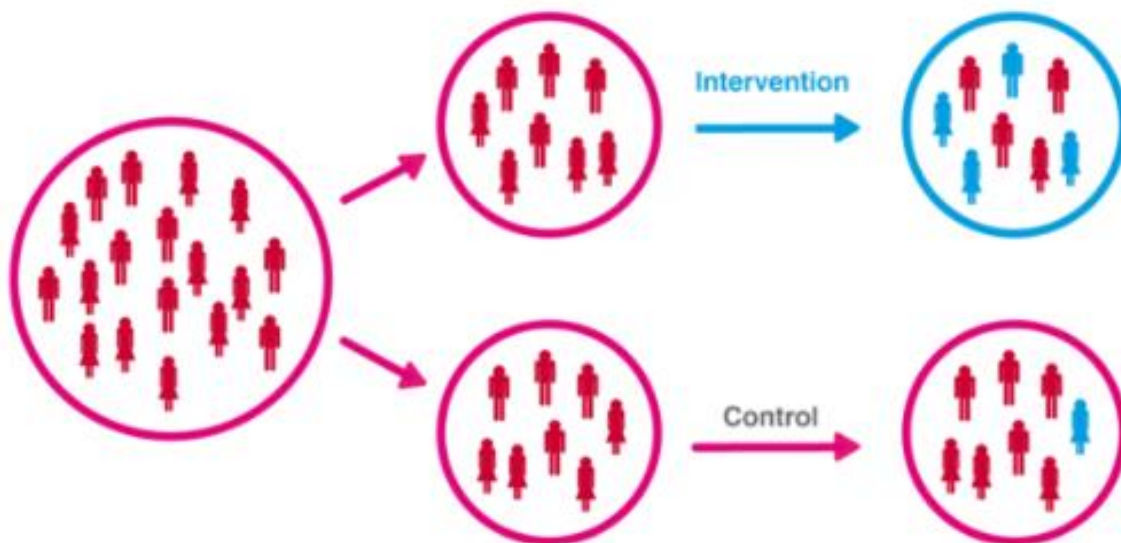


Metodické usmernenie Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky zo dňa 13.06.2023 č. 018572/2023/oBI k realizácii RCT v prostredí verejnej správy Slovenskej republiky



Dátum vytvorenia: 10. 11. 2022

Vypracovali: Anna Makarová, Jaroslava Blažíčková

Dátum aktualizácie: 16. 05. 2023

Obsah

Cieľ.....	3
Randomizované kontrolné štúdie – definícia metódy	3
1 Teoretické východiská	5
1.1 Čo je RCT	5
1.2 Limitujúce faktory.....	7
1.3 Vhodnosť využitia.....	7
1.4 Iné výskumné metódy a RCT	8
1.5 Zdroje, výskumný tím a časový harmonogram.....	11
2 Prípravná fáza RCT.....	13
2.1 Dizajn experimentu	13
2.2 Veľkosť vzorky a pilotné testovanie	19
2.3 Náhodné hodnotenie (randomizácia)	23
2.4 Riziká RCT, najčastejšie problémy a možnosti ich riešenia	24
3 Realizácia RCT.....	27
4 Analýza RCT.....	27
4.1 Zber a spracovanie dát	27
4.1.1 Zber dát.....	27
4.1.2 Spracovanie dát.....	28
4.2 Meranie efektov	29
4.3 Chyby merania	31
4.4 Interpretácia výsledkov	33
5 Realizovanie RCT v prostredí verejnej správy Slovenskej republiky	34
5.1 Realizovanie RCT v rámci projektového riadenia vo verejnej správe SR.....	34
5.2 Realizovanie RCT bez projektového riadenia vo verejnej správe SR	36
6 Prípadové štúdie.....	37
Slovník pojmov	38
Použitá literatúra	41
Príloha: Porovnanie RCT a iných výskumných metód na zhodnotenie efektov intervencie	43

Cieľ

Metodické usmernenie popisuje základné teoretické východiská, procedurálne a technické postupy pre orgány verejnej moci pri návrhu a implementácii nového riešenia alebo inovácii existujúceho riešenia v súlade so zásadami behaviorálnych inovácií.

Cieľom dokumentu je popísať základné teoretické východiská procesu testovania a zadefinovať základný rámec hľadania najvhodnejšej alternatívy návrhu behaviorálnej inovácie.

Cieľom metódy randomizované hodnotenia alebo randomizované kontrolné štúdie (angl. *Randomized control trials*, ďalej len „RCT“) je najmä testovanie kauzálnych súvislostí. Môže výrazne pomôcť pri testovaní dopadov zavádzaných verejných politík, zmien existujúcich verejných politík a porozumeniu prečo a ako sa občania v kontakte s nimi rozhodujú.

Randomizované kontrolné štúdie – definícia metódy

RTC získavajú v posledných rokoch v oblasti verejných politík čoraz väčšiu dôležitosť. V medicínskych vedách ide o štandardizovanú a dlhodobo využívanú metodológiu. V oblasti skúmania ekonomického rozhodovania sa občanov ide o pomerne novú metódu s rastúcou popularitou.

RCT sú formou experimentu, ktorého účastníci sú náhodne rozdelení, a následne zaradení do jednej kontrolnej a jednej alebo viacerých testovaných skupín. Jeho hlavnou výhodou je, že umožňuje výskumníkom a tvorcom politík prispôbiť ich dizajn tak, aby odpovedal na konkrétne otázky o účinnosti politiky alebo programu.

RCT je **výskumná metóda** založená na hlbokom poznaní mechanizmov správania sa inštitúcií, a najmä jednotlivcov, konkrétneho praktického problému a predchádzajúcich štúdií zameraných na podobný typ problému. Experiment predstavuje najprísnejší spôsob určenia existencie vzťahu príčiny a následku.

Použitie RCT je náročné na čas, preto je táto metóda vhodná najmä pre projekty s dostatočným časovým rámcom.

Pre použitie RCT je potrebná dostatočne veľká cieľová skupina, ktorú je možné náhodne rozdeliť. Pred použitím RCT je veľmi dôležitá identifikácia a definícia jasného, reálneho a merateľného javu. Náhodné pridelovanie do intervenčných skupín zabezpečí minimalizáciu ovplyvňujúcich faktorov, čím sa predíde skresľovaniu výsledkov.

Záver RCT sú postavené na podrobnej analýze získaných údajov najmä preto, že jednotlivé javy, problémy, politiky a miestny kontext sa navzájom veľmi odlišujú.

RCT teda **predstavujú komplexný nástroj** využiteľný na skúmanie a hľadanie riešení rôznorodých problémov vo verejnom sektore.

V tejto príručke poskytujeme len stručné oboznámenie sa s RCT, schematický prehľad procesu prípravy, realizácie a vyhodnotenia RCT. Vzhľadom na rozsah a komplexnosť problematiky môže slúžiť ako prvotné oboznámenie sa s RCT. Pre záujemcov o hlbšie pochopenie problematiky uvádzame pri viacerých témach odkazy na literatúru, ktorú je možné na tento účel využiť.

RCT **nie sú univerzálne** a nemôžeme ich jednoducho replikovať. Sú presne prispôsobené tak, aby sa minimalizovali ich presahy na ostaté komponenty sledovaných programov realizovaných napr. viacerými poskytovateľmi služieb alebo zahŕňajúcich viacero modelov služieb.

RCT predstavujú v prvom rade výskumnú metódu. Na ich realizácii sa podieľa tím zložený z týchto členov; tu sú ich pozície, úlohy a funkcie:

- **koordinátor z konkrétnej inštitúcie** - dohliada na realizáciu štúdie, ako aj na zástupcov inštitúcie, ktorí veľmi dobre poznajú danú problematiku, vzťahy v rámci inštitúcie štátnej správy alebo samosprávy,
- **skúsení výskumníci** - sú schopní analyzovať vedeckú literatúru, majú skúsenosť s metodikou vedeckého výskumu a navrhnu na základe vedomostí o správaní dizajn experimentu,
- **zástupcovia inštitúcie** – tí, ktorí majú prístup k dátam a vedia s nimi pracovať,
- **analytici** - majú vedomosti na spracovanie získaných údajov,
- prípadne ďalší členovia s ohľadom na špecifický charakter konkrétnej štúdie (napr. programátori a pod.).

1 Teoretické východiská

Správne a premyslene navrhnuté a implementované náhodné testy môžu zodpovedať otázky ako: *Aký efektívny je tento program? Vyskytli a zaznamenali sme nejaké neúmyselné vedľajšie efekty? Kto mal z intervencie najväčší úžitok? Ktoré súčasti programu fungujú alebo nefungujú? Ako nákladovo efektívny je navrhovaný program? Ako si stojí navrhovaný program v porovnaní z inými navrhnutými alternatívami na dosiahnutie podobných cieľov?*

1.1 Čo je RCT

RCT je metóda hodnotenia **dopadov** politík, kde jednotlivé alternatívne návrhy

- **nových politík** (napr. zavedenie novej dávky, poskytnutie novej služby, zavedenie nového spôsobu hodnotenia výsledkov žiakov základných škôl, vytvorenie registra darcov krvi a pod.) alebo
- **návrhy ich zmien** pomocou konkrétnych **nástrojov** (napr. rozšírenie okruhu oprávnených osôb pre sociálnu dávku, zvýšiť periodicitu prehliadok u lekára pre vybrané skupiny, znížiť maximálny počet žiakov v škole a pod)

môžeme testovať prostredníctvom experimentálnej metódy náhodného rozdelenia všetkých potenciálnych beneficentov politiky (občanov, ktorí z nej majú potenciálny prospech) do viacerých skupín.

Základom realizácie RCT je rozdelenie účastníkov do dvoch typov skupín:

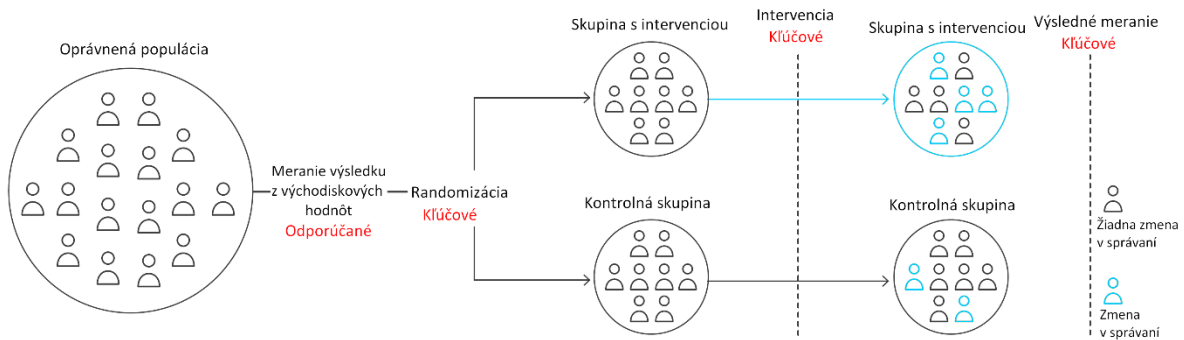
- kontrolná skupina,
- testovaná skupina (skupiny).

Skupiny, v ktorých účastníci dostávajú rôzne typy intervencie označujeme ako **testovaná skupina** - skupina/skupiny s intervenciou (angl. *Treatment group*) a skupina, ktorá nie je vystavená žiadnej intervencii je **kontrolná skupina** (angl. *Control group*).¹⁾ Výskumníci merajú rozdiely v správaní sa účastníkov zaradených do jednotlivých skupín (testovaných a kontrolnej). Porovnaním zmien v správaní sa účastníkov náhodné testy umožňujú získať presný a nezaujatý odhad **dopadu** uskutočnenej intervencie - teda aké konkrétne zmeny v živote účastníkov možno pripísať konkrétnej zmene programu alebo politiky.

¹⁾ Anglickú terminológiu uvádzame preto, že ide o zaužívané pojmy, s ktorými sa čitateľ môže stretnúť pri hlbšom štúdiu danej problematiky či pri bežnej praxi.

Zjednodušene zobrazuje priebeh RCT nasledujúci obrázok.

Obrázok 1: Schematický priebeh RCT



Zdroj: Edovald, T., & Firpo, T. (2016)

Cieľom RCT je:

- štatisticky overiť platnosť hypotézy. Zhodnotiť, či **hypotéza** platí alebo nie. *Hypotézy o dopade jednotlivých zmien politik definujeme na základe teoretických východísk alebo na základe predchádzajúcich uskutočnených štúdií. Výhodou RCT je, že dávajú tvorcom politik predstavu o efektívnosti a dopade konkurenčných alternatív navrhovaných politik prostredníctvom dát;*
- pochopiť **kauzálne vzťahy**. *Pri tvorbe dizajnu náhodných testov manipulujeme pre testované skupiny vždy iba jednu dimenziu alebo faktor s tým, že všetko ostatné zostane nezmenené. Týmto spôsobom môžeme odhaliť závislosť medzi konkrétnou črtou politiky a jej dopadom na zmenu správania sa.*

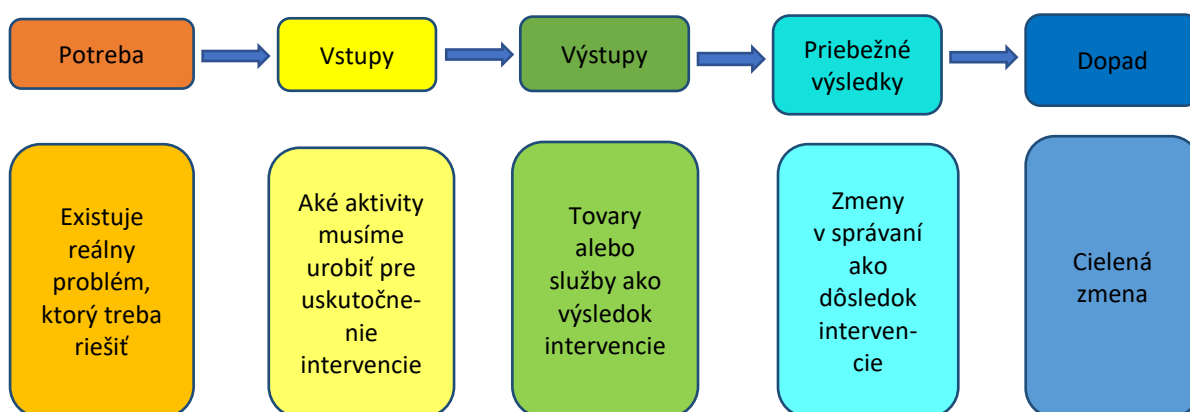
Náhodný test neodpovedá len na otázku čo funguje, ale aj **ako** a **prečo** to funguje. Kombináciou výsledkov jedného alebo viacerých uskutočnených RCT s poznatkami z ekonomickej teórie, deskriptívnymi empirickými dôkazmi a konkrétnymi znalosťami problému, môžeme lepšie pochopiť vplyv intervencie na zmenu správania sa.

Priebeh RCT je možné rozdeliť na tri základné fázy - prípravnú fázu, realizačnú fázu a fázu vyhodnotenia (obrázok č. 2):

- **prípravná fáza** zahŕňa zhodnotenie potrieb a vstupov, vrátane procesu plánovania a prípravy dizajnu experimentu,

- **realizačná fáza** je zameraná na konkrétnu realizáciu navrhnutého dizajnu vrátane zberu výstupov a priebežných výsledkov,
- **fáza vyhodnotenia** využíva získané dáta na zhodnotenie dopadov RCT a interpretáciu výsledkov.

Obrázok 2: Priebeh RCT



Zdroj: J-Pal Action Lab (2021a)

1.2 Limitujúce faktory

RCT **nie je** správnou metódou zisťovania dopadov politik a programov, ak:

- je pravdepodobné, že **externé faktory** (napr. vláda prostredníctvom zmeny zákonov) budú počas náhodného testu **zasahovať do programu**, alebo ak je program výrazne zmenený počas hodnotenia,
- výsledok experimentu **nebude mať vplyv na výsledný dizajn** politiky,
- **nemožno priradiť subjekty** do kontrolnej a testovaných skupín náhodne, ako napríklad pri makroekonomickej politike,
- je **neetické** poskytnúť intervenciu testovanej skupine alebo ju odoprieť kontrolnej skupine,
- RCT nie je nákladovo efektívne, je príliš **časovo náročné** (napr. čas realizácie RCT presahuje termín na zavedenie politiky) **alebo nemá politickú podporu**,
- je príliš **ťažké kontrolovať** možné **ohrozenia** fungovania navrhnutého **dizajnu** štúdie, (napr. prelievanie dopadov medzi kontrolnou a testovanými skupinami).

1.3 Vhodnosť využitia

Ak sú politiky a nástroje vhodne navrhnuté a implementované, RCT sú výborným nástrojom na **meranie ich vplyvu**. Ich význam sa zvyšuje ak plánujeme do nového programu investovať veľa zdrojov.

Ak je ťažké na prvý pohľad rozhodnúť, či je RCT uskutočniteľným a vhodným spôsobom hodnotenia konkrétnej intervencie, je možné využiť otázky z kontrolného listu (angl. *checklistu*). Ak sú vaše odpovede na všetky položené otázky „áno“, potom je RCT pravdepodobne vhodným nástrojom na riešenie identifikovaného problému.

Kontrolný list vhodnosti využitia RCT pre testovanie politiky

1. Existuje reálny **problém**, ktorý treba riešiť zavedením/zmenou politiky?
2. Má subjekt, ktorý tvorí, implementuje alebo upravuje politiku alebo jej nástroje záujem problém **riešiť**?
3. Je **vzorka** oprávnených účastníkov štúdie dostatočne veľká?
4. Máme dostatok **zdrojov** (materiálnych aj personálnych) na realizáciu testovania?
5. Existuje dostatočne veľká **cieľová skupina** intervencie?
6. Možno cieľovú skupinu **náhodne rozdeliť** na testovanú a kontrolnú?
7. Máme prístup k **dátam** na meranie dopadu prípadne ďalším informáciám o účastníkoch experimentu?
8. Máme personálne kapacity a dostatočné znalosti na **vyhodnotenie dát** pre výber uplatňovanej politiky?

1.4 Iné výskumné metódy a RCT

RCT predstavuje najsilnejší experimentálny dizajn, ktorý identifikuje príčinnú súvislosť vyplývajúcu z intervencie. RCT však nie je možné/vhodné vždy využiť.

Okrem RCT existujú aj iné **experimentálne metódy** využiteľné v ekonomickom výskume. Ide o komplexné metódy so samostatnou metodologickou základňou:

- **Kvázi-experiment** (alebo prirodzený experiment) – ide o typ observačnej štúdie (pozorujeme existujúce údaje). Nastáva v prípade, ak bola konkrétna intervencia implementovaná, ale okolnosti súvisiace s implementáciou nie sú pod kontrolou výskumníkov. Výhodou tohto typu experimentu je, že jeho realizácia nie je podmienená participáciou výskumníkov na jeho príprave a vplyva skôr z potreby spoločnosti (napr. znižovania závislosti od fajčenia). Vzhľadom na to, že naturálne

experimenty vznikajú prirodzene pri zavádzaní nových politík alebo ich úpravách, výhodou sú aj obmedzené náklady na strane výskumníka. Jeho nevýhodou je chýbajúca kontrola nad priebehom experimentu a vplyv externých faktorov. Pre analýzu získaných údajov je často potrebné využiť sofistikované ekonometrické metódy s cieľom očistenia výsledkov od uvedených externých vplyvov.

PRÍKLAD

Príkladom využitia kvázi-experimentu je štúdia o účinkoch letného času na nehodovosť v cestnej premávke, ktorá sa uskutočnila v USA. Autori použili údaje o dopravných nehodách so smrteľnými následkami za obdobie rokov 1976-2003 a využili prirodzený experiment vyplývajúci zo zmeny legislatívy v roku 1986, ktorá zmenila čas, kedy jednotlivé štáty USA prešli na letný čas.²⁾

- **Laboratórny experiment** – predstavuje realizáciu ekonomického experimentu v laboratórnych podmienkach za plnej kontroly vplyvu externých faktorov zo strany experimentátora. Pre realizácii ekonomického laboratórneho experimentu je dôležitým prvkom finančná motivácia účastníka. Táto odmena je diferencovaná s ohľadom na správanie sa účastníkov experimentu, teda každý z účastníkov môže získať inú finančnú odmenu. Dizajn týchto štúdií je navyše schematický, často až abstraktný. Neviaže sa na konkrétnu životnú situáciu, ale skúmajú sa ním hlbšie pohnutky ekonomických subjektov. Táto schematickosť predstavuje aj nevýhodu laboratórnych experimentov, keďže dizajn týchto experimentov je špecifický. Mnohé výskumy preukázali, že dobre nadizajnovaný experiment v laboratóriu má pomerne veľkú mieru vonkajšej platnosti výsledkov.

PRÍKLAD

Príkladmi využitia laboratórneho experimentu je skúmanie nastavovania ponuky a dopytu ekonomických subjektov, rozhodnutia ekonomických subjektov v aukciách alebo pri recipročnom správaní. Známou je tzv. diktátorská hra zameraná na testovanie štandardného ekonomického predpokladu, že jednotlivci budú konať výlučne vo vlastnom záujme.³⁾

²⁾ Pre viac informácií pozri Sood and Ghosh (2007).

³⁾ Pre viac informácií o metodológii a podstate laboratórnych ekonomických experimentov pozri napríklad Kagel a Roth (2020).

- **Laboratórium v teréne** (angl. *Lab-in-field experiment*) – predstavuje kombináciu terénneho a laboratórneho experimentu pri zachovaní schematickosti laboratórneho dizajnu avšak s relevantnou cieľovou populáciou⁴⁾ (vytvorenou napríklad na základe určitej charakteristiky), ktorá je pre dané rozhodnutia motivovaná. Výhodou tohto typu experimentu je možnosť skúmania rozhodnutí subjektov, ktoré sú skúmaným problémom reálne zasiahnuté, čo môže napomôcť napríklad pri verejných politikách lepšie pochopenie konkrétnej intervencie. Nevýhodou je už spomínaný schematický charakter dizajnu ale aj náklady spojené s finančnou motiváciou účastníkov experimentu.

PRÍKLAD

Príkladom využitia experimentu typu laboratórium v teréne je skúmanie rozhodnutí o cenotvorbe na vzorke skutočných predajcov, ktorí v danej oblasti majú skúsenosti, prípadne skúmanie rizikových rozhodnutí u manželov a podobne.⁵⁾

Okrem experimentálnych metód ako je RCT, je v ekonomickom výskume možné využiť aj **neexperimentálne metódy**, ako napríklad metódu rozdielov v rozdieloch (angl. *diff-in-diff*) alebo metódu štatistického párovania. Každá z týchto metód má svoje základné predpoklady a voľba ich využitia vyžaduje pokročilejšie ekonometrické vedomosti. Výber z týchto metód spolu s ich popisom, predpokladmi a požadovanými údajmi uvádzame v prílohe 1 tohto dokumentu.

Voľba medzi konkrétnou metódou výskumu vyplýva z charakteru oblasti záujmu, stanovenej výskumnej otázky a problému, ktorý je potrebné riešiť, ako aj dostupnosti skúmanej vzorky a ďalších dát, dostupnosti zdrojov (materiálnych a finančných, časových).

⁴⁾ Pri laboratórnych experimentoch sú zväčša účastníkmi experimentov študenti s ohľadom na ich dostupnosť na akademickej pôde, kde sú alokované experimentálne laboratória ako aj nižšou cenou ich práce, čo ovplyvňuje náklady na experimentálnu štúdiu.

⁵⁾ Viac informácií o experimentoch typu laboratórium v teréne nájdete napríklad v Gneezy a Imas (2017).

1.5 Zdroje, výskumný tím a časový harmonogram

Pri plánovaní RCT je potrebné uvažovať o potrebe **materiálnych i personálnych zdrojov** nadväzujúcich na jednotlivé fázy realizácie RCT.

V prvej fáze realizácie RCT je potrebné počítať najmä s nákladmi súvisiacimi s vytvorením výskumného tímu.

V ďalších fázach (plánovanie vstupov, výstupov a merania priebežných výsledkov) vstupujú do procesu materiálne náklady súvisiace s implementáciou intervencie (napr. náklady na poštovné, náklady na vytvorenie softvéru, dátový priestor, ale aj terénnych pracovníkov a pod.).

V prípravnej fáze RCT je potrebné vytvoriť **rozpis finančných nákladov realizácie RCT** a uistiť sa, že týmito zdrojmi disponujeme. Možnosťou financovania RCT okrem vlastného rozpočtu sú aj výskumné granty v spolupráci s výskumnými inštitúciami.

Realizácia RCT prináša viaceré personálne a finančné výzvy:

Personálne výzvy – dôležité je uzatváranie **partnerstiev** naprieč výskumnými inštitúciami a vládnymi jednotkami. Tieto si vyžadujú vybudovanie pracovného vzťahu s implementačným partnerom už na začiatku tvorby dizajnu RCT. Formulácia hypotéz, tvorba prehľadu literatúry, ako aj dizajn experimentu a vyhodnotenie výsledných efektov vyžaduje skúsenosť s realizáciou výskumných aktivít. Zloženie **výskumného tímu** predpokladá zapojenie:

- hlavného výskumníka - zväčša experta na behaviorálnu a experimentálnu ekonómiu,
- RCT manažéra – resp. manažéra terénnych aktivít,
- dátových analytikov/ka - prípadne experta na ekonometriu na strane vládnych jednotiek a tvorcov politík,
- experta na implementáciu politiky.

Partnerstvo musí byť založené na pochopení príslušných úloh a zodpovedností oboch tímov (tvorcov dizajnu aj realizátorov experimentu), spoločnom záväzku k hodnoteniu a akceptácii motívov členov tímov.

Finančné výzvy - rozpočtovanie realizácie RCT musí zahŕňať:

- **mzdové náklady,**
- **finančné a materiálne náklady** spojené s realizáciou konkrétnej intervencie.

PRÍKLAD

V prípade listového experimentu je potrebné kalkulovať so zmenou poštových nákladov, nákladov na dodatočnú tlač a pod. V prípade experimentov, kde je potrebné účastníkov informovať alebo pripomenúť povinnosť/aktivitu s využitím SMS/telefonickým kontaktovaním je potrebné rátať aj s týmito nákladmi. Pri terénnych experimentoch je zasa potrebné počítať s nákladmi na školenie terénnych pracovníkov a nákup technického vybavenia potrebného na realizáciu experimentu (napr. tablet, telefón, a pod.).

Časový harmonogram realizácie RCT nadväzuje na jednotlivé fázy experimentu. **Trvanie experimentu** je individuálne v závislosti od konkrétneho dizajnu experimentu. Dôležitý je aj očakávaný čas potrebný na to, aby program ovplyvnil výsledky a povahu výstupov, ktoré sledujeme. Načasovanie následného zberu údajov musí brať do úvahy, koľko času je potrebného po implementácii programu, **aby sa výsledky prejavili** (napr. termín splatnosti platobného príkazu, termín podávania daňového priznania, termín zverejnenia výsledkov hodnotenia PISA⁶⁾ a pod.).

Postupnosť získavania výsledkov programu pomáha identifikovať vhodné ukazovatele výsledkov a vhodný čas na ich meranie. Niektoré programy (napr. programy na zvyšovanie príjmu) sa zameriavajú na poskytovanie krátkodobých výhod, zatiaľ čo iné (napr. programy základného vzdelávania) sa zameriavajú na dlhodobé efekty. Pri niektorých výsledkoch z ich podstaty trvá dlhšie, kým sa dostavia (napr. zmeny v očakávanej dĺžke života alebo plodnosti v dôsledku reformy zdravotníctva) ako iné (napr. úžitok z absolvovania školiaceho programu). **Časový harmonogram, ako aj trvanie jednotlivých fáz experimentu je potrebné plánovať vopred.** Priamo totiž ovplyvňuje výšku nákladov na realizáciu RCT.

⁶⁾ Program PISA (**Programme for International Student Assessment**) je celosvetová štúdia Organizácie pre hospodársku spoluprácu a rozvoj (OECD) v členských a nečlenských krajinách zameraná na hodnotenie vzdelávacích systémov

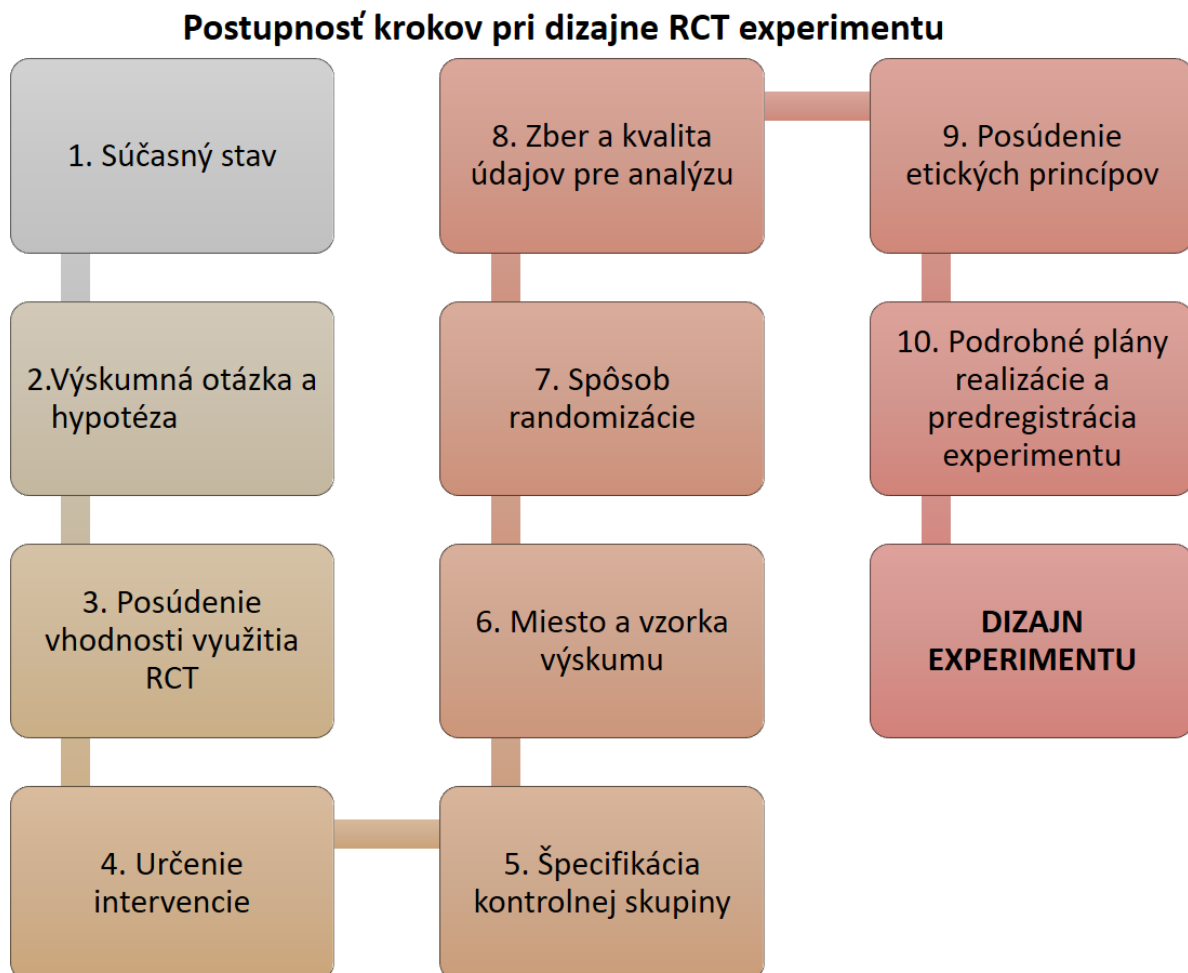
2 Prípravná fáza RCT

V rámci prípravnej fázy začíname s tvorbou dizajnu experimentu. Tvorba vhodného dizajnu je pre úspešnú realizáciu RCT veľmi dôležitá. Chyby v dizajne, realizácii, veľkosti vzorky môžu viesť k skresleniu alebo k znehodnoteniu výsledkov.

2.1 Dizajn experimentu

Dobre navrhnutý dizajn experimentu si vyžaduje v prvom rade dobré **plánovanie**, hlboký **prehľad** existujúcich informácií o skúmanom probléme a formulovanie vhodnej výskumnej **otázky**, návrh intervencií a zhodnotenie významu výstupov pre meranie skúmanej premennej.

Obrázok 3: Dizajn experimentu



Príprava dizajnu RCT:

- 1. Preskúmanie v súčasnosti nastavených cieľov programu alebo politiky** a spôsobu, akým sa tieto ciele dosahujú. Analýza súčasného stavu využitím procesných máp, logického rámca, modelu teórie zmeny,⁷⁾ hodnotenia potrieb alebo hodnotenia podľa teórie programu rozkladá teoretický postup fungovania programu a popisuje logické kroky, ktorými navrhovaná politika napĺňa identifikovanú potrebu.
- 2. Stanovenie výskumnej otázky**, ktorú chceme položiť a formulácia **hypotézy** na základe výsledkov existujúcej vedeckej literatúry (o uskutočnených RCT).

Stanovenie výskumnej otázky a hypotéz závisí od potrieb a záujmu realizátora experimentu. V prvom kroku je dôležité **preskúmať existujúcu vedeckú literatúru**.

Ak na túto tému existuje predchádzajúci výskum, je vhodné nadviazať na výsledky týchto štúdií. V prípade, ak otázka nebola skúmaná alebo nebol použitý RCT, je potrebné vyhodnotiť, či je RCT v tomto prípade vhodnou výskumnou metódou.

Pri verejných politikách existuje široké spektrum otázok, pre využitie RCT je však potrebné pri formulovaní výskumnej otázky zamerať sa na **otázky o vplyve**: „*Fungoval program alebo politika?*“ RCT však môžu ísť aj nad rámec toho, čo funguje, a skúmať otázky, ako napríklad:

- *Ktoré zložky intervencie sú najdôležitejšie pre dosiahnutie účinku?*
- *Ktorá verzia alternatívnych intervencií má najväčší vplyv na správanie účastníkov?*
- *Sú výsledky dosiahnuté v jednom kontexte replikovateľné v inom kontexte?*
- *Aké sú základné procesy na dosiahnutie (alebo nedosiahnutie) vplyvu?*

Pri formulovaní konkrétnej otázky musíme uvažovať o tom, aký vplyv by mali výsledky získané pomocou RCT pri prijímaní rozhodnutí v politike a praxi a tiež, či táto otázka môže byť zodpovedaná:

- *dobre (napr. Možno merať výsledky?),*
- *presne (napr. Máte dostatočnú veľkosť vzorky?),*

⁷⁾ Teória zmeny modeluje, ako konkrétne intervencie vytvárajú výsledky vedúce k dosiahnutiu očakávaných dopadov. Môže byť navrhnutá pre akúkoľvek úroveň intervencie – udalosť, projekt, program, politiku, stratégiu alebo organizáciu. Teória zmeny má svoje miesto najmä počas fázy plánovania experimentu, ale môže byť užitočná aj na monitorovanie a hodnotenie.

- *reprezentatívne (napr. Je možné zásah rozšíriť neskôr?),*
- *eticky.*

Formulovanie výskumnej otázky je **zásadným krokom pre vnútornú i vonkajšiu platnosť výsledkov výskumu.**⁸⁾ Existujú nástroje, ktoré pomáhajú výskumníkom tým, že poskytujú podrobné pokyny na formuláciu výskumnej otázky (napr. rámec PICO - *Populácia, Intervencia, Kontrolná skupina a Výsledok* - na formulovanie výskumnej otázky).

Následne musíme vyhodnotiť, či sa dá na túto otázku odpovedať testovaním kauzálneho vzťahu alebo hypotézy a položiť si ďalšie otázky: *Ktoré komponenty môžu byť randomizované (napr. Kto dostane program, kto dodá program alebo kedy bude program doručený)? Ako program a teória zmeny mapujú výskumnú otázku? Aké údaje môžu výskumníci zhromaždiť, aby odpovedali na výskumnú otázku?* a pod.

Po zedefinovaní otázky, definujeme hypotézy. **Hypotézy** predstavujú tvrdenia, ktoré je možné testovať s využitím štatistických testov. Testy hypotéz pozostávajú z:

- špecifikácie nulovej hypotézy H_0 (súčasný stav),
- špecifikácie alternatívnej hypotézy H_1 ,
- špecifikácie testovacej štatistiky a jej rozdelenia podľa nulovej hypotézy,
- výberu hladiny významnosti α na určenie oblasti odmietnutia,
- výpočtu testovacej štatistiky zo vzorky údajov.

3. Návrh výskumnej metódy - posúdenie vhodnosti RCT na zodpovedanie stanovenej výskumnej otázky. RCT je vhodné využiť najmä na testovanie dopadov. V tomto kroku je potrebné kritické posúdenie realizovateľnosti RCT s ohľadom na stanovený výskumný záujem, ako aj s zhodnotením vhodnosti očakávaných výsledkov pre meranie dopadu.

4. Určenie intervencie - aký zásah je potrebný na zodpovedanie výskumnej otázky prípadne testovanie hypotézy. V zásade určenie vhodnej intervencie vyplýva z teórie, modelu alebo konkrétnej politiky, ktorú chce výskumník otestovať. V praxi je však

⁸⁾ Vnútorná platnosť (angl. *Internal validity*) je miera istoty, že kauzálny vzťah, ktorý je v experimente testovaný, nie je ovplyvnený inými faktormi alebo premennými. Vonkajšia platnosť (angl. *External validity*) je rozsah, v akom možno zistené výsledky zovšeobecniť aj na iné kontexty. Vnútorná i vonkajšia platnosť výsledkov závisí od konkrétneho dizajnu experimentu.

určenie konkrétneho zásahu komplikovaným procesom, pretože existujú obmedzenia týkajúce sa toho, čo môže organizácia uskutočniť/poskytnúť a tiež praktické limity v dizajne (napr. koľko textu sa zmestí do formulára alebo aký má experimentátor prístup k cieľovej skupine - občanom).

Intervencia by mala v ideálnom prípade **testovať** iba **jeden mechanizmus alebo prvok** a mala by byť rovnakým spôsobom uskutočňovaná vo všetkých náhodne pridelených jednotkách. Ak chceme napríklad otestovať nástroje trhu práce je niekedy potrebné otestovať kombináciu intervencií, ktoré nemožno od seba oddeliť.

Súčasne s prípravou intervencií je dôležité zväziť **počet testovaných skupín**, čo úzko **súvisí s veľkosťou vzorky**. Čím viac je testovaných skupín pre danú populáciu, tým menšia je veľkosť merateľného účinku v porovnaní s kontrolnou skupinou, čo môže viesť k tomu, že nebude možné otestovať hypotézu obsiahnutú v každej intervencii.

- 5. Špecifikácia kontrolnej skupiny**, označujeme ju aj porovnávacía alebo skupina bez intervencie. Kontrolná skupina je veľmi dôležitá, pretože vytvára kontrafaktuálny⁹⁾ obraz toho, čo by sa stalo bez zásahu.

Rozhodovanie o tom, čo predstavuje bežný stav, však môže byť **problematické**. Môže sa totiž líšiť z hľadiska lokality (rôzne pobočky tej istej organizácie môžu mať rôzne postupy) a tiež v priebehu času. Podmienkou úspešnosti je zabezpečiť tzv. oslepenie jednotlivých skupín (s intervenciou a bez intervencie) – teda aby sa členovia navzájom nedozvedeli či sú/nie sú intervenovaní a nesnažili sa ovplyvniť kompetentných.

- 6. Rozhodnutie o mieste výskumu (lokality) a vzorke (účastníkoch experimentu)**. Lokalita výskum závisí od skúmaného problému. Napríklad, ak skúmame intervencie na podporu vzdelávania nízkopríjmových skupín obyvateľstva, môžeme RCT realizovať v regiónoch s vysokou koncentráciou nízkopríjmových domácností. Dostupnosť dostatočného počtu ľudí alebo jednotiek zasiahnutých konkrétnou politikou alebo programom je veľmi dôležitá: ešte **predtým, ako bude odsúhlasená realizácia RCT** musíme vedieť, či existuje dostatok jednotlivcov alebo iných jednotiek (domácností, firiem, pacientov a pod.) na zachytenie dopadov testovania. **Výpočet veľkosti vzorky** závisí od predpokladanej veľkosti účinku intervencie a variability vzorky.

⁹⁾ Podmienečné vyhlásenie o tom, čo „by sa stalo“, keby sa stalo niečo, čo sa v skutočnosti nestalo.

7. Rozhodnutie o postupe náhodného pridelovania do skupín (tzv. randomizácie).

Randomizácia je jadrom RCT, takže je dôležité, aby bol tento postup správny, inak RCT zlyhá.

Pri randomizácii je potrebné zabezpečiť, aby prebehla naozaj náhodne, bez ovplyvňovania členov tímu pracujúcich na experimente. Iba dôslednou minimalizáciou ovplyvňujúcich faktorov je možné získať jasný a interpretovateľný výsledok.

8. Zber údajov, ktoré sú potrebné na vykonanie analýzy – t. j. údajov o výsledku, ale aj údajov o iných opatreniach.

RCT sú náročné na údaje a kvalita údajov musí byť veľmi vysoká. Vo fáze plánovania je dôležité zabezpečiť, aby sa zbierali správne údaje.

Môže byť k dispozícii niekoľko druhov údajov, ktoré možno merať rôznymi spôsobmi, buď pozorovaním alebo prieskumom. V programe môže existovať celý rad výsledkov a je dôležitý výber tých, ktoré je potrebné zbierať. Zbieranie zbytočne širokého množstva údajov zvyšuje náklady RCT. Na druhej strane nedostatočné množstvo údajov a chýbajúce údaje môžu mať vplyv na validitu a robustnosť získaných výsledkov.

9. Posúdenie etických princípov,

ktoré musia byť zabudované do všetkých aspektov dizajnu. Významnou súčasťou prípravy dizajnu experimentu je zabezpečenie dodržiavania etických princípov. Napr. výskumná štúdia s ľudskými účastníkmi by mala byť navrhnutá tak, aby maximalizovala úžitok z výskumu a minimalizovala riziká pre účastníkov.

Existuje široká odborná zhoda v tom, že náhodné hodnotenie by sa nemalo používať na testovanie účinkov nároku, ktorý majú právo získať všetci členovia cieľovej populácie. V tomto prípade by bolo neetické vytvoriť porovnávaciu skupinu, ktorá by nemohla prijať program, aj keď by bolo možné získať ďalšie poznatky z vykonania náhodného hodnotenia. Z rovnakého dôvodu výskumníci nemôžu testovať zásahy, ktoré s vysokou pravdepodobnosťou spôsobia poškodenie zdravia alebo života účastníkov. V iných prípadoch môže byť užitočné porovnať dve rôzne verzie zásahu, napríklad existujúcu verziu a verziu s pridaným novým komponentom. Tieto problémy je možné riešiť úpravou návrhu dizajnu.

Informovaný súhlas účastníka o jeho zapojení do experimentu - proces, ktorým subjekt dobrovoľne potvrdí svoju ochotu zúčastniť sa konkrétneho experimentu po tom, čo bol informovaný o všetkých aspektoch, ktoré sú relevantné pre rozhodnutie účastníka zúčastniť sa. V niektorých prípadoch individuálny súhlas môže nahradiť skupinový súhlas (napríklad súhlas riaditeľa školy za všetkých žiakov danej školy) alebo nemusí byť takýto súhlas

vyžadovaný, a to najmä z logistických dôvodov. **Posúdenie etickej roviny navrhovaného experimentu je v kompetencii etickej komisie inštitúcie, ktorá daný experiment vykonáva, ak taká existuje, alebo etickej komisie vyššej inštalácie.**

Posúdenie etickej roviny je komplexná úloha, keďže každý RCT je špecifický a nie je možné uplatniť univerzálny postup na všetky RCT. Formuláre, ktoré je potrebné pre posúdenie dodržania etických princípov pri realizácii experimentu používať, obsahujú celé spektrum otázok. Odpoveď áno na konkrétnu otázku ešte nevytvára predpoklad, že daný RCT je neetický. Iba prináša do hodnotenia ďalšiu dimenziu, ktorú musí komisia po oboznámení sa s detailmi štúdie posúdiť.

PRÍKLAD

Príklady otázok z formulára na posúdenie etickosti štúdie:

- *Majú podporovatelia/sponzori vplyv na dizajn štúdie, zber a analýzu údajov, rozhodovanie o publikovaní alebo príprave záverečnej správy?*
- *Zahŕňa skúmaná vzorka deti (vo veku menej ako 18 rokov), mentálne nespôsobilé osoby, pacientov, príslušníkov etnických menšín, jednotlivcov, ktorí sú vo väzbe alebo žiakov a študenti v škole prípadne osoby, ktoré sú v profesionálnom alebo klientskom vzťahu s výskumníkmi zapojenými do tohto projektu?*
- *Zahŕňa navrhovaný výskum spracovanie citlivých údajov (napr. poloha, zdravie, životný štýl, etnická príslušnosť a politické presvedčenie, náboženské alebo filozofické presvedčenie)?*
- *Spôsobuje štúdia pre účastníkov riziko, ujmu alebo negatívne dôsledky, ako napríklad (ale nie výlučne) možnosť psychickej, sociálnej, ekonomickej alebo právnej ujmy pre účastníka?*
- *Zahŕňa navrhovaný výskum privodenie bolesti, viac ako mierneho nepohodlia, alebo vyvoláva u účastníkov psychický stres či úzkosť?*
- *Zahŕňa navrhovaný výskum podvod?¹⁰⁾*
- *Spolieha sa výskum na platformu, kde údaje respondentov môže monitorovať tretia strana (napríklad Amazon Turk, eBay, ...) alebo je pravdepodobné, že akékoľvek informácie zo*

¹⁰⁾ Klamanie znamená zavádzanie alebo nesprávne informovanie účastníkov o skutočnej povahe situácie. Inými slovami, neexistuje žiadny podvod, ak čokoľvek, čo povieme subjektom v experimente, je pravda. V ekonómii je klamanie alebo zavádzanie účastníkov počas experimentu považované za neetické správanie. Je však prirodzené, že o niektorých informáciách účastníci nie sú informovaní, napríklad o existencii viacerých testovaných skupín s rôznymi intervenciami.

štúdie budú v priebehu realizácie RCT postúpenia externým spoločnostiam alebo organizáciám?

- *Odpiera výskum osobám nezapojeným sa do experimentu prístup k službe, na ktorú by za bežných okolností mali nárok?*

10. Príprava podrobných plánov realizácie RCT vrátane predregistrácie experimentu.

Príprava výskumného plánu pomáha utriediť rôzne prvky experimentu a ich interakcie.

Súčasťou plánu je aj analýza nákladovej efektívnosti (tzv. *Cost efficiency analysis*, skr. CEA), ktorá hodnotí jednotlivé programy z hľadiska jednoduchého pomeru nákladov a ich dopadov a umožňuje nám porovnanie programov v čase a mieste. CEA sama osebe nemusí poskytovať dostatočné informácie na všetky politické alebo investičné rozhodnutia, ale môže byť užitočným východiskovým bodom pri výbere medzi rôznymi programami, ktorých cieľom je dosiahnuť rovnaký alebo podobný výsledok.

Čoraz bežnejšia je tiež predregistrácia RCT, jeho hypotéz a dizajnu. Výhodou predregistrácie je zabezpečenie transparentnosti a dodržania výskumných postupov.¹¹⁾

2.2 Veľkosť vzorky a pilotné testovanie

Všetky RCT vyžadujú dostatočnú veľkosť vzorky na zistenie vplyvu intervencie. Potrebnú veľkosť vzorky je potrebné vypočítať vo fáze tvorby dizajnu, teda ešte pred realizáciou RCT. Od veľkosti vzorky závisí pridelenie množstva účastníkov do kontrolnej skupiny a testovaných skupín, ako aj počet možných testovaných skupín.

Keď je **vzorka príliš malá**, prirodzené výkyvy vo výsledku nám nemusia umožniť presvedčivo (štatisticky významne) zmerať vplyv intervencie. Nevieme potom povedať, či rozdiel medzi testovanou skupinou a kontrolnou skupinou je výsledkom konkrétnej intervencie alebo čistou náhodou. RCT v tejto situácii by zbytočne zaťažilo účastníkov štúdie a spotrebovalo zdroje, ktoré by sa dali produktívnejšie využiť inde.

Pri RCT je cieľom mať vzorku, ktorá je **schopná s primeraným stupňom istoty odhaliť účinok intervencie meraný rozdielom vo výsledkoch medzi testovanou skupinou a kontrolou skupinou alebo medzi viacerými skupinami s intervenciou.**

¹¹⁾ Pre viac informácií o predregistrácii pozri napríklad tu <https://www.socialscienceregistry.org/>.

Vzťah medzi veľkosťou vzorky a jej schopnosťou odhaliť účinok je možné vymedziť nasledovne:

- **čím väčšia** je vzorka, tým skôr **odhalíme aj menší** účinok intervencie (za predpokladu nezmenených ďalších podmienok),
- tento vzťah závisí aj od **rozdelenia pozorovaní** výslednej premennej - **širšia** distribúcia **zväčšuje** veľkosť vzorky potrebnej na testovanie danej veľkosti účinku. To znamená, že ak výsledky skúmanej intervencie nadobúdajú veľké množstvo hodnôt, je potrebná väčšia vzorka na zhodnotenie veľkosti dopadu.

Na odhadnutie veľkosti vzorky potrebujeme predbežné informácie o **veľkosti očakávaného účinku a štandardnej odchýlke**. Najčastejšie použijeme účinok z vedeckého článku, z ktorého vychádzame, alebo vypočítaný predpoklad nášho teoretického modelu. Smerodajnú odchýlku možno zistiť ľahšie ako veľkosť účinku, pretože je pravdepodobné, že sú k dispozícii údaje o existujúcich výsledkoch, z ktorých je možné vypočítať rozdelenie – strednú hodnotu i štandardnú odchýlku.

Určenie veľkosti efektu môže byť pri experimentoch vo verejnom sektore náročné. Namiesto výpočtu veľkosti vzorky z predpokladanej veľkosti účinku je možné na výpočet veľkosti vzorky použiť **minimálnu prijateľnú alebo nákladovo efektívnu veľkosť**. odhadnutím výšky minimálneho detekovateľného efektu (angl. minimum detectable effect, skr. MDE).

Súčasťou výpočtu vzorky je aj posúdenie hodnoty štatistickej sily. **Štatistická sila** hovorí o **veľkosti pravdepodobnosti odmietnutia falošnej nulovej hypotézy**. Keď zostavujeme experiment, chceme ho vytvoriť taký, aby dokázal odhaliť efekt s nízkou mierou rizika, že ide o chybu. Je dôležité vyhnúť sa falošne pozitívnemu výsledku.

PRÍKLAD

Skúmame ako nastaviť nástroje týkajúce sa podpory efektivity zdravotnej starostlivosti, ktorých prínosom je, ak sa zníži počet hospitalizácií o 10 percent, ale v nadväznosti na nízku početnosť skúmanej vzorky ste schopní zistiť zmeny v počte hospitalizácii len o 20 percent alebo viac. Ak sa počet hospitalizácií po zrealizovaní výskumu pre účastníkov štúdie zníži o (štatisticky nevýznamných) 15 percent, nemôžete si byť istí, že intervencia znížila počet hospitalizácií o 10 percent, ale tiež si nemôžete byť istí, že intervencia znížila počet hospitalizácií o 0 percent (t. j. nemal žiadny vplyv). Výsledky RCT teda ponúkajú nepresvedčivé dôkazy o tom, či by takto nastavený program mal alebo nemal pokračovať. Nezistenie štatisticky významného účinku môže byť nesprávne interpretované ako zlyhanie programu, a nie ako zlyhanie RCT. Ak je výskum realizovaný na nedostatočnej vzorke a s nedostatočnou štatistickou silou hovoríme, že sme nenašli žiadny dôkaz účinku, ale to neznamená, že sme našli dôkaz žiadneho účinku. Médiá a široká verejnosť však môžu ľahko spojiť pojmy „nenájdenie

dôkazov o účinku“ so „zistením bez účinku“. Výsledkom je, že nepresvedčivé zistenia môžu poškodiť reputáciu organizácie alebo programu takmer rovnako ako zistenia bez účinku.

Pri **jednoduchších dizajnoch** je možné využiť aj webové kalkulačky, ktoré sú ľahko dostupné a relatívne ľahko použiteľné.¹²⁾ Pre komplexnejšie dizajny existujú špeciálne softvérové aplikácie (napr. *Optimal Design* (<http://hlmssoft.net/od/>) a *GPower* (<http://www.gpower.hhu.de>)).

Šesť pomocných pravidiel na určenie veľkosti vzorky a štatistickej sily

- *Pravidlo č. 1: Väčšia vzorka zvyšuje štatistickú silu hodnotenia.*
- *Pravidlo č. 2: Ak je veľkosť efektu programu nízka, je potrebné mať väčšiu vzorku na dosiahnutie robustného výsledku.*
- *Pravidlo č. 3: RCT programu s nízkou mierou využitia si vyžaduje väčšiu vzorku.*
- *Pravidlo č. 4: Ak sú u skúmanej populácie veľké rozdiely vo výsledkoch, pre dosiahnutie platných výsledkov potrebujete väčšiu vzorku.*
- *Pravidlo č. 5: Sila vzorky je maximalizovaná, ak je vzorka rovnomerne rozdelená medzi testovaciu a kontrolnú skupinu.*
- *Pravidlo č. 6: Randomizácia na úrovni klastra na rozdiel od individuálnej úrovne znižuje štatistickú silu RCT. Čím podobnejšie sú výsledky jednotlivcov v rámci klastrov, tým väčšia musí byť vzorka.*

Výhodou veľkého súboru subjektov je, že môžeme analyzovať aj podskupiny s vysokou mierou platnosti zistených výsledkov. Vyššia veľkosť vzorky je spojená aj s vyššími nákladmi na jej získanie. Limitujúcim faktorom sú aj stropy veľkosti vzorky, ktoré môžu vyplývať tak z limitov týkajúcich sa finančných a personálnych zdrojov alebo z predmetu zamerania výskumu: počet samosprávnych alebo geografických oblastí (okresov, regiónov, a pod.) je limitovaný alebo kapacitné obmedzenia týkajúce sa kontaktovania účastníkov v určitom časovom období (v rámci ktorého splnia podmienku na zaradenie do experimentu – napr. nezaplatia daň načas).

Pilotné testovanie RCT umožňuje **overenie vhodnosti dizajnu** s ohľadom na riešený problém i stanovenú výskumnú otázku ako aj **testovanie využiteľnosti získaných dát** pri hodnotení dopadov (obrázok 4).

¹²⁾ Tieto kalkulačky fungujú tak, že do príslušných políček zadáme očakávanú úroveň významnosti, veľkosť účinku a smerodajnú odchýlku a kalkulačka vypočíta požadovanú veľkosť vzorky v určenom políčku. Ešte raz zdôrazňujeme, že tento typ nástroja je vhodný len pre jednoduché dizajny RCT. Napríklad kalkulačka Austrálskeho štatistického úradu (dostupná tu <https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator>).

Obrátok 4: Kroky prípravy dizajnu RCT



Zdroj: Getler et al. (2016)

Pilotné testovanie dizajnu experimentu je realizáciou navrhnutého RCT na menšej vzorke s cieľom:

- identifikovať chyby,
- pripraviť sa na výzvy spojené s hodnotením intervencie.

Pilotné testovanie sa uskutočňuje podľa plánu pre plánovaný RCT v dostatočnom časovom predstihu pred ostrou realizáciou konkrétneho RCT. Počas pilotného testovania zároveň získavame dáta, čo nám umožňuje verifikovať, či dáta spĺňajú požadovanú štruktúru, a či je vykazovanie dát v súlade s našimi požiadavkami. Ešte pred realizáciou RCT tak môžeme implementovať opatrenia a úpravy pri zistených problémoch so zberom dát počas RCT, čím šetríme náklady na realizáciu experimentu.

Pilotné testovanie je užitočné najmä vtedy, keď je hodnotený program (nová dávka, dotácia a pod.) **nový** a stále prebieha tvorba jeho logistiky, alebo v prípadoch, keď si RCT vyžaduje **významné zmeny** vo fungovaní programu.

PRÍKLAD

Príklady otázok, na ktoré môže pilotné testovanie odpovedať:

- **Nábor účastníkov** - Je program schopný pritiahnúť dostatočný počet jednotlivcov, aby zaplnili dostupné miesta v testovanej skupine a kontrolnej skupine?
- **Aktivita účastníkov** - Aké percento ľudí zaradených do testovanej skupiny sa skutočne zapája do programu (napr. využívajú skúmanú dávku /službu/vyšetrenie, na ktorú majú nárok)? Poznámka: Nízka miera využitia môže negatívne ovplyvniť štatistickú silu hodnotenia.
- **Prechody** - Zúčastňujú sa programu ľudia, ktorí boli pôvodne zaradení do kontrolnej skupiny? Poznámka: Veľký počet prechodov môže sťažiť odhad dopadu programu.

- **Zdieľanie údajov** - Máme k dispozícii systémy na bezpečné zdieľanie údajov? Vyskytli sa problémy s chýbajúcimi údajmi alebo zlou kvalitou údajov?

2.3 Náhodné hodnotenie (randomizácia)

Randomizácia a anonymizácia sú dve metódy na **zníženie skreslenia výberu**¹³⁾ vo výskumných štúdiách.

Randomizácia poskytuje **mechanizmus pre nestranné a náhodné pridelovanie intervencie a zabezpečuje, aby skupiny vytvorené na uskutočňovanie experimentov boli navzájom čo najviac podobné a výsledky, čo najpresnejšie.**

Mechanika randomizácie:

1. potrebujeme zoznam všetkých účastníkov,
2. každému pridáme náhodné číslo,
3. rozhodneme, ktoré číslo (poradie/rozsah/párne/nepárne) bude patriť kontrolnej skupine a testovanej skupine.

Túto úlohu je možné vykonať pomocou softvéru, ktorý obsahuje funkciu na generovanie náhodných čísel alebo na náhodný výber čísel zo súboru dát. Účastníkov potom zoradíme podľa týchto náhodných čísel a v samostatnom stĺpci podľa dohodnutej veľkosti pre každú skupinu štúdie pridáme identifikátor kontrolnej alebo testovanej skupiny (napr. 1, 2, 3, 4 atď.) Následne zoradíme účastníkov naspäť podľa poradia identifikátorov. Ekonometrické softvérové balíky, ako napríklad R a Stata, majú príkazy, ktoré vykonávajú randomizáciu automaticky a rovnakým spôsobom.

Randomizáciu je možné urobiť aj podľa špecifických **charakteristík účastníkov experimentu**, napríklad randomizácia na úrovni regionálnych oblastí. Konkrétny proces randomizácie závisí od konkrétneho výskumného zámeru a cieľa experimentu. Ak napríklad skúmame zavedenie online služby a pokrytie internetovým signálom sa v regiónoch výrazne odlišuje, musia byť obyvatelia regiónov zastúpení v kontrolnej a testovanej skupine v rovnakej miere.

PRÍKLAD

¹³⁾ Pod pojmom skreslenie výberu chápeme systematickú chybu, ktorá vzniká, keď výskumník nie je schopný náhodne zaradiť účastníkov do intervenčných a kontrolných skupín. Ak dôjde k skresleniu výberu, skupiny sa budú pravdepodobne líšiť v meraných alebo nameraných základných charakteristikách kvôli spôsobu, akým boli účastníci vybraní alebo priradení.

Niekedy randomizáciu nie je možné vykonať. Napríklad nemôžeme vykonať náhodné hodnotenie vplyvu pohlavia na výsledky vzdelávania, keďže nie je možné použiť randomizáciu na zistenie, ako by sa inak „identickéj“ osobe darilo v škole, keby bola iného pohlavia. Môžeme ale pomocou RCT študovať napríklad diskrimináciu na trhu práce. Existujú štúdie, ktoré vytvorili hypotetické životopisy, kde sa zmenilo iba meno osoby, aby naznačovalo jej pohlavie alebo etnický pôvod, a potom skúmali počty odpovedí zamestnávateľov a pozvanie na pohovor v každom variante.

2.4 Riziká RCT, najčastejšie problémy a možnosti ich riešenia

Medzi problematické oblasti realizácie RCT môžeme zaradiť:

1. Etické otázky

Problém: V niektorých prípadoch môžu existovať etické obavy z uskutočnenia RCT -ak nemá napríklad realizátor zdroje na pokrytie celej vzorky štúdie.

Riešenie: Urobíme RCT bez obmedzenia prístupu k intervencii. Napríklad, náhodne vyberieme ľudí, ktorí dostanú poštuchnutie/pripomenutie (e-maily s pripomienkou alebo telefonáty), aby sa zapísali do programu bez toho, aby sme zamietli prístup akýmkoľvek zainteresovaným účastníkom.

Problém: Ak má implementujúci partner zdroje na pokrytie celej vzorky, ale ešte nevie, či je intervencia účinná.

Riešenie: Použijeme randomizovaný návrh zaradenia na počiatočné zapojenie do programu iba pre časť vzorky a zistíme tak účinky intervencie pred rozšírením testovania na celú vzorku.

2. Realizovateľnosť experimentu

Problém: V randomizovanom dizajne výskumníci náhodne rozhodujú o tom, kto program dostane a kto nie. Niekedy je to neuskutočniteľné.

Riešenie: Upravíme dizajn a napríklad randomizujeme naprieč lokalitami, kde účastníci skupiny dostávajú intervenciu alebo naprieč poskytovateľmi programu/služby.

Problém: Náklady a úsilie na vykonávanie prieskumov a implementáciu programu majú vplyv na dizajn a veľkosť vzorky RCT. Ak existujú zdroje na rozšírenie programu na všetky subjekty v sledovanej oblasti.

Riešenie: Použijeme postupný dizajn (vo fázach) a náhodne priradíme poradie, v ktorom sú skupiny alebo jednotlivci zaradení do testovanej skupiny s intervenciou.

Problém: Program má príliš prísne kritériá oprávnenosti.

Riešenie: Uvoľníme kritériá oprávnenosti a náhodne vyberieme medzi jednotlivcami na hranici oprávnenosti. Ak sa do programu môžu zapojiť tí, ktorí majú nárok, ale musia sa prihlásiť, náhodne distribuujem poštuchnutie/pripomenutie, aby sa subjekty zapojili do programu.

3. Veľkosť vzorky

Problém: Aby výsledky RCT boli robustné a platné, je potrebné zabezpečiť, aby skúmaná mala vzorka dostatočnú mieru štatistickej sily – identifikovať koľko jednotiek (osôb, domácností, firiem, inštitúcií) zahrnúť do štúdie, aký je pomer jednotiek pridelených do každej skupiny a na akej úrovni randomizovať (napr. študent verzus trieda študentov verzus celá škola).

Riešenie: Ak je veľkosť vzorky malá, randomizujeme na nižšej úrovni a/alebo použijeme stratifikované náhodné priradenie.¹⁴⁾ Obmedzením tohto je riešenia je, že zníženie úrovne randomizácie môže zvýšiť efekty prelievania a stratifikácia na príliš veľa premenných môže vytvoriť nevyvážené podskupiny.

¹⁴⁾ Stratifikovaná randomizácia je dvojstupňový postup, pri ktorom sú účastníci RCT najprv rozdelení do skupín (nazývaných vrstvy, angl. strata) podľa určitých vlastností, ktoré môžu ovplyvniť výsledné správanie. V rámci každej z týchto skupín sú potom účastníci priradení do testovanej skupiny podľa samostatných randomizačných schém. Pre viac informácií o stratifikovanej randomizácii pozri Gerber a Green (2012).

4. Ohrozenie vnútornej validity

Problém: Rovnako ako všetky metódy hodnotenia, aj RCT čelia ohrozeniu svojej vnútornej platnosti.

Riešenie: Použijeme postupy aplikácie intervencie a zberu údajov na minimalizáciu alebo aspoň meranie takýchto hrozieb. Napríklad výskumníci, ktorí sa zaoberajú prelievaním, môžu randomizovať na vyššej úrovni (napr. úroveň školy verzus úroveň študentov), aby sa vyhli prelievaniu efektu na kontrolnú skupinu a zmerali úplný účinok v skupine intervenciou.

3 Realizácia RCT

Po úspešnom nadizajnovaní experimentu a prípadnom úspešnom pilotnom otestovaní môžeme pristúpiť k samotnej realizácii experimentu. Realizáciou experimentu dostávame odpovede na hypotézy a ich analýzou informáciu o dosiahnutých výsledkoch.

Je nevyhnutné, aby každá osoba zapojená do procesu pochopila a poznala svoju úlohu pri realizácii experimentálnych krokov.

Najčastejšie príčiny ovplyvňujúce priebeh realizácie experimentu sú:

- neúplné zadanie,
- nedodržovanie dizajnu experimentu,
- nepresne určené požiadavky,
- nedostatočne odhadnuté náklady,
- nedostatočná kvalifikácia pracovníkov pracujúcich na experimente.

4 Analýza RCT

Posledným štádiom realizácie RCT je analýza získaných dát, interpretácia a reportovanie výsledkov v zrozumiteľnej forme. Aj v tejto fáze si musíme dať pozor, pretože RCT môže byť ovplyvnené množstvom faktorov, čo prináša skreslenia tak vo fáze analýzy výsledkov, ako aj ich reportovania.

4.1 Zber a spracovanie dát

Zber a spracovanie dát prebiehajú v dvoch odlišných fázach realizácie experimentu. Zber dát zväčša prebieha počas RCT, zatiaľ čo spracovanie dát sa realizuje po ukončení experimentu.

4.1.1 Zber dát

Plánovanie zberu dát je potrebné už pri vypracovávaní dizajnu experimentu, pričom sa zameriavame na dve základné skupiny údajov.

- **Základné (primárne) údaje** - stanovili sme si hypotézu a na jej potvrdenie zbierame dáta tak, aby bolo možné pozorovať vplyv konkrétnej zavedenej intervencie dostatočne robustne. Ide o základné **informácie o zapojení účastníka do kontrolnej skupiny alebo testovanej skupiny** s intervenciou a jeho správaní. Zber týchto údajov je základným predpokladom vyhodnotenia dopadov intervencie.

- **Doplňkové (sekundárne) údaje** - do tejto kategórie údajov spadajú najmä kovariáty, ktoré môžu zahŕňať rozličné charakteristiky účastníkov, ktoré získavame jednak pre to, aby sme mohli otestovať správnosť randomizácie, ako aj preto, aby sme mohli kontrolovať heterogénne účinky intervencie pri analýze zistených výsledkov. Na lepšie pochopenie výsledkov je tiež užitočné zbierať aj informácie o kontexte, nákladovej efektívnosti a kvalitatívnych informáciách („prečo“ a „ako“). Doplňkové údaje môžu zahŕňať aj ďalšie administratívne údaje alebo údaje zo sčítania.

Mnohé RCT sa pri meraní výsledkov spoliehajú na uskutočňovanie prieskumov. Vďaka rozvoju technológií sa zbierajú inovatívne a je možné realizovať veľmi špecifické merania. Môžeme merať silu individuálnych preferencií, povedzme, ak ide o riskovanie alebo oplatenie veľkorysého gesta; podrobné údaje o polohe alebo GPS; informácie o sociálnych sieťach a finančných vzťahoch v obci; kvalite a množstve nakupovaného a spotrebovaného tovaru, atď.

Okrem údajov z prieskumov sa čoraz častejšie využívajú aj **administratívne údaje**, ktoré umožňujú odpovedať na otázky o charakteristikách veľkého množstva subjektov za relatívne nízke náklady (napr. údaje z verejných registrov). Administratívne údaje sú mimoriadne užitočné napríklad na uskutočnenie dlhodobých následných kontrol na vzorke účastníkov štúdie. Dôležitým prvkom je aj kontrola **kvality údajov** pre odhalenie chýb v zaznamenávaných dátach.

UŽITOČNÁ POZNÁMKA

Ak plánujeme použiť sekundárne údaje, nesmieme zabudnúť pri navrhovaní prieskumu premyslieť, ako tieto dva zdroje údajov integrujeme (pre analýzu potrebujeme vytvoriť jeden konzistentný dataset).

4.1.2 Spracovanie dát

Pre spracovanie dát RCT sa štandardne využívajú ekonometrické metódy spracovania dát. V prvom kroku je možné využiť **jednoduchú grafickú analýzu, ktorá je následne dopĺňaná štatistickými testami hypotéz a regresnými modelmi**. Ak je dizajn RCT správne nastavený, pre spracovanie dát je možné využiť menej náročné ekonometrické metódy (napríklad metódu najmenších štvorcov).

Pri spracúvaní získaných dát využívame konečné údaje o správaní účastníkov zaradených v jednotlivých skupinách na porovnanie priemerného výsledku testovanej skupiny s

priemerným výsledkom kontrolnej skupiny po intervencii. Tento rozdiel predstavuje vplyv implementovanej intervencie.

Efekt intervencie = priemer výsledku testovanej skupiny – (mínus) priemer výsledku kontrolnej skupiny

Podľa typu skúmaných premenných (spojité, kategorické) je možné použiť na testovanie hypotéz parametrické testy štatistických hypotéz (napr. *Studentov t-test*) alebo neparametrické testy štatistických hypotéz (napr. *Mann–Whitney U test*). Komplexnejšie hypotézy je možné testovať s využitím regresných modelov (napr. metódou najmenších štvorcov alebo logistickou regresiou).¹⁵⁾ Komplikovanejšie regresné modely zvyšujú presnosť získaných výsledkov, nakoľko umožňujú kontrolovať vplyv doplnkových charakteristík skúmanej populácie. Tieto charakteristiky môžu korelovať s výsledným správaním a tak ovplyvniť výsledky vplyvy testovanej intervencie.

4.2 Meranie efektov

Pri meraní efektov sa opierame o odpovede na základné otázky, ktoré sme si položili už vo fáze prípravy experimentu:

1. *Čo potrebujeme merať?*
2. *Aký typ údajov je potrebný na meranie?*

Pri analýze efektov je veľmi dôležité rozlišovanie medzi **kauzalitou (príčinnosťou) a koreláciou**. RCT sa zameriava na skúmanie **príčinnosti**. Experimentom sa snažíme zistiť, či sa zvyšovaním/znižovaním hodnoty jednej premennej zvyšuje/znižuje hodnota druhej premennej (zmena môže byť aj v opačnom smere). Teda či medzi skúmanými premennými (v našom prípade intervencia – správanie sa) existuje príčinný vzťah a jedna udalosť je výsledkom výskytu druhej udalosti; t. j. medzi týmito dvoma udalosťami existuje príčinná súvislosť (vieme zdefinovať príčinu a jej dôsledok).

Korelácia je štatistická miera (vyjadrená ako číslo), ktorá popisuje veľkosť a smer vzťahu medzi dvoma alebo viacerými premennými. V prípade korelácie sa dve premenné vyvíjajú rovnako, avšak táto miera nehovorí nič o príčinnosti vzťahu týchto premenných (napríklad priemerná dĺžka sukni v USA a ceny ropy sa v sledovanom období môžu vyvíjať rovnako a teda sú korelované, ale neexistuje tu príčinný vzťah).

¹⁵⁾ Viac o testovaní štatistických hypotéz sa môžete dočítať napríklad v Lehmann a Romano (2022).

DÔLEŽITÁ POZNÁMKA

*Pri meraní efektov intervencie pomocou vzťahov medzi skúmanými premennými je pochopenie rozdielov medzi kauzálnym a korelačným vzťahom zásadné. Príčinný vzťah znamená, že jedna udalosť je výsledkom výskytu druhej udalosti a teda **efekt intervencie je dôsledkom kauzality!** Korelácia medzi premennými však automaticky neznamená, že zmena jednej premennej je príčinou zmeny hodnôt druhej premennej a teda **korelácia nie je efektom intervencie!***

V závislosti od stanovenej výskumnej otázky musíme zbierať nasledujúce údaje:

- **Premenné** - základné ako aj doplnkové. Ak sa domnievame, že účinok intervencie sa môže líšiť v závislosti od určitých charakteristík účastníkov alebo miesta štúdie (ako je pohlavie, mestská alebo vidiecka poloha, vzdialenosť k najbližšej nemocnici atď.), musíme zbierať aj údaje o týchto doplnkových charakteristikách.
- **Prediktory súladu a miery skutočného súladu s intervenciou** - na individuálnej a skupinovej úrovni. Usilujeme sa získať údaje o správaní cieľovej skupiny účastníkov (objektívne – napr. pravidelnosť pri platení preddavkov, úhrada pokút; subjektívne – napr. dodržiavanie odporúčaní lekára, dodržiavanie nariadení a pod.), ktoré nám pomôžu odhadnúť správanie. Napríklad pri lekárskej intervencii zahŕňajúcej užívanie lieku chceme na začiatku RCT získať informácie o tom, či respondent v súčasnosti užíva nejaké lieky denne, ako často to zabudne urobiť a potom na konci štúdie chceme informáciu o tom, ako často užívali svoje lieky v rámci štúdie.
- **Proxy indikátory, teda zástupné indikátory** - používame ich vtedy, keď je ťažké zmerať hlavné koncepty, ktoré sa skúmajú napríklad dôsledky právneho správania, zisťovanie celkového príjmu alebo bohatstvo. Vyberáme také, ktoré vykazujú koreláciu s indikátorom očakávaného správania sa.

DÔLEŽITÁ POZNÁMKA - Čím vyššia korelácia, tým lepšia proxy.

- **Kontext** - pre externú platnosť/zovšeobecniteľnosť. Pri určovaní zdrojov údajov hľadáme príklady, ktoré zahŕňajú vystavenie sa iným ale podobným programom, ak je to relevantné, atď.
- **Náklady** - na výpočet nákladovej efektívnosti sú potrebné údaje o nákladoch na samotnú intervenciu, ako aj údaje o cene prípadných ďalších alternatívnych riešení, úsporách/nákladoch, ktoré respondenti v dôsledku intervencie získajú/vynaložia a pod.
- **Kvalitatívne informácie** - „prečo“ a „ako“. Ak napríklad respondent v prieskume odpovie, že jeho dieťa navštevovalo školu častejšie ako v minulosti, následná otázka by sa mala pýtať na hlavný dôvod tejto zmeny z jeho pohľadu.

Väčšina RCT používa panelové údaje, pretože cieľom experimentu je sledovať výsledky a zbierať údaje na úrovni jednotlivcov, ktorí podstúpili/nedostali špecifickú intervenciu. Ak predpokladáme, že účastníci štúdie zmenia svoje správanie sa kvôli účasti v štúdiu, použijeme prierezové údaje.

Kvalita údajov pre meranie efektov

Cieľom merania je získať **spoľahlivé údaje**, pomocou ktorých je možné zodpovedať stanovené výskumné otázky a posúdiť dopad danej intervencie. Nepresné meranie môže viesť k nespoľahlivým údajom, z ktorých je ťažké vyvodiť platné závery.

Merania a zber dát sa v RCT zvyčajne uskutočňuje v dvoch kľúčových fázach:

1. základné meranie - používame ho na zhromažďovanie popisných charakteristík o vzorke účastníkov štúdie (ako je priemerný vek, príjem domácnosti alebo zloženie vzorky podľa pohlavia),

2. konečné meranie – používame ho na odhad účinku intervencie. Najmä pri technicky alebo časovo náročných RCT zbierame údaje aj uprostred intervencie na monitorovanie implementácie intervencie.

4.3 Chyby merania

Základné problémy:

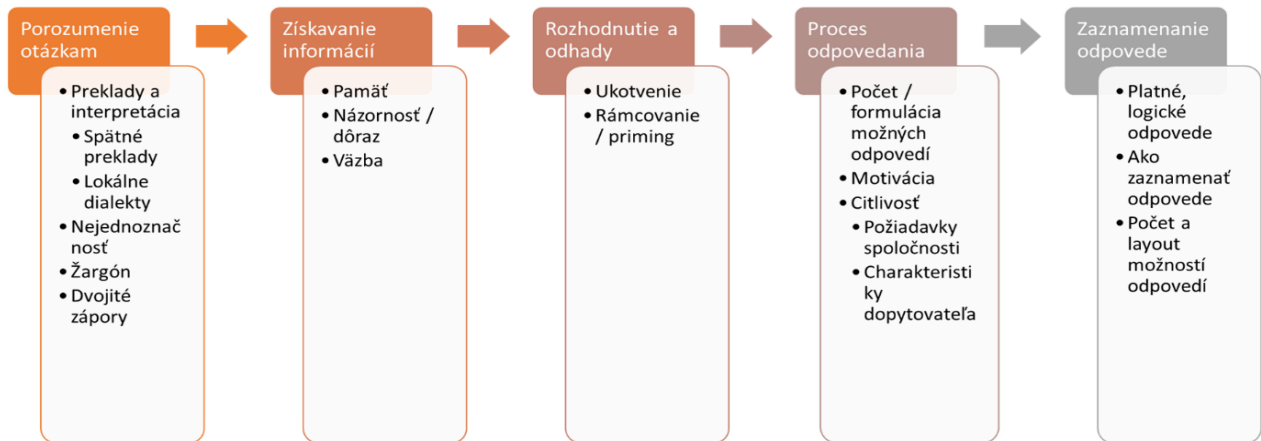
1. nesprávna formulácia základných premenných – príliš široko, zle alebo vágne definované,
2. odpoveď poskytnutá respondentom sa líši od skutočnosti alebo skutočnej hodnoty,
3. vznik systematickej chyby,¹⁶⁾
4. schopnosť merania vyjadriť vplyv skúmanej intervencie nemusí byť dostatočná - *chyba validity*,¹⁷⁾

¹⁶⁾ Chyby merania môžu byť náhodné alebo systematické. Náhodná chyba bude v priemere nula, zatiaľ čo systematická chyba nie.

¹⁷⁾ Je potrebné rozlišovať medzi chybou merania, ku ktorej dochádza počas procesu zberu údajov, od chyby validity, ku ktorej dochádza, keď ukazovatele nezodpovedajú príslušnému konceptu.

5. skreslenie reakcií účastníkov experimentu napríklad rozptýlením, chorobou, zlým porozumením atď.

Obrázok 5: Zdroje chýb merania



Zdroj: World bank (n.d.)

Dôležitá je tiež kontrola vstupných dát. Chyby ľudského faktora alebo technického zariadenia môžu totiž viesť k chybným hodnotám a chybnému vyhodnoteniu efektu. Tieto situácie môžeme zachytiť v rámci pilotného testovania.

Chyby merania sa týkajú všetkých fáz realizácie experimentu. Základné zdroje chýb merania je možné zoskupiť do troch kategórií:

- 1) **problémy s otázkami** - vznikajú v dôsledku zle vytvorených otázok. Problematické otázky zahŕňajú vágnosť, negatíva (napr. Neprestali by ste využívať....?), dvojité otázky, domnienky, a pod.;
- 2) **problémy s odpoveďami** - predstavujú najmä prekrývajúce sa kategórie odpovedí alebo odpovede, ktoré sú ovplyvnené efektom rámčovania;
- 3) **problémy s respondentmi** – vyskytujú sa v dôsledku prirodzenej zaujatosti respondenta pri odpovedaní na otázky. Zahŕňajú skreslenia pripomínania, ukotvenie, skreslenie o spoločenskej potrebe a pod. Respondenti sa môžu rozhodnúť uviesť nesprávny odhad (možno kvôli zaujatosti – napr. subjekty môžu považovať ryžu za nezdravú, a preto budú tvrdiť, že jej jedia menej).

Závažnosť chyby merania závisí od typu a rozsahu chyby, ako aj od toho, či odchýlka koreluje s intervenciou. Skreslenie nekoreluje s intervenciou, ovplyvňuje testovanú aj kontrolnú skupinu rovnako, a preto nebude skresľovať odhad rozdielu medzi dvoma skupinami na konci. **Skreslenie, ktoré koreluje s intervenciou, je závažnejšie:** ovplyvňuje testovanú a kontrolnú

skupinu odlišne, čo znamená, že odhad rozdielu medzi skupinami je v priemere skreslený. To môže viesť k chybnému záveru o smere (v analýze znamienku) a rozsahu účinku intervencie.

4.4 Interpretácia výsledkov

RCT vyžaduje zabezpečenie kvality realizácie, aby sa zaručila kvalita štúdie. Je dôležité poskytnúť zadávateľovi, ale aj odbornej verejnosti dostatočné podrobnosti pri opisovaní metodológie a zistení.

Zverejňovanie výsledkov RCT sa líši v závislosti od konkrétneho informačného kanála. Zatiaľ, čo správa o realizovanom RCT pre zadávateľa môže byť detailná a menej technická, pri publikovaní akademickej štúdie sa zameriavame na priblíženie technických detailov a teoretických predpokladov. Na druhej strane pri informáciách do médií sa zameriame na hlavný zistený výsledok a jednoduchý popis prínosov.

DÔLEŽITÁ POZNÁMKA

Údaje o plánovanej aj skutočnej realizácii RCT vždy zverejňujeme až po ukončení RCT, aby sme predišli ovplyvneniu správania sa účastníkov experimentu.

V nadväznosti na typ konkrétneho informačného kanála môžeme zverejňovať:

- **Základné údaje** - pri opise RCT je dôležité okrem zistených efektov zamerať sa aj na vysvetlenie hodnotenej intervencie a motiváciu tohto výskumu. Pri zverejňovaní zistení RCT by sa mal poskytnúť aj podrobný popis teórie zmeny, pretože podrobný popis intervencie umožňuje prepojiť teóriu zmeny s analýzou zistení. Pri podrobnom popise *metodológie* opíšeme výber vzorky pre RCT, ako aj metódu náhodného pridelovania. V tomto opise je dôležité uviesť aj počet účastníkov (napr. domácností a/alebo jednotlivcov) v testovanej a kontrolnej skupine.
- **Ďalšie analýzy** - správa o RCT musí obsahovať aj údaje o **testovaní úspešnosti randomizácie** (najčastejšie vo forme prehľadnej tabuľky). Odhady vplyvu intervencie možno podporiť pomocou ďalších analýz (napr. s využitím metódy „diff-in-diff“, na analýzu heterogénnych účinkov intervencie pre celú vzorku, ako aj pre podskupiny). Tieto zistenia následne prepojíme s konkrétnou teóriou zmeny. Cieľom je odpovedať na otázku „*Podporujú výsledky analýzy teóriu zmeny?*“ **Ak nie, ktorý predpoklad stojaci za teóriou zmeny nebol naplnený? Ktoré ďalšie zistenia z tejto štúdie by sa mali vysvetliť? Aké možné dôvody, či už v rámci teórie zmeny alebo mimo nej, mohli viesť k výsledkom? Odpovede na tieto otázky pomôže hodnotiteľom identifikovať konkrétne**

zistenia relevantné pre politiku. Uvádzame ich v závere správy a musia byť napojené na analýzu údajov. V závere správy o výsledkoch je tiež dôležité zahrnúť diskusiu o tom, či je možné výsledky extrapolovať na rôzne nastavenia, a **ak áno**, tak je potrebné aj uviesť na ktoré.

5 Realizovanie RCT v prostredí verejnej správy Slovenskej republiky

Implementáciou vyhlášky Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 85/2020 Z. z. o riadení projektov a vyhlášky Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky č. 547/2021 Z. z. o elektronizácii agendy verejnej správy (ďalej len „Vyhláška“) do praxe v poskytovaní elektronických služieb v oblasti verejnej správy zabezpečíme, aby služby boli vytvorené alebo zmenené na základe identifikovaných potrieb a očakávaní koncových používateľov. **Zohľadňujeme tým používateľskú skúsenosť reálnych koncových užívateľov.**

Využitím behaviorálnych inovácií a experimentov však dokážeme rozšíriť možnosti zlepšenia služieb občanom v mnohých iných situáciách. Jedným z cieľov tohto dokumentu je poskytnúť metodický rámec pre implementáciu experimentálneho testovania inovácií a všeobecne tvorby verejnej politiky/služby, pretože meraný a dobre interpretovaný efekt intervencie je základom zlepšovania služieb.

Realizácia experimentu je časovo náročná a vyžaduje odbornú skúsenosť. V podmienkach verejnej správy Slovenskej republiky (ďalej len „SR“) bude preto spravidla takýto typ testovania realizovaný v rámci projektu.

5.1 Realizovanie RCT v rámci projektového riadenia vo verejnej správe SR

V projektoch verejnej správy sa aplikuje metodika projektového riadenia PRINCE2™.

Metóda jasne určuje zodpovednosti za rôzne aspekty projektu vo svojich referenčných metodikách doplnených o smernice, ktoré upravujú špecifiká jednotlivých biznis vlastníkov služieb.

Pod pojmom „projekt“ rozumieme jednorazový proces zameraný na dosiahnutie definovaného cieľa, pozostávajúci zo súboru zosúladených, riadených a časovo ohraničených činností, ktorý:

- a) súvisí so zavádzaním, správou alebo podporou informačných technológií a týka sa tvorby a úpravy informačných technológií verejnej správy,
- b) je pre objednávateľa jedinečný, pričom to nie je pravidelná činnosť,
- c) má presne určený začiatok a koniec trvania projektu a harmonogram realizácie jednotlivých projektových fáz,
- d) má definované konkrétne finančné zdroje, ľudské zdroje, merateľné kvalitatívne a kvantitatívne prínosy,
- e) **vyžaduje analýzu súčasného stavu, špecifikáciu cieľovej skupiny, špecifikáciu cieľového stavu a spôsobu jeho dosiahnutia,**
- f) má definovanú analýzu nákladov a prínosov, ktorej účelom je preukázať, že navrhovaný projekt je spoločensky návratný a v najlepšej možnej miere naplňa stanovené ciele verejnej politiky, a to na základe porovnania hmotných a nehmotných prínosov a nákladov.

Projekt vrátane veľkej zmenovej požiadavky sa realizuje vo fázach, ktoré spolu tvoria životný cyklus projektu:

Prípravná fáza - definuje rámcový projektový zámer, rámcovú funkčnú špecifikáciu, zoznam rizík. Vytvára komplexné predpoklady na iniciáciu projektu.

Iniciačná fáza - vypracováva komplexný projektový produkt – detailizuje výstupy z prípravnej fázy.

Realizačná fáza - navrhuje, vytvára, testuje, dodáva, nasadzuje špecializované produkty v etapách:

- a) Analýza a dizajn
- b) Nákup technických a programových prostriedkov a služieb
- c) Implementácia a testovanie
- d) Nasadenie a postimplementačná podpora

Dokončovacia fáza - administratívne a technicky uzatvára celý projekt.

...

Plánovanie testov realizačných produktov projektu **je v plnej kompetencii projektového manažera objednávateľa**, ktorý ho zabezpečuje v spolupráci s projektovým manažérom dodávateľa vo forme projektového produktu R1-2 Plán testov/TC.

Proces RCT experimentu začína tvorbou dizajnu experimentu. V rámci prípravy dizajnu preskúmavame súčasný stav nastavených cieľov programu alebo politiky využívaním rôznych nástrojov – prieskumy, procesné (používateľské) mapy, teórie zmeny, atď.

V zmysle platnej metodiky projektového riadenia sa teda jeho **využitie naplánuje hneď v prípravnej fáze projektu a presne špecifikuje počas iniciačnej fázy.**

Používateľský prieskum zbiera potreby a očakávania koncových používateľov, analyzuje ich motiváciu, problémy, znalosti a potreby.

Používateľská cesta zobrazuje postup koncového užívateľa pri využívaní programu alebo politiky. Jej mapovaním získame proces jednotlivých krokov a identifikujeme možnosti optimalizácie.

Podľa Vyhlášky je výstup záverov používateľského prieskumu súčasťou projektového zámeru alebo projektového iniciálneho dokumentu. **Výstup záverov používateľského prieskumu je súčasťou detailného návrhu riešenia. Práve tento výstup použijeme ako vstupné dáta pre dizajn RCT experimentu.**

5.2 Realizovanie RCT bez projektového riadenia vo verejnej správe SR

RCT experiment je možné realizovať aj bez projektového riadenia.

V prípade, že **klúčový používateľ služby** (zo strany orgánu verejnej moci, **identifikuje možnosť úpravy užívateľského procesu** a zmena nespadá pod projekt, alebo nevyžaduje projektové riadenie, **postupuje v súlade s aktuálnym platným predpisom.** Do procesu plánovania/prípravy zmeny zaradíme dizajn RCT experimentu. Je potrebné v rámci plánovania rátať najmä s časovou náročnosťou experimentu.

Zodpovednosť za realizáciu experimentu preberá biznis vlastník služby.

Cieľom využitia behaviorálnych inovácií v procese zavádzania/zmien služieb/politik je zabezpečiť, aby služby/politiky a aj weby verejnej správy boli zrozumiteľné a užívateľsky prívetivé. Pomoc a podporu pri implementovaní behaviorálnych inovácií, dizajne zmien najmä v oblasti psychologickú a ekonomickej poskytne tím behaviorálnych inovácií na Ministerstve investícií regionálneho rozvoja a informatizácie SR - [Lepšie služby – Tvoríme lepšie služby. \(lepsiesluzby.digital\).](#)

6 Prípadové štúdie

V prostredí verejnej správy SR bol realizovaný experiment s využitím RCT testovania s názvom **Zvýšenie platobnej disciplíny klientov Sociálnej poisťovne** (nových klientov, klientov s úpravou výšky poistného, klientov opakovane zaradených medzi poistencov).

Prípadová štúdia z tohto experimentu je publikovaná ako samostatný dokument a zobrazuje „*end-to-end*“ proces využitia možnosti merania vplyvu intervencií pri úprave existujúceho procesu prostredníctvom behaviorálnych intervencií.

Niekoľko realizovaných RCT experimentov v prostredí verejnej správy mimo SR popisuje samostatný dokument s názvom „**Metodika realizácie RCT v prostredí verejnej správy Slovenskej Republiky – časť „Prípadové štúdie“**“.

Slovník pojmov

Analýza efektívnosti nákladov (angl. *Cost efficiency analysis*) - Technika, ktorá porovnáva relatívne náklady s výsledkami (účinkami) dvoch alebo viacerých intervencií. Pri porovnávaní dvoch vzájomne sa vylučujúcich alternatív sa odporúča prijať intervencia s najnižším pomerom nákladovej efektívnosti.

Experimentálna intervencia (angl. *Experimental intervention*) - Nová intervencia, ktorá sa skúma, aby sa zistilo, či má vplyv na výsledok.

Chyba merania (angl. *Measurement error*) - Rozdiel medzi tým, čo ukazuje hodnota získaná meraním (napr. výsledok testu študenta) a skutočnou hodnotou (napr. vedomosti a schopnosti študenta), ktorá sa objavuje v dôsledku chýb techniky, ľudských chýb alebo jednoducho kvôli vplyvu náhodných faktorov. Keďže dokonalá presnosť nie je možná, chyba merania je nevyhnutná, je ale dôležité poznať jej rozsah.

Informovaný súhlas (angl. *Informed consent*) - Proces, ktorým subjekt/účastník štúdie dobrovoľne potvrdí svoju ochotu zúčastniť sa konkrétneho experimentu po tom, čo bol informovaný o všetkých aspektoch experimentu, ktoré sú relevantné pre jeho rozhodnutie zúčastniť sa.

Intervencia (angl. *Intervention*) - Program, politika, projekt alebo akýkoľvek iný postup, ktorého cieľom je zmeniť správanie sa účastníkov.

Kontrafaktuálny (angl. *Contrafactual*) - Podmienečné vyhlásenie o tom, čo „by sa stalo“, keby sa stalo niečo, čo sa v skutočnosti nestalo. Kontrafakty majú ústrednú úlohu pri teoretizovaní o príčine zmeny. V RCT sa výkon kontrolnej skupiny používa na odhadnutie toho, čo by sa stalo v intervenčnej skupine, keby nedostali intervenciu.

Kontrolná skupina (angl. *Control group*) - Skupina účastníkov štúdie, ktorá nedostáva skúmanú intervenciu. Namiesto toho môžu dostať štandardnú intervenciu (niekedy nazývanú „obvyklý postup“) alebo fiktívnu intervenciu (napr. placebo; používa sa najmä vo výskume v oblasti medicíny). Výsledky z kontrolnej skupiny sa porovnávajú s výsledkami zo skupiny, ktorá podstúpila skúmanú intervenciu. Účelom je identifikovať rozdiely medzi týmito skupinami a zistiť, či ich možno pripísať intervencii.

Kovariát (angl. *Covariate*) - Premenná iná ako nezávislá premenná alebo prediktor, ktorá koreluje so závislou premennou. Typicky sa výskumníci snažia kontrolovať kovariát pomocou štatistických techník.

Kritéria zaradenia (angl. *Inclusion criteria*) - Kritériá zaradenia sú zoznam charakteristík (napr. faktorov alebo dôvodov), ktoré musia mať účastníci výskumu, aby sa mohli zúčastniť na experimente.

Nezávislá premenná (angl. *Independent variable*) - Predpokladaná príčina v štúdiu. Inými slovami, premenná manipulovaná výskumníkom (t. j. intervencia), ktorý predpovedá, že manipulácia bude mať vplyv na inú premennú, ako je napríklad závislá premenná.

Pilotné testovanie (angl. *Pilot study*) - Typ štúdie, ktorý predchádza realizácii RCT v plnej podobe. Údaje z pilotnej štúdie možno analyzovať, ale nemožno ich použiť vo výsledkoch hlavnej štúdie.

Náhodná kontrolná štúdia (angl. *Randomized control trial*) - Návrh výskumu alebo hodnotenia s dvoma alebo viacerými náhodne vybranými skupinami (testovaná skupina a kontrolná skupina), v ktorých výskumník kontroluje alebo zavádza intervenciu (napríklad nový program alebo politiku) a meria jej vplyv na závislú premennú aspoň dvakrát (merania pred a po teste).

Skreslenie výberu (angl. *Selection bias*) - Systematická chyba, ktorá vzniká, keď výskumník nie je schopný náhodne zaradiť účastníkov do intervenčných a kontrolných skupín. Ak dôjde k skresleniu výberu, skupiny sa budú pravdepodobne líšiť v nameraných alebo nameraných základných charakteristikách kvôli spôsobu, akým boli účastníci vybraní alebo priradení.

Testovaná skupina (angl. *Treatment group*) - Subjekty/účastníci sú vystavení vplyvu nezávislej premennej; nazýva sa aj experimentálna alebo intervenčná skupina.

Štatistická sila (angl. *Statistical power*) - Pravdepodobnosť, že štúdia odhalí, ako štatisticky významný je intervenčný efekt špecifikovanej veľkosti vzhľadom na konkrétnu veľkosť vzorky štúdie. Vopred špecifikovaná veľkosť intervencie sa často volí tak, aby intervencii/experimentu poskytol požadovanú štatistickú významnosť.

Štatistická významnosť (angl. *Statistical significance*) - Pravdepodobnosť, že akékoľvek rozdiely pozorované medzi skúmanými skupinami sú „skutočné“ alebo spôsobené náhodou. Ak je dostatočne nepravdepodobné, že k rozdielu vo výsledkoch došlo len náhodou, rozdiel sa určí ako „štatisticky významný“. Je dôležité poznamenať, že štatistická významnosť nemusí vždy znamenať vecnú alebo praktickú významnosť.

Teória zmeny (angl. *Theory of change*) - Vysvetľuje, ako sa aktivity chápu tak, že prinášajú sériu výsledkov, ktoré prispievajú k dosiahnutiu konečných zamýšľaných dopadov. Môže byť vyvinutá pre akúkoľvek úroveň intervencii – udalosť, projekt, program, politiku, stratégiu alebo organizáciu.

Účastníci experimentu (angl. *Participants*) - Termín používaný na označenie výskumného subjektu alebo respondenta, ktorý bol vybraný na účasť v RCT.

Veľkosť účinku (angl. *Treatment effect*) - Pozorovaná súvislosť medzi intervenciami a výsledkami a sila pozorovanej asociácie. V teste je rozdiel medzi dvoma skupinami (veľkosť účinku) opísaný v jednotkách štandardnej odchýlky (t. j. rozdiel delený štandardnou odchýlkou).

Vnútoraná platnosť (angl. *Internal validity*) - Rozsah, v akom možno výsledky štúdie pripísať intervenciám (zásahom) a nie chybám v dizajne výskumu. Inými slovami, do akej miery ste schopní povedať, že žiadne iné premenné okrem tej, ktorú študujete (t. j. intervencia), nespôsobili výsledok.

Vonkajšia platnosť (angl. *External validity*) - Rozsah, v akom sa zistenia štúdie uplatňujú nad rámec tejto štúdie. Inými slovami, do akej miery sú zistenia zovšeobecniteľné alebo použiteľné na iné okolnosti.

Východiskové údaje (angl. *Baseline*) - Súbor meraní na začiatku štúdie, s ktorými sa často porovnávajú následné výsledky.

Použitá literatúra

Angrist, J., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Banerjee, A., Duflo, E., & Sharima, G. (2020). Long-term Effects of the Targeting the Ultra Poor Program. NBER Working Paper 28074.

Bates, M. A., & Glennerster, R. (2017). The Generalizability Puzzle. *Stanford Social Innovation Review*.

Behagel, L., Crépon, B., & Le Barbanchon, T. (2014). Unintended Effects of Anonymous Resumes. IZA Discussion Paper Series No. 8517.

Behrman, J. (2007). Policy-Oriented Research Impact Assessment (PORIA) Case Study on the International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash Transfer Program. International Food Policy Research Institute.

Bertrand, M., & Mullainathan, S. (2004). Are Emily And Greg More Employable Than Lakisha And Jamal? A Field Experiment On Labor Market Discrimination. *American Economic Review*, 94(4), 991-1013.

Brokešová, Z., Deck, C., & Péliová, J. (2017). Comparing a risky choice in the field and across lab procedures. *Journal of Economic Psychology*, 61, 203-212.

Cunha, J., M., De Giorgi, G., & Jayachandran, S. (2019). The Price Effects of Cash Versus In-Kind Transfers. *The Review of Economic Studies*, 86(1), 240–281.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), 1739-74.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results*. Cambridge University Press.

Edovald, T., & Firpo, T. (2016). *Running randomised controlled trials in innovation, entrepreneurship and growth: An introductory guide*. Innovation Growth Lab, Nesta, London.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

Gneezy, U., & Imas, A. (2017). Lab in the field: Measuring preferences in the wild. In Handbook of economic field experiments (pp. 439-464). North-Holland.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4), 855-902.

John, P. (2017). *Field experiments in political science and public policy: Practical lessons in design and delivery*. Routledge.

J-PAL. (2021a). *Case Study: Cognitive Behavioral Therapy in Chicago Schools: Theory of Change and Measuring Outcomes*. Abdul Latif Jameel Poverty Action Lab. Cambridge, MA.

J-PAL. (2021b). *Case Study: Get Out the Vote: Why Randomize?* Abdul Latif Jameel Poverty Action Lab. 2019. Cambridge, MA.

J-PAL. (2021c). *Case Study: Extra Teacher Program: How to Randomize*. Abdul Latif Jameel Poverty Action Lab. 2019. Cambridge, MA.

J-PAL. (2021d). *Quick guide to power calculations*. Abdul Latif Jameel Poverty Action Lab. 2019. Cambridge, MA.

Kagel, J. H., & Roth, A. E. (Eds.). (2020). *The Handbook of Experimental Economics*. Princeton university press.

Lamanna, F. (2014). *A Model from Mexico for the World*. The World Bank.

Lehmann, E. L. a Romano, J.P. (2022). *Testing Statistical Hypotheses*. Springer Texts in Statistics.

Muralidharan, K., & Niehaus, P. (2017). *Experimentation at Scale*. NBER Working Paper 23957.

Sacarny, A., Yokum, D., Finkelstein, A., & Agrawal, S. (2016). Medicare Letters to Curb Overprescribing of Controlled Substances Had No Detectable Effect on Providers. *Health Affairs*, 35(3).

Sood, N., & Ghosh, A. (2007). The short and long run effects of daylight saving time on fatal automobile crashes. *The BE Journal of Economic Analysis & Policy*, 7(1).

White, H., Sabarwal, S., & de Hoop, T. (2014). *Randomized controlled trials (RCTs)*. Methodological Briefs, Impact Evaluation, (7).

Príloha: Porovnanie RCT a iných výskumných metód na zhodnotenie efektov intervencie

	Metóda	Popis	Aké predpoklady sú potrebné a aké náročné sú predpoklady?	Požadované údaje
Randomizácia	Randomizované náhodné hodnotenie/RCT	Meria rozdiely vo výsledkoch medzi náhodne priradenými účastníkmi programu do testovacej a kontrolnej skupiny.	Výsledná premenná je ovplyvnená iba samotnou účasťou v programe, nie pridelením k účasti v programe alebo účasťou na náhodnom hodnotení. Ako pri iných metódach, veľkosť vzorky musí byť dostatočne veľká, aby boli tieto dve skupiny štatisticky porovnateľné; veľkosť vzorky sa vypočítava vo fáze návrhu výskumu.	Údaje o výsledkoch pre náhodne priradených účastníkov do testovacích a kontrolných skupín (testovacie a kontrolné skupiny).
Základné neexperimentálne porovnávacie metódy	Pred-po (angl. <i>Pre-Post</i>)	Meria rozdiely vo výsledkoch pre účastníkov programu pred programom a po tom, ako program nadobudol účinnosť.	Okrem programu neexistujú žiadne ďalšie faktory (vrátane vonkajších faktorov, snahy o zmenu samotnými účastníkmi, zmenených ekonomických podmienok atď.), ktoré by časom zmenili zistený výsledok pre účastníkov. V stabilnom, statickom prostredí a v krátkom časovom horizonte môže predpoklad platiť, ale nie je možné ho overiť.	Údaje o výsledkoch, týkajúcich sa správania sa účastníkov programu pred začiatkom programu a po nadobudnutí účinnosti programu.
	Jednoduchý rozdiel	Meria rozdiely vo výsledkoch medzi účastníkmi programu po tom, čo program nadobudol účinnosť, a inou skupinou, ktorá sa programu nezúčastnila.	Neexistujú žiadne rozdiely vo výsledkoch účastníkov a nezúčastnených osôb s výnimkou účasti v programe. Obe skupiny mali rovnakú pravdepodobnosť, že vstúpia do programu pred jeho začiatkom. Nezúčastnené osoby nemusia spĺňať kritériá oprávnenosti, môžu bývať na inom mieste alebo jednoducho vidia v programe menšiu hodnotu (vlastný výber). Akékoľvek takéto	Údaje o výsledkoch týkajúce sa správania sa účastníkov programu, ako aj ďalšej skupiny nezúčastnených osôb po

			faktory môžu byť spojené s rozdielmi vo výsledkoch nezávislých od účasti v programe.	nadobudnutí účinnosti programu.
	Rozdiely v rozdieloch (angl. <i>diff-in-diff</i>)	Meria rozdiely vo výsledkoch pre účastníkov programu pred a po programe v porovnaní s nezúčastnenými.	Akékoľvek ďalšie faktory, ktoré mohli ovplyvniť meraný výsledok v priebehu času, sú rovnaké pre účastníkov aj nezúčastnených, takže bez programu ich situácia bola rovnaká. V krátkom časovom horizonte a pri primerane podobných skupinách môže byť tento predpoklad hodnoverný.	Údaje o výsledkoch, týkajúce sa správania sa účastníkov programu, ako aj o inej skupine nezúčastnených osôb pred začiatkom programu a po nadobudnutí účinnosti programu.
Pokročilejšie štatistické neexperimentálne metódy	Viacrozmerná regresia/metóda najmenších štvorcov	Prístup „jednoduchého rozdielu“ môže byť – a v praxi je takmer vždy– realizovaný pomocou multivariačnej regresie. Umožňuje to zohľadniť ďalšie pozorovateľné faktory, ktoré môžu tiež ovplyvniť výsledok, často nazývané „kontrolné premenné“ alebo „kovariáty“. Regresia odfiltruje účinky týchto kovariátov a meria rozdiely vo výsledkoch medzi účastníkmi a nezúčastnenými, pričom účinok kovariantov zostáva konštantný.	Okrem účinkov kontrolných premenných neexistujú žiadne ďalšie rozdiely medzi účastníkmi a nezúčastnenými, ktoré ovplyvňujú meraný výsledok. To znamená, že akékoľvek nepozorovateľné alebo namerané faktory, ktoré ovplyvňujú výsledok, musia byť rovnaké pre účastníkov aj nezúčastnených. Zatiaľ čo pridanie kovariátov môže zmierniť určité obavy pri prijímaní jednoduchých rozdielov, obmedzené dostupné údaje v praxi a nepozorovateľné faktory znamenajú, že metóda má podobné problémy ako jednoduchý rozdiel.	Výstupné údaje o účastníkoch programu, ako aj ďalšiu skupinu nezúčastnených osôb, ako aj „kontrolné premenné“ pre obe skupiny.

	Štatistické párovanie	Presná zhoda: účastníci sú priradení k nezúčastneným, ktorí sú identickí na základe „rovnakých premenných“, aby sa zmerali rozdiely vo výsledkoch.	Podobne ako pri viacrozmernej regresii: medzi účastníkmi a nezúčastnenými s rovnakými zhodnými premennými, ktoré ovplyvňujú meraný výsledok neexistujú žiadne rozdiely. Pri párovaní sú hlavným problémom nepozorovateľné rozdiely. Predpoklady, ktoré je potrebné dodržiavať, aby bolo možné vyvodiť platné závery sú dosť náročné.	Údaje o výsledkoch o správaní sa účastníkov programu, ako aj ďalšiu skupinu nezúčastnených osôb, ako aj „zhodné premenné“ pre obe skupiny.
	Inštrumentálne premenné	Návrh používa „inštrumentálnu premennú“, ktorá je prediktorom účasti v programe. Metóda potom porovnáva jednotlivcov podľa ich predpokladanej účasti a nie podľa skutočnej účasti.	Inštrumentálna premenná nemá priamy vplyv na výslednú premennú. Jej jediným účinkom je účasť jednotlivca v programe. Návrh inštrumentálnej premennej vyžaduje nástroj, ktorý nemá žiadny vzťah s výslednou premennou. Problém je v tom, že väčšina faktorov, ktoré ovplyvňujú účasť na programe pre inak podobných jednotlivcov, tiež nejakým spôsobom priamo súvisí s výslednou premennou.	Údaje o výsledkoch pre účastníkov a osôb nezúčastnených na programe, ako aj „inštrumentálna premenná“.

Zdroj: J-Pal Action Lab (n.d.)