

Oto Potluka

Management dopadových evaluací

Tato publikace vznikla díky podpoře projektu č. P403/12/P350 „Kontrafaktuální hodnocení dopadů strukturálních fondů v podnikové sféře“, financovaného Grantovou agenturou České republiky na Katedře managementu VŠE v Praze.

Autor: Oto Potluka

Technická spolupráce: Michal Kučera, Jana Čamrová, Martin Špaček
Vydal: IREAS centrum, s. r. o., Mařákova 292/9, Praha 6, 160 00
tel. : +420 725 068 902
e-mail: ireas@ireas.cz
<http://www.ireas.cz>

Místo a termín vydání: Praha, 2014
Vydání: první
Počet vydaných kusů: 300
Počet stran: 56
Tisk a vazba: Tiskárna JDS s. r. o., Ke Hřišti 507 , 252 28 Černošice
ISBN: 978-80-905987-0-6

© 2014 IREAS centrum, s.r.o.

OBSAH

1	Úvod.....	7
2	Dopadové evaluace a jejich využití.....	11
2.1	Stávající vývoj v oblasti dopadových evaluací	11
2.2	Zahraníční příklady dopadových evaluací v oblasti zaměstnanosti	13
3	Metody.....	17
3.1	Obecné podmínky metod	17
3.2	Experimentální design evaluace - randomizace	21
3.2.1	Náhodný výběr podpořených subjektů	23
3.2.2	Výzva k účasti náhodným způsobem	23
3.2.3	Postupné přidělování podpory.....	24
3.3	Kvazi experimentální metody	24
3.3.1	Rozdíly v rozdílech	24
3.3.2	Párování podle propensity skóre (Propensity score matching).....	29
3.3.3	Regresní diskontinuita (Regression discontinuity design).....	35
3.3.4	Instrumentální proměnné (instrumental variables).....	40
4	Shrnutí příkladu a postupu zpracování dopadových evaluací	45
4.1	Zkušenosti z realizovaných kontrafaktuálních dopadových evaluací.....	45
4.2	Dostupnost dat jako determinant pro realizaci dopadové evaluace.....	46
5	Závěry.....	49
6	Seznam použité literatury	51

1 ÚVOD

Pro správné řízení programů a zaměření projektů je nutná znalost toho, jak tyto programy fungují a jaký mechanismus změny za jejich intervencemi stojí. Manažeři programů proto využívají dopadových evaluací, které jim poskytují informace o tom, které intervence fungují a také jak fungují. Tato znalost pak manažerům umožňuje provádět korekce programů tak, aby jejich realizace byla efektivní ve vztahu k cílovým skupinám a potřebám, na které příslušný program reaguje svými intervencemi. Následující text je pak shrnutím toho, jak, kdy a proč dělat dopadové evaluace.

Obecně se evaluace liší svým provedením i účelem, podle toho, v jaké fázi programového cyklu jsou prováděny. Různé metody jsou vhodné k realizaci evaluačních studií v určitých fázích implementace programů, ale nehodí se pro jiné. v evaluační praxi se proto setkáváme s různými typy evaluací podle účelu jejich realizace a použití jejich výsledků (Batterbury, 2006) nebo způsobu jejich provedení (Morra-Imas a Rist, 2009). Na tomto místě se budeme zabývat primárně kontrafaktuálními¹ dopadovými evaluacemi veřejných výdajových programů, které jsou realizovány téměř na samém konci programového cyklu, nebo po skončení intervencí. Toto časové určení je však pouze zdánlivé. Datová náročnost dopadových evaluací vyžaduje jejich plánování již od samotného začátku příprav programů. To platí, jak pro programy financované z veřejných prostředků, tak pro programy financované soukromými zdroji (např. nadacemi, či firmami).

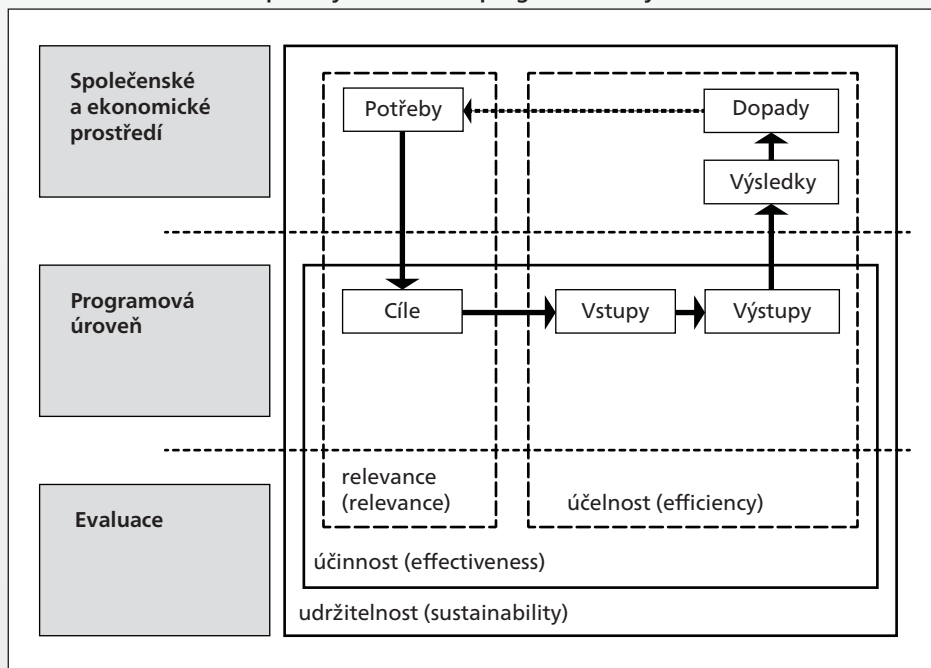
Současná praxe je však také částečně ovlivněna programovým cyklem (Evropská komise, 2004). Programoví manažeři dle tohoto přístupu očekávají, že evaluace probíhají na konci samotného cyklu. Nejprve nastupují fáze programování, identifikace, formulace, implementace, a teprve potom evaluace a audit (viz Evropská komise, 2004, str. 16). v zásadě však příprava dopadových evaluací musí začít již ve fázích pro-

¹ V angličtině je používán pojem counterfactual impact evaluation. Dle MMR-NOK (2013) je v českém jazyce používán pojem kontrafaktuální dopadová evaluace.

gramování a identifikace a musí být významně zastoupena ve fázi formulace, kdy se připravuje design dopadových evaluací.

Tento fakt je významný jak pro manažery programů, tak i pro samotné evaluátory. Pro první z nich je důležité již samotné plánování termínů provedení evaluací a dostupnosti potřebných dat v pravý čas a jejich dostupnost vůbec, ale také plánování vhodných metod. Pro evaluátory jsou významné stejné faktory, aby bylo možné dopadové evaluace realizovat. Schematické zařazení dopadů v programovém cyklu a jejich evaluace je patrné z obrázku č. 1.

Obrázek č. 1: Zařazení dopadových evaluací v programovém cyklu



Zdroj: Upraveno podle Potluka a kol. (2010), s. 91

Význam dopadových evaluací spočívá hlavně v tom, že jejich výsledky odpoví na otázku, zda hodnocený program měl efekt na cílovou skupinu a jak velký tento efekt byl. Ve své podstatě je každá dopadová evaluace postavena na kombinaci dvou otázek: otázce, jaký dopad měla zkoumaná intervence a otázce, na koho měla intervence dopad. Dále uváděné příklady se budou primárně věnovat příkladům dopadů in-

tervencí Evropského sociálního fondu a Evropského fondu pro regionální rozvoj na zaměstnanost v podpořených firmách.

Již několik desetiletí jsou akademické obci známy evaluační postupy využívající kvantitativních metod (například Rosenbaum a Rubin, 1983; Rubin, 1977). Zejména jde o jejich praktické využití v USA či v rozvojové pomoci. Tyto postupy nejsou evropským vědcům a evaluátorům neznámé, ale obecně nebyla v evropském prostoru doposud příliš velká poptávka po podobném typu evaluací ze strany programových manažerů. V několika posledních letech se ale tyto metody začínají uplatňovat více i v evropském kontextu při hodnocení dopadů veřejných výdajových programů.

Další text je členěn následovně. Nejprve jsou diskutovány dopadové evaluace a jejich stávající využití v managementu veřejných výdajových programů. Třetí kapitola je věnována diskusi jednotlivých používaných metod, včetně podmínek, které musí být splněny pro úspěšnou realizaci dopadových evaluací. Tato kapitola také uvádí praktické příklady. Poslední kapitola shrnuje obsah předešlého textu.

2

**DOPADOVÉ EVALUACE
A JEJICH VYUŽITÍ****2.1 Stávající vývoj v oblasti dopadových evaluací**

Stávající praxe v evropských evaluacích je postavena primárně na využití kvalitativních metod výzkumu. Teprve v několika posledních letech se začaly objevovat evaluační studie využívající kvantitativní metody. Obvykle se jedná o evaluace zaměřené na podporu firem či podporu zaměstnatelnosti nezaměstnaných osob. Takovými příklady jsou dopadové evaluace podpory investic do hmotného kapitálu firem v ASVAPP (2012) či hodnocení dopadů kohezní politiky na inovační aktivity ve firmách ve studii Czarnitzki, Bento a Doherr (2011).

Velmi významný impuls pro využití potenciálu rozvoje dopadových evaluací v Evropské unii přinesla iniciativa posledních několika let ze strany Evropské komise. Ta totiž začala prosazovat širší využívání dopadových evaluací (více například prezentace z konferencí Gaffey, 2009; Martini, 2009; posléze také Gaffey, 2011; Mouqué, 2011a či Mouqué, 2011b, propagační charakter má i shrnutí zkušeností v publikaci Mouqué, 2012). Problematikou dopadových evaluací se nezabývá pouze generální ředitelství pro regionální politiku, ale také generální ředitelství pro zaměstnanost (viz například publikace se shrnutím zkušeností z hodnocení podpor Evropského sociálního fondu Morris, Tödtling-Schönhofer a Wiseman (2013). Mimoto je organizována i řada vzdělávacích aktivit. v Evropě jsou dostupné například školy kontrafaktuálních dopadových evaluací Summer school in counterfactual impact evaluation organizovaná společností Prova v Itálii či FBK-IRVAPP Winter School o Fundamentals and Methods for Impact Evaluation of Public Policies v italských Benátkách nebo Causal Inference and Program Evaluation organizované CEMFI v Madridu.

Jak již bylo zmíněno, používání ekonometrie za účelem dopadových evaluací v hospodářské politice nebylo nejen příliš obvyklé, ale řadě aktérů nebylo vůbec známo, přestože byly tyto metody publikovány již před delší dobou. Použití těchto metod se sice objevuje, ale obvykle jde o hodnocení intervencí veřejných politik, ve kterých jsou vědcům nějakým způsobem dostupná data pro analýzy. Paradoxně takové studie vznikly z popudu samotných vědců a jejich zájmu o tyto metody, nikoli za účelem zlepšení implementace veřejných výdajových programů a efektivnosti využití veřejných prostředků. Proto se dopadové evaluace teprve postupně dostávají do povědomí evaluační komunity (Kváča a Potluka, 2011 či Mouqué, 2011a).

V období hospodářské krize vzrostla potřeba efektivního využití veřejných prostředků, kvůli klesajícím příjmům těchto rozpočtů. Potřeba jejich efektivního využití následně vyvolala potřebu změřit efekty vynakládaných programových finančních prostředků. Proto se začaly objevovat zmínky o kontrafaktuálních dopadových evaluacích i na úrovni Evropské komise. Takovou zmínkou je například Pátá zpráva o hospodářské, sociální a územní soudržnosti (Evropská komise, 2010) nebo tzv. Barcova zpráva (Barca, 2009).

Je logické, že se snaha zjišťovat efekty veřejných politik zároveň projevila přímo i v legislativě pro programovací období 2014 – 2020. Konkrétně evaluaci dopadů nařizuje řídicím orgánům nařízení č. 1303/2013 ve svých člancích 54 a 56 o obecných ustanoveních ohledně Evropského fondu pro regionální rozvoj, Evropského sociálního fondu a Fondu soudržnosti na roky 2014 – 2020. Povinnost zpracovat dopadovou evaluaci za každou prioritní osu potvrdil také pracovník generálního ředitelství pro regionální politiku Stryczynski (2014) či je patrná z podmínek přípravy monitorovacích systémů, které pracují s daty na mikroúrovni, a tedy i s možností realizovat kontrafaktuální dopadové evaluace (Evropská komise 2013a). Lze tudíž předpokládat, že kontrafaktuální dopadové evaluace budou i nadále rozvíjeny a používány. Požadavky na dopadové evaluace se ale objevují i jinde. Například Evropská komise (2013b) metody kontrafaktuálních dopadových evaluací zmiňuje i v jiných oblastech.

Na druhou stranu je však nutné varovat před automatickým používáním metod kontrafaktuálních dopadových evaluací bez znalosti intervence a úskalí těchto metod. Existuje totiž řada metodologických omezení, jak zmiňují Kváča a Potluka (2011): *„V kontextu strukturálních fondů těmito metodami nelze hodnotit programy jako celek*

(což ale pro jejich komplexitu často nejde rozumně ani jinými metodami), ale velmi dobře úroveň jednotlivých výzev nebo oblastí podpory.“

Jelikož jsou kontrafaktuální dopadové evaluace u kohezní politiky poměrně nové, existuje o nich zatím jen malé povědomí a hlavně malá reálná schopnost dopadové evaluace skutečně zpracovat. Tato situace je jednak na straně řídicích orgánů odpovědných za management programů financovaných z prostředků Evropské unie, ale také obecně v evaluační komunitě, tedy i na straně zpracovatelů evaluací.

2.2 Zahraniční příklady dopadových evaluací v oblasti zaměstnanosti

V mimoevropském prostoru existuje velká řada dopadových evaluací. Například zkušenosti z oblasti rozvojové spolupráce jsou velmi inspirativní a lze v nich čerpat inspiraci pro design a realizaci dopadových evaluací i v Evropské unii. Repertoár příkladů z realizovaných dopadových evaluací je možné vyhledat například v Impact Evaluation Repository.²

V evropském kontextu je možné se setkat zejména s vědeckými studii v oblasti dopadů veřejných podpor na zaměstnanost či investic do vědy a výzkumu. Nicméně výsledky neposkytují jednoznačný obrázek o tom, zda veřejné intervence pomáhají nebo ne.

Dle dostupných studií v oblasti politik na trhu práce nelze nalézt jednoznačný pohled na to, zda podle dopadových evaluací programy podpor skutečně pomáhají či nikoli. Uvedme například studie aktivní politiky zaměstnanosti Wunsch a Lechner (2008), Lechner, Wunsch a Scioch (2013) a Lechner, Miquel a Wunsch (2011). První ze jmenovaných studií ukazuje poměrně pesimistický obraz dopadů reforem trhu práce ve Spolkové republice Německo po roce 1998, kdy analýza ukazuje dopady různých typů podpor zaměstnatelnosti podpořených osob. Druhý zmiňovaný výzkum dokonce odhaluje negativní dopady na velikost firem (měřeno obratem) a jejich ziskovost. Jde o situaci, kdy jsou pracovníci podpořených firem zapojeni do programů aktivní politiky trhu práce. S ohledem na objektivitu je vhodné doplnit i výsledky třetí zmiňované studie, která nachází dlouhodobé pozitivní dopady na

² Dostupné na http://www.campbellcollaboration.org/IDCG_updates/3ie_launches_Impact_Evaluation_Repository2.php

zaměstnanost podpořených osob. Pozitivní dopady na zaměstnanost ukazuje i studie Criscuolo et al. (2009) v případě Velké Británie a programu Regional Selective Assistance.

Abramovsky et al. (2011) neprokázali vliv vzdělávání pracovníků s nízkou kvalifikací na jejich zaměstnatelnost. Studie Hamersma (2008) sice prokázala krátkodobé pozitivní efekty na zaměstnanost v případě podpor firem, tyto pozitivní efekty však studie v dlouhodobém horizontu nepotvrdila.

Obecným problémem je zkreslení dat v momentě, kdy firmy ohlašují změny zaměstnanosti díky realizaci podpořených projektů. Podpořené firmy chtějí ukázat dobré výsledky, a tudíž nadsazují ohlašované stavy zaměstnanosti, či jako výsledek podpory oznamují i ta pracovní místa, která by vznikla i bez podpory. Proto například Betcherman, Daysal a Pagés (2009) upozorňují na tento fakt a nutnost určité rezervovanosti při analýzách dat, která poskytují samotné podpořené firmy evaluátorům či manažerům programů. Proto je vhodnější použít data, jejichž sběr přímo nesouvisí s realizovanou evaluací. Taková data pak nejsou zkreslena pro účely dopadové evaluace.

Nezávisle získaná data pak mohou být klíčem k určení skutečného efektu podpor. S tím je spojen i další klíčový fakt, a sice dlouhodobá udržitelnost takto vytvořených pracovních míst. Girma, Görg, Strobl a Walsh (2008) ukazují na irském příkladu, že místa vzniklá díky podpoře v průměru existují čtyři roky po ukončení podporovaného projektu. Poté nejsou firmy dále motivovány tato místa udržet, protože už jim nehrozí žádná případná sankce spojená se smlouvou o podpoře. To má velký význam pro přípravu dopadových evaluací, které by se tedy neměly zaměřovat pouze na situaci během nebo těsně po ukončení realizace projektů, ale i na dobu po uplynutí tzv. udržitelnosti.

Byť je zaměstnanost ostře sledovaným ukazatelem, existují i studie, které se vzdělávání věnují i z pohledu jiných hospodářských ukazatelů. Například se této tematice nepřímou věnují studie Holzer et al. (1993), Lynch a Black (1995) a Barrett a O'Connell (1999). V těchto studiích sice nejde přímo o zaměstnanost, ale o produktivitu práce. Holzer et al. (1993) ve své studii vyzkoumali, že vzdělávání má pozitivní vliv na produktivitu práce. Ve studiích se pak hledá způsob, jakým působí vzdělávání na produktivitu (například, jak působí obecné či specifické vzdělávání zaměstnanců).

V každém případě je však na místě uvážlivé zhodnocení situace. Kluve a Schmidt (2002) v souvislosti s heterogenitou trhů práce upozorňují na to, že není možné automaticky přejímat závěry dopadových evaluací z jedné země do druhé, ale je nutné brát v úvahu národní specifika jednotlivých ekonomik.

3

METODY

3.1 Obecné podmínky metod

V této kapitole jsou představeny metody, které jsou používány pro kontrafaktuální dopadové evaluace. Základ představovaných metod tvoří ekonometrické metody, při jejichž použití a splnění předem daných metodických podmínek dostaneme relativně přesné odhady dopadů zkoumaných intervencí. Pro tyto metody je však třeba použít kvalitní data v podobě poměrně velkého množství zkoumaných případů a také jasně definovat evaluační otázku (dopady čeho na co zkoumáme).

Kontrafaktuální dopadové evaluace jsou postaveny na principu srovnání pozorované skutečnosti (v angličtině používané pojmy *observable - factual*) po poskytnutí podpory či intervence a situací stejných subjektů, která by hypoteticky nastala bez realizace této podpory (*non-observable – counterfactual*). Rozdíl mezi těmito situacemi pak vyjadřuje dopad zkoumaného programu či politiky. Tímto způsobem se tedy měří efekt, který lze přičíst intervenci. Tím, že není možné sledovat zkoumané subjekty jak s podporou, tak bez ní, je nutné nalézt jiný způsob výpočtu dopadu. V tomto případě se musí využít zákona velkých čísel. S rostoucím počtem měření se zvyšuje i pravděpodobnost, že vypočtené průměry jsou průměry nezskreslenými, tedy skutečnými. Pak je použita kontrolní skupina (u experimentálních metod), nebo srovnávací skupina (u kvazi-experimentálních metod), která zastupuje situaci podpořených subjektů bez podpory.

Tím vstoupil do hry první významný předpoklad těchto metod – nutnost mít dostatečně velký vzorek podpořených i nepodpořených subjektů. Zároveň s tím ale vstupuje do hry i druhý požadavek – v charakteristikách musí být obě srovnávané skupiny statisticky podobné, aby bylo možné použít nepodpořené subjekty pro modelování situace podpořených subjektů, ale bez podpory.

Při zpracování dopadových evaluací se musí dbát na to, že odhadovaný dopad nelze spočítat rovnou – není totiž pozorovatelný – je totiž součtem tří různých efektů (viz Morgan a Winship (2007, kap. 2). Prosté srovnání průměrů zahrnuje:

1. Skutečný dopad podpory, o jehož výpočet máme primárně zájem;
2. Vychýlení odhadu dopadu kvůli tomu, že podpořené subjekty by dosáhly jiného výsledku než nepodpořené subjekty ve srovnávací skupině, i kdyby podporu nedostaly;
3. Vychýlení kvůli tomu, že nepodpořené subjekty v kontrolní skupině by měly odlišný výsledek než podpořené subjekty, i v případě obdržení podpory.

Tato situace je způsobena tím, že výsledné odhady jsou ovlivňovány pozorovatelnými i nepozorovatelnými charakteristikami zkoumaných subjektů. Výsledek je pak ovlivněn tzv. výběrovým vychýlením (*selection bias*), kdy některé subjekty o podporu požádají, i když ji nepotřebují a jiné subjekty o ni nežádají nikdy, přestože by jim potenciálně pomohla. To je dáno tím, že rozhodnutí o účasti v programu je dáno samotnými žadateli.

U pozorovatelných charakteristik jsme schopni statisticky porovnat, zda jsou zkoumaná a srovnávací skupina statisticky shodné, a tudíž ověřit, že je možné dopady počítat (kdyby se skupiny lišily, nebyla by splněna jedna ze základních podmínek – tj. že obě skupiny jsou shodné ve své výchozí situaci před startem programu). U nepozorovatelných charakteristik toto udělat nemůžeme.

V takovém případě lze situaci řešit náhodným přidělením podpory. Náhodné přidělení totiž rozdělí výše zmíněné skupiny subjektů, které vychylují výsledky z řad (ne) příjemců podpory rovnoměrně do obou skupin a tudíž nedojde k vychýlení odhadu dopadu. Náhodné rozdělení totiž zařadí všechny charakteristiky subjektů do obou srovnávaných skupin rovnoměrně. Z těchto důvodů se věnujeme v další části jak experimentálním přístupům k designu evaluací, tak i tzv. kvazi-experimentálním přístupům.

Výše uvedený postup také podtrhuje jednu základní charakteristiku jakékoliv dopadové evaluace, což je fakt, že jde vždy jen o odhady skutečného dopadu. Vhodným použitím metod podle realizované intervence se můžeme s větší či menší přesností přiblížit ke skutečnému efektu intervence, vždy ale půjde jen o odhad. V dalším textu jsou uvedeny podmínky, které musí být splněny, aby byl dopad odhadován co nejpřesněji.

Statistické metody umožňují překonat bariéru v podobě nemožnosti získat data o situaci po podpoře a současně o situaci téhož subjektu bez podpory. Při použití těchto metod se nahrazuje nezměřitelná situace, jak by vypadaly podpořené subjekty bez podpory, změřitelnou situací nepodpořených subjektů. Je k tomu však zapotřebí mít dostatečně velký vzorek, kdy nepozorovaná skutečnost je nahrazena pozorovanou skutečností jiné (velmi podobné) skupiny subjektů.

Pro aplikaci popisovaných metod je nutné splnit následující podmínky:

- **Statistická reprezentativnost.** Počet zahrnutých případů musí splňovat základní podmínky statistické reprezentativnosti dané velikostí celkového výskytu případů (např. všech podpořených firem, všech firem v ekonomice apod.).
- **Pro zkoumanou i srovnávací skupinu musí být k dispozici velký počet měření.** Tato podmínka je postavena na zákonu velkých čísel, kdy s rostoucím počtem zkoumaných případů můžeme říci i to, zda se zjišťované průměrné charakteristiky blíží skutečným průměrným charakteristikám v populaci (firmy, regiony, obce, jednotlivci, apod.). Splnění této podmínky zajistí, že velký počet měření poskytne statistickou významnost odhadů, a tím přispěje k důvěryhodnosti výsledků (minimální množství zkoumaných případů musí splňovat výše zmíněnou statistickou reprezentativnost). Tato podmínka platí pro obě skupiny případů, se kterými se pracuje, tj. jak pro zkoumanou skupinu, tak i srovnávací skupinu. Pokud tato podmínka není splněna, může dojít k vychýlení výsledků.
- **Homogenita případů.** Pro použití metod kontrafaktuálních dopadových evaluací je potřeba mít k dispozici pouze jeden typ poskytované podpory, který je zkoumán. Komplexní podpory takto nelze vyhodnocovat. Jsou-li jedny subjekty v rámci stejného programu podpořeny například investiční podporou a jiné formou vzdělávání, není tento program možné vyhodnocovat jako celek. Mechanismus změny a vzniku dopadů se u takových případů liší. Z toho plyne, že těmito metodami není možné odhadovat dopady heterogenních programů či nejasně definovaných intervencí. Tuto podmínku lze splnit například tím, že se pro jeden komplexnější program uskuteční více dílčích dopadových evaluací, kdy je pro každou z nich tato podmínka splněna (a splněny jsou ovšem i ostatní zmiňované podmínky). Ideálním případem je situace, kdy jsou obě skupiny – podpořená i nepodpořená – statisticky identické a liší se pouze v tom, že jedna skupina byla podpořena a druhá ne.
- **Dostupná kvalitní data.** Dostupnost a kvalita dat jsou základní a nutné podmínky realizace dopadových evaluací. Pro splnění těchto podmínek je nutné, aby dato-

vé soubory splňovaly několik charakteristik (podrobně se této problematice věnují Potluka a Brůha, 2013):

- Pro zpracování kvalitní dopadové evaluace je nutné mít k dispozici data, která nejsou zkreslena například tím, že subjekty vědí, že účelem dat je ověření dopadů podpor. V takovém případě může snadno nastat situace, že jsou výsledky úmyslně nadsazovány, jak již bylo zmíněno výše na základě zahraničních zkušeností.
 - Dalším aspektem dat je tzv. non-response bias, tj. vychýlení výsledků tím, že data nereprezentují charakteristiky celé populace, ale jen její určité části (například v datech z šetření mezi těhotnými ženami se objevilo silné zkreslení, protože na výzvu k poskytnutí dat reagovaly ve velké většině pouze vysokoškolsky vzdělané ženy, proto byla ve výzkumu Potluková et al. (2012) použita jiná než původně plánovaná data).
 - Dostupnost dat je klíčová, nicméně v současné praxi budou evaluátoři narážet na právní bariéry v podobě ochrany osobních údajů a neochoty úřadů poskytnout vhodná data (viz například zkušenosti Itálie popsané v ASVAPP (2012), Litvy dle Zvalionyte (2011) i České republiky dle Potluka a Brůha (2013). Paradoxně existuje velké množství panelových dat pro kontrafaktuální dopadové evaluace, ke kterým se však evaluační týmy nejsou schopny dostat kvůli různým právním a technickým omezením.
 - Kvalita dat může být také poměrně rizikovým faktorem. Při sběru dat musí být splněny podmínky použití jednotné metodiky sběru. V opačném případě mohou být výsledné odhady zkresleny tím, že pro některé roky byla použita jiná metodika sběru dat či definice proměnných. Pokud o této skutečnosti evaluační tým neví, velice snadno udělá zásadní interpretační chybu.
- **Proporcionalita intervence.** Další z podmínek je existence změřitelného dopadu zkoumané intervence. Změna způsobená realizovaným programem či intervencí musí být odlišitelná od dalších obecných vlivů (je-li podpora marginální, vystavujeme se riziku, že evaluace nepovede k žádným výsledkům).

Dále v textu jsou u jednotlivých metod uváděny i další metodologické poznámky a předpoklady, pokud takové existují.

Terminologicky je nutné také uvést, že v případech, kdy jsou skupiny podpořených a nepodpořených subjektů vytvořeny pomocí experimentu, mluvíme o skupině nepodpořených subjektů jako o kontrolní skupině. Pokud však tato skupina nevznikla prostřednictvím experimentu, mluví se o srovnávací skupině.

3.2 Experimentální design evaluace - randomizace

Obvyklým zlatým standardem v přírodních nebo lékařských vědách je použití randomizovaných přístupů (*Randomized Control Trials*). Příkladem je situace, kdy je zkoumaný lék podáván náhodně vybrané skupině pacientů. v takovém případě se kromě podávaného testovaného léku jedné skupině podává jiné skupině i lék bez účinné látky (placebo), a tím se odliší efekt reakce lidí na to, že je jim podáván lék, byť neobsahuje žádné účinné léčivo. Tím dojde k očištění výsledků o tzv. placebo efekt. V humanitních oborech není tento přístup tak častý. V České republice ani v Evropě není randomizovaný výběr podpořených subjektů ve veřejných výdajových programech používán. Schází k tomu dostatečná politická podpora.

Tento postup v sobě skrývá řadu výhod, například není nutné platit hodnotitele žádostí o podporu, losování zvyšuje transparentnost (pokud se losuje větší množství „výherců“ nehrozí situace „karlovarské losovačky“, tj. „náhodného“ výběru jednoho předem vybraného vítěze), zároveň je výhodou i možnost zpracování kvalitní dopadové evaluace a zjištění, zda program má či nemá efekty na cílovou skupinu.

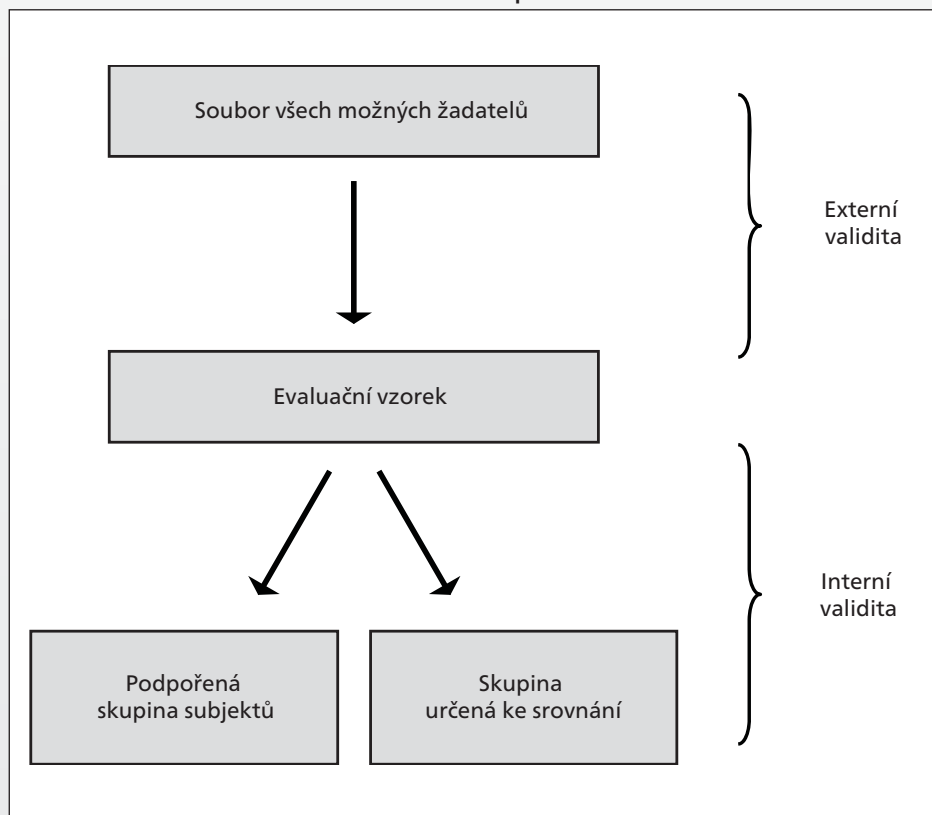
Jak uvádí Potluka a Špaček (2013, s. 6): *„Vzhledem k současné české praxi procesu hodnocení předložených projektových žádostí zřejmě nelze předpokládat brzké přistoupení k čistým randomizačním postupům. Nicméně si lze představit situaci, kdy by předložené projektové žádosti prošly hodnotícím procesem a byly jim přiděleny body proto, aby byla stanovena hranice základní kvality, která by vyřadila všechny opravdu špatné projekty. Následně by se dala využít randomizace mezi projekty, které úspěšně prošly hodnocením, tj. z množiny všech uznatelných projektů jich náhodným výběrem vybrat tolik, na kolik stačí finanční alokace dotačního programu, bez ohledu na konkrétní počet bodů přidělený hodnotiteli. Nedošlo by tak k velké úspoře za práci hodnotitelů, ale byl by eliminován problém nevyváženosti posudků jednotlivých hodnotitelů a vznikla by srovnávací skupina využitelná pro evaluaci.“*

Světová praxe však s randomizací stále častěji pracuje (příklady takových evaluací uvádějí například Khandker, Koolwal a Samad 2010 či Gertler a kol. 2011). Jak již bylo zmíněno, poskytují tyto postupy výhodu transparentního výběru tehdy, jsou-li finanční zdroje programu omezené (finanční zdroje jsou omezené vždy). Navíc se tím řeší problém výběrového zkreslení.

Další z aspektů, na který je nutné dát si pozor, je interní a externí validita výsledků. Interní validita se týká pouze vybraného vzorku, na kterém je realizován odhad dopadu. Randomizace zabezpečí validitu tím, že statistické charakteristiky jsou stejné pro obě skupiny, pro tu s podporou i tu bez ní. Externí validita se oproti interní vztahuje na celou populaci. Pokud je vypočtený odhad dopadu externě validní, je možné ho generalizovat na celou populaci, nejen na testovaný vzorek (interní validita). V tom tkví výhoda randomizovaného výběru podpořených subjektů – náhodný výběr zabezpečí shodné statistické charakteristiky obou skupin a tím i externí validitu výsledků.

Schematicky vypadá experimentální design, jak je uvedeno na obrázku č. 2.

Obrázek č. 2: Interní a externí validita odhadů dopadů



Zdroj: Upraveno podle Gertler a kol. (2011)

Randomizaci lze provést několika různými způsoby. Některé z nich nejsou tak „nápadné“, a tudíž by bylo možné říci, že jsou politicky průchodnější.

3.2.1 Náhodný výběr podpořených subjektů

Při omezených zdrojích programu, kdy počet žádostí převyšuje možnosti samotného programu, je nutné některé z žadatelů odmítnout. Transparentním procesem je pak například los, který určí žadatele, kteří podporu dostanou (nebo budou naopak odmítnuti).

Tato metoda umožní realizátorovi evaluace dosáhnout situace, kdy je eliminován vliv vnějších faktorů, protože vnější vlivy jsou náhodně, ale rovnoměrně zastoupeny v obou skupinách.

3.2.2 Výzva k účasti náhodným způsobem

Druhým způsobem, jak realizovat experiment, je výzva k účasti náhodným výběrem. Výzvu k účasti náhodným způsobem lze provést tak, že někteří z potenciálních žadatelů dostanou informaci o existenci a možnostech podpor v programu. Toto náhodně zvolené informování potenciálních žadatelů umožňuje odlišit od sebe potenciální žadatele, kteří se programu zúčastní pouze tehdy, jsou-li osloveni, od dalších skupin osob, ať už oslovených nebo ne. Mezi další skupiny patří potenciální žadatelé, kteří se za žádných okolností programu nezúčastní (tedy ti, kteří se nezúčastní, i když dostanou informaci – *never takers*) a těch potenciálních žadatelů, kteří se zúčastní za všech okolností (tj. ti, kdo se do programu přihlásí, i když o něm nebyli přímo informováni – *always takers*). Tato skutečnost je pak využívána u některých metod k identifikaci dopadů a očištění o výběrové zkreslení (*selection bias*). Metoda, které tento design využívá, je dvoustupňová metoda nejmenších čtverců (2SLS – *two stage least squares*, nebo také známá pod názvem metoda instrumentálních proměnných).

Způsobů, jak takové náhodné oslovení realizovat, je opět celá řada. Oslovení lze například udělat rozesláním zvacích dopisů osobám, jejichž rodné číslo končí na sudá či lichá čísla (výběr však musí být náhodný). Tím se neeliminuje možnost přihlášení všech potenciálně možných žadatelů, kteří se o programu dozvědí, tedy ani těch, kteří zvací dopis neobdrželi. Metoda je vůči potenciálním žadatelům spravedlivá v tom, že na začátku nezvýhodňuje nikoho, všichni mají stejnou pravděpodobnost, že budou osloveni.

3.2.3 Postupné přidělování podpory

I v případě, kdy je „poptávka“ po podpoře z programu nižší než možnosti samotného programu, a tudíž se dostane na všechny žadatele o podporu, je možné přistoupit k randomizaci. V takovém případě však není uvolněna podpora všem podpořeným subjektům najednou, nýbrž postupně.

Tím vznikne mezi podpořenými subjekty skupina, která již reálně podporu obdržela a skupina, které je má sice přislíbena, ale teprve ji obdrží. Vznikne tak skupina podpořených subjektů a k ní skupina kontrolní. Podmínkou ovšem je, že výběr subjektů v první fázi podpory proběhne náhodně.

Určitou nevýhodou tohoto způsobu je, že je nutné pečlivě zvážit, jak dlouhé má být zpoždění poskytnutí podpory pro kontrolní skupinu. Příliš dlouhé zdržení může vést u této druhé skupiny podpořených osob k deziluzi a neochotě spolupracovat při evaluaci. Situaci výrazně ovlivňuje i to, jaká data jsou k dispozici. Jinak bude vypadat zpoždění, máme-li měsíční, kvartální či jen roční data. Zpoždění v řádu měsíců či kvartálů je snáze odůvodnitelné a přijatelné pro cílovou skupinu. U zpoždění v řádu let může část cílové skupiny rezignovat na účast.

3.3 Kvazi experimentální metody

Není-li možné z jakýchkoliv důvodů realizovat experiment, lze realizovat dopadovou evaluaci pomocí kvazi experimentálních metod. Tyto metody sice nenahrazují sílu experimentu, ale částečně se mu přibližují. Vystává u nich ale řada omezení a předpokladů, bez jejichž naplnění poskytnou tyto metody zkreslené výsledky.

V následujících řádcích budou uvedeny základní z nich – rozdíly v rozdílech, párování pomocí propensity skóre, regresní nespojitost a instrumentální proměnné.

3.3.1 Rozdíly v rozdílech

Metoda rozdílů v rozdílech je založena na principu srovnání vývoje situace mezi dvěma skupinami minimálně ve dvou obdobích. Srovnání výsledků mezi dvěma skupinami, kdy jedna skupina subjektů podporu obdrží a druhá ne, ukazuje sice rozdíl v hodnotách zkoumaných proměnných mezi těmito skupinami, ale není jasné, do jaké míry je tento rozdíl způsoben samotnou intervencí. Proto je do analýzy zapojen ještě

časový rozměr. Tento druhý rozdíl (rozdíl v čase) pak ukazuje, „jak by se pravděpodobně“ vyvíjela skupina podpořených subjektů, kdyby byla vystavena stejným podmínkám jako nepodpořené subjekty. Rozdíl těchto rozdílů (rozdíl v čase a rozdíl mezi subjekty) pak ukazuje dopad, který intervence měla na podpořené subjekty.

Základním předpokladem této metody je, že vývoj obou skupin by byl bez intervence shodný, tudíž křivka vývoje by byla paralelní. S rostoucím časovým odstupem mezi předintervenčním a postintervenčním obdobím bude nutné tuto podmínku více ověřovat. Pokud by neplatil předpokládaný paralelní vývoj způsobený ostatními faktory, docházelo by ke zkreslení odhadovaných dopadů. Podle situace by docházelo k jejich podhodnocování či naopak nadhodnocování.

Preciznější odhady dopadu vycházejí v případech, kdy je k dispozici delší časová řada dat pro zkoumanou proměnnou, jak pro podpořené, tak i nepodpořené subjekty. Samozřejmě je možné udělat analýzu jen na jednom měření před intervencí a jedním měření po intervenci. Nicméně se v tom skrývá určité riziko, že tato dvě období mohou být něčím výjimečná a vychylovat se od normálního vývoje. Delší časová řada tak může potvrdit, že dlouhodobě se obě skupiny skutečně vyvíjejí se stejnou dynamikou. Metoda rozdílů v rozdílech pak ověřuje, zda poskytnutá podpora změnila dynamiku u podpořené skupiny.

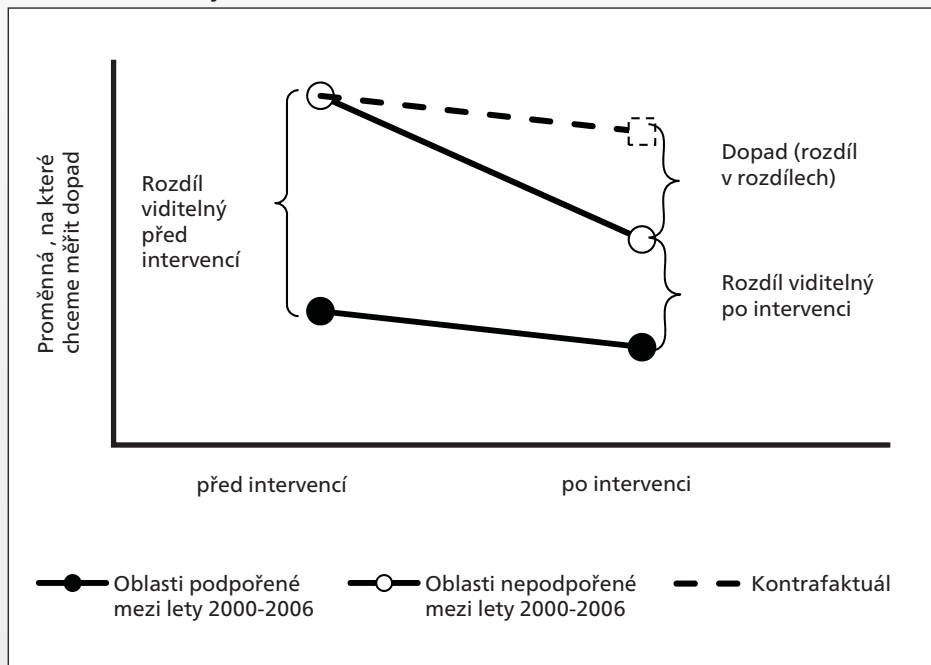
Nejjednodušší formu aplikace této metody znázorňuje obrázek č. 3, na kterém je zachycena situace dvou skupin subjektů a dvou období. Čárkovaná linie představuje onu podmínku paralelního vývoje v obou skupinách. Pozorovat ji ovšem můžeme pouze u nepodpořených subjektů.

Evropská komise (2009) uvádí existenci dvou přístupů k vysvětlení odhadu dopadu na podpořenou skupinu subjektů.

1) Prvním z přístupů je změření rozdílů mezi srovnávanými skupinami po skončení poskytnuté podpory. To je však teprve první krok. Ještě nejde o dopad, protože výsledný rozdíl nemusí být ovlivněn pouze intervencí. Druhým krokem je porovnání rozdílů mezi oběma skupinami ještě před zahájením intervence. Rozdíl mezi oběma výsledky nám pak poskytne přesnější odhad dopadu, protože se nám podaří eliminovat alespoň část výběrového zkreslení. Za předpokladu, že se nepozorované rozdíly mezi oběma srovnávanými skupinami nemění (to je zásadní podmínka této metody), získáme poměrně věrohodný obrázek o dopadech podpory.

- 2) Jiný způsob získání odhadu dopadu je, že nejprve porovnáme situaci podpořených subjektů před intervencí a po ní. Tím získáme informaci o tom, jak dynamicky se změnila jejich situace. Totéž pak učiníme i pro skupinu nepodpořených subjektů. Opět tedy odečteme od situace po intervenci hodnotu zkoumané proměnné před intervencí. Dalším krokem je pak odečtení těchto dvou výsledků od sebe. Získáme tím odhadovaný dopad intervence na zkoumanou proměnnou.

Obrázek č. 3: Rozdíly v rozdílech



Zdroj: Evropská komise (2009), vlastní úpravy

Oběma způsoby se dostaneme ke stejnému výsledku. Jediný rozdíl mezi nimi spočívá hlavně v tom, že jinak přistupují k datům. První postup poukazuje na výběrové zkreslení a rozdílné charakteristiky obou zkoumaných skupin. Druhý způsob pak poukazuje na zkreslení výsledků prostřednictvím přirozené dynamiky zkoumané proměnné.

Základní podmínkou této metody je dostupnost dat alespoň ve dvou obdobích a pro dvě skupiny subjektů, a to nezávisle na realizované zkoumané intervenci. To se ně-

kdy opomíjí při přípravě programů. Obvykle jsou pak k dispozici pouze data pro před a post-intervenční období pro podpořené subjekty, případně data před intervencí pro obě skupiny (získaná například z projektových žádostí všech žadatelů a posléze pak ze zpráv od podpořených subjektů). Obvykle se však opomíjí to, že získat data od nepodpořených subjektů bývá obtížný úkol. Tyto subjekty jsou frustrovány tím, že neuspěly se svou žádostí o podporu, tudíž nespátřují žádný prospěch v poskytnutí dat (spíše naopak, je pro ně satisfakcí data neposkytnout). Retrospektivní dotazování a sběr dat je sice určitým řešením, které ale skýtá velkou řadu nepřesností a úskalí, například v tom, že si respondenti budou vědomi, že data se sbírají za konkrétním účelem dopadové evaluace (více o retrospektivním získávání dat viz Ravallion, 2012). Metoda je také obtížně použitelná pro intervence, které mají kontinuální charakter, kdy je podpora poskytována opakovaně.

Při realizaci dopadové evaluace pomocí této metody uvádí literatura (Evropská komise, 2009) pořadí doporučených kroků:

- 1) **Definování proměnné pro měření dopadu.** Analýzu lze zpracovat pro více proměnných. Je vždy ale nutné zvážit, má-li taková dopadová evaluace smysl a co nám její výsledky řeknou. Zároveň však musí být dostupná data.
- 2) **Definování časového rozsahu.** Jak již bylo zmíněno, tato metoda se hodí pro diskrétní časové ohraničení podpory. Proto je design tohoto typu dopadové evaluace volen s ohledem na délku realizace podpor a také dostupnost dat v určitých časových obdobích. Větší množství dat v širším časovém pásmu sice ukáže trendy v datech, ale zároveň je také těžší taková data získat.
- 3) **Samotná realizace výpočtů.** Minimálně pro dvě období a dvě skupiny se spočtou rozdíly, jak je již popsáno výše v této kapitole. Srovnáním rozdílů pak dojdeme k odhadu dopadu.
- 4) **Předpoklad paralelismu.** Jde o zásadní předpoklad metody. Kdyby taková situace nenastala, pak budou odhady dopadu vychýleny. Právě získání většího rozsahu dat před intervencí umožňuje přesněji odhadnout, jaký je dlouhodobý vývoj v obou skupinách a zda skutečně lze předpokládat paralelní vývoj. Bez tohoto předpokladu nemá metoda rozdílů v rozdílech smysl.
- 5) **Použití regrese k replikaci výsledků metody.** Regresní model má oproti výše popsanému intuitivnímu přístupu výhodu, že k nezávislým proměnným na pravou stranu rovnice lze přidávat další proměnné, které máme k dispozici a které mohou ovlivňovat výsledek. Tím se zmenšuje síla předpokladu paralelního vývoje (začleněné proměnné postihnou případný další vliv).

6) **Zahrnutí kovariancí do regrese.** Do regresního modelu je také možné přidávat proměnné, které mohou mít vliv na směrnici výsledné odhadnuté křivky v regresi. Jde o součiny proměnných (angl. *interaction terms*). Přidání těchto proměnných může vést ke zpřesnění odhadů dopadů intervence.

Hlavní výhoda metody rozdílů v rozdílech spočívá v tom, že je relativně intuitivní a snadno použitelná. Není k ní zapotřebí velkých datových souborů zahrnujících komplexní charakteristiky subjektů a stačí jen čtyři typy měření.

Nevýhodou může být naopak situace, kdy metoda nebyla plánována při přípravě programu, a nejsou k dispozici data z období před intervencí. Nevýhodou je také nutný předpoklad paralelního vývoje obou srovnávaných skupin.

Zkušenosti dle Potluka a kol. (2013a) ukazují na kombinaci této metody s metodou párování pomocí propensity skóre. Kombinace obou metod byla zvolena kvůli tomu, že se tím řeší heterogenita vzorku. Tato studie zkoumala vliv podpor Evropského sociálního fondu na rozvoj podniků. Do studie byly zahrnuty proměnné jako zaměstnanost, tržby a hospodářský výsledek, na kterých byl testován dopad. V tabulce č. 1 je uveden souhrn výsledků pro různé velikosti firem. Z výsledků je patrné, že pouze pro velké firmy byl odhad dopadu v rámci dostačující statistické signifikance. Pro celý zkoumaný vzorek sice vyšel pozitivní dopad v průměrné výši čtyř pracovních míst v každé podpořené firmě, ale není dostatečně statisticky signifikantní.

Z uvedeného příkladu je zřejmé, že je vhodné použít kombinaci obou přístupů, jednak intuitivní grafické provedení, ale také regresní model. Z výsledků v tabulce č. 1 je patrné, že bez provedení statistického testu by mohl evaluátor s manažerem programu dojít k závěru, že dopady jsou pozitivní pro všechny firmy, ačkoli to metoda rozdílů v rozdílech nepotvrzuje.

Studie Potluka a kol. (2013a) pracovala se třemi skupinami firem: a) firmy, které žádaly o podporu a skutečně ji dostaly; b) firmy, které o podporu zažádaly, ale nedostaly ji a konečně c) firmy, které o podporu vůbec nežádaly. Nejvhodnější pro odhad dopadu je srovnání podpořených žadatelů s odmítnutými žadateli. V obou případech totiž žadatelé projeví stejnou ochotu získat podporu a připravili žádost.

To vyžadovalo mít dostatečný počet firem v obou skupinách. Ve výzvách 35, 39 a 60 Operačního programu Lidské zdroje a zaměstnanost tak autoři této studie pracovali

s 1261 zamítnutou žádostí a 1350 podpořenými žádostmi. To jsou dostatečné počty, aby byly splněny základní podmínky použité metody.

Tabulka č 1: Výsledky odhadu dopadu na zaměstnanost ve firmách dle DiD pouze pro výzvu 35 (srovnání podpořených a odmítnutých žadatelů)

Velikost firem	Bodový odhad	p-value	Na jedno místo (tis. Kč)
Všechny firmy	4,0	0,27	318,01
Mikrofirmy	1,4	0,35	189,56
Malé firmy	1,8	0,41	671,87
Střední firmy	27,5	0,27	31,59
Velké firmy	25,5	0,08	63,36

Zdroj: Potluka a kol. (2013a)

3.3.2 Párování podle propensity skóre (Propensity score matching)

Metoda párování (matchingu) je založena na principu, že srovnávací skupina je seřazena tak, aby byla se zkoumanou skupinou shodná, s tím rozdílem, že neobdržela podporu ze zkoumaného programu. V obou skupinách jsou tak zastoupeny subjekty se shodnými charakteristikami, nebo alespoň s co nejpodobnějšími charakteristikami.

Párování lze v jednodušších případech s méně sledovanými charakteristikami udělat tak, že se nejprve celková skupina rozdělí podle jedné charakteristiky a poté podle další, dokud jsou k dispozici charakteristiky, podle kterých chceme dělit celkovou skupinu. Například u firem vznikne podle kraje sídla firmy pro Českou republiku hned 14 podskupin. Když se pak dále rozdělí podle velikosti (malé, střední a velké firmy), získáme už 42 podskupin. Z toho je patrné, že dalším dělením už získáváme příliš velké množství dílčích podskupin. Provést srovnání pro každou takovou podskupinu podpořených subjektů s podskupinou složenou z nepodpořených subjektů je pak velmi náročné. S největší pravděpodobností se u řady podskupin ukáže, že obsahují jen velmi malé množství subjektů. Tím se pak snižuje věrohodnost takto provedeného srovnání.

Metoda párování pomocí propensity skóre toto velké množství dělení nahrazuje jedním skóre, které zastupuje všechny dílčí charakteristiky.

Mohlo by se zdát, že čím více charakteristik do výpočtu zahrneme, tím lépe. Přidání dalších a dalších charakteristik sice zvyšuje přesnost modelu, ale další charakteristiky už nepřidávají tak velkou hodnotu pro nejlepší spárování subjektů. Sice se s rostoucím množstvím sledovaných charakteristik zvyšuje přesnost párování, ale sběr dalších dat je časově a finančně náročný. Navíc jde vždy o charakteristiky, které jsme schopni měřit, tudíž o charakteristiky pozorovatelné. Nepozorovatelné charakteristiky však nemáme k dispozici.

Nepozorovatelné charakteristiky však nejsou do výpočtu propensity skóre zahrnuty. Tudíž, liší-li se ony dvě skupiny v nepozorovatelných charakteristikách, „kazí“ výsledný výpočet. Sebelepší párování tak nedokáže nahradit experiment, který nepozorovatelné charakteristiky rozmístí shodným dílem do obou skupin.

Velké množství charakteristik, podle kterých se páruje, navíc způsobuje, že se pro spárování nalezne málo párů tak, aby byly jednotlivé charakteristiky shodné. Právě tento problém vícerozměrného párování se snaží řešit metoda párování pomocí propensity skóre.

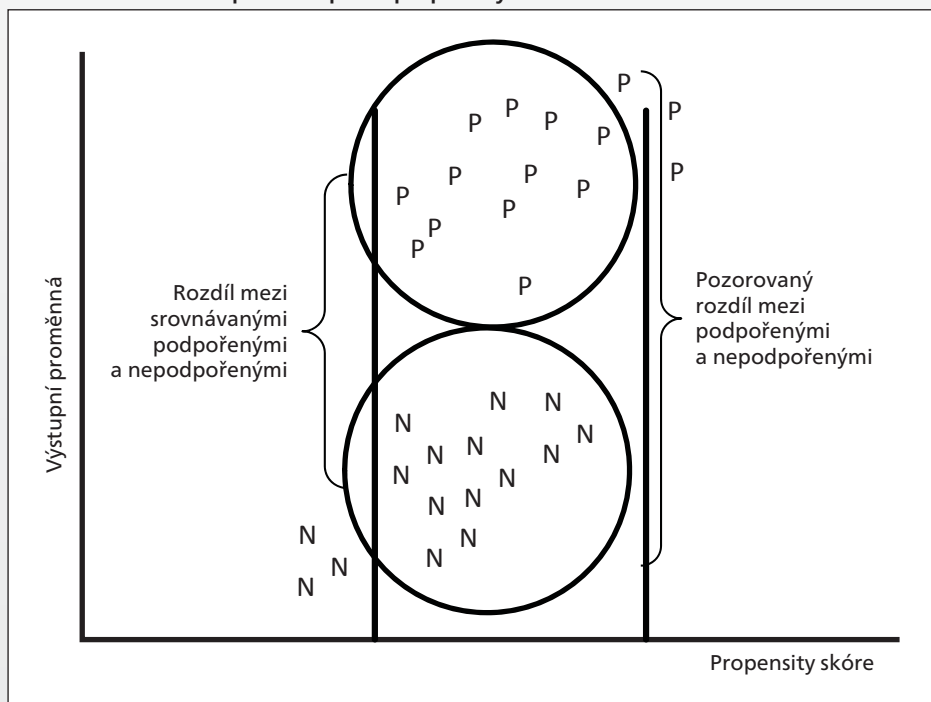
- Zmiňované charakteristiky zároveň poslouží k výpočtu propensity skóre, se kterým se poté počítá při párování jednotlivých subjektů mezi skupinami. Dojde tak vlastně k nahrazení celé řady proměnných jednou jedinou proměnnou. Ve své podstatě se tak snažíme spočítat pravděpodobnost obdržení podpory na základě vstupních charakteristik. V příkladu Potluka a kol. (2013) jsou použity tyto:
 - Velikost firmy (formou počtu zaměstnanců – přepočtené úvazky);
 - Právní forma (formou dummy proměnné za jednotlivé typy);
 - OKEČ (formou dummy proměnné za jednotlivé OKEČ);
 - Region působnosti (formou dummy proměnné za jednotlivé regiony);
 - Interakce velikosti firmy s výše uvedenými charakteristikami (násobek velikosti firmy s výše uvedenými charakteristikami);
 - Změna dlouhodobého majetku firmy (vypočteno jako procentní rozdíl mezi lety 2010 a 2008);
 - Podpora z jiných veřejných zdrojů (veřejná podpora dle dat ČSÚ a podpora OP LZZ).

V ideálním případě, pokud jsme schopni zjistit všechny charakteristiky, které mají vliv na dopad, je možné považovat výstupy ze srovnávací skupiny za kontrafaktuální situaci a dopad lze získat prostým srovnáním průměrů. V reálné situaci však nejsme nikdy schopni ověřit, jestli tomu tak je a zda neexistují nepozorované charakteristiky, které obě skupiny odlišují.

Předpokladem metody je, že jsou k dispozici data o pozorovaných charakteristikách obou skupin (se stejnými definičními znaky a shodnou metodou sběru). Dále je pak nutné, aby existoval přesah sledovaných charakteristik (při použití skóre pak přesah skóre) v obou skupinách (pro tento přesah se v angličtině používá pojem *common support*). Tento přesah zajišťuje, že alespoň nějaké subjekty z obou skupin bude možné spárovat.

Výše popisovanou situaci lze snadno znázornit graficky (viz obrázek č. 4). Mějme podpořené (P) a nepodpořené (N) subjekty, pro které jsme spočítali propensity skóre. Metodu pak můžeme použít, pokud pro jedno skóre nalezneme alespoň jeden subjekt v každé skupině. Ty poté můžeme párovat.

Obrázek č. 4: Metoda párování podle propensity skóre



Zdroj: Evropská komise (2009), vlastní úprava

Kdybychom měli ideální data, pak bychom pro každý subjekt ze skupiny podpořených subjektů našli právě jeden subjekt ze skupiny nepodpořených s naprosto

identickými charakteristikami. V reálném světě jde však o ideální a nedosažitelnou variantu. Právě pro tento případ se pak hodí nahrazení více charakteristik jednou – propensity skóre.

Toto skóre lze získat prostřednictvím logistické či probit regrese. Párování pak probíhá na základě takto získaného skóre. Zkoumaný dopad je pak spočítán jako rozdíl průměrů obou srovnávaných skupin.

Základní podmínkou této metody je dostupnost dat pro obě srovnávané skupiny ještě před zahájením intervence. Pokud by byla k dispozici pouze data z období intervence, párování by bylo zkresleno právě existencí podpory a možným ovlivněním některých charakteristik.

Při použití metody párování pomocí propensity skóre se sleduje několik posloupných kroků (Evropská komise 2009):

1) **Výpočet propensity skóre.** V tomto kroku je jednak rozhodnuto o tom, jaký model bude použit pro výpočet propensity skóre pro každý případ v datovém souboru (pomocí logit nebo probit regrese, pro všechny subjekty však stejná metoda), ale také o tom, které proměnné charakterizující případy budou zahrnuty do výpočtu propensity skóre. Takovými jsou zejména charakteristiky, které ovlivňují to, zda je subjekt oprávněn zapojit se do zkoumaného programu. Také je nutné, aby byly do výpočtu zahrnuty ty proměnné, které nejsou ovlivněny programem, tj. buď proměnné měřené před zahájením programu, nebo proměnné, které zůstávají nezávislé na programu v průběhu jeho realizace, či jsou ovlivněny očekáváním účasti v programu (u osob jde například o pohlaví, či věk, u firem například o právní formu, region působnosti apod.).

Rozdělení charakteristik v obou skupinách je takové, že se překrývají (*common support*). Tím neexistuje systematické zkreslení mezi oběma skupinami.

2) **Spárování subjektů pomocí propensity skóre.** Po spočítání propensity skóre pro každý subjekt následuje spárování subjektů ze skupiny podpořených a nepodpořených subjektů. Obvykle jsou v literatuře a studiích uváděny čtyři používané postupy, jak spárovat zkoumané subjekty (další metody uvádí například Khandker, Koolwal a Samand 2010):

a. **Párování s nejbližším sousedem (*Nearest neighbour matching*).** Toto párování páruje ty subjekty, jejichž propensity skóre má k sobě nejbližší. Tato metoda může být provedena „s vrácením“, kdy nepodpořené subjekty mohou být

použity pro párování vícekrát podle toho, jaké je rozložení propensity skóre v obou skupinách. Podpořené subjekty jsou však používány pouze jednou. V případě párování bez vracení jsou i nepodpořené subjekty zařazeny do páru s podpořenými subjekty pouze jednou a další párování se jich netýká.

- b. **Kaliper / párování rádiusem (*Calliper / radius matching*)**. Kaliper párování je postaveno na tom, že zkoumaný subjekt je přiřazen podle svého propensity skóre k nejbližšímu nepodpořenému subjektu, ale párování je podmíněno tím, že je dán interval, do jaké vzdálenosti se mohou lišit propensity skóre dvou párovaných subjektů. Tímto způsobem by se měl řešit problém, že při párování mohou být použity dva subjekty, které jsou ve své podstatě od sebe vzdálené. Nevýhodou ovšem je, že bude pravděpodobně spárováno méně subjektů, tudíž se vydatnost datového vzorku sníží. Párování rádiusem je variantou, kdy jsou jednomu zkoumanému projektu přiřazeny všechny vyhovující subjekty ze srovnávací skupiny, které leží uvnitř intervalu.
 - c. **Stratifikační či intervalové párování (*Stratification or interval matching*)**. Tento způsob párování se provádí tak, že jsou všechny subjekty v obou skupinách seřazeny podle velikosti propensity skóre a podle něj do několika intervalů. Tyto intervaly se liší počtem subjektů, které se v nich nacházejí. Poté je odhadnut dopad intervence pro každý interval zvlášť. Celkový dopad je pak získán jako vážený průměr takto získaných dílčích odhadů dopadu. U této metody je nutné provést zjištění, zda se tyto intervalové skupiny systematicky neliší (např. velký podíl nepodpořených subjektů s nízkým skóre a obráceně velký podíl podpořených subjektů v intervalech s vyšším skóre).
 - d. **Kernel matching**. Tato metoda párování nepáruje pouze dva subjekty mezi sebou, ale pro odhad výsledku pracuje se všemi subjekty ze srovnávací skupiny. Vypočítá se pomocí váženého průměru odhadu dopadu, kde váhami jsou rozdíly mezi propensity skóre v jednotlivých párech. Tím pak mají větší váhu odhady dopadu u těch subjektů, jejichž propensity skóre se lišilo jen málo, zatímco ty s velkým rozdílem (a tedy i horšími vlastnostmi spárování) mají váhu menší – čím horší spárování, tím menší váha v celkovém odhadu dopadu.
- 3) **Posouzení kvality párování**. Nyní se ověřuje existence společné podpory (*common support*). Existuje několik způsobů, jak to udělat. Například prostým vizuálním srovnáním v grafu či srovnáním minimálních a maximálních hodnot propensity skóre. Ty subjekty, které leží mimo interval společné podpory, by měly být z analýzy odstraněny, protože by stejně zhoršovaly přesnost odhadu dopadu, neboť k sobě nemají správný subjekt z druhé skupiny.

4) **Odhad dopadu.** Po provedení výše uvedených kroků už stačí srovnat průměry mezi oběma skupinami a otestovat statistickou signifikanci (t-testem) výsledků.

Nevýhodou párování ve srovnání s experimentálními metodami je, že musíme předpokládat, že výběrové zkeslení je řešeno pomocí charakteristik, které máme k dispozici. Dále je pak nutné, aby se charakteristiky mezi oběma skupinami překrývaly. Pokud se nepřekrývají, jsou skupiny rozdílné a není možné zpracovat odhady dopadu intervence.

V literatuře lze nalézt i podmíněné párování (*conditional matching*), které lze použít tam, kde lze předpokládat, že existuje nepozorovatelná významná charakteristika, jež není rovnoměrně zastoupena v obou srovnávaných skupinách.

Byť je metoda párování pomocí propensity skóre mezi vědci poměrně hojně používaná, nelze ji použít automaticky. Problém vícerozměrné analýzy dle dostupných charakteristik srovnávaných skupin je skryt do výpočtu propensity skóre, ale není eliminován zcela.

Tato metoda je oblíbená i proto, že je poměrně snadno aplikovatelná a nevyžaduje například dělicí hranici (jako metoda regresní nespojitosti) nebo existenci instrumentální proměnné (jako metoda instrumentálních proměnných).

Výhodou této metody oproti ostatním metodám využívajícím regresi je, že pracuje s překryvem charakteristik mezi podpořenými a nepodpořenými subjekty. Používá se také funkční závislost (většinou lineární), což obvykle nebývá zdůvodněno v datech ani v teorii (Angrist a Pischke 2009).

V české praxi byla tato metoda použita například ve studiích Hora a kol. (2009) či Hora a Sirovátka (2012), ve kterých autoři zkoumali dopady aktivní politiky zaměstnanosti na zaměstnanost osob, které byly registrovány na úřadech práce jako nezaměstnané. Dalším příkladem je studie Potluka a kol. (2013a), která zkoumá dopady podpor firmám na zaměstnanost, hospodářský výsledek a tržby. Částečně již byly v předchozím textu prezentovány výsledky kombinace této metody s metodou rozdílů v rozdílech.

Z již zmiňovaného příkladu Potluka a kol. (2013a) bylo propensity skóre možné použít pro párování podpořených a nepodpořených firem. Jinou alternativou je regres-

ní model či odhad na základě stratifikace propensity skóre, viz např. Steiner a Cook (2013). Tato literatura udává (a získané zkušenosti z uvedené studie to podporují), že párování je velmi náročné na data. Občas jsou výsledky neintuitivní a je lepší zvolit regresní model či stratifikaci pomocí propensity skóre. U regresní specifikace je třeba určit řád polynomu propensity skóre, který vysvětluje zkoumaný indikátor. Také bylo využito Akaiovo informačního kritérium (blíže viz Burnham a Anderson, 2002) pro výběr vhodného modelu. Oba zvolené přístupy poskytly autorům studie stejné doporučení pro výběr polynomu druhého stupně.

Princip zpracování vícerozměrných charakteristik zkoumaných subjektů využívá například i metoda synthetic control method. Abadie, Diamond a Hainmueller (2010 a 2014) ji používají například k testování předpokládaného hospodářského vývoje západní části Německa v případě, že by nedošlo ke sjednocení. Metoda pracuje s tím, že na základě znalosti jiných regionů modeluje umělý subjekt pro srovnání (zde například je to země, která se svými charakteristikami pokud možno co nejvíce podobá západní části Německa před sjednocením).

3.3.3 Regresní diskontinuita (Regression discontinuity design)

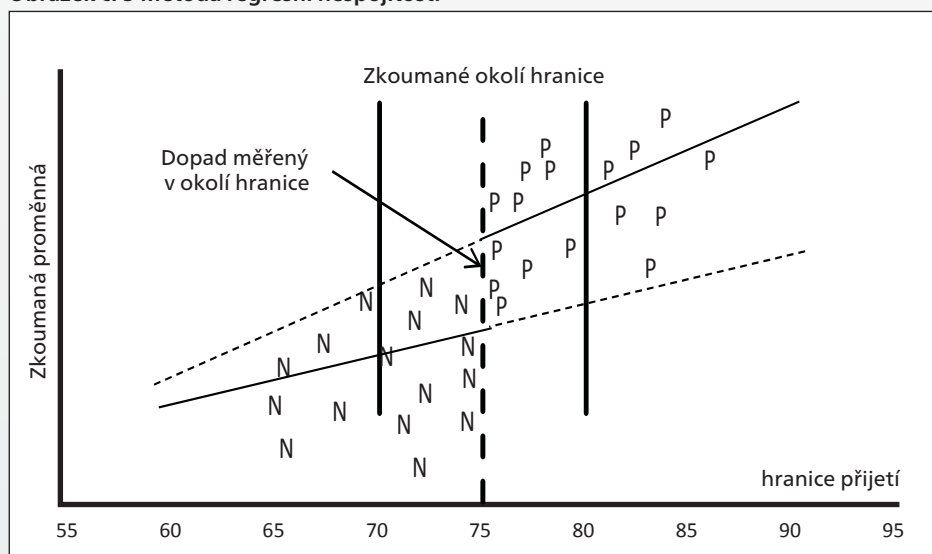
Další z používaných metod je metoda regresní diskontinuity. Její princip spočívá v tom, že existuje určitá hranice, která definuje rozdíl mezi podpořenými a nepodpořenými subjekty. Na jedné straně této hranice jsou nepodpořené subjekty a na druhé straně jsou subjekty podpořené. Podmínkou je, že je tato hranice zvolena u proměnné, která má kontinuální charakter (může jít o počet zaměstnanců, počet bodů v testu, výše výplaty, počet bodů u projektové žádosti, věk žadatelů apod.).

Těsné okolí této hranice lze považovat za situaci, která je velmi podobná náhodnému výběru přidělení podpory. Subjekty jsou shodné ve svých charakteristikách kromě té, že jedny subjekty jsou nad a druhé pod rozdělovací hranicí. Například v podstatě můžeme říci, že se podniky s 250 zaměstnanci neliší od podniků, které mají 251 zaměstnance, přesto bývají již v jiné kategorii (střední vs. velké podniky). Nicméně je nutné upozornit, že volba hranice pro uznání podpory musí být volena náhodně. Je-li tato hranice známa předem, může u potenciálních žadatelů vyvolat strategické chování, kterým tyto subjekty upravují své charakteristiky, aby spadaly definičně mezi podpořitelné subjekty (například s náborem zaměstnanců počkají až na období, kdy je spuštěn program nebo politika, o kterou se zajímají).

Výhoda této metody dopadové evaluace spočívá v tom, že není potřeba stanovovat další předpoklady pro její úspěšné použití. Při splnění základních podmínek pak stačí jen porovnat regresní odhady v místě hranice. Problémem této metody je, že odhad dopadu je pouze s interní validitou a nelze jej dost dobře generalizovat (problém s externí validitou odhadu). Důvodem je, že testované subjekty můžeme považovat za „shodné“ pouze v těsném okolí hranice, která rozděluje subjekty na podpořené a nepodpořené. Podrobněji diskutují problematiku metodologického přístupu k regresní nespojitosti například Imbens a Kalyanaraman (2009) nebo Angrist a Pischke (2009).

Na obrázku č. 5 je vidět grafické znázornění principu metody regresní nespojitosti.

Obrázek č. 5 Metoda regresní nespojitosti



Zdroj: Evropská komise (2009), vlastní úpravy

Také tato metoda má několik vstupních podmínek, aby bylo možné ji aplikovat. Zkoumané subjekty jsou u stanovené hranice a nejsou schopny upravovat hodnotu proměnné, podle které se určuje, zda podporu získají či ne. Ani to však ještě nemusí pro správné použití této metody stačit. Příklad (Potluka a kol. 2013a) ukazuje, že je metoda použita tak, že 65 bodů je hraničních pro přijetí nebo zamítnutí projektové žádosti. Výsledky jsou ale s největší pravděpodobností ovlivněny tím, že hodnotitelé dopředu věděli, že hranice 65 bodů je rozhodující pro definitivní zamítnutí žádosti

firem o podporu vzdělávání. Z tohoto pohledu pak mohli své rozhodování kolem této hranice upravovat. Pokud by program neměl dostatečný rozpočet na pokrytí všech žadatelů s dostatečně kvalitní žádostí, bylo by použití metody regresní nespojitosti přesnější – nikdo by až do konce hodnotícího procesu nevěděl, kolik bodů je zapotřebí k získání podpory.

Jinou podmínkou je, že na zvolené dělicí hranici se mění jen to, že jedny subjekty podporu dostanou a druhé ne. Problémem je, pokud se na této hranici mění i nějaká jiná charakteristika subjektů. Pokud takové charakteristiky existují, mohou ovlivňovat odhadovaný dopad intervence, a tím znehodnocovat dopadovou evaluaci.

Tato metoda také pracuje s tím, zda jde o ostrý přechod hranice (sharp), kde se pravděpodobnost získání podpory liší skokově kolem zvolené hranice. Taková situace nastává, když na jedné straně hranice je pravděpodobnost získání podpory 100% a na druhé straně je pravděpodobnost 100 %, že subjekty podporu nedostanou. V praxi taková situace ovšem nastat nemusí. Pak mluvíme o tzv. neostré variantě regresní nespojitosti.

V uváděném příkladu (Potluka a kol. 2013b) jde o neostrou verzi použití metody, kdy pod hranicí 65 bodů je 100% pravděpodobnost zamítnutí žádosti, ale nad touto hranicí není pravděpodobnost 100 % pro získání podpory a realizaci projektu, protože některé firmy nakonec projekty nerealizují (důvody mohou být různé).

Jinou obtíží při použití této metody je zvolení šířky pásma, ve kterém se nacházejí subjekty použité pro analýzu. Čím blíže k dělicí hranici, tím přesnější odhad, ale také méně pozorování. Se zvyšující se vzdáleností od dělicí hranice klesá přesnost, ale zvyšuje se počet měření. Při zpracování dopadové evaluace je tak nutné zvolit kompromis mezi množstvím měření v analýze a tím, jak přesné chceme mít výsledky.

Jiným problémem může být mylné identifikování dopadu tam, kde žádný není. Například tím, že je použita lineární podoba křivek, zatímco ve skutečnosti je vazba polynommická a lom křivky prochází právě hranicí rozdělující obě skupiny (takový příklad ukazují například Angrist a Pischke, 2009).

Metoda regresní nespojitosti je zpracována v několika dále popsanych krocích:

- 1) **Výběr proměnné a hraniční hodnoty pro rozdělení podpory.** Zvolená (spojitá) proměnná je taková, podle níž jsou subjekty rozděleny na podpořené a nepod-

pořené. Hranicí je pak taková hodnota, která rozděluje subjekty na podpořené a nepodpořené.

- 2) **Výběr hraničního intervalu.** Pro aplikaci metody a její přesnost je důležitým faktorem výběr intervalu kolem hranice, jenž zahrnuje subjekty, které budou do analýzy zahrnuty a které naopak ne. Kompromis mezi množstvím subjektů v analýze (vydatnost) a šíří zvoleného pásma (přesnost analýzy) byl již diskutován výše.
- 3) **Vymezení zkoumané proměnné.** Pro analýzu je nutné také vybrat spojitou proměnnou, na které budeme měřit dopady intervence.
- 4) **Rozhodnutí o ostré či neostré verzi metody.** Toto rozhodnutí je závislé na tom, jaké je rozdělení pravděpodobnosti přidělení podpory.
- 5) **Odhad dopadu intervence.** Pozorovaný rozdíl mezi hodnotami zkoumané proměnné mezi podpořenými a nepodpořenými subjekty na rozdělovací hranici pak určuje odhadovaný dopad.

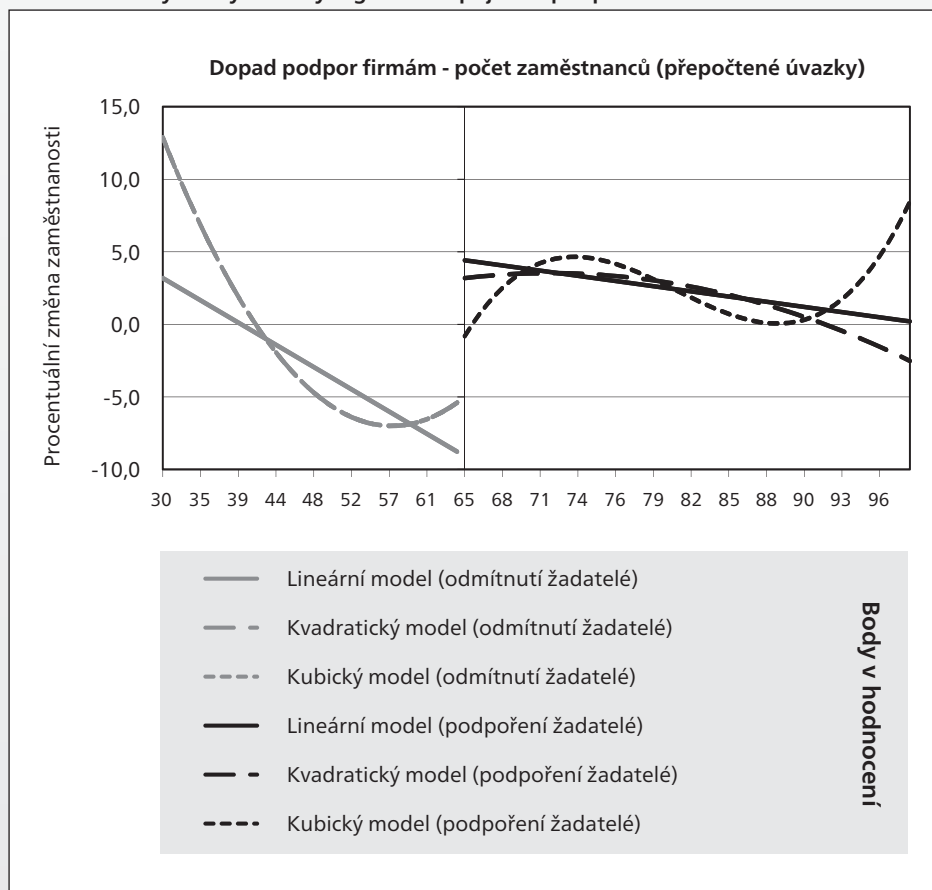
Metoda regresní nespojitosti umožňuje identifikovat příčinný účinek programu, aniž by bylo nutné brát v potaz různá arbitrární vylučující omezení, předpoklady výběru, funkční formy nebo alokační předpoklady. Metoda regresní nespojitosti je nejlepší alternativou k intervencím na bázi náhodných výběrů. Cennou vlastností této metody je, že měření, na základě něhož jsou subjekty rozděleny na oprávněné a neoprávněné žadatele, nemusí být shodné s měřením zjišťujícím výstupy intervence.

Jako každá metoda má i tato určitá omezení. Prvním z nich je omezení na případy a jejich výběr dle pozorovatelných charakteristik. Druhým omezením je, že metoda má jen interní validitu a odhadnutý dopad se týká pouze subjektů v bezprostřední blízkosti dělicí hranice. Výsledky této metody tedy nelze vztáhnout na celou populaci a není dosaženo externí validity.

Příkladem aplikace této metody v České republice je studie Potluka a kol. (2013b), ve které autoři využili toho, že projektové žádosti firem o podporu vzdělávání z Evropského sociálního fondu jsou hodnoceny a na základě hodnocení je podpora buď přidělena, nebo ne. Obrázek č. 6 ukazuje vizuální prezentaci výsledků této studie. Interval spolehlivosti nejsou v grafu znázorněny, nicméně v samotné studii Potluka a kol. (2013b) již prezentovány jsou. Dle nich pak lze říci, že odhady dopadu podpor firmám z Evropského sociálního fondu na zaměstnanost ve firmách jsou statisticky signifikantní.

Tato metoda by byla vhodná, kdyby nebyla hranice stanovena ve výši 65 bodů v hodnocení projektové žádosti. Tato hranice je totiž arbitrárně stanovena v operačním programu. Pokud žádost dostane méně než 65 bodů, je zamítnuta. Hodnotitelé žádostí tedy vědí, že dají-li žádosti méně než 65 bodů, nebude podpořena. To může ovlivňovat jejich chování. Vhodnější by byla situace, kdy je prostředků v programu méně, a tudíž se nedostane podpora pro všechny žadatele, kteří získají více než 65 bodů. Tím by se do hry zapojil jistý prvek náhody, protože nikdo nedokáže předem odhadnout, jak kvalitní projekty do soutěže přijdou a jaká bude finální bodová hranice, která rozdělí žádosti na podpořené a nepodpořené.

Obrázek č. 6: Výsledky metody regresní nespojitosti podpor vzdělávání ve firmách



Zdroj: Potluka a kol. (2013b)

Potluka a Brůha (2013) uvádějí další možnost, jak realizovat evaluaci pomocí regresní nespojitosti. Jde o situaci, kdy existuje více dělicích hranic. V takovém případě je možné použít vícerozměrnou regresní nespojitost, jak ji uvádějí Papay, Willett a Murname (2011). Nicméně v jejich případě studie Potluka a Brůha (2013) obsahoval datový soubor pouze 17 firem, které podaly více projektových žádostí a získaly podporu alespoň pro dvě z nich.

3.3.4 Instrumentální proměnné (instrumental variables)

S touto metodou se lze setkat i pod názvem metoda dvoustupňových nejmenších čtverců (2SLS – Two stage least squares). Tato metoda obsahuje určité prvky experimentu. Využívá se tehdy, když je sice na rozhodnutí potenciálních žadatelů, zda se programu zúčastní nebo ne, ale existuje ještě nějaký externí faktor ovlivňující pravděpodobnost získání skutečné podpory. Tento externí faktor ovlivní skutečnou účast bez možnosti ovlivnění ze strany žadatelů. Tím, že žadatelé nemají možnost tyto faktory ovlivnit, umožňuje tato metoda odstranit výběrové zkreslení. Nicméně pro instrumentální proměnnou musí platit dvě základní podmínky, jinak ji nelze použít a uvažovaný instrument není validní. Těmito podmínkami jsou:

- 1) Proměnná uvažovaná jako instrument musí významným způsobem ovlivňovat pravděpodobnost získání podpory
- 2) Tato proměnná ale nesmí ovlivňovat zkoumanou proměnnou dopadu, jediň skrze pravděpodobnost získání podpory

Je-li takový instrument nalezen (pro použití této metody bývá právě nejtěžší najít v hodný instrument), pak ho lze použít k odfiltrování výběrového vychýlení a k odhadu dopadu intervence. Druhá vlastnost není ověřitelná a je brána jako předpoklad.

Při splnění těchto podmínek poskytuje metoda instrumentálních proměnných poměrně dobré výsledky při odhadování dopadu intervencí a autoři studie Potluka a kol. (2013a) ji z jimi použitých metod považují za nejměrohodnější k aplikaci.

V evaluačních studiích se lze setkat s různými příklady instrumentálních proměnných, což jsou například neočekávané změny právních předpisů, geografické faktory, náhodné (neočekávané) výpadky ve financování programů (někdo tudíž podporu dostane a někdo ne). Také je možné využít přístup, který byl popsán s informováním potenciálních uživatelů programu.

Dopadovou evaluaci prostřednictvím metody instrumentálních proměnných je možné provést dvěma způsoby. Prvním z nich je metoda Waldova estimátoru, druhým přístupem je dvoustupňová metoda nejmenších čtverců.

Waldův estimátor bývá používán v jednodušších situacích s jedním instrumentem, který nabývá dvou hodnot (buď 0, nebo 1 – příkladem je využití při odhadování dopadů podle toho, zda například jsou regiony způsobilé pro nějakou podporu či nikoli). Navíc do modelu nevstupují další charakteristiky. Jde o jednoduchý model pro pochopení způsobu práce s touto metodou.

Postup zpracování je následující:

- 1) **Definování základních proměnných vstupujících do modelu.** Těmito proměnnými jsou jednak:
 - výstupní proměnná Y , na které sledujeme dopad podpor,
 - dále vstupuje do modelu binární proměnná T (subjekt s dotací má hodnotu proměnné $T=1$, pokud nikoliv, pak $T=0$),
 - proměnná U představující nepozorované charakteristiky zapojených subjektů,
 - proměnná Z představující vnější sílu, která ovlivňuje T , ale nekoreluje s U (nabývá hodnoty $Z=1$, je-li subjekt způsobilý pro podporu, a pokud není, nabývá hodnoty $Z=0$),
 - symbol $E(\cdot)$ představuje průměr a symbol $E(\cdot | \cdot)$ představuje střední hodnoty.
- 2) **Určení toho, jak instrument ovlivňuje vystavení subjektů intervenci.** V tomto kroku je vysvětleno, jak vnější faktory (instrument) ovlivňují pravděpodobnost účasti ve zkoumaném programu. Toto je možné vypočítat jako odečet podílu zapojení se do programu subjekty nevystavených působení instrumentu ($P(T=1|Z=0)$) od míry zapojení se do programu subjekty vystavených působení instrumentu ($P(T=1|Z=1)$). Rozdíl pak udává čistý efekt instrumentu Z .
- 3) **Použití instrumentu pro odhad dopadu.** Kdybychom dopad na zkoumanou proměnnou měřili přímo jako rozdíl mezi podpořenými subjekty ($E(Y|T=1)$) a nepodpořenými subjekty ($E(Y|T=0)$), získali bychom dopad politiky zkruslený výběrovým vychýlením. Pro odhad dopadu intervence bez tohoto vychýlení je nezbytné použít instrument. Pro odhad skutečného dopadu programu je nutné porovnávat hodnoty zkoumané proměnné u subjektů, které byly ($E(Y|Z=1)$) a které nebyly ($E(Y|Z=0)$) vystaveny působení instrumentu. Také je nutné brát v úvahu rozdíl-

né míry účasti v daném programu. Odhad dopadu programu je vyjádřen tzv. Waldovým estimátorem, zapsaným následovně:

$$\delta = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{(T = 1 | Z = 1) - P(T = 1 | Z = 0)}$$

Waldův estimátor se používá pouze pro binární instrumenty. V praxi se proto častěji využívá postup dvoustupňové metody nejmenších čtverců. K největším omezením této metody patří, že nalezení vhodného instrumentu není úplně triviálním postupem. Posouzení vnějšího vlivu instrumentu bývá často velmi subjektivní a zkreslené.

Naopak hlavní výhodou této metody je, že je velmi blízká experimentálním metodám s náhodným výběrem podpořených subjektů. V některých experimentálních přístupech je tato metoda používána, není-li například možné z etických důvodů jednoznačně rozdělit účastníky na podpořenou a nepodpořenou skupinu (někdo odstoupí z účasti a naopak někdo jiný se přidá).

Autoři studie Potluka a kol. (2013a) postupovali tak, že jako vhodný instrument zvolili pravděpodobnost podpory projektové žádosti podle toho, kteří hodnotitelé měli za úkol žádost ohodnotit. Nejprve odhadli model diskrétní volby. Pravděpodobnost, že žádost o podporu bude přijata a projekt podpořen, vysvětlují autoři pomocí proměnné, kterou nazvali „kombinace osobních vychýlení expertů“. Používají pro to anglickou zkratku „sum of experts' personal biases – SEPB“. Dále autoři používají proměnné vyjadřující charakteristiky zkoumaných firem. Mezi ně jsou zahrnuty obor působnosti firmy (NACE); region, ve kterém firma působí; právní formu a velikost firmy. Zpracování analýzy proběhlo tak, že ze všech hodnocení všech projektů všemi hodnotiteli se spočítal průměrný počet bodů (XN) a průměrné hodnocení každého hodnotitele (XH). Rozdíl těchto dvou proměnných (XH – XN) je hodnotitelovo osobní vychýlení (EPB). Pro každý projekt pak sečetli EPB všech hodnotitelů a získali tím hodnotu proměnné SEPB.

Před vlastními výpočty autoři ověřili, že proměnná SEPB je významným určovatelem pravděpodobnosti získání podpory. Tím je tedy splněna první (nutná) podmínka pro použití metody instrumentálních proměnných. Autoři pak dále uvádějí, že druhá podmínka, že vypočtená proměnná SEPB nemá vliv na zkoumané proměnné hospodaření firmy jinak než skrze pravděpodobnost získání podpory, není statisticky testovatelná a autoři ji vzali jako nutný předpoklad.

V dalším kroku autoři použili dummy proměnné vztažené k velikosti firem, obor působnosti a součin proměnné podpory s dummy proměnnými velikosti firmy a oboru. Odhadnuté koeficienty u těchto proměnných pak udávají odhad dopadu podpory.

Bodové odhady, včetně statistické signifikance výsledků použití této metody, jsou znázorněny v tabulce č. 2.

Tabulka č. 2 Odhad dopadu podpor Evropského sociálního fondu na zaměstnanost ve firmách pomocí instrumentálních proměnných

Typ firmy	Bodový odhad podpory	p-value	Kč na jedno místo
Všechny firmy	10,38	0,09	184 189
Mikrofirmy	0,56	0,49	1 374 781
Malé firmy	1,13	0,41	415 075
Střední firmy	7,95	0,04	121 328
Velké firmy	15,15	0,00	104 037
Firmy v NACE C (relativně vůči ostatním firmám)	9,10	0,26	-
Firmy v NACE D (relativně vůči ostatním firmám)	3,06	0,07	-

Zdroj: Potluka a kol. (2013a)

4

SHRUTÍ PŘÍKLADU A POSTUPU ZPRACOVÁNÍ DOPADOVÝCH EVALUACÍ

4.1 Zkušenosti z realizovaných kontrafaktuálních dopadových evaluací

Z výše uvedených předpokladů jednotlivých metod a výsledků příkladu je patrné, že autoři Potluka a kol. (2013a) našli pozitivní dopady podpor ve firmách. Určitá nekonzistentnost mezi výsledky různých metod (viz příklady uváděné výše) je způsobena tím, že některé metody jsou pro určitý typ evaluace méně vhodné. Závěrem srovnání použitých metod je, že by řídicí orgány i evaluátoři měli ke kontrafaktuálním dopadovým evaluacím přistupovat zodpovědně, nejprve prostudovat zkoumanou intervenci a poté zvolit design, který je nejvhodnější pro danou evaluaci, nikoli zkoušet všechny možné metody.

V uváděném příkladu Potluka a kol. (2013a) je patrné, že metoda regresní nespojitosti byla patrně ovlivněna tím, že do metody do jisté míry vstupovaly proměnné, které pravděpodobně nebyly úplně náhodné, ve smyslu toho, že se hodnotitelé žádostí patrně chovali určitým taktickým způsobem kolem hranice 65 bodů. Sami autoři pak metody párování pomocí propensity skóre uvedli, že určitá nesourodost výsledků je dána kvalitou párování v této metodě. Závěrem se autoři Potluka a kol. (2013a) přiklání k výsledkům hodnocení dopadů pomocí metody instrumentálních proměnných, protože dle jejich názoru jde o metodu, která je pro daný příklad nejvíce validní.

Dalším navazujícím krokem dopadových evaluací musí být i vysvětlení kauzálních vztahů stojících v pozadí změn. Velmi častým jevem při prezentaci výsledků je, že se autoři spokojí s odpovědí na to, zda zkoumaná intervence má, nebo nemá dopad na

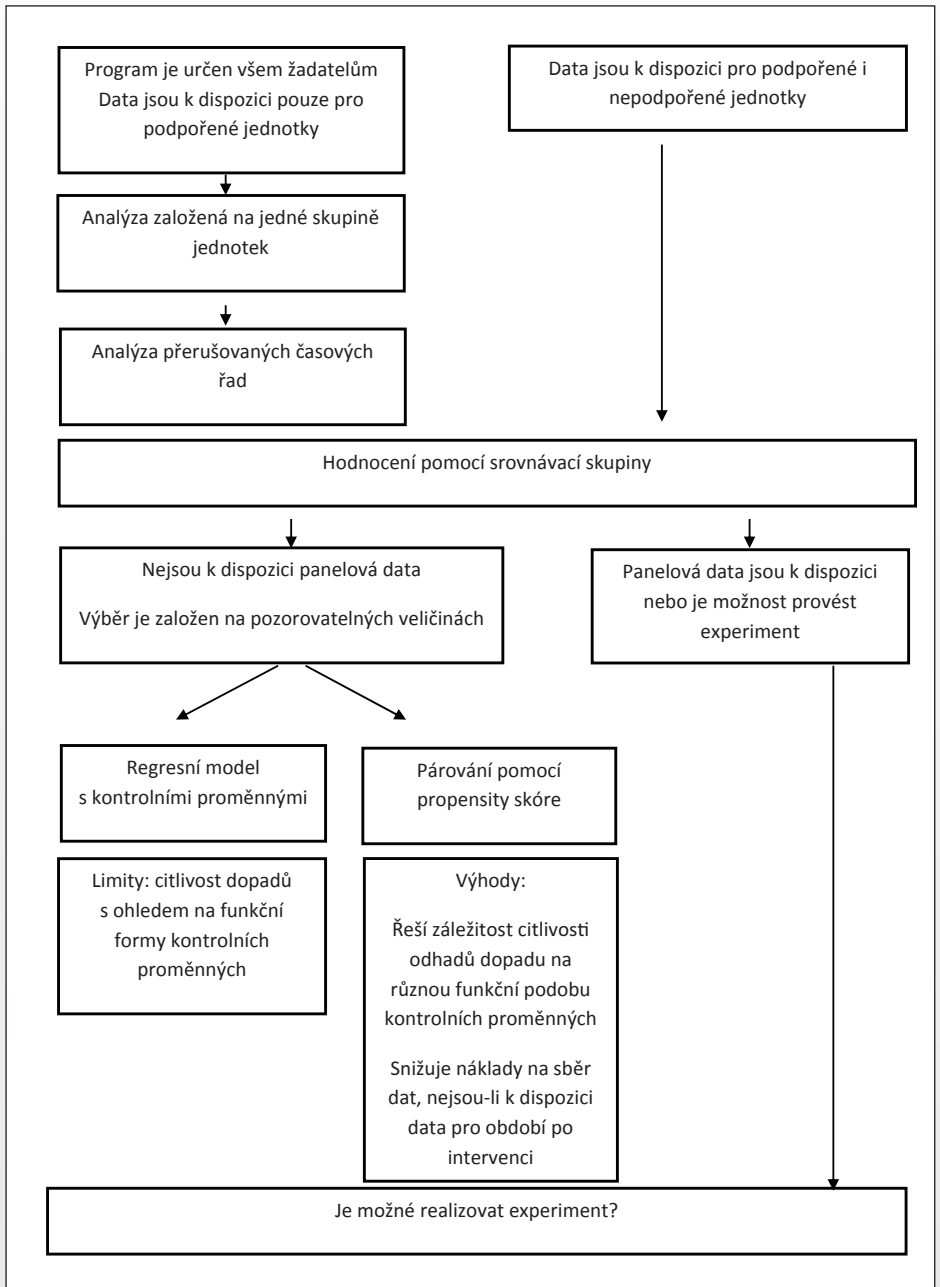
zkoumanou cílovou skupinu. To není dostačující pro praktické využití výsledků evaluací. Ve výsledcích nebývá prezentována cost-benefit analýza, podle které by bylo možné odvodit, zda je program efektivní nebo ne. Také často schází vysvětlení teorie změny stojící za řetězcem příčin a následků vedoucích ke kýženým změnám u cílové skupiny (k teorii změny blíže viz Morra-Imas a Rist, 2009). Kromě otázky, jaký dopad měla zkoumaná intervence a otázky, na koho měla intervence dopad, je nutné doplnit i otázku, jaký je mechanismus, kterým bylo tohoto dopadu dosaženo. Pokud by odpovědí na první otázku bylo, že dopad není pozitivní, pak je jasně patrný význam třetí otázky pro další směřování programu. Třetí otázka má svůj význam i v případě nalezení pozitivních dopadů, protože může odhalit, že lze program případně implementovat efektivněji.

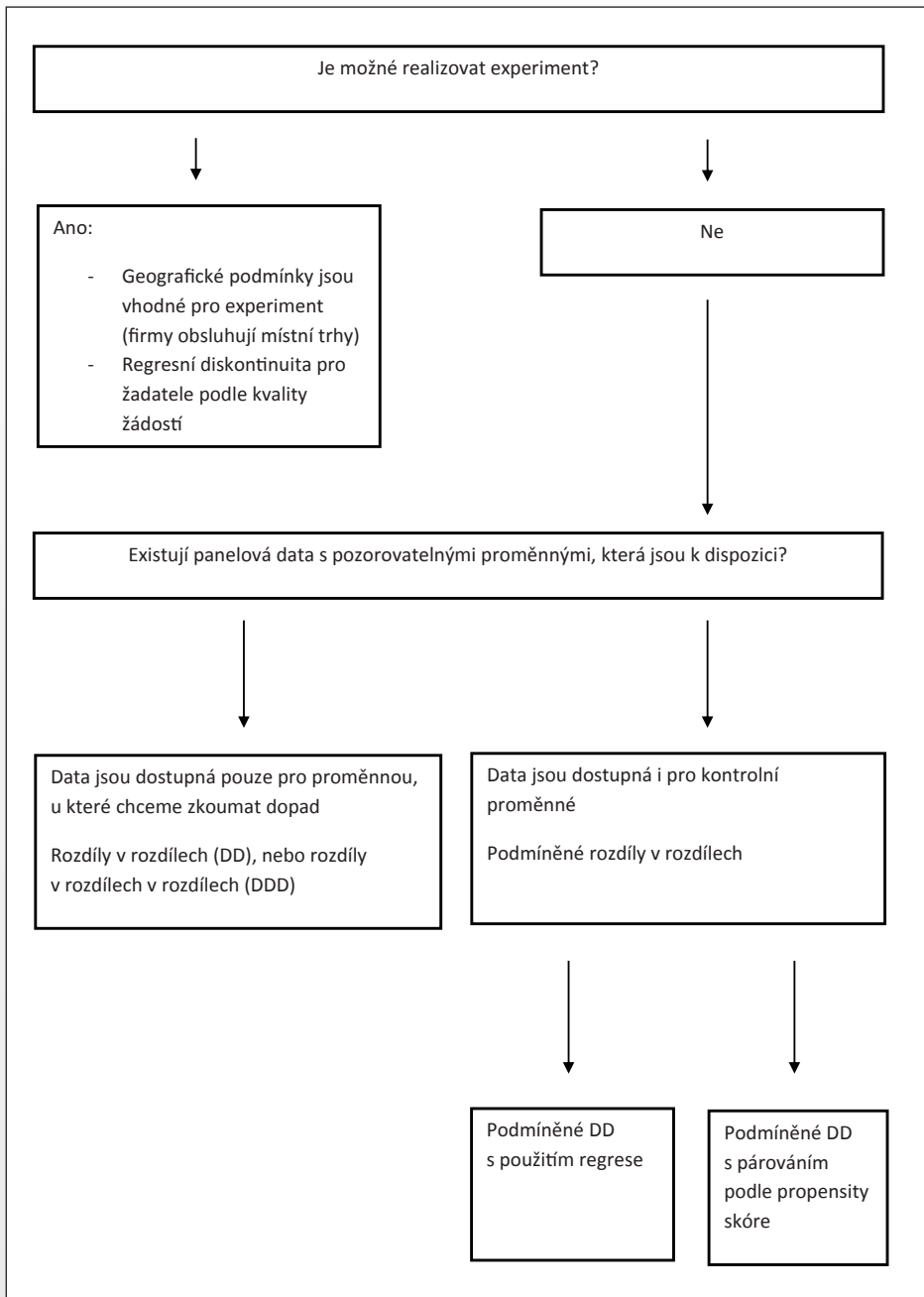
Sebelepší metody a odhady dopadů programů však nezajistí skutečnou aplikaci závěrů v praxi. Ta leží primárně v rukou politiků, případně manažerů zkoumaných programů. Například z uváděné evaluace Potluka a kol. (2013a) bylo patrné, že i na základě zpracování pomocí metody instrumentálních proměnných se potvrdilo, že hodnotitelé projektových žádostí aplikují stejná kritéria hodnocení různým způsobem. Řídicí orgán na to reagoval tím, že zahájil sérii školení, na kterých se hodnotitelé školili v aplikaci kritérií a jejich jednotném použití. Dalším efektem je snaha standardizovat žádosti tam, kde je to možné, aby nebylo zapotřebí práce tolika hodnotitelů, a tím se snížily i náklady na realizaci výzev a zjednodušil se proces hodnocení.

4.2 Dostupnost dat jako determinant pro realizaci dopadové evaluace

Obrázek č.7 znázorňuje, jak postupovat při dopadové evaluaci podle dostupnosti dat pro analýzu.

Obrázek č. 7





Zdroj: Zpracováno dle Bondonio (2010), použito dle Potluka a Špaček (2013)

5

ZÁVĚRY

Při použití kontrafaktuálních dopadových evaluací pro efektivní plánování a implementaci veřejných výdajových programů je nutné upozornit na fakt, že jde o programy realizované v politickém prostředí. Určující je tudíž míra ochoty politiků naslouchat tomu, co výsledky evaluací říkají a co to znamená pro realizované programy. V některých případech to může znamenat i negativní zprávy, pokud realizované programy nedosahují pozitivních výsledků. V programech realizovaných firmami či nadacemi bývá situace odlišná, protože tyto subjekty se primárně zodpovídají svým vlastníkům či zakladatelům a kontrolní mechanismy jsou přímočařejší.

Správné naplánování a pilotní odzkoušení programů může však přinést obrovské úspory a efekty, jsou-li výsledky dopadových evaluací použity skutečně v rámci programového cyklu. Někdy to může znamenat i zastavení neefektivních programů – a tím ušetření rozpočtových prostředků. K tomu však musí mít příslušná politická reprezentace dostatečnou politickou odvahu. Tu však v současné době v Evropě mnohdy ani nemá k tomu, aby dopadové evaluace proběhly (co kdyby se náhodou ukázalo, že program nemá deklarované efekty). Natožpak, aby byl program realizován pomocí náhodného výběru účastníků.

Výběr vhodných metod je nutný už na začátku přípravy programu, i když není nezbytně nutný. Velmi záleží na tom, zda a jaká data jsou dostupná pro dopadové evaluace programu. Tam, kde je to vhodné, lze použít i kombinaci metod. Například lze využít metody párování podle propensity skóre s metodou rozdílů v rozdílech. První zmiňovaná metoda pomůže zajistit shodné skupiny pro srovnání a na těchto datech je pak aplikována druhá zmiňovaná metoda. Kombinace pak snižuje riziko výběrového zkreslení. K tomu je ale nutné, aby manažeři programů měli alespoň základní znalost o použitelných metodách nebo o jejich předpokladech a úskalích.

Při koncipování dopadových evaluací by manažeři programů měli brát v úvahu i termín, na který dopadovou evaluaci naplánovat. Jde o to, které okamžiky jsou pro implementaci programu významné. Jde zejména o dobu těsně po skončení realizace projektů, ale i doba po skončení tzv. udržitelnosti (u projektů, kde se to vyžaduje – u komerčních projektů tato situace není – projekty mají dobu životnosti). Mimoto je nutné začít plánovat dopadové evaluace již během přípravy samotného programu. V současné praxi programových manažerů v Evropské unii se to u programů financovaných z fondů Evropské unie často opomíjí. Může za to částečně i koncept programového cyklu, který evaluace řadí až do finální části celého cyklu a tím ponechává manažery v klidu. Ti pak řeší evaluace až ke konci, kdy již není čas ani prostředky k tomu, aby zajistili dostatečně kvalitní kontrafaktuální dopadové evaluace, zejména kvůli tomu, že nemají k dispozici vhodná data.

Z výše uvedeného textu je patrné, že použití zmiňovaných metod skrývá řadu předpokladů a je možné, že zpočátku se budou jak evaluátoři, tak manažeři programů dopouštět chyb. Nicméně lze očekávat, že se bude zvyšovat kvalita zpracovaných dopadových evaluací s rostoucím povědomím o dopadových evaluacích a zkušenostmi jak zadavatelů, tak zpracovatelů.

Zároveň by manažeři programů neměli brát dopadové evaluace jako formální prvek, jak se to jeví nyní (zejména proto, že pro operační programy fondů EU je dopadová evaluace zmíněna v nařízení Evropské unie jako povinnost), ale jako smysluplný a užitečný nástroj ke zvýšení efektů a efektivity jimi řízených programů. Pokud tomu tak bude, zvýší se efektivnost veřejných výdajových programů a tím se ušetří i prostředky daňových poplatníků.

6

**SEZNAM POUŽITÉ
LITERATURY**

- [1] Abadie, A.; Diamond, A.; Hainmueller, J. (2010): Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, *Journal of the American Statistical Association*, Volume 105, Issue 490, DOI: 10.1198/jasa.2009.ap08746
- [2] Abadie, A.; Diamond, A.; Hainmueller, J. (2014, forthcoming): Comparative Politics and the Synthetic Control Method, *American Journal of Political Science*, DOI: 10.1111/ajps.12116
- [3] Abramovsky, L.; Battistin, E.; Fitzsimons, E.; Goodman, A.; Simpson, H. (2011): Providing Employers with Incentives to Train Low-Skilled Workers: Evidence from the UK Employer Training Pilots, *Journal of Labour Economics*, Vol. 29, No. 1, pp. 153 – 193
- [4] Angrist, J. D.; Pischke, J. (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, ISBN 978-0-691-12035-5
- [5] ASVAPP (2012): Counterfactual impact of cohesion policy: impact and cost-effectiveness of investment subsidies in Italy, staženo 6. března 2013 z http://ec.europa.eu/regional_policy/information/evaluations/impact_evaluation_en.cfm#1
- [6] Barca, F. (2009): An agenda for a reformed Cohesion Policy: a place-based approach to meeting European Union challenges and expectations, staženo 29. listopadu 2012 z http://ec.europa.eu/regional_policy/archive/policy/future/pdf/report_barca_v0306.pdf
- [7] Barrett, A.; O'Connell, P., J. (1999): Does Training Generally Work? The Returns to In-Company Training, IZA Discussion Paper No. 51, August 1999, staženo 15. listopadu 2014 z <http://www.econstor.eu/bitstream/10419/20887/1/dp51.pdf>
- [8] Batterbury, S. C. E. (2006): Evaluation Concepts and Methods – Principles and Purposes of EU Cohesion Policy Evaluation, *Regional Studies*, Vol. 40. 2, pp. 179-188, April 2006
- [9] Betcherman, G.; Daysal, N. M.; Pagés C. (2009): Do employment subsidies work? Evidence from regionally targeted subsidies in Turkey; *Labour Economics*; Vol. 17, pp. 710–722
- [10] Bondonio, D. (2010): Přednášky na Counterfactual Impact Evaluation Summer School v Pracatinat

- [11] Burnham, K. P.; Anderson, D. R. (2002): Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach (2nd ed.), Springer-Verlag, ISBN 0-387-95364-7
- [12] Criscuolo C.; Ralf, M.; Overman, H.; Van Reenen, J. (2009): The causal effects of an industrial policy, staženo 3. června 2013 z <http://www.nber.org/papers/w17842>
- [13] Czarnitzki, D.; Bento, C. L.; Doherr T. (2011): Counterfactual impact evaluation of cohesion policy: Examples from Support to Innovation and Research, staženo 3. června 2013 z http://ec.europa.eu/regional_policy/information/evaluations/pdf/impact/ciewp2_final.pdf
- [14] Evropská komise (2004): Aid Delivery Methods Volume 1: Program Cycle Management Guidelines, staženo 26. února 2012 z http://ec.europa.eu/europeaid/multimedia/publications/publications/manuals-tools/t101_en.htm
- [15] Evropská komise (2009): EVALSED: The resource for the evaluation of Socio-Economic Development: Guide a Evaluation sourcebook: Methods and Techniques, staženo 15. listopadu 2014 z http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf a http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/difference-in-differences/difference-in-differences_en.htm#3
- [16] Evropská komise (2010): Fifth Report on Economic, Social and Territorial Cohesion: Investing in Europe's future, staženo 12. dubna 2012 z http://ec.europa.eu/regional_policy/sources/docoffic/official/reports/cohesion5/pdf/5cr_en.pdf
- [17] Evropská komise (2013a): Programming Period 2014-2020 Monitoring and Evaluation of European Cohesion Policy European Social Fund Guidance
- [18] Evropská komise (2013b): Evaluation in the field of State aid: Draft Methodological guidance paper - concepts and recommendations
- [19] Gaffey, V. (2009): Evaluating Cohesion Policy at EU Level - Balancing Breadth and Rigour: in RSA Annual Conference: Understanding and Shaping Regions: Spatial, Social and Economic Futures, Leuven, Belgium
- [20] Gaffey, V. (2011): Plenary Session: After the Fifth Cohesion Report, in What Future for Cohesion Policy? An Academic and Policy Debate, Bled, Slovenia
- [21] Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., Vermeersch, Ch. M. J., (2011), Impact Evaluation in Practice, The World Bank, ISBN 978-0-8213-8541-8
- [22] Girma, Sourafel; Görg, Holger; Strobl, Eric; Walsch, Frank (2008): Creating jobs through public subsidies: An empirical analysis, Labour Economics 15, pp. 1179 - 1199
- [23] Hamersma, S. (2008), The Effects of an Employer Subsidy on Employment Outcomes: a Study of the Work Opportunity and Welfare-to Work Tax Credits, Journal of Policy Analysis and Management, Vol. 27, No. 3, pp. 498 – 520

- [24] Holzer, H.; Block, R.; Cheatham, M.; Knott, J. (1993): Are Training Subsidies for Firms Effective? The Michigan Experience, *Industrial and Labour Relations Review*, Vol. 46 pp. 625 - 636
- [25] Hora, O.; Sirovátka, T.; Tomešová-Bartáková, H.; Vyhliđal, J. (2009): Hodnocení programů aktivní politiky zaměstnanosti realizovaných v roce 2007 se zaměřením na rekvalifikace (struktura, cílenost, krátkodobé a střednědobé efekty na opuštění evidence), staženo 15. listopadu 2014 z http://praha.vupsv.cz/Fulltext/vz_300.pdf
- [26] Hora, O.; Sirovátka, T. (2012): Srovnání efektů aktivní politiky zaměstnanosti v České republice v období růstu (2007) a během první fáze krize (2009), Výzkumný ústav práce a sociálních věcí, v.v.i., ISBN 978-80-7416-106-3
- [27] Imbens, G; Kalyanaraman, K. (2009). Optimal Bandwidth Choice for the Regression Discontinuity Estimator, NBER Working Papers 14726, National Bureau of Economic Research, staženo 15. listopadu 2014 z <http://www.nber.org/papers/w14726.pdf>
- [28] Khandker, S. R., Koolwal, G. B., Samad, H. A. (2010), Handbook on impact evaluation: Quantitative methods and practices, The World Bank, ISBN 978-0-8213-8028-4
- [29] Kluge, J.; Schmidt, Ch., M. (2002) Can training and employment subsidies combat European unemployment?; CEPR, CES, MSH, Economic Policy October 2002, pp. 411 – 448
- [30] Kváča, V., Potluka, O. (2011): Pilotní evaluace srovnávací metodou (counterfactual) v OP Lidské zdroje a zaměstnanost. Konference ČES Evaluace na rozcestí – trendy a praxe, 2. 6. 2011
- [31] Lechner, M.; Miquel, R.; Wunsch, C. (2011): Long-run effects of public sector sponsored training in West Germany, *Journal of the European Economic Association*, Vol. 9(4), pp. 742 – 784
- [32] Lechner, M.; Wunsch, C.; Scioch, P. (2013): Do Firms Benefit from Active Labour Market Policies?, WWZ Discussion Paper 2013/11, September 2013, Center of Business and Economics (WWZ), University of Basel, staženo 15. listopadu 2014 z https://wwz.unibas.ch/uploads/tx_x4epublication/Do_Firms_benefit_DiscussionPaper_2013-11_Wunsch.pdf
- [33] Lynch, L.; Black, S. (1995): Beyond the Incidence of Training: Evidence from a National Employers Survey. National Bureau of Economic Research Working paper No 5231, Cambridge, staženo 15. listopadu 2014 z <http://www.nber.org/papers/w5231>
- [34] Martini, A. (2009): Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy, 6th European Conference on Evaluation of Cohesion Policy, Warsaw, staženo 15. listopadu 2014 z http://ec.europa.eu/regional_policy/archive/conferences/evaluation2009/index_en.htm
- [35] Morgan, S. L.; Winship, C. (2010): Counterfactuals and causal inference: methods and principles for social research, Cambridge University Press, ISBN 978-0-521-67193-4

- [36] Morra-Imas, L. G., Rist, C. R. (2009): The road to results: designing and conducting effective development evaluations, The International Bank for Reconstruction and Development / The World Bank, ISBN 978. 0-8213-7891-5
- [37] Morris, S.; Tödtling-Schönhofer, H.; Wiseman, M. (2013): Designing and commissioning of counterfactual impact evaluation: a practical guidance for ESF managing authorities, ISBN 978-92-79-28238-6, staženo 15. listopadu 2014 z <http://ec.europa.eu/social/main.jsp?catId=738&langId=cs&pubId=7646>
- [38] Mouqué, D. (2011a): Presentation at workshop: How to capture the effects of EU funding? Bringing together qualitative and quantitative methods, Open Days
- [39] Mouqué, D. (2011b): Conclusion of the conference, Conference on counterfactual impact evaluation, Warsaw
- [40] Mouqué, D. (2012): What are counterfactual impact evaluations teaching us about enterprise and innovation support? staženo 15. listopadu 2014 z http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_02_counterfactual.pdf
- [41] Papay, J. P.; Willett J. B.; Murnane, R. J. (2011): Extending the Regression-Discontinuity Approach to Multiple Assignment Variables. *Journal of Econometrics*. 2011, roč. 161, č. 2, s. 203-207. ISSN 03044076
- [42] Potluka, O.; Derlukiewicz, N.; Gombitová, D.; Horáková, J.; Jílková, J.; Koczisky, G.; Korenik, S.; Košťál, C.; Kunze, C.; Kuttor, D.; Louda, J.; Makarevich, T.; Nemeč, J.; Pisár, P.; Rogowska, M.; Slintáková, B.; Špaček, M.; Švecová, L.; Woitek, F. (2010): Impact of EU Cohesion Policy in Central Europe. Lipsko: Leipziger Universitätsverlag GmbH, 2010. 141 s. ISBN 978-3-86583-541-3
- [43] Potluka, O.; Brůha, J. (2013): Zkušenosti s kontrafaktuální dopadovou evaluací v České republice, *Evaluační teorie a praxe*. [online], 2013, roč. 1, č. 1, s. 53–68. ISSN 2336-114X. staženo 15. listopadu 2014 z URL: <http://www.evaltep.cz/inpage/archiv/>
- [44] Potluka, O.; Brůha, J.; Pělucha, M.; Květoň, V.; Vrbová, L.; Špaček, M.; Loun, J. (2013a): Pilotní counterfactual impact evaluation OP LZZ, oblast podpory 1. 1; Příloha č. 1 Hodnocení dopadu ESF na zaměstnanost ve firmách, staženo 15. listopadu 2014 z <http://km.vse.cz/wp-content/uploads/2011/06/2PZ-Priloha01-zamestnanost.pdf>
- [45] Potluka, O.; Brůha, J.; Vozár, O.; Špaček, M.; Loun, J. (2013b): The Impacts of Subsidies on Czech Firms. *Statistika* [online], 2013, roč. 50, č. 4, s. 56–62. ISSN 0322-788X. URL: [http://www.czso.cz/csu/2013edicniplan.nsf/engc/4600438B63/\\$File/180213q4056.pdf](http://www.czso.cz/csu/2013edicniplan.nsf/engc/4600438B63/$File/180213q4056.pdf). ISSN 1804-8765 (online)
- [46] Potluka, O.; Špaček, M. (2013): Postupy a metody kontrafaktuálních dopadových evaluací pro Operační program Zaměstnanost v období 2014 – 2020, staženo 15. listopadu 2014 z http://www.mpsv.cz/files/clanky/17051/Metodika_CIE_MPSV_131015.pdf

- [47] Potluková, E.; Potluka, O.; Jiskra, J.; Límanová, Z.; Telička, Z.; Bartáková, J.; Springer, D. (2012): Is Age a Risk Factor for Hypothyroidism in Pregnancy? An Analysis of 5223 Pregnant Women. *The Journal of Clinical Endocrinology & Metabolism* [online], 2012, roč. 97, č. 6, s. 1–8. ISSN 0021-972X. URL: <http://jcem.endojournals.org/content/early/2012/03/15/jc.2011-3275.abstract?sid=1ccd3e7d-e1d8-4957-8037-02b56b36101a>
- [48] Ravallion, M. (2012): Can We Trust Shoestring Evaluations? Policy Research Working Paper 5983, The World Bank, staženo 15. listopadu 2014 z <http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-5983>
- [49] Rosenbaum P., Rubin D. (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, 70, s. 41-50
- [50] Rubin, D. B. (1977): Assignment to Treatment Group on the Basis of a Covariate, *Journal of Educational Statistics*, vol. 2, 1-26
- [51] Steiner P., Cook D. (2013): Matching and Propensity Scores, kapitola připravovaná pro *The Oxford Handbook of Quantitative Methods*, staženo 15. listopadu 2014 z <http://www.ipr.northwestern.edu/workshops/past-workshops/quasi-experimental-design-and-analysis-in-education/recent-articles/docs/QE2010-MatchingPS.pdf>
- [52] Stryczynski, K. (2014): Future development of evaluation practice: Expectations from a DG REGIO perspective, presentation at the Annual conference of the Czech Evaluation Society, 10th June 2014
- [53] MMR-NOK (2013): Metodický pokyn MMR-NOK Zásady tvorby a používání indikátorů v programovém období 2014–2020
- [54] Wunsch, C.; Lechner, M. (2008): What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes, *Kyklos*, Vol. 61, No. 1, pp. 134 – 174
- [55] Žvalionytė, D. (2011): Counterfactual Impact Evaluation of Social Integration Services for Socially Vulnerable Groups: experience of Lithuania, in Conference on counterfactual impact evaluation - Warsaw 12 December 2011, staženo 15. listopadu 2014 z http://ec.europa.eu/regional_policy/impact/evaluation/conf_doc/warsaw_12122011/zvalionyte_ppmi_lithuania.pdf

