & Juarez, L. P. (2015). Intuition is not evidence: Prescriptions for behavioral interventions from social psychology. *Behavioral Science & Policy*, 1(1), 13-20.
- Zakani, Z., Ebrahim Kani, M.H., Morovvati, F., Givzad, A., & Ghaderi, A. (2021). Using text-based nudges designed by behavioral science to increase COVID-19 vaccination rates. *Iran Behavioral Insights Center [in persian]*