



Univerzita Karlova v Praze
Fakulta sociálních věd
Institut sociologických studií
Katedra sociologie

Diplomová práce:

Chyby měření v průzkumu trhu a v sociologickém výzkumu

Pavel Kohout

Praha 2006

Konzultant: Mgr. Jiří Remr

Poděkování

Děkuji Mgr. Jiřímu Remrovi za velmi cenné připomínky k této práci.

Věnování

Tuto práci věnuji Jirkovi Svobodovi za šanci vyzkoušet si, jak průzkum trhu funguje v praxi a Darkovi Pikálkovi za trpělivost a zkušenosti, které mi dosud stihl předat.

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a použil jen uvedených pramenů a literatury.

V Praze, dne 12. ledna 2006



.....
Pavel Kohout

„Beginners in our trade are apt to make the same mistakes we did. If market research was a science, this would not happen so much because people would do things in the same prescribed way.”

Harry Henry

*zakládající člen British
Market Research Society*

Obsah

1.	Abstrakt	5
2.	Metoda zpracování a cíl textu	5
3.	Co jsou průzkum trhu a sociologický výzkum?	6
4.	Základní druhy průzkumu trhu	7
5.	Etika a kvalita výzkumu	9
6.	Plánování výzkumných projektů	12
7.	Metody výzkumu	14
7.1	Výhody a nevýhody - osobní dotazování	17
7.2	Výhody a nevýhody - telefonické dotazování	20
7.3	Výhody a nevýhody – dotazníky poštou a emailem, internetové výzkumy	22
7.4	Smíšené módy dotazování (mixed mode surveys)	24
8.	Chyby ve výzkumech	28
9.	K čemu jsou dobré dotazníky	36
9.1	Výhody a nevýhody dotazníku	37
9.2	Vyvíjení dotazníku	38
9.3	Určení relevantních otázek	38
9.4	Pilotáž, testování dotazníku	39
9.4.1	Následné hodnocení rozhovorů	41
9.4.2	Behaviorální testování	41
9.4.3	Dodatečné zkoušení pochopení významu (probing)	45
9.4.5	Kognitivní rozhovory	47
9.4.6	Využití skupinových diskusí při hodnocení kvality otázek	49
9.5	Design dotazníku	51
10.	Chyby při tvorbě dotazníku	52
10.1	Uzavřené a otevřené otázky	54
10.2	Otázky směřující do minulosti	59
10.3	Senzitivní otázky	62
10.4	Pozitivně a negativně orientované otázky, opačné výroky	65
10.5	Vliv jedné otázky na jinou otázku	68
10.6	Zavedení nového konceptu do otázky	69
10.7	Obrácení pořadí výroku v otázce	69
10.8	Některé další zásady bezpečné práce s dotazníkem	70
11.	Chyby na straně tazatelů	73
11.1	Minimalizace a měření velikosti chyby tazatele	76
12.	Chyby non-response	78
12.1	Minimalizace a napravování efektu nr	79
13.	Závěr	81
14.	Resume	83
15.	Přílohy	84
15.1	Použité zdroje	84
15.2	Grafy, schémata, tabulky a dotazníky	87
15.3	Projekt diplomové práce	96

1. Abstrakt

Text se zabývá problémy, se kterými se lze setkat při plánování, realizaci a vyhodnocování marketingových průzkumů, dále průzkumů spotřebitelských preferencí, životního stylu nebo veřejného mínění. V textu se popisují postupy vedoucí k úspěšnému řešení takových situací a eliminující chyby výzkumu, ať již se jedná o chyby zaviněné či nezaviněné. Největší pozornost je věnována důležitosti preventivních opatření výskytu chyb. Text podtrhuje důležitost opakovaného testování dotazníků různými metodami a shrnuje dosavadní zkušenosti s prací s tímto nástrojem. Na metodologické analýze uskutečněných případových studií se ukazuje důležitost uplatňování opatření vedoucí k zajištění vyšší validity a reliability výzkumů. Práce na několika příkladech dokládá, že jediný způsob, jak vytvořit fungující soubor pravidel stavby otázek je jejich testování v různých obměnách, v různém pořadí, v různých kontextech a to až do doby, kdy budeme schopni s jistotou říci, která pravidla mají obecnou platnost a která jsou specifická pro různé situace.

2. Metoda zpracování a cíl textu

Při zpracování této diplomové práce vycházím především z dostupné odborné tematické literatury. V odborné literatuře, především zahraničního původu, lze najít závěry četných metodologických projektů, které se zabývaly vysvětlováním příčin poškození validity výsledků výzkumů. Cílem je provést smysluplnou syntézu závěrů již publikovaných výzkumných zpráv. Práci chci pojmout jako průřez průběhem výzkumu od fáze plánování a výběru metody, přes výběr výzkumného instrumentu, definování populace, sběru dat, dohledem nad terénními pracemi a vyhodnocením výzkumu konče. V každé fázi výzkumu se setkáváme s problémy, které je třeba zvažovat, kontrolovat a případně napravovat. Důraz tohoto textu je kladen především na moment, kdy problém ještě nenastal – čili na způsoby prevence vzniku a kontroly problémových situací.

3. Co jsou průzkum trhu a sociologický výzkum?

V roce 2004 v České republice různí zadavatelé za výzkum trhu a sociologické výzkumy utratili 64 milionů USD¹ [Esomar World Research Report. *Industry study on 2004*]. Výzkum trhu (nebo také marketingový výzkum) je dnes specializovaným oborem činnosti, která vyžaduje pokročilou míru znalostí hned z několika oblastí lidského vědění, ať už se jedná o sociologii, psychologii, matematiku, statistiku nebo ekonomii. Je těžké přesně zjistit, kdy můžeme poprvé hovořit o něčem takovém jako výzkum trhu, nicméně je pravděpodobné, že prapůvod činností s ním spojených lze hledat již v době rozvoje obchodu a směny. Již tehdy prodávající muselo zajímat, kdo má o co zájem, kolik prostředků nebo jakou kvalitu zboží je ochoten nabídnout protihodnotou, jaké jsou spotřební návyky populace žijící na určitém území apod. O tom, kdy lze nalézt počátky skutečně systematického marketingového výzkumu bychom mohli debatovat. Jeden zdroj² uvádí, že za první marketingový výzkum lze považovat šetření americké reklamní agentury N. W Ayer & Son z roku 1879 v oblasti produkce obilovin na národní úrovni. Smyslem této studie bylo získat podklady pro účinnou realizaci reklamní kampaně společnosti vyvíjející zemědělskou mechanizaci. Stejný zdroj uvádí, že za první řádný sociologický výzkum v dnešním smyslu slova lze považovat výzkum zorganizovaný Georgem Gallupem³, který v roce 1932 v USA uskutečnil průzkum veřejného mínění (odtud Gallup Poll).

Ale, co je to vlastně marketingový průzkum v dnešním smyslu slova? Zaměstnanci reklamní agentury N. W Ayer & Son v roce 1879 nebo George Gallup o 53 let později hledali informace. Informace o tom, jak má být konstruován výrobek, který bude co nejlépe odpovídat potřebám soukromých zemědělců na americkém venkově, nebo kdo s největší pravděpodobností vyhraje příští prezidentské volby. Dnes informace potřebují nejen výrobci těžké techniky a členové volebních týmů, ale také prodejní řetězce, banky, výrobní podniky, půjčovny automobilů, úředníci Evropská komise nebo zastupitelstva obcí. Všechny tyto subjekty si mohou objednat průzkum, protože chtějí zjistit, jak a kde nakupujeme, jaké produkty finančního trhu využíváme nebo plánujeme využívat, kde leží cenová hranice, kterou bychom zaplatili za vypůjčení konkrétního modelu osobního vozu z půjčovny, kolik % obyvatel ve věku 15 a více let v České republice je v tuto chvíli nakloněno přijetí jednotné měny Eura a tak bychom mohli pokračovat. Z toho, co bylo až doposud řečeno, by mělo být zřejmé, že zadavatelé výzkumů zjednodušeně řečeno vydávají prostředky za účelem získání

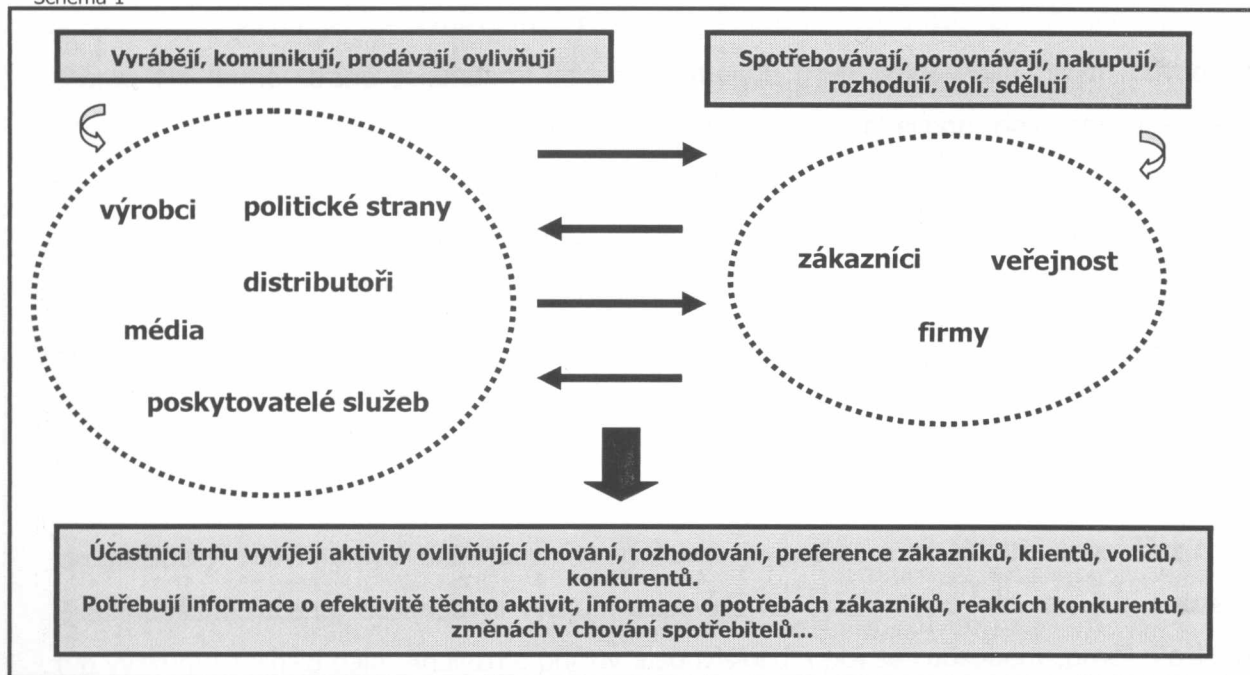
¹ nárůst 10% oproti hodnotě z roku 2003

² <http://www.exodusresearch.com/resources.htm>

³ Legenda mezi odborníky pohybujiícími se v odvětví výzkumu trhu George Gallup, zakladatel mezinárodní sítě výzkumných agentur Gallup. Po jeho smrti vede George H. Gallup International Institute jeho syn. Podle některých názorů znamenal G. Gallup pro odvětví výzkumu trhu tolik jako Leo Burnett (zakladatel sítě reklamních agentur Leo Burnett advertising) pro reklamní průmysl.

informací, které nemají, které si svépomocí nemohou opatřit a které potřebují pro své rozhodování a fungování. Následující schéma je pouze jedním z představitelných modelů pro zobrazení popisované situace.

Schéma 1



4. Základní druhy průzkumu trhu

Ačkoliv se marketingový průzkum jako vlastní tržní odvětví rozvíjí v plné míře teprve několik desetiletí, typologie jeho variant je členitá. První způsob dělení druhů marketingového výzkumu může být podle toho, od koho tyto informace získáváme. Z tohoto pohledu rozpoznáváme výzkumy *spotřebitelské* (někdy označovaného pouze jako B2C – Business to Customer), ve kterých jsou zdrojem informací domácnosti a jednotlivci. Spotřebitelské výzkumy se zaměřují především na nákupní zvyky a preference konečných uživatelů produktů. Předmětem výzkumu *business to business* (zkráceně B2B) jsou naopak právnické osoby – organizace, firmy. Zatímco v případě spotřebitelských výzkumů vypovídá respondent pouze za sebe (hospodyně za domácnost), v B2B výzkumech dotazujeme zástupce oslovovaných subjektů.

Odpovědí na otázku od koho chceme informace získat jsme ještě rozhodnutí o jejich původu nevyřešili zcela. Před rozhodnutím o tom, zda vstoupit do spolupráce s některou z agentur poskytující výzkumy a platit za služby zjišťování a analýzy informací, je na místě zamyslet se nad tím, zda požadované informace v použitelné formě již někde neexistují a zda nelze svépomocí zorganizovat to, čemu v marketingové terminologii říkáme *sekundární* výzkum. Sekundární výzkum spočívá ve sběru informací v již zpracované podobě. Tyto

informace nelze dále třídit, jsou standardizované a univerzálně použitelné. Může se jednat o různě podrobné statistiky, výroční zprávy, publikované výzkumné zprávy, informace z vnitřních zdrojů organizací, internetových stránek firem, informačních linek, atd. Společným jmenovatelem a výhodou všech sekundárních informací je především nízká cena a rychlost, s jakou je lze vzhledem k časové a finanční náročnosti *primárních* výzkumů sebrat. Na druhé straně sekundární informace mohou trpět „zanesením“ svým původním účelem, pro který byly určeny. Svým charakterem – všeobecnou dostupností – také nepřinášejí žádnou nebo jen malou konkurenční výhodu.

Zadavatelé výzkumů se rekrutují z různých sektorů trhu. Kategorie zboží nebo služeb, které tvoří meritum jejich podnikání, zároveň definují další dělení výzkumů. Kromě výzkumů v sektoru *telekomunikací* se dále můžeme setkat s výzkumy v oblastech *rychloobrátkového zboží (FMCG⁴)*, *financí*, *automotive* (výrobci automobilů, náhradních dílů, přepravci), *retailu* a dalších.

V jakémkoliv sektoru trhu může probíhat výzkum *deskriptivní*, *kauzální* nebo *prognostický*. Hlavní smysl deskriptivního výzkumu je popsat a změřit trh nebo část trhu. Zjednodušeně řečeno, deskriptivní výzkum odpovídá na otázky typu kdo, co, jak. Kauzální typ výzkumu zachází dále. Analyzuje příčiny a souvislosti a ptá se způsobem „proč“. Poslední z nastíněné triády typu výzkumů – *prognostický* – se pokouší co nejpřesněji předpovědět budoucí chování, předpovědět podobu klientova rozhodnutí, odhadnout jeho preference, predikovat míru úspěšnosti nového výrobku měřeno objemem prodeje, apod.

Stále složitější metody zabezpečení odpovídající kvality výzkumu vedou k nárůstu finanční náročnosti projektů. Není proto neobvyklé, že se o náklady na výzkum podělí několik zadavatelů najednou. Řeč je o *syndikovaném* výzkumem jehož tradičními zástupci jsou panelový výzkum a omnibus. *Panelový* výzkum je charakteristický svým způsobem výběru respondentů. Ti nejsou rekrutováni pro každou vlnu zvlášť, ale účastní se výzkumu pravidelně. Všechny panely trpí jistou mírou amortizace, odchodem části respondentů. Protože panel musí vždy splňovat proporce populace, kterou má reprezentovat, probíhá také doplňování respondentů podle přesně stanoveného klíče s přihlédnutím k případným kvótám. Důvodů, proč se klient výzkumné agentury rozhodne pro *omnibusový* výzkum, může být několik. Například snadná a především rychlá dostupnost univerzálních a obecně potřebných informací (trendy ve společnosti, socio-demografické trendy) relevantních pro všechny kategorie. Dotazován je reprezentativní vzorek populace v dané zemi. Omnibus může dobře posloužit pro odhad penetrace určitého jevu/vlastnosti/znaku v populaci, pro zjištění příjmů obyvatelstva, vybavenosti domácností, vzdělání různých sociálních skupin.

⁴ Zkratka Fast Moving Consumer Goods

Omnibus není vhodný využívat pro úlohy na cílových skupinách, které jsou čímkoliv specifické a odlišují se od obecné populace. Jistě, je zde možnost tzv. boostování, navýšení počtu rozhovorů v těchto zvláštních kategoriích, ale takový nástroj opět navyšuje cenu výzkumu. Nakonec výčtu nevýhod omnibusu jsem si nechal situaci, kdy se pro omnibusové šetření rozhodnou najednou dva zadavatelé z jiných zájmových skupin nebo sektorů trhu a jejich otázky lze těžko kombinovat v jedné vlně. Opakem syndikovaného výzkumu je výzkum *exklusivní*. Jak napovídá sám název, jde o výzkum prováděný výhradně pro jediného klienta, který je zároveň majitelem shromážděných dat a analýz.

Opakované výzkumy (retail audit, household panel, aj.) mají přinést efekt porovnání v čase, sledování vývoje trendů a slouží pro dlouhodobější strategické plánování spíše než pro okamžitá rozhodnutí. Okamžitá taktická rozhodnutí – se odehrávají na základě *ad-hoc* výzkumných projektů. Výzkumný tým při zadání pro ad-hoc projekt projektuje vždy nový design výzkumu, nastavuje vždy nově parametry studie a porovnává pro a proti použití různých výzkumných metod. Z hlediska frekvence opakování rozlišujeme ještě jeden typ výzkumů, tzv. *trackingy*. Tyto zásadně kontinuální výzkumy v přesných intervalech měří zadané cíle.

Považuji za nutné zdůraznit, že jednotlivé kategorie vzniklé při možných rozděleních typů výzkumů se mohou překrývat. Běžně se tak můžeme v praxi setkat se syndikovaným, popisným, spotřebitelským výzkumem, který se navíc každoročně opakuje.

Ne náhodou jsem si nechal zmínku o asi vůbec nejzákladnějším dělení typů výzkumů vůbec nakonec. Řeč je o dualitě kvalitativního a kvantitativního výzkumu. Ve své práci se budu zabývat explicitně kvantitativními výzkumy.

5. Etika a kvalita výzkumu

Pokud zacházíme s informacemi, máme povinnost řídit se profesními, morálními, ale i zákonnými mantinely práce s nimi. Informace získáváme od respondentů, kteří nám je vždy poskytují dobrovolně, ve svém volném čase a většinou bezúplatně. Ivor Shalofsky se ve svém článku *Quality and ethics in market in market research* [in Mc Donald, Vangelder, 1998:1105] zamýšlí nad pohnutkami, které k tomu respondenta mohou vést a nabízí pět možných vysvětlení.

1. respondent věří, že jeho odpovědi nějakým způsobem pomohou zlepšit kvalitu produktů nebo služeb tomu, kdo si výzkum objednal
2. respondenta zajímá téma výzkumu a rád se o něm bude bavit
3. respondent je potěšen, že se někdo zajímá zrovna o jeho názory

4. respondent považuje za samozřejmou součást civilizovaného chování odpovídat na otázky týkající se zajímavých produktů nebo služeb
5. respondent může předpokládat, že tazatel je pověřen nějakou vyšší autoritou k provedení rozhovoru a proto rozhovor neodmítne

Respondent čistě na základě vztahu důvěry poskytuje tazateli své odpovědi na mnohdy velmi osobní otázky (především v sociologických výzkumech). Výzkumné agentury, které jsou členy některého z existujících profesních sdružení (o kterých se zmíním dále) mu dávají slib, že s jeho daty bude zacházeno přísně důvěrně, že jeho odpovědi nebudou nikdy spojovány s jeho jménem ani že nebudou předány třetí osobě.

Kromě toho, že při své práci musí marketingoví pracovníci dbát na etiku své práce, nesmí přitom zapomínat na kvalitativní standardy, které se vyvíjely s tím, jak toto odvětví roste. Četl jsem několik textů, které se zabývaly pouze otázkou co lze považovat za kvalitní výzkum. Stále nejsem, stejně jako autoři těchto textů, schopný poskytnout jedinou správnou definici. Co má na kvalitu větší váhu? Je to správně provedený výběr respondentů, je to přísný a důsledný dohled nad terénním sběrem dat, je to bezchybně sestrojený dotazník? Nebo jsou to všechny dohromady? A co z nich má potom větší váhu? Když se dále v práci budu zabývat chybami měření při výzkumech, budu se snažit kvalitu výzkumů definovat opačně. Tedy zvýrazňováním toho, co kvalita není. Jedno ze specifík zaměstnání v průzkumu trhu je takové, že k němu nevede jen jedna cesta. Pionýři marketingových výzkumů mnohdy formální vzdělání neměli. Dnes již existují studijní obory, ve kterých se marketingový průzkum vyučuje, existují speciální výukové programy jako např. kurzy pořádané The Market Research Society (MRS) ve Velké Británii nebo Council of American Research Survey Organisation (CASRO) v USA. Kromě toho je možné se účastnit seminářů typu ADETEM (Francie), BVM (Německo), AISM (Itálie) nebo NVvM (Holandsko) [in Mc Donald, Vangelder, 1998:1106].

Dnes máme k dispozici důkladně zpracované směrnice pro práci s respondenty, tazateli, daty a výsledky výzkumů. Dohled nad jejich dodržováním byl svěřen účelově ustanoveným organizacím, které fungují na národní i mezinárodní úrovni. V České republice působí sdružení SIMAR. Jedná se o „výběrové neziskové sdružení předních agentur výzkumu trhu a veřejného mínění působících v České republice“ [<http://www.simar.cz/default.aspx>]. V současné době má SIMAR 13 členů⁵. Na internetových stránkách sdružení je možné nahlédnout do kvalitativních standardů týkající se doporučení pro práci tazatelů (včetně způsobů jejich školení, kontroly práce), dále doporučení pro procesy pořizování dat, metody

⁵ GfK Praha, Millward Brown CR, Tambor, Median, STEM/MARK, Factum Invenio, Incoma Research, Opinion Window Research International, IPSOS ČR, AVE Marketing, Confess Research, Network Media Service a XUXA

dotazování Mystery Shopping (utajené nakupování), telefonické dotazování atd. SIMAR ještě před několika lety prosazoval kontroly práce členských agentur navzájem. V roce 2003 ale došlo ke změně a je spekulativní tvrdit, že se tak stalo kvůli odchodu agentury TNS AISA (tehdy NFO AISA), která je podle deklarovaného obratu největší na českém trhu marketingového výzkumu. Nyní již SIMARovské agentury vzájemnou kontrolu činnosti neprovádí. Zavázaly se, že budou na dodržování základních etických a kvalitativních principů práce dohlížet vlastními silami. SIMAR je členem jiné, panevropské profesní skupiny, která sdružuje právě národní asociace. Řeč je o EFAMRu. European Federation of Market Research Organisations byla založena v roce 1992 pro potřeby unifikace kvalitativních standardů uplatňovaných na úrovni jednotlivých zemí.

Počátky vývoje unifikovaných kvalitativních standardů práce v marketingovém výzkumu se datují do 70. let minulého století, kdy ve Velké Británii vznikl tzv. Interviewer Quality Control Scheme (IQCS). IQCS byl inovativní v tom, že zohlednil rozličné požadavky kladené na práci s tazateli. V té době se o jeho šíření a propagaci staral British Standards Institute. Až do roku 1998. Tehdy předsednictvo EFAMRa rozhodlo o tom, že je standardy třeba ještě více unifikovat. Výsledkem harmonizace různých dostupných předpisů je kodex EMQRS. Nejnovějším přírůstkem do rodiny těchto oficiálních opatření je osvědčení ISO. Přesněji certifikát ISO 9001. Z vlastní zkušenosti mohu říci, že spíše než souhrnem doporučení pro dodržení kvality výzkumů je toto osvědčení vydávané Mezinárodní organizací pro standardizaci⁶ indikátorem, zda cerifikovaná společnost postupuje ve svých operacích standardizovaným způsobem, zda je možné zpětně dohledat výsledky práce zaměstnanců na všech úrovních řízení, za jakých podmínek probíhá komunikace s klienty, jestli je bezpečně archivována projektová dokumentace, korespondence, kontrakty atp. K podpoře tvrzení, že certifikáty ISO (pravidelně auditované) samy o sobě nemohou zabezpečit skutečnou kvalitu výzkumů připomínám, že stejné osvědčení, které v České republice zatím získalo několik výzkumných agentur, drží také společnosti fungující v České republice v jiných oborech podnikání⁷.

V současné době v radě EFAMRa zasedají představitelé asociací třinácti evropských zemí⁸. EFAMRO, na rozdíl od nejvýznamější celosvětové organizace pro dohled nad dodržováním zásad práce marketingových agentur ESOMARu, založeného v roce 1948, se zabývá legislativním pozadím průzkumů. ESOMAR dohlíží na metodologii výzkumů a plní funkci hlasatele nejnovějších poznatků a trendů.

⁶ <http://www.iso.org/iso/en/iso9000-14000/index.html>

⁷ <http://www.iso.cz/hledat.asp>

⁸ Belgie, Itálie, Francie, Česká republika, Německo, Švédsko, Holandsko, Polsko, Španělsko, Portugalsko, Velká Británie, Turecko a Maďarsko

6. Plánování výzkumných projektů

Když výzkumný projekt nesplní očekávání, můžeme hledat příčinu hned na jeho začátku – ve fázi plánování projektu a zejména v porozumění a definici cílů výzkumu. Design výzkumu a zejména stanovení cílů výzkumu je kritická část projektu. Je-li v této fázi cokoliv špatně, ani kvalitní sběr dat a bezchybné zpracování nepřinese žádný výsledek. Problém bude nejen s vágně definovaným cílem výzkumu, ale na druhé straně se zadáním, který je příliš komplexní. Pokud se setkáme se zadáním, které chce najednou vyřešit více nesourodých cílů, je namístě mít na zřeteli technický průběh projektu. Není možné zadavateli potvrdit provedení studie a v jeho průběhu zjistit, že výzkumný nástroj není schopen obsáhnout takto široce pojatý problém nebo při vyhodnocování zjišťovat, že se respondenti v problematice projektu nedokázali orientovat a na otázky odpovídali způsobem, který nám nepomůže nebo neodpovídali vůbec. Cíle výzkumu se musí odrazit ve znění otázek výzkumného nástroje a nezáleží na tom, zda jdeme dotazovat s tištěnými dotazníky, zda respondentům telefonujeme nebo zda s nimi vedeme intenzivní hloubkové rozhovory v kvalitativním výzkumu. Zjednodušeně můžeme říci, že každá otázka, která nevede k dílčímu porozumění řešeného problému, nemá v dotazníku nebo scénáři rozhovoru místo. Nejenže se irelevantními otázkami prodlužuje čas dotazování se všemi negativními důsledky, které to s sebou přináší, navíc ale plýtváme respondentovou ostražitostí klesající s postupujícím časem dotazování. Ne náhodou bývají otázky, na kterých nám nejvíce záleží řazeny do přednějších sekcí dotazníku, samozřejmě za podmínky dodržení logiky a tématické smysluplnosti dotazníku.

Důležitosti detailně vymezeného výzkumného problému a potřeby stanovení hypotéz bych ilustroval příkladem z praxe. Nejmenovaná agentura byla požádána o kontinuální měření kvality služeb poskytovaných zákazníkům společnosti provozující síť čerpacích stanic v České republice. Tato společnost provozuje čerpací stanice v několika dalších evropských zemích a můžeme říci, že je významným hráčem na petrochemickém poli. Zadavatel výzkumu, kterým je ústředí této mezinárodní firmy, se rozhodl pro financování trackingu⁹. V zadání k výzkumu se zaměstnanci výzkumné agentury mohli dočíst, že studie má za cíl změřit výkonnost provozovatelů sítí na rovině jednotlivých zemí a za druhé získat tímto mezinárodní srovnání. Design výzkumu zachází tak daleko, že zahrnuje kontrolní testování části poboček konkurenčního provozovatele. Předmětů hodnocení je několik. Výzkumníci začínají sledováním čistoty příslušenství čerpací stanice, hodnotí se ale také ochota personálu, jeho vystupování, způsob oslovení a obsluhy zákazníků, pokračujeme způsobem

⁹ trackingem se označují longitudinální, opakované studie měření určeného subjektu

prezentování výrobků na celé prodejní ploše obchodu čerpací stanice a končíme kontrolou funkčnosti dalšího příslušenství stanice jako myčka, restaurace, prostory pro vysavače nečistot atd. Hodnotící kritéria jsou definována naprosto jednoznačně jako série výroků, tazatelé procházejí před každou vlnou důkladným školením vedeným samotným klientem a kontinuita optiky hodnocení je zaručena opakovaným zapojováním identických tazatelů. Navíc v každé vlně s hodnocením přibližně 200 poboček pracujeme s maximálně dvěma až třemi tazateli. Jestliže v tradičních F2F výzkumech usilujeme o to, aby jeden tazatel prováděl co nejméně standardizovaných rozhovorů¹⁰, je logika metodologie Mystery Shopping opačná. Zvláště ve srovnávacích výzkumech usilujeme o co největší uniformitu hodnocení¹¹.

Kamenem úrazu jinak dobře připraveného výzkumu je skutečnost, že české ústředí informuje jednotlivé oblastní zástupce o tom, že v inkriminovaném období probíhá hodnocení čerpacích stanic. Personál čerpacích stanic, stejně tak jako oblastní zástupci, jsou do určité míry odměňováni také na základě toho, jak dopadne hodnocení jimi poskytovaných služeb. Z rozhovorů s tazateli i z vlastní zkušenosti mohou říci, že případy nápadné servilnosti a úklidu příslušenství stanic v časech, kdy to od personálu pracovní směrnice rozhodně nevyžadují, nejsou výjimkou. Nicméně hodnocení stanice proběhne i v těchto případech. Svým prozrazením si nemůžeme být jisti, dokud se nás personál přímo nezeptá na skutečný důvod naší návštěvy, což se s největší pravděpodobností nestane. Je přinejmenším pravděpodobné, že chování personálu stanice je v tento moment posunuté. Otázkou není jakým směrem, ale do jaké míry. Lze namítnout, že zaměstnavatel má zákonnou povinnost informovat zaměstnance o tom, že po čas výkonu práce mohou být předmětem hodnocení. Nikde ale není řečeno, že zaměstnavatel musí specifikovat dobu, ve kterou se tomu děje. Skutečný důvod takového nastavení podmínek výzkumu je nasnadě. Jestliže jedním z cílů studie poptávané a financované evropskou centrálou zadavatele je mezinárodní srovnání výkonnosti a standardu provozoven společnosti na úrovni jednotlivých zemí, chovají se čeští manažeři při informování oblastních zástupců ze svého pohledu naprosto racionálně.

Ze zmíněného příkladu je možné vyvodit několik závěrů. Ukázal jsem, že cíle studie jsou závislé na exaktní formulaci výzkumného problému, že jedna studie může sledovat několik různých cílů (v tomto případě výlučně první nebo druhý) a že i při sebelepším definování

¹⁰ z důvodu minimalizace ovlivnění výsledků případným výskytem dotazníků vyplněných tazatelem samotným. Moser s Kalotonem (1971:283-284) jako další důvody pro zapojení vyššího počtu tazatelů při sociologických výzkumech uvádějí také lepší výsledky výzkumu při výběrech respondentů z co nejšířší sítě výchozích výběrových bodů (sapling points). Dále je při vyšším počtu pracujících tazatelů možno dosáhnout nižší míry odmítnutí rozhovorů. Tazatelé pracují v prostředí, které znají a tím, že nemusí za respondenty cestovat, mohou ušetřený čas využít pro více pokusů k oslovení vybraného respondenta (včetně možnosti nabídnout respondentovi schůzku v termínu, který mu vyhovuje v delším časovém období). Možnost provést výzkum v kratším období při zapojení vyššího počtu tazatelů je banálně samozřejmá, ale skutečná výhoda.

¹¹ Každý z nás má mírně odlišné představy o tom, jak vypadá zdvořilé přivítání, jak vypadají čerstvé pečivo, jak jsou cítit pečlivě uklizené toalety atd.

výzkumného problému a měřitelných hodnotících parametrů mohou být výsledky studie vychýlené od výsledku dosažitelného za standardních podmínek. V posledním případě záměrným ovlivněním objektivních podmínek studovaného jevu.

Každý výzkum má své respondenty. V každém šetření se obracíme na určitý počet lidí, kteří splňují určité charakteristiky a od kterých chceme získat odpovědi na otázky, které nás zajímají. Relevantní skupinu respondentů¹² se vymezuje dvěma základními způsoby – geograficky a demograficky. Vždy se rozhodujeme o tom, zda dotazujeme celou populaci nebo jen její část. V takovém případě tedy mluvíme o výběru a dostáváme se do oblasti teorie samplování. Ta se zabývá samotným způsobem výběru respondentů, volbou jednotky vzorku (jednotliví respondenti, domácnost, obec, ...), oporou výběru (proti jakému celku vzorek vybíráme), jak velký vzorek musí být aby dobře odrážel charakteristiky cílové populace (jednotky, na kterou lze vztáhnout výsledky výzkumu). Problematika výběrů je velmi obsáhlé téma, které nelze obsáhnout stručně a není tématem mé práce. O samplování bylo publikováno nepřehledné množství literatury¹³. V textu se výběrem respondentů dále nabudeme zabývat. S jednou výjimkou. Je jí fenomén „non-response“, dílčí téma okruhu samplování zasahující do tematiky napravování nedokonalostí dotazníků, které je věnovaná část práce.

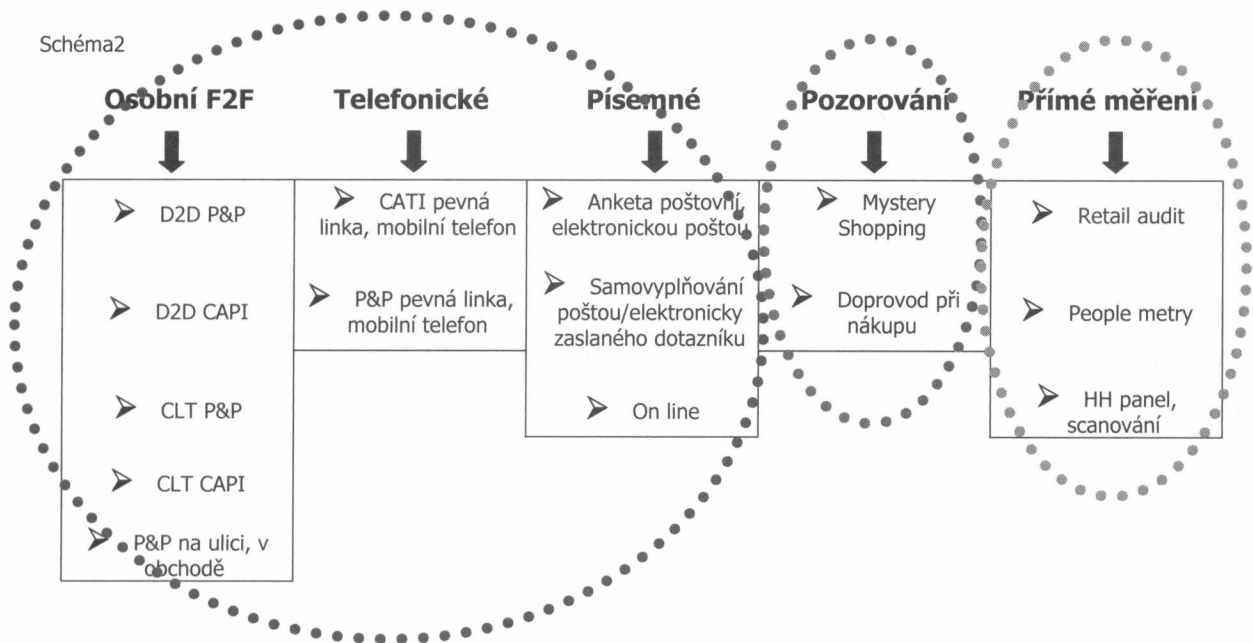
7. Metody výzkumu

Ve výzkumu trhu stejně jako v sociologických studiích máme k dispozici několik metod s jakými můžeme přistoupit k respondentům a žádat po nich odpovědi nebo jiným způsobem získávat podklady pro zodpovězení cílů výzkumu. Volba konkrétní metody se odvíjí od předmětu studie, dotazované jednotce a rozsahu výzkumu. Schéma 2 na následující straně je přehledem známých metod výzkumu, se kterými se můžeme setkat. Představuje několik větších kategorií metod, kde první z nich jsou metody dotazování.

¹² respondent od slova respond=odpovídat

¹³ Cochran, W. G. (1963) *Sampling Techniques*. New York: Wiley and Sons, Inc.; Deming, William E. (1950) *Some Theory of Sampling*. New York: John Wiley and Sons, Inc.; Kish, Leslie. (1965) *Survey Sampling*. New York: John Wiley and Sons, Inc.; Parten, M. (1950). *Surveys, Polls, and Samples: Practical Procedures*. New York: Harper and Brothers. a další zdroje

Schéma2



Metody dotazování se dále člení podle volby technického provedení studie a podle místa provedení výzkumu. Zkratka D2D¹⁴ symbolizuje obecně výzkumy, při kterých tazatelé provádějí rozhovory v domácnostech respondentů nebo vyhledají firemního respondenta v místě jeho pracoviště. Zkratka P&P¹⁵ symbolizuje výzkumy, při kterých tazatelé zaznamenávají odpovědi do papírového dotazníku. CAPI¹⁶ je zkratka pro počítačové polo-automatizované systémy osobního dotazování s využitím počítače, ať už jde o přenosné laptopy nebo palmtopy – počítače do dlaně. CLT¹⁷ znamená dotazování koncentrované v jednom místě, typicky studiu. Jak je patrné z prvního sloupce tabulky, uvedené technologie lze nalézt i v kombinacích.

Další metoda dotazování je s využitím telefonu a to buď pomocí programů vytáčejících čísla respondentů z databází/seznamů nebo za asistence generátorů náhodných telefonních čísel. Tazatelé odpovědi zapisují do předem naprogramovaných skriptů – elektronických dotazníků uložených v paměti stolního počítače tazatele. Opět z vlastní zkušenosti mohu potvrdit, že stále ještě probíhají výzkumy, ve kterých tazatel volí telefonní číslo respondenta ručně a stejně tak zapisuje ručně do papírového dotazníku jeho odpovědi (varianta P&P pevná linka, mobilní telefon).

Posledním typem je písemné dotazování. Děje se tak jednak starší metodou zaslání vytištěného dotazníku respondentovi k vyplnění klasickou poštou a od 90. let minulého

¹⁴ Door To Door

¹⁵ Pen and Paper

¹⁶ Computer Assisted/Aided Personal Interviewing

¹⁷ Central Location test

století také zasíláním elektronických dotazníků na e-mailové soukromé účty. V případě on-line výzkumu agentura dotazník nerozesílá každému respondentovi jednotlivě, nýbrž umístí dotazník na fixním místě v internetové síti a respondenti se k němu připojují. Anketa se používá například při hodnocení prodejen a poboček. Bezpochyby jde o metodu ekonomickou. Anketní lístky (jednodušeji koncipované dotazníky) jsou k dispozici na viditelném místě. Zákazníci jsou vyzváni, aby lístek vyplnili. Je zcela zřejmé, že míra odpovědí je zde nižší. Existuje předpoklad, že anketní lístek vyplní skupina zákazníků, která se značně liší od průměrného cílového respondenta. Pro aktivní účast na průzkumu totiž musí mít respondent určitý důvod. To je v pořádku, pokud nám jde například pouze o shromáždění možných negativních zkušeností s nově zavedeným výrobkem. Chceme-li však například objektivně měřit spokojenost zákazníků, je tato metoda neuspokojivá. Anketa dosahuje lepších výsledků, pokud jsou anketní lístky rozdávány zaměstnancem společnosti. Ten je také k dispozici, pokud některý ze zákazníků potřebuje s vyplněním lístku pomoci. Příklad výzkumu Mystery Shoppingu jako typu pozorování jsem uváděl v minulé kapitole. Mimo utajených nakupování, které vyžadují fyzickou přítomnost tazatele, do rodiny výzkumů založených na pozorování spadají ještě výzkumy Mystery Calls. Cílem utajených telefonátů může být příkladně hodnocení chování personálu na zákaznických linkách. Doprovod při nákupu zvolí plánovači výzkumu v případě, kdy je zajímavá specifické nákupní chování spotřebitele. Respondent-spotřebitel na skutečnost, že je sledován, může a nemusí být upozorněn. Výhoda první varianty spočívá v opakované možnosti položení doplňující otázky osvětlující motiv pro konkrétní volbu mezi nabízenými možnostmi. Na rozdíl od druhé možnosti, kdy nakupující(ho) pozorujeme utajeně, musíme počítat s možným zkreslením skutečného nákupního chování. Nakupující hospodyně si pravděpodobně uvědomuje, že je sledována a častěji proto vloží v případě nerozhodnosti do košíku výrobek prémiové značky za vyšší cenu, který by si za normálních okolností nekoupila.

Všechny metody přímého měření se vyznačují tím, že nevyužívají respondentů v takovém smyslu jako metody dotazování nebo pozorování. Jádrem Retail auditu spočívá ve velmi podrobném monitoringu prodejů vybraného vzorku prodejen s následnou extrapolací na celou skupinu maloobchodních jednotek. To vše pro účely odhadu tržních podílů výrobců. Peplemetrový výzkum se také obejde bez klasických tazatelů. V současné době má přibližně 1200 domácností k dispozici aparát, který dokáže rozpoznat bez vnějšího zásahu jaká stanice je v určitý moment naladěna na televizním přijímači. Data pravidelně zpracovává a odesílá k centrálnímu zpracování. Zadavatelé reklamy teoreticky již druhý den po odvysílání spotu zařazeného do vysílání mohou mít dobrý odhad o tom kolik procent populace reklamu shlédlo. V České republice ale probíhá také jiný druh měření. Jedná se o

monitoring poslechovosti rozhlasových stanic, čtenosti deníků a sledovanosti televizních stanic. Jediným a největším projektem tohoto druhu na našem území je výzkum Mediaprojekt. Na jeho realizaci se podílí dvě výzkumné agentury – GFK Praha a Median. Ty každý rok klasickým F2F způsobem osloví kolem 30.000 respondentů a zjišťují, jakou rozhlasovou stanicí minulý den poslouchali a nebo jaké noviny zrovna četli.

Každá z metod výzkumu má své výhody i nevýhody a na ty se, pokud se týče metod dotazování, nyní zaměřím¹⁸. Je dobré mít přehled o limitech každé z nich. Na konci kapitoly se zmíním ještě o speciálním případě designu výzkumu využívající kombinaci několika metod.

7.1 Výhody a nevýhody - osobní dotazování

Osobní dotazování spočívá v přímém kontaktu respondenta s tazatelem. Blankenship a Breen se domnívají, že ve 30. letech minulého století mělo osobní dotazování mezi ostatními metodami výsadní postavení. Alespoň co se týče osobního dotazování v domácnostech. V té době byly ženy zaměstnané jen zřídka. Bylo tím zajištěno, že když tazatel navštívil náhodně vybranou domácnost, většinou zde někoho zastihl. Tehdejší úroveň vybavení domácnosti telefonem v USA dosahovala přibližně padesáti procent. Metodou telefonického dotazování nebylo možné zajistit reprezentativnost výzkumu. A výzkumy poštou v té době dosahovaly návratnosti přibližně patnácti procent [Blankenship et al, 1992: 130].

Dnes je osobní dotazování v jiné situaci. Podle údajů Českého statistického úřadu bylo v roce 2002 vybaveno pevnou telefonní linkou 63,4% domácností a na sto domácností připadalo 99,5 jednotek mobilních telefonů¹⁹. Dovolím si tvrdit, že vlivem politiky tehdejšího dominantního poskytovatele telefonních služeb pevné linky je dnes vybavenost tímto zařízením v domácnostech nižší, na rozdíl od penetrace mobilních telefonů. Podíl SIM karet na jednoho obyvatele v ČR překročil na konci prvního pololetí roku 2005 hranici 106%²⁰. Telefonická dosažitelnost respondentů ve vybraných věkových kategoriích již není problémem a na poli výzkumů zasílaných poštou se již také velmi pokročilo co do znalostí cest ke zvýšení návratnosti. Dobře provedené výzkumy poštou se mohou pochlubit návratností v rozmezí sedmdesáti až odmdesáti procent při aplikaci průvodních dopisů a vln upomínek.

¹⁸ metody pozorování a měření nejsou předmětem práce a byly uvedeny pro úplnost

¹⁹ [http://www.czso.cz/csu/edicniplan.nsf/t/7E002FD06F/\\$File/tab10.pdf](http://www.czso.cz/csu/edicniplan.nsf/t/7E002FD06F/$File/tab10.pdf)

²⁰ http://www.telecom.cz/infocentrum/tiskove_centrum

Osobní rozhovory mají svou nezastupitelnou úlohu když potřebujeme, aby respondent mohl pracovat s nápovědou (karty s možnostmi odpovědí, fotografie, koncepty), aby hodnotil cokoliv na základě vizuálního nebo chuťového vjemu, když hledáme velmi specifickou populaci, kterou bychom při využití jiných metod dotazování nezasáhli (nebo nezasáhli efektivně), např. když potřebujeme uskutečnit rozhovor se skupinou klientů konkrétního, méně významného bankovního domu nebo když potřebujeme mluvit s návštěvníky určité prodejny atd. Pro osobní dotazování se rozhodneme, když vidíme, že délka rozhovoru pravděpodobně překročí dobu 20 – 30 minut. Snadněji přemluvíme unaveného respondenta k dalším 10-ti minutám rozhovoru při osobním styku než po telefonu. Skutečnost, že tazatel je v osobním kontaktu s respondentem, mu umožňuje okamžitě reagovat na změny v chování respondenta. Bystrý tazatel zvládne kontrolovat nonverbální projevy komunikace respondenta. Může měnit rychlost, s jakou pokládá otázky; v případě, že u respondenta rozpozná únavu z opakujícího se nabízení škál odpovědí, může se operativně rozhodnout frekvenci nabídky snížit. Může dotazování kdykoliv pozastavit a domluvit se v pokračování v rozhovoru i v jiný den. V tomto ohledu je osobní dotazování velmi flexibilní. Tazatel má plnou vizuální kontrolu nad dotazováním. Nejenže může kontrolovat, že na otázky odpovídá vybraný respondent, navíc ale může v případě rozhovorů prováděných v domácnosti respondenta provést nezávislé zhodnocení socio-ekonomické kategorizace na základě míry vybavení domácnosti. Dobře vyškolení a poctiví tazatelé zaručí, že otázky budou zodpovídaný v přesně stanoveném pořadí, což nemůžeme kontrolovat u dotazníků zasílaných poštou. V neposlední řadě můžeme v regionech s lokálními specifiky využít ve svůj prospěch znalosti prostředí místních tazatelů, jak potvrzuje Ward [in Mc Donald a Vangelder, 1998:151] a má osobní zkušenost. Před časem jsem měl jako člen koordinačního týmu příležitost zadávat výzkum B2B na Kypru. Cílem deskriptivní mezinárodní studie bylo získat co nejpodrobnější informace o rozšíření využívání technologií e-Business. I když byl výzkum původně koncipován pro telefonické dotazování, přesvědčili nás kolegové z Kyperské agentury o nevhodnosti nasazení této metody v místních podmínkách. Když jsme byli varováni, že Kyperští podnikatelé se na toto téma po telefonu s tazateli bavit nebudou a navíc že budou vyžadovat sjednání schůzky k rozhovoru v předstihu, rozhodli jsme se to respektovat a vyhovět jejich argumentům pro osobní dotazování.

Ward [in Mc Donald a Vangelder, 1998:172-175] uvádí další příklady národních specifíků výzkumů, například nutností počítat s množstvím jazykových dialektů v zemích Jižní Afriky, nemožností provádět B2B CATI výzkumy v Japonsku plynoucí z prostorového uspořádání kanceláří, potřebou získat souhlas manžela Indické ženy pokud s ní chceme

provádět rozhovor nebo neochotou jihoamerických tazatelů dotazovat ve večerních a nočních hodinách kvůli strachu z násilí.

Hlavním negativem osobního dotazování je jeho finanční náročnost. Cena za rozhovor osobního dotazování je určena jednak složitostí vyhledání dotazované populace a délkou rozhovoru. V případě delších rozhovorů je dnes již slušností kompenzovat respondentův čas buď finanční nebo jinou odměnou právě v závislosti na typu populace. Tisíc korun má pro běžného respondenta jinou hodnotu než pro majitele společnosti s obratem desítek milionů korun. Tazatele platíme nejen za uskutečněné rozhovory, ale musíme myslet na to, že vyhledávání a přesun mezi místy dotazování se děje v reálném čase. V jejich pracovním čase. Náklady na osobní dotazování tedy navyšují také kompenzace cestovních nákladů nebo poplatků za poštovní služby – meziposlání a odeslání vyplněných dotazníků. Dále, jakmile jsou dotazníky již jednou vytištěné a na cestě k tazatelům, máme již jen velmi malý manévrovací prostor ke změnám v materiálech, ať jde o dotazník samotný nebo karty. Hrozbou dat pořízených osobním dotazováním je samozřejmě také jejich věrohodnost. Mezi pozitivy osobního dotazování jsem jmenoval plnou vizuální kontrolu tazatele nad respondentem, což je současně dvojsečná charakteristika. Tím, že se tazatelé pohybují v terénu, nemůže nad nimi mít plnou kontrolu výzkumný tým. Nemůžeme kontrolovat, jak svědomitě tazatelé dotazníky vyplňují. Navíc máme vždy co do činění s dalšími vlivy tazatelů, například s problémem vlivu tazatelů, kdy stejně znějící otázka položená dvěma tazateli může dosáhnout jiných odpovědí. Ale to bych předbíhal.

Na praktickém příkladu z vlastní praxe demonstruji hned několik výše zmíněných negativ osobního dotazování. Již zmíněný projekt EuroBarometr probíhá také na Slovensku, koordinačním centrem pro obě země je Praha. Průměrná délka dotazníku v minutách činí 40 minut. Je naprosto nevhodné dotazovat tento výzkum telefonicky, kvůli své délce a nelze uvažovat ani o dotazování poštou, protože dotazník je velmi složitý. Pravidelně se zde objevuje množství odskoků a filtrů. V letošním roce proběhlo již 8 vln výzkumů, v každé vlně jsme se zavázali získat odpovědi od 1000 respondentů. Metodologií výběru je náhodná procházka s oslovením toho respondenta v domácnosti na zadané adrese, který v domácnosti oslavil poslední narozeniny. Aby byly uspokojeny nároky na studii z hlediska výběru, musí tazatel před odchodem k jiné adrese provést 4 návštěvy-pokusy uskutečnit s respondentem rozhovor. To prodlužuje čas terénního sběru dat a způsobuje navýšení cestovních nákladů tazatelům pokud nedotazují přímo v místě svého bydliště. I když ke každé vlně existuje velmi podrobný harmogram dílčích činností, minimálně v polovině letošních vln jsme terénní fázi výzkumu museli oddalovat kvůli finalizaci a schválení dotazníku zadavatelem. Tazatelé jsou ale dopředu instruováni, aby s plánovaným objemem

práce počítali a mají na to vyhrazen svůj čas. Pokud na výzkumu nemohou začít pracovat v předem určený den, jsou nevyužití, protože operativní přidělení jiného výzkumu by trvalo déle než obvyklé 1-2 dny neplánovaného zdržení. Pokud by tito tazatelé pracovali jako operátoři CATI, mohli by projít několika hodinovým školením k jinému výzkumu a odpoledne již provádět výzkum.

I když poslední vlna začala přesně podle harmonogramu, vyskytl se jiný problém. Třetí den výzkumu ohlásil jeden z tazatelů, že v dotazníku našel pod jiným označením dvě otázky stejného znění. Jednalo se o chybu formátování dotazníku, která musela být opravena. Oblastní zástupci ještě tentýž den kontaktovali telefonicky přes 200 slovenských tazatelů zapojených do výzkumu a na chybu je upozornili. I před okamžitou reakcí bylo v tu chvíli již přibližně 300 dotazníků vyplněných, bez možnosti opravy.

Během terénní fáze výzkumu probíhá zpravidla 2-3 krát za vlnu nákladný telefonický monitoring počtu uskutečněných rozhovorů. V České republice probíhá dotazování F2F, ale s využitím CAPI, takže máme aktuální data k dispozici každý den a stejně tak můžeme sledovat plnění počtu rozhovorů v jednotlivých krajích, podle velikosti místa bydliště, pohlaví a věku respondenta. Tak můžeme předcházet situacím nedosáhnutí stanoveného počtu rozhovorů nebo případu výrazného narušení rozložení rozhovorů podle regionů atp. Po ukončení terénu musí v obou zemích proběhnout kontrola, zda tazatelé na uvedených adresách respondentů v uvedený čas a datum rozhovor skutečně provedli. Každá agentura si rozsah kontrol práce svých tazatelů stanovuje sama. Předpisem pro studii EuroBarometr je kontrola 100% sebraných rozhovorů obesláním korespondenčními lístky a čekáním na jejich potvrzení. Takové opatření rozhodně prospívá k důvěryhodnosti studie, na druhé straně ji dále prodražuje. Pokud bychom výzkum prováděli na CATI, mohli bychom tazatele čistě teoreticky kontrolovat ve 100% případů ještě spolehlivějším způsobem náslechnů, nemuseli bychom respondenty obtěžovat vyplňováním konfirmace a provádět další kontroly dat – například testy podobnosti odpovědí stejných tazatelů nebo zkoumání mezičasů průběhu rozhovorů vedených českými tazateli.

7.2 Výhody a nevýhody - telefonické dotazování

Ve Spojených státech je dnes dotazování s využitím telefonní sítě – CATI – nejvyužívanějším cestou oslovení respondentů. 96% domácností je zde vybavena telefonní linkou²¹. Zda je CATI nejfrekventovanější metodou dotazování také v České republice si netroufám tvrdit. Každopádně je díky téměř 100% vybavenosti českých firem telefonní linkou a vysokému

²¹ <http://www.surveysystems.com/online.html>

podílu vybavenosti obecné populace pevnou nebo mobilní telefonní linkou hojně využívána. Když mluvím o telefonických výzkumech, mám na mysli telefonické výzkumy s využitím počítačů. Již dříve jsem se zmínil o telefonických výzkumech se zaznamenáváním odpovědí do papírového dotazníku, ale ty jsou dnes spíše výjimečné a dále se jim nebudeme věnovat.

Stejně tak jako osobní dotazování, mají i telefonické výzkumy své klady a zápory. CATI výzkum je postaven na naprogramovaném dotazníku, který se tazateli objevuje na obrazovce. Změny v dotazníku lze provést prakticky kdykoliv během terénní fáze. Několikrát jsem byl svědkem situace, kdy zadavatel výzkumu požadoval operativní změnu v dotazníku během prvního dne dotazování, když zjistí, že by uvítal jiné znění otázky, když má přání rozšířit okruh možností odpovědí, když si přeje doplnit nebo naopak odstranit otázku nebo pokyn tazatelům atp. Odpovědi respondenta tazatel zapisuje do počítače. Dobře naprogramovaný skript dokáže během rozhovoru v pozadí provádět výpočty, pracovat s informacemi o respondentovi známými z databáze nebo nově zjištěnými během rozhovoru a s těmito informacemi metodou dosazování údajů nebo filtrování irelevantních otázek v průběhu rozhovoru dále pracovat. Tato metoda je levná²² a rychlá. Její rychlost spočívá ve velkém množství rozhovorů, daném velmi krátkou dobou nutnou na rekrutaci respondenta. Především však v tom, že v okamžiku dokončení rozhovoru jsou již data v elektronické formě v podstatě k dispozici (pomineme-li kódování otevřených otázek). Dalším, velmi důležitým aspektem, je její vysoká kvalita. Tazatelé totiž pracují přímo v agentuře a jsou pod neustálým dohledem. Kontrola rozhovorů probíhá tak, že vedoucí pracovník se napojuje na jednotlivé tazatele, odposlouchává jejich rozhovory a současně vidí na svém monitoru, jak zapisují odpovědi. Přirozeně, tazatelé nemají žádnou možnost zjistit, zda jsou v konkrétním okamžiku odposloucháváni. CATI výzkum můžeme realizovat na základě databáze zadavatele výzkumu nebo na základě již zmiňovaného generátoru telefonních čísel. Touto cestou je zajišťován náhodný výběr. Snadno lze vyřešit situace, kdy se respondentovi rozhovor momentálně nehodí. Tazatel s respondentem dohodne schůzku a automatizovaný systém ve stanovený termín respondenta automaticky kontaktuje. Dětskou nemocí některých CATI pracovišť je nemožnost zaručit, aby respondentovi volal přesně stejný tazatel, který s ním schůzku domlouval.

Někteří metodologové argumentují, že metoda CATI je nevhodná v případě, když chceme respondenta za jeho čas odměnit. Ve skutečnosti, chceme-li respondentovi za rozhovor dát dárek, nezpůsobuje telefonická metoda žádné problémy. Dárek je prostě později poslán poštou anebo, jde-li o peníze, mohou být převedeny na respondentův účet.

²² Metoda je levná relativně. Při vysokém % odmítnutých rozhovorů, špatně dostupné cílové skupině, zastaralé nebo špatně postavené databázi se telefonický výzkum může velice prodražit. Jsme nuceni vytáčet více telefonních čísel a strávíme více času na telefonních linkách.

Na konec výčtu pozitiv telefonického dotazování jsem si nechal prakticky neomezený dosah této metody. Abych uvedl příklad, je pouze na rozhodnutí výzkumného týmu, kterého zajímá mezinárodní srovnání odpovědí na výzkumné otázky, zda výzkumem pověří tamní agenturu nebo zda najme cizojazyčné tazatele a výzkum provede z centrálního studia CATI vlastní země.

Limitem pro využití této metody je délka rozhovoru a nemožnost použít vizuální pomůcky. K negativům lze zařadit také naprostá absence kontroly samotného respondenta. Nejen, že si nemůžeme být nikdy jistí, zda rozhovor provádíme s tím, s kým si myslíme, ale nevidíme, jak respondent fyzicky reaguje na tazatelské otázky, jaké emoce při odpovědích projevuje. Na druhou stranu právě CATI dotazování může slavit úspěch v případech dotazování senzitivních témat, na které by se respondenti mohli tvářit v tvář tazateli stydět odpovídat.

Spíše pro zajímavost bych varoval před možným zneužitím metody CATI. Dobrého jména agentury mohou jiné subjekty zneužít pro své telemarketingové aktivity. Ale to spíše než s metodologickými úskalími techniky souvisí s etikou podnikání.

7.3 Výhody a nevýhody – dotazníky poštou a emailem, internetové výzkumy

Další hojně využívanou metodou je zasílání dotazníků poštou. Respondent vyplňuje dotazník sám, bez asistence tazatele. Tato metoda je často jmenována jako nejlevnější metoda výzkumu trhu. Pokud bereme v úvahu pouze dotazníky zasílané kamenou poštou, je toto hodnocení opět relativní jako v případě výzkumů CATI. V případě nízkého zájmu respondentů o výzkum, když chceme dosáhnout cílového počtu 500 dotazníků, musíme při nízké očekávané response rate²³ deseti procent rozeslat 5000 dotazníků. Samozřejmě existují výzkumy, které se chlubí response rate 75%, ale to jsou spíše výjimečné případy. Navíc výzkumy v této kategorii lze jen těžko vzájemně srovnávat. Response rate je závislá na několika faktorech. Přinejmenším na tématu výzkumu, délce dotazníků, způsobu distribuce dotazníků, způsobu prezentace výzkumu, frekvenci zasílání upomínek (pokud je o nich rozhodnuto), znění děkovného dopisu atd.

Nespornou předností výzkumů poštou je možnost použití delších dotazníků. Při přijetí takového dotazníku je klíčový první moment, kdy se respondent rozhoduje jak s ním naloží. Pokud se rozhodne na výzkumu spolupracovat, poskytuje tato metoda respondentovi naprostou volnost co se týče volby času odpovídání i době odpovídáním strávené. To je

²³ Termín response rate obvykle nebývá překládán. Vyjadřuje se jím míra atraktivity dotazníku měřená jeho návratností=přijetí/odmítnutí spolupráce respondentů na výzkumu.

nezbytné u dotazníků, které zjišťují respondentovo chování v určitém časovém rozmezí (například týdenní režim stravování), u otázek, které vyžadují odpověď okamžitě nedostupnou nebo takovou, kterou respondent musí ověřovat. Jak již bylo řečeno, anonymita situace při vyplňování dotazníku přiměje respondenta odpovědět na senzitivní otázky prokazatelně pravdivěji než při dotazování F2F. Tomu bude věnována jedna z dalších kapitol.

Tato metoda je velmi zdlouhavá, protože v zájmu kvalitních dat je třeba trvat na co nejvyšší odpovědní míře. Dotazníkové šetření poštou spojené s několika vlnami upomínek se může snadno protáhnout až na několik měsíců. Metoda není vhodná pro zkoumání všech druhů respondentů. Response rate se liší u různých skupin respondentů v závislosti na dosaženém vzdělání a míře orientace v problematice. Navíc různí lidé mají jinou míru zájmu na účasti na výzkumu. Smith [in Mc Donald, Vangelder, 1998:54] uvádí krásný příklad takového jevu. Představme si poštovní anketu mezi absolventy vysoké školy po několika letech jejich absolutoria. Všech, bez ohledu na jejich výsledky během studia, se ptáme jak zpětně hodnotí kvalitu přednášek. Je pravděpodobné, že vyšší response rate naměříme u těch úspěšnějších. Řekl bych, že za studijně méně úspěšné budou odpovídat ti, kterým se daří profesně. Ti méně úspěšní a nespokojení s kvalitou přednášek pravděpodobně podle Smithe své negativní vzpomínky dobrovolně oživovat nebudou. V případě dotazníků zasílaných emailem si musíme uvědomit, že příkladně v České republice je skladba uživatelů internetu a majitelů emailových účtů specifická také z demografického a socio-ekonomického hlediska. Dotazníku zaslánému respondentovi poštou hrozí hned několik rizik. Zaprvé nevíme, zda ho bude vyplňovat správná osoba, což je stejný problém jako u klasicky rozesílaných tištěných dotazníků. Elektronické poště s příloženým dotazníkem hrozí vážné riziko, že bude zaměněn s nevyžádanou poštou nebo reklamním emailem a jako takový bude bez přečtení smazán. Také musíme kontrolovat, že se z jedné adresy vrátí maximálně jeden dotazník (kontrola duplicit). Dotazník určený pro účely samovyplňování musí splňovat jiná kritéria než dotazník určený pro F2F výzkum. Je naprosto nevhodný pro otázky pracující se spontánní znalostí kombinovanou se znalostí dotazovanou, vyžaduje přísnou grafickou úpravu atd. Dále se nedoporučuje spoléhat na respondenty v poštovním dotazování v tom smyslu, že dotazník vyplňují v předepsané sekvenci. Více o tomto budu mluvit v kapitole Chyby ve výzkumech.

Výzkumy po internetu mají některá pozitiva společná s výzkumy založenými na CAPI metodologii. Společné výhody pramení z využití počítačové logiky programovaných dotazníků. Kromě toho, že dotazník (umístěný na vzdáleném serveru, ke kterému se respondenti připojují nebo si jej stahují a po vyplnění data odesílají) zvládá techniku předem naprogramovaných odskoků a doplňování otázek podle dříve zadaných odpovědí, můžeme

do dotazníku na internetu vkládat různé multimediální přílohy. Můžeme kontrolovat časování plnění výzkumu, pokud se ale nejedná o veřejně přístupnou anketu, na kterou jsou respondenti upozorňováni pop-up²⁴ navigací. Pokud jde například o zaměstnanecký B2B výzkum, dokážeme podle kompletního přehledu IP adres uživatelů firemní sítě, např. intranetu, kontrolovat nejen kdy jim umožníme zobrazení dotazníku, ale také snadno, právě kvůli identifikátoru IP adresy, odhalíme případy opakovaného vyplnění dotazníku. Asi bych se opakoval, kdybych znovu podtrhoval lepší schopnost samovyplňovacích, a tím tedy i internetových výzkumů dosahovat validnějších výsledků v případě témat s emocionálním nábojem. Anonymita internetového prostředí, pokud odlédneme od B2B výzkumů, způsobuje také to, že respondenti v porovnání s dotazníky zasílanými poštou, daleko podrobněji odpovídají na otevřené otázky²⁵.

Největší náklady v internetových výzkumech s trochou nadsázky představuje práce programátorů ve fázi přípravy skriptu. Od chvíle umístění na síť už dotazník prakticky žije vlastním životem, což paradoxně vede k některým úskalím jinak velmi rychlé a levné metody. Již jsem řekl, že internetová populace neodráží věrně socio-demografické charakteristiky obecné populace. Dále, pokud se respondent v jakékoliv z fází vyplňování rozhodne, že již nechce ve spolupráci pokračovat nebo potřebuje proces odpovídání urychlit, může dotazníkem „proběhnout“ aniž by nutně musel číst znění otázek a znát všechny možnosti odpovědí. Opět, s výjimkou B2B výzkumů, i v dotaznících, ve kterých zjišťujeme demografické nebo socio-ekonomické pozadí respondenta, se musíme spoléhat na pravdivost uváděných informací. U internetových anket se respondentem velmi snadno může stát občan jiného státu. Odtud by mělo být zjevné, že i když v sobě internetové výzkumy bezesporu skrývají obrovský potenciál, musíme zatím jejich výsledky brát s rezervou.

7.4 Smíšené módy dotazování (mixed mode surveys)

Sdílím názor A.Dillmana, který něvěří tomu, že nové výzkumné metody zcela nahradí ty tradiční. Předpovídá využívání různých metod pro různé typy studií, což je vlastně zachování současného stavu [Dillman, 2002:7]. Současně se zamýšlí nad důvody pro rozvoj využívání smíšených módů výzkumů, které kombinují dvě a více výzkumných metod v jednom šetření. Dillman si myslí, že rostoucí popularita takových výzkumů spočívá v klesající response rate obecně, ale že agentury ke kombinování designu sahají také z úsporných důvodů. Tazatel se například telefonicky spojí s respondentem, vysvětlí mu okolnosti výzkumu a pokud

²⁴ Informační okna, která se uživatelům internetu náhodně zobrazují na monitoru

²⁵ <http://www.surveysystems.com/online.html>

respondent o výzkum projeví zájem, přepojí ho tazatel na automatizovaný systém dotazování IVR²⁶ a věnuje se rekrutaci dalšího respondenta [Dillman, 2002:8-9]. Sporným faktorem je tvrzení některých výzkumníků, že možnost volby způsobu dotazování pro respondenta dokáže navýšit response rate [dtto]. Hlídání hladiny response rate by ale nemělo být rozhodujícím faktorem pro volbu výzkumné metody, případně pro volbu setu výzkumných metod²⁷. Známe omezení CATI nebo IVR, ve kterých respondent nevidí celou škálu odpovědí. Když chceme tyto módy kombinovat s osobním dotazováním nebo dotazováním na internetu, musíme na to myslet již při tvorbě dotazníku. Ukazuje se, že stejné škály mohou vést k výrazně jiným výsledkům za použití různých metod. Dillmanův tým naměřil v jedné případové studii 39%-ní podíl odpovědí pro nejkrajnější bod škály spokojenosti (zcela spokojen) v případě metodologie IVR. Na stejnou otázku, v dotazníku zasláném poštou, se jako zcela spokojeno ve stejné otázce profilovalo 21% respondentů. V internetové verzi výzkumu 26%. Možným vysvětlením je rovnoměrnější váha různých kódů odpovědí při možnosti je vidět pohromadě, než slyšet a pamatovat si nejlépe jen varianty na opačných koncích škál [Dillman, 2002:18].

Dalším zajímavým a do jisté míry varovným příkladem toho, jak může dopadnout kombinace více metod ve sběru dat je řešení situace, ve které se kdysi ocitli kolegové v agentuře. Zde nemáme možnost srovnání výkonů škálových otázek, ale zato můžeme porovnat výsledky multiple response otázky. Pro nejmenovaného exklienta ze sektoru finančních služeb jsme v několika omnibusových, F2F vlnách měřili povědomí respondentů o produktech finančního trhu nabízených českými finančními institucemi, využívání produktů, purchase intention těchto produktů a znalost institucí nabízejících krátkodobé úvěry. Agentura se zavázala, že dodá data z celkem 2000 rozhovorů omnibusu²⁸, tzn. ze čtyř vln. Chyba v plánování výzkumu způsobila, že v první vlně omnibusu se příslušné otázky

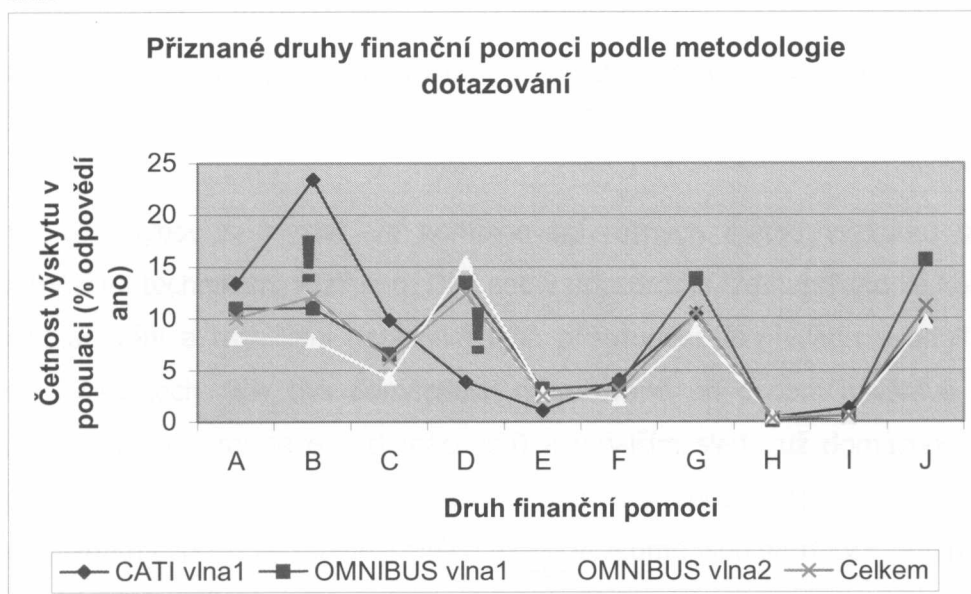
²⁶ Interactive Voice Response je technologie umožňující telefonujícímu zadávat odpovědi pomocí zkrácené volby na klávesnici telefonu.

²⁷ Na úroveň response rate každého výzkumu má vliv celá řada faktorů. Pokud se rozhodujeme pro využití klasických metod výzkumu, ve kterých hrají roli tazatelé, musíme kromě ostatních faktorů ovlivňujících hladinu response rate myslet na to, jak dobře jsou tazatelé seznámeni s výzkumem a jak dobře se orientují v zásadách práce tazatelů. Jak prokázali Durbin a Stuart [1951:173] in Moser a Kalton [1971:289] zkušenost tazatelů může mít na response rate dokonce velmi významný vliv. V experimentu, kterého se ve Velké Británii zúčastnili tři skupiny po pěti stech tazatelích – profesionální tazatelé ze Státního Centra pro Sociologické Výzkumy a Britského Institutu pro Veřejné Mínění a studenti Univerzity pro Ekonomii v Londýně – Durbin se Stuartem naměřili na hladině spolehlivosti 99% signifikantní rozdíly v úspěšnosti kontaktování respondentů. Úspěšnost byla měřena získáním plně nebo částečně vyplněného dotazníku. Tyto rozdíly zůstávají signifikantní i při pohledu na data za regiony, ve kterých výzkum proběhl i při pohledu na data za různé věkové skupiny respondentů i podle pohlaví. Zatímco obě skupiny profesionálních tazatelů dosáhly slušných 3,8%, respektive 3,2% odmítnutých rozhovorů, tazatelé-studenti byli odmítnuti ve 13,5 procentech případů.

²⁸ Omnibus je zaměřený na obecnou populaci, cílovou skupinou jsou obyvatelé ve věku 15–69 let. V rámci jedné vlny je v ČR dotazováno 500 respondentů. Respondenti jsou vybíráni podle předem stanovených kvót z hlediska věku, pohlaví, regionu a velikosti místa bydliště tak, aby struktura výběrového souboru odpovídala struktuře celkové populace. Dotazování se provádí metodou osobních rozhovorů školených tazatelů TNS AISA s respondenty v domácnostech [<http://www.tns-aisa.cz/omnibus.asp?lg=cz>].

neobjevily a vedoucí projektu operativně rozhodl o tom, že chybějící rozhovory budou včas zajištěny metodologií CATI. V grafu 3 se můžeme podívat na rozložení výsledků na jednu z dotazovaných otázek, která zjišťovala, zda některý ze členů domácnosti čerpá půjčku a pokud ano, jakým způsobem (viz Dotazník1 přílohy). Z důvodu exkluzivity výsledků nejsou položky A-J v grafu řazeny ve stejném pořadí jako při dotazování (položka A neznamená půjčku od přátel nebo rodiny atd.). V ideálním případě by se měly výsledky v rámci vlny 1 omnibusu v obou metodologiích shodovat, protože kvótní výběr respondentů na CATI byl plánován podle metodologie výběru respondentů v omnibusu. Jak ale pozorujeme, minimálně finanční pomoc typu B a D je v obou metodologiích zmiňována výrazně jinak. Možným vysvětlením u produktu B je psychologický efekt anonymního dotazování po telefonu, kdy respondet CATI telefonistovi spíše než při osobním rozhovoru sdělí, že musí využívat půjčku. Jak ale vysvětlit takovou disproporci u produktu D? Hodnoty pro omnibus vlnu 2 jsou počítány ze 3. a 4. vlny klasického omnibusu, stejnou metodologií jako omnibus vlna 1. Rozdíly oproti vlně omnibusu 1 zde nejsou tak dramatické jako v případě rozdílu mezi daty z CATI vlny1 a omnibusu vlny 1 a připisují je skutečnému vývoji situace spotřebitelů mezi vlnami.

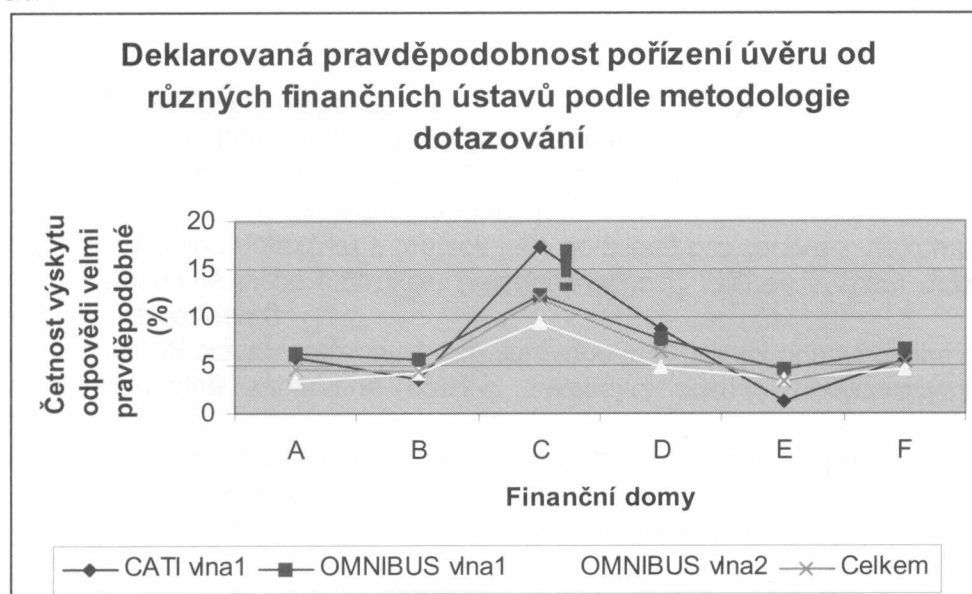
Graf3



N CATI vlna1=470, N omnibus vlna1=536, N omnibus vlna2=1097, N celkem=2103

Bezradný jsem při interpretaci rozdílu naměřených výsledků pro otázku zjišťující pravděpodobnost pořízení úvěru u vybraných bankovních domů (plné znění otázky viz Dotazník1 přílohy). V grafu 4 nenalezneme závažné rozdíly výsledků mezi 1. a 2. vlnou omnibusu, ale v případě jediného bankovního domu jsou počty odpovědí kategorie „velmi pravděpodobně“ rozhovorů za metodologii CATI významně četnější. V CATI sice nebylo možné předložit respondentům kartu se škálou odpovědí, ale přesto bychom se měli ptát, proč nejsou četnější u všech nabízených bank?!

Graf4



N CATI vlna1=470, N omnibus vlna1=532, N omnibus vlna2=1097, N celkem=2099

Bylo by chybou si myslet, že kombinování různých metod výzkumu se objevilo s nejmodernějšími technikami výzkumu. De Leeuw upozorňuje, že například ke kombinování osobního dotazování a rozesílání dotazníků bylo přistoupeno kupříkladu v longitudinálních panelových výzkumech, kdy byli domácnosti rekrutované při osobní návštěvě (současně tazatel zjistil základní informace o domácnosti) a v dalším sledu už domácnost dostávala dotazníky poštou [de Leeuw in Journal of Official Statistics, 2005:234].

Kompletní možnosti využití smíšených módů výzkumů jsou shrnuty v originálním znění ve schématu 3, přílohy.

8. Chyby ve výzkumech

V roce 1944 publikoval Edwards Deming v *American Sociological Review* článek s názvem „On errors in surveys“. Předložil zde svou verzi kompletní klasifikace faktorů (celkem 13 větších okruhů), které se podílejí na kvalitě výstupů výzkumu obecně. Deming se o ni vyjadřuje tak, že je možné navrhnout typologii mírně odlišnou a že pro různé výzkumy mohou různé kategorie faktorů být nanejvýš důležité, nicméně všechny tyto faktory do jisté míry podepisují na celkové kvalitě jakéhokoliv výzkumu. Ona klasifikace chyb, kterou ve své době představil, vypadala následovně:

1. Variabilita odpovědí respondentů
2. Rozdíly mezi jednotlivými výzkumnými metodami
3. Odchytky a variabilita odpovědí způsobené tazatelem
4. Odchytky způsobené zaujetím respondenta²⁹
5. Chyby v designu dotazníku a tabulek jako podkladů pro zprávu z výzkumu zahrnující
 - a) nedostatečně jasné instrukce, nejednoznačnost, lišící se význam otázky pro jiné skupiny respondentů
 - b) opomenutí otázek potřebných pro správnou interpretaci odpovědí jiné otázky
 - c) emocionálně zabarvené otázky, zavádějící otázky, nevyčerpávající nabídka odpovědí
 - d) nerozpoznání relevantních a irrelevantních tabulkových výstupů
 - e) posilování efektu non-response
6. Změny týkající se cílové populace v době od ukončení fáze sběru dat a přípravy zprávy z výzkumu
7. Odchytky způsobené non-response
8. Odchytky způsobené opožděným hlášením problémů v některé z fází výzkumu
9. Odchytky způsobené špatným výběrem termínu dotazování
10. Odchytky způsobené nereprezentativním výběrem respondentů
11. Chyby výběru
12. Chyby zpracování dat
13. Chyby interpretace dat

²⁹Odchytky na straně respondenta (bias of the auspices) mohou vznikat vědomě či nevědomě vlivem zaujetí určitého postoje ke zkoumané problematice nebo organizaci, jež výzkum provádí. Respondent se zaujatým může stát také tehdy, když se znalostí toho, kdo výzkum financuje a kdo bude jeho výsledky využívat upravuje své odpovědi za účelem ochrany svých zájmů. [http://cmg.wkac.ac.uk/courses/fyp/FYP_QUIK.DOC] Takových příkladů, dokreslujících jak může respondent korigovat své odpovědi předtím, než je sdělí tazateli, aby tím sebe nepoškodil, je celá řada. Představme si respondenta účastníka se cenového testu nápojů (například piva), který má říci, za jakou cenu by ještě kupoval určitou značku tohoto nápoje, při které ceně by se začal zajímat o suplement jiného původu a při jaké ceně by pivo přestal konzumovat. Test cenové elasticity Van Westendorp je jen jedním z mnoha nástrojů, které se využívají pro určení (klienty, v tomto případě konzumenty) akceptovatelné cenové hladiny výrobku, který je nově uváděn na trh nebo u výrobků inovovaných. Zda naměřené výsledky nejsou zatíženy vlivem zaujetí respondenta se ale musíme ptát i při vyhodnocování výsledků jiných než čistě cenových otázek. Například při interpretaci výsledků výzkumů veřejného mínění. Dejme si opět zcela konkrétní příklad. Nejmenovaná společnost působící na trhu poskytovatelů elektrické energie si opakovaně objednala provedení výzkumu veřejného mínění obyvatel žijících v regionu sousedícím s provozovnou zadavatele výzkumu. Respondenti měli mimo jiné vyjádřit míru souhlasu s výroky týkajícími se života v obci. V grafu 1 přílohy si můžeme všimnout, že většina obyvatel obce se již čtvrtý rok po sobě vyjadřuje ve smyslu, že se životní podmínky v obci zlepšily. Později, kdy jim již bylo sděleno, kdo si výzkum objednal, na zcela transparentní otevřenou otázku, co jim ze strany tohoto zadavatele chybí, jmenují v tomto pořadí „snížení cen energie pro svou a okolní obce“, „více sponzorských darů“ a až později vyšší zabezpečení, informovanost aj. (graf 2 přílohy).

Těchto 13 okruhů rozřídil do dvou základních kategorií, které nazval „errors of observation“ a „errors of non-observation“. Když Deming zmiňuje „errors of observation“, říká, že je možné je chápat jako chyby, které mají svůj původ ve fázi získávání dat³⁰, naopak „errors of non-observation“ vznikají především ve fázi výběru a definování výzkumného vzorku [Deming, 1944: 335-350 in De Vauss, 2002].

O více než 40 let později napsal Robert M. Groves do časopisu *Public opinion quarterly* stať s názvem *Research On Survey Data Quality*. Od té doby se stala typologie chyb ve výzkumu složitější minimálně z toho důvodu, že různé skupiny lidí pohybující se v oblasti výzkumu „vynalezli“ pro některé již známé chyby nová označení. Statistickí vidí v celkové chybě výzkumu³¹ především rozdíl mezi střední hodnotou proměnné ve výběrovém souboru a hodnotou stejné proměnné v celé populaci³². Psychologizující výzkumníci se na chybovost výzkumu dívají z hlediska validity a reliability [Groves, 1987: 350-363 in De Vauss, 2002].

Srozumitelné vysvětlení významu obou pojmů podává Disman. Velmi stručně vystihuje validní měření jako takové, „...které jsme skutečně zamýšleli měřit“ [Disman, 2002: 62].

Existují tři základní skupiny typů validity a Disman ukazuje několik způsobů, jakými lze validitu měřit. V první skupině nalezneme validitu založenou na členství ve známé skupině. Mám za to, že stejný typ validity je v zahraniční literatuře označován jako *discriminant validity*. Disman uvádí příklad, kdy potřebujeme vyvinout dotazník měřící rasistické postoje. Validaci dotazníku bychom mohli provést rozhovory jednak běžnými respondenty a pro kontrolu také se zástupci skupiny, která bude spíše vykazovat takové postoje svých členů, které jsme definovali jako rasistické. Respondenti ze skupiny hlásící se k rasismu by v ideálním případě měli dosahovat vyšších skóre v otázkách měřících koncept rasismu. Tato technika předpokládá existenci kontrolní skupiny, což by v případě skupiny rasistů neměl být problém. Potíž by mohla nastat kdybychom se rozhodli měřit méně transparentní postoje, konstrukty měřitelné projektivními technikami a kontrolní skupina by byla hůře dostupná.

Prediktivní validita spočívá v porovnání výsledku měření se skutečnými, tvrdými daty. Kdybychom zůstali u problematiky měření rasistických postojů a chtěli zjistit, zda náš nástroj dokáže odhalit rasistické sklony u respondenta, chtěli bychom si teoreticky vyžádat výpis

³⁰ measurement error

³¹ total survey error

³² řeč je o „mean square error“

z rejstříku trestů a hledali zde záznamy o tom, zda respondent byl někdy stíhán za činy označitelné jako rasistické³³.

Třetím a posledním typem validity v této skupině je validita souběžná (concurrent validity). Souběžně testujeme jeden koncept více metodami, například nepřímými, přímými nebo projektivní otázkami a zjišťujeme, zda dosahují stejných výsledků.

Hendl klasifikuje výše jmenované typy validit jako validity kriteriální [Hendl, 2004: 48]. Jiným typem je validita konstruktová³⁴. Konstruktová validita pracuje s testováním souboru hypotéz „...logicky spojujících zkoumanou vlastnost s technikou měření, kterou používáme“ [Disman, 2002: 65]. „Důkazy o konstruktové validitě mohou mít konvergentní charakter (test prokazuje vztahy k těm proměnným, jež podle teorie očekáváme) nebo diskriminační charakter (naopak nemá vztah k proměnným, když tento vztah neočekáváme). V této etapě přezkušování nové procedury je důležité, zda výsledky predikují stavy, které podle teorie očekáváme. Například test výkonnosti by měl nabývat nízké hodnoty, pokud sportovec nedosahuje dobrých výsledků v soutěžích“ [Hendl, 2004: 49]. Hendl pod konstruktovou validitu řadí i zjevnou validitu, ale zde se mírně liší od Dismana, který zjevnou (face) validitu řadí pod testy založené na obsahu testovaného jevu. Kromě zjevné validity, která nevyžaduje žádný test a je založena na posouzení dostatečnosti shody mezi tím, co chceme měřit a jakými indikátory to měříme, jmenuje Disman ještě validitu založenou na mínění skupiny soudců a validitu testovanou výčtem obsahu (domain of meaning). První z posledně jmenovaných validit spoléhá na takové expertní znalosti profesionální skupiny hodnotitelů, které pomohou rozpoznat relevanci a vhodnost zvoleného konstruktů k měření vybraného jevu. Test výčtem obsahu je výsledkem uvědomění si toho, že různí lidé různě myslí a pod jedním pojmem si mohou představit různé věci.

Před časem dostali kolegové ve výzkumné agentuře zadání změřit spokojenost zákazníků se službami zadavatele výzkumu, který je dodavatelem stavebních hmot. S podobným požadavkem jsme se doposud nesetkali a museli jsme přiznat, že nedokážeme navrhnout takový dotazník, který by spokojenost změřil věrohodným způsobem. Přesvědčili jsme zadavatele o tom, že by bylo nanejvýš vhodné, před samotným hodnocením spokojenosti, provést kvalitativní šetření ve všech zemích, kde měl být výzkum proveden. V tomto případě byl výzkum veden kromě České republiky také v Maďarsku a Rakousku. Po sériích skupinových diskusí byly v každé zemi odděleně navrženy sady atributů hodnocených respondenty jako nejdůležitější. My si samozřejmě nebudeme ukazovat tři varianty celého

³³ Disman uvádí příklad s měřením úspěšnosti studentů během studia na vysoké škole. Stejný příklad je rozebírán také v metodické příručce „Essentials of survey research and analysis“ z roku 1998, kterou publikoval R.J.Pollard.

³⁴ Disman používá pojem konstruovaná validita

dotazníku, ale v tabulce 1 se podíváme, jaké atributy se v různých zemích ukázaly jako důležité v charakteristice hodnocení, kterou jsem vybral zcela náhodně jako jeden z osmi oddílů dotazníku. Je jím kvalita obchodního zástupce.

Tabulka 1

	<i>Česká republika</i>	<i>Maďarsko</i>	<i>Rakousko</i>
1.	Dosažitelnost	Dosažitelnost	Dosažitelnost
2.	Otevřenost ke stížnostem	Otevřenost ke stížnostem	Otevřenost ke stížnostem
3.	Znalost produktu, který odebíráme	Znalost produktu, který odebíráme	Znalost produktu, který odebíráme
4.	Pro-aktivní přístup k zákazníkům (informování o novinkách, změnách cen)	Pro-aktivní přístup k zákazníkům (informování o novinkách, změnách cen)	Pro-aktivní přístup k zákazníkům (informování o novinkách, změnách cen)
5.	Rozumí potřebám mého podnikání a dokáže se podle nich přizpůsobit	Rozumí potřebám mého podnikání a dokáže se podle nich přizpůsobit	Prodejce nabízí jednu kontaktní osobu od okamžiku objednávky po dodání zboží
6.	Nápomocnost	Organizování setkání s klientem	Dobry osobní vztah s prodejcem
7.	Profesionalita	Celková kompetentnost	Rozumí potřebám mého podnikání a i když se jedná o technické specifikace
8.		Přátelskost	Přátelskost
9.		Frekvence návštěv, která mi vyhovuje	Frekvence návštěv, která mi vyhovuje
10.		Rychlá pomoc v případě problému	Rychlá pomoc v případě problému
11.		Otevřenost k novým nápadům a doporučením	Schopnost společně vyvíjet řešení šitá na míru
12.			Dokonalá znalost cen
13.			Představení nabízeného produktu prodejcem
14.			Individuální podpora a pomoc

Tmavě zelené buňky obsahují atributy, které se ve stejné podobě objevily alespoň ve dvou zemích. Světle zeleně jsou zabarvena pole, která lze srovnat jen přibližně. Dále vidíme, že v Rakousku výzkumný tým našel takové atributy, které podnikatelé ve zbývajících zemích v daném tržním segmentu zřejmě nepovažují důležité, protože se o nich nezmínili. Proto jsme například atribut „Prodejce nabízí jednu kontaktní osobu od okamžiku objednávky po dodání zboží“ do českého a maďarského dotazníku nezařadili, protože zde podle nás neměří spokojenost s dodavatelem stavebních směsí. Výzkum nebyl v podstatě ničím jiným, než rozsáhlým testem výčtu obsahu ke zjištění toho, za jsme v plné míře vystihli body hodnocení spokojenosti zákazníků³⁵ Po této fázi navazoval kvantitativní CATI výzkum na populaci zákazníků zadavatele. Design výzkumu a teď mám na mysli především výběr atributů v různých zemích dále dokládá, že test validity v jednom případě a zemi neposkytuje univerzálně platný výsledek validity setu atributů v různých kulturních prostředích.

³⁵ Disman uvádí jiný příklad s výzkumem volného času respondentů.

Validita výzkumů do určité míry souvisí s jejich další charakteristikou – reliabilitou. Výzkum je reliabilní, pokud za použití stejných metod a nástrojů na definované populaci přináší nezměněné výsledky. Disman do definice reliability vnáší navíc časovou dimenzi, když říká, že „reliabilita je takové měření, které nám při opakované aplikaci dává shodné výsledky, pokud se ovšem stav pozorovaného objektu nezměnil“ [Disman, 2002: 62]. Reliabilita a validita jsou spojené nádoby. Pokud výzkum není reliabilní, není ani validní. Ale nelze říci, že prokázaná reliabilita je zárukou validity. Můžeme donekonečna provádět sice stejně vycházející opakování testu, ale test nemusí měřit koncept, který měřit má. Nereliabilita může mít několik zdrojů, kterými se budeme zabývat dále v textu až budeme mluvit o chybách na straně tazatelů a respondentů.

Stejně tak jako validitu můžeme měřit i reliabilitu. Nejčastěji používanými metodami jsou testy dependability a vnitřní konzistence. Pokud jde o dependabilitu, provádí se nejčastěji tzv. test-retest. Jde o minimálně dvě opakované měření téhož konceptu a to buď a nebo bez časového odstavu. Pro vyhodnocení reliability se vytvářejí koeficienty korelace naměřených výsledků³⁶. Typem měření vnitřní konzistence může být metoda půlení testů, známá jako split-half. Jestliže jsme se rozhodli pro metodu split half, chceme zjistit míru korelace položek v rámci jedné tematické kategorie, například již zmiňované atributy prodejního zástupce. Podrobně se o způsobu výpočtu korelačních koeficientů můžeme dočíst u Hendla [Hendl, 2004: 263-266]. Nejvíce pozornosti je zde věnováno výpočtu Pearsonsova koeficientu, nicméně k posouzení korelace se používají i jiné ukazatele - koeficient konzistence - Cronbachovo α a Spearman-Brownův vzorec³⁷. Tím ale možnosti měření reliability nejsou vyčerpány. Použit můžeme také metodu paralelních testů, kdy stejný konstrukt měříme dvěma rovnocennými nástroji. Tato metoda má opět dvě alternativy – můžeme ji využít s nebo bez časového odstavu.

Nyní bych se rád vrátil ke Grovesově klasifikaci chyb ve výzkumu. Z psychologizujícího pohledu jsme definovali validitu a reliabilitu jako charakteristiky kvality výzkumu. Ze statistického hlediska lze přesnost měření hodnotit charakteristikami bias a variance. Bias je zkreslení hodnoty měření proti skutečné hodnotě proměnné v populaci. Toto zkreslení je stabilně vysoké při opakovaných měřeních. Variací má Groves na mysli s největší pravděpodobností směrodatnou odchylku, neboli kolísání hodnoty měření ve

³⁶ „Ve studiích s kvantitativními daty se jako míra spolehlivosti často používá koeficient vnitrotřídní korelace. Vyjadřuje, jaký díl celkové variability ve skórech (která je způsobena jak rozdíly mezi jedinci, tak intra-individuálními rozdíly při opakovaných měřeních) může být přisouzen různosti charakteristik jedinců. Spolehlivost ovlivňují: zobecnitelnost výsledků vzhledem k času, jednotnost administrace a skórování, vnitřní konzistence, homogenita subjektů a v neposlední řadě počet otázek.“ [http://www.copn.cz/clan/]

³⁷ [http://www.copn.cz/clan/]

směru a intenzitě při různých měřeních. Bias i variance jsou vyjádřením možnosti replikovatelnosti výsledků výzkumu. Pro ty zajímavější se spíše o chyby výběru, je bias vyjádřením zkreslení, které vznikne v datech ke stejnému výzkumnému problému sebraných po určitém čase od jiných respondentů a které pramení právě ze způsobu jejich výběru [Groves, 1987: 351 in De Vauss, 2002]. Pro někoho jako jsem já, kterého zajímají chyby ve fázi sběru dat, má smysl sledovat bias jako například opakované podhodnocení proměnné hrubý příjem domácnosti při zachování jinak stejných podmínek výzkumu³⁸. Zajímá mě i variance, ale spíše jako důsledek vlivu práce tazatelů. Bias a variance mohou tedy vznikat jak ve fázi před terénním sběrem dat tak i během něj. Do první kategorie chyb patří chyba pokrytí, non-response a chyba výběru. Chyba pokrytí vzniká z mylného nedefinování části populace jako cílové. Při výzkumu automobilistů v ČR bychom chtěli mít poměrně zastoupeny řidiče všech značek jezdících na českých silnicích a nevytvořili bychom kvótu pro majitele vozů Škoda s významným podílem na trhu. Non-response již byla v dříve zmiňována stejně jako chyba výběru.

Druhá kategorie chyb je širší. Spadá sem vše od špatného definování měřeného konceptu, přes nezvládnutí použití vybrané metody dotazování, chyby tazatelů, respondentů, těch, kteří pořizují data, těch, kteří kódují otevřené otázky, sezónní vlivy až po chyby ve zpracování dat. Jak si všímá Bailar, na celkové chybě výzkumu se může podílet několik negativních vlivů najednou. Na základě již provedených případových studií víme obecné poučky, že krátké a jednoduché dotazníky mají šanci na dosažení vyššího response rate, že lidé si při vzpomínání dokáží vybavit události intenzivněji pokud otázka směřuje pouze velmi blízko do minulosti nebo že lidé v deníčcích nadhodnocují frekvenci činností, které měli ve sledovaném období dělat [Bailar, 1987: 323-325 in Journal of Official Statistics]. Už ale víme jen přibližně, jak by lidé odpovídali při mírně změněném znění otázky než v provedené studii, pro kterou je závěr publikován. Chyby výběru dokážeme kontrolovat a dostatečně přesně měřit. Můžeme se spolehnout na tabulky odhadů výběrové chyby, které jsou platné za všech okolností a do hry vstupují zjednodušeně řečeno „pouze“ velikost výběrového souboru a hladina významnosti, na které hledáme velikost intervalu hodnot měřené charakteristiky. Nevýběrové chyby se dají odhadovat o poznání hůře. Existují samozřejmě modely, které se odhadem nevýběrové chyby zabývají, ale jsou do značné míry závislé na výběru kontrolních proměnných.

První cílené iniciativy o měření a kontrolu chyb vznikaly ve 40. letech minulého století. Výsledkem byly tzv. modely smíšeného měření chyb, které měly měřit všechny zdroje chyb a odhadovat celkovou chybu výzkumu. Později se jim začalo říkat výzkumné modely a měly

³⁸ předpokladem je, že mezi první a druhou vlnou výzkumu se i přes časový posun nic nezměnilo

minimálně tři oblasti využití. Modely umí zjistit, jaké typy chyb výzkum zatěžují. Dále musí odhadnout podíl jednotlivých chyb na celkové chybě výzkumu. Nakonec tyto modely odhadují vliv jednotlivých komponentů druhu chyb. Pokud například zjistí přítomnost chyby respondentů, zkoumají dále bias i varianci [Forsman, 1989: 42 in Journal of Official Statistics]. Forsman popisuje dva modely. Protože si nejsem jistý, zda pro jejich názvy existuje český ekvivalent, uvádím je v původním znění. Tím prvním je *mean square error decomposition* model. Tím druhým je *linear model approach*. Forsman píše, že tyto modely nikdy příliš nezakořenily v jiných zemích než tam, kde byly vyvinuty, tj. v USA a Kanadě. Příčin může být několik. Modely v první řadě nedokáží zahrnout všechny druhy chyb. Dále předpokládají, že validační rozhovory (re-interviews) jsou nezávislé na odpovědích v řádně vlně výzkumu³⁹. Jiné dva modely popisuje Tortora. *General nonsampling error and bias* model lze použít tam, kde známe skutečnou hodnotu proměnné v populaci. Tu nebývá vždy lehké zjistit. Využívají se pro tento účel záznamy lékařských ústavů, pojišťoven, sociální správy, informace od zaměstnavatelů atd. [Tortora, 1987: 339-342 in Journal of Official Statistics].

Sociologické průzkumy nepracují pouze s demografickými proměnnými, ale měří také postoje, míru znalosti čehokoliv nebo životní hodnoty. To jsou proměnné bez reálné opory v jakýchkoliv statistikách. Na stránkách Českého statistického úřadu nebo Ministerstva zdravotnictví můžeme zjistit, jaká je skutečná průměrná výška nebo váha českých mužů a žen ve věku 15 let a starších a porovnat tak námi naměřenou statistiku s tímto údajem. Nikde už ale nezjistíme, kolik % z nich skutečně rozhodně souhlasí, spíše souhlasí, spíše nesouhlasí nebo rozhodně nesouhlasí s výrokem, že geneticky modifikované potraviny jsou morálně akceptovatelné nebo že nabídka cizích jazyků na českých školách je dostatečně bohatá⁴⁰.

Jabine si pokládá otázku, nakolik se zadavatelé výzkumů, uživatelé výsledků nebo veřejnost skutečně zajímají o velikost chyby měření. Provokuje námitkou, že ti, kteří potřebují výsledky výzkumu pro podložení svých, třeba nepopulárních rozhodnutí nebo dokumentování výsledků svého úsilí už zpětně nemají zájem na přiznání zkreslení výsledků. Jako příklad uvádí census pracovní síly v USA z roku 1990. Jeho výsledky byly ovlivněny nestejnou mírou response rate u různých socio-ekonomických skupin obyvatelstva. Po zveřejnění výsledků, které uváděly podhodnocenou úroveň nezaměstnanosti, byly provedeny opakované rozhovory, které chybu potvrdily. I přesto byly výsledky vydávány za

³⁹ Srovnej s [Forsman, 1989: 43-53 in Journal of Official Statistics]

⁴⁰ Tyto otázky byly zařazeny do dotazníku k projektu EuroBarometru v jedné z podzimmích vln roku 2005 (7/2005).

směrodatné i odpovědní vládní činitelé projevili jen malou ochotu k revizi výsledků výzkumu [Jabine, 1987: 335-338 in Journal of Official Statistics].

Mám tento pocit poté, co jsem prošel velkým množstvím textů odkazujícím se ke měření chyb vzniklých jinak než výběrem respondentů. Neexistuje konsenzus v tom, jak měřit nevýběrovou chybu; ví se, že nevýběrové chyby se na celkové chybě výzkumu podílejí v kombinaci, ale neumíme určit, jak jeden typ chyby ovlivňuje druhý; neexistuje shoda o jakémsi pořadí závažnosti těchto chyb a závěry uskutečněných případových studií se nedají zobecnit. Opatření vedoucí k zajištění vyšší validity a reliability výzkumů představují v každém případě extra náklady k výzkumu a zadavatelé si tento luxus v devíti z deseti případů nedopřejí. Připomíná to začarovaný kruh. Pokud se rozhodneme například testovat validitu položek v dotazníku tak, že zařadíme vybrané atributy do dvou verzí dotazníku, dotazník se tím tak celkově prodlouží, prodlouží se délka dotazování. To může zvyšovat výskyt případu odmítnutí rozhovoru, v CATI výzkumech se může prodražovat terénní sběr dat. Nebo ještě lepší příklad. Už víme, že chyba způsobená vlivem tazatelů na respondenta⁴¹ lze zmenšit zaměstnáním většího počtu tazatelů pracujících na výzkumu. Současně povolání většího počtu tazatelů vyvolává vyšší nároky na jejich školení a komunikaci s nimi.

To vše mě vede k tomu, abych své úsilí zaměřil na bod, od kterého se odvíjí kvalita dat všech výzkumů na jakémkoliv téma a s jakoukoliv populací. Od něj se odvíjí chybovost práce tazatelů nebo vychýlenost odpovědí respondentů. Tvorba dotazníku je proces, na jehož konci musí být výzkumný nástroj, který je nejen dobře postavený (ve smyslu splnění metodologický požadavků), ale také atraktivní pro respondenta, tzn. že působí profesionálním dojmem, ptá se srozumitelně, je zajímavý atd. Výzkum musí být pokud možno pro respondenta zajímavá a nestresující zkušenost. Pokud tomu tak není a respondent je zklamaný nebo dokonce rozrušený, taková skutečnost v každém případě ovlivní naměřené výsledky. Mohou být vychýlené, nejednoznačně interpretovatelné nebo dokonce zavádějící.

⁴¹ dva tazatelé dotazující stejného respondenta mohou získat lišící se odpovědi

9. K čemu jsou dobré dotazníky

Přibližně před 40 lety, zakladatel dnes již mezinárodní sítě agentur pro výzkum trhu a veřejného mínění G.H. Gallup podotkl, že výsledky výzkumů jsou spíše než kvalitou výzkumného vzorku ovlivněny kvalitou výzkumného nástroje. Dotazníky jsou dnes používány jako přímé či nepřímé indikátory názorů veřejnosti, mohou přímo či nepřímo zjišťovat spotřebitelské návyky různých společenských skupin, jsou používány k pochopení rozhodování určitých skupin populace, měří spokojenost s hmotnými i nehmotnými atributy našeho každodenního života. Takto bychom jistě mohli ještě chvíli pokračovat.

Nehledě na to, jaké informace ten který dotazník získává, existuje několik závazných metodologických principů, na které by jeho tvůrci neměli zapomínat předtím, než se pro použití tohoto bezesporu velmi rozšířeného výzkumného nástroje rozhodnou. Podoba každého dotazníku závisí na tom, jakému respondentovi je určen, závisí na charakteru zjišťovaných údajů a na hloubce poznání, o které usilujeme.

Poněkud zjednodušeně můžeme říci, že dotazník slouží k tomu, abychom poměrně standardizovaným způsobem v relativně omezeném čase zasáhli takový počet respondentů, který nám povolí úsudky na základě známých statistických zákonnitostí. Dobře navržený a efektivně využitý dotazník umožní získat nejen celkový obraz o fenoménu, který je v centru naší pozornosti, ale také o jeho komponentech a pilířích. Pokud ve výzkumu zjišťujeme informace sociodemografické povahy, můžeme se na základě získaných dat pokusit také o analýzu na úrovni jednotlivých skupin populace. V praxi bývají výsledkem takové analýzy závěry porovnávající situaci nestejných příjmových skupin obyvatelstva, různých věkových kohort, přinášejí srovnání analýzou podle proměnných jako jsou pohlaví, velikost místa bydliště, počet dětí, členství v profesních nebo zájmových skupinách, klientství v soukromých/veřejných (finančních/nefinančních popř.) institucích, vlastnictví produktů nebo využívání služeb apod.

Z několika typů výzkumných formátů se dále zaměřuji především na úskalí kompilace a použití dotazníků pro výzkumy F2F a to jak B2B tak B2C. Shrnuji dosavadní zkušenosti s prací s tímto zatím nepostradatelným nástrojem, ukazuji správné cesty jeho navrhování a testování. Seznamuji se základními typy otázek, se kterými operuje, na konkrétních příkladech ukazuji před čím je třeba se mít při jeho navrhování na pozoru a v neposlední řadě v jistém smyslu odhaluji myšlenkové pochody těch, pro které jsou dotazníky určeny.

9.1 Výhody a nevýhody dotazníku

Existují různé metody pro výzkum veřejného mínění nebo výzkum trhu. Zde uvedu jen některé důvody proč bychom si mohli vybrat právě standardizované dotazníkové šetření.

1. Data získaná dotazníkovým šetřením lze spolehlivě a bez problémů hromadně zpracovat některým ze statistických programů, které byly pro tento účel napsány a které procházejí neustálým vývojem.
2. Dotazník je nástroj, se kterým se již pravděpodobně setkala významná část členů vybrané populace. I když tato skutečnost může být při některých studiích spíše přítěžující okolností, lze předpokládat, že do výzkumu tímto nástrojem mohou být právě díky jeho relativní jednoduchosti zahrnuty všechny skupiny populace.
3. Standardizované dotazníky mohou při dodržení zásad práce s dotazníkem minimalizovat odchylky způsobené i jen nepatrně odlišným přístupem různých výzkumných týmů a tazatelů. Odpověď respondenta není ovlivněna odlišným uvažováním výzkumníka.

Naopak existuje řada důvodů, proč někteří výzkumníci u některých výzkumných zadání volí odlišnou metodologii než dotazníkové šetření. Mohli je k tomu přivést například tyto příčiny:

1. U strukturovaných dotazníků odpadá možnost ověření nebo lépe do-vysvětlení, popř. rozvedení respondentových odpovědí. Strukturované dotazníky jsou v tomto směru velmi nepružné. Přicházíme o velkou část informací v případě, že respondent cítí potřebu doplnit svou odpověď nebo jeho odpověď jednoduše nezapadá do předpřipraveného schématu. Tomuto jevu můžeme částečně předejít ponecháním prostoru na poznámky tazatele, ale ani tato cesta úplně neřeší toto úskalí. Nemůžeme totiž v reálném čase kontrolovat práci všech tazatelů.
2. Udává se, že přibližně 90% veškeré komunikace je vizuální. Gesta a jiné vizuálně komunikační symboly samozřejmě běžný standardizovaný dotazník zachytit nedokáže. Záleží na tématu výzkumu, jak moc nás tato skutečnost může mrzet. Nezachycená nonverbální komunikace bude mít jiný význam v případě výzkumu na téma frekvence využívání čisticích prostředků a jiný ve výzkumu zjišťující postoje k rómské problematice.
3. Konečně, standardizované dotazníky pro F2F rozhovory se nehodí pro výzkum velmi specifických populací. Těmi mohou být těžko dosažitelné skupiny zaměstnanců, např. obchodní zástupci.

9.2 Vyvíjení dotazníku

Předtím, než začneme tvořit, měli bychom si jasně specifikovat výzkumný problém - uvědomit si, na co má výzkum odpovědět. Tato úvaha by nás měla nasměrovat k těm opravdu relevantním otázkám a naopak by nás mohla zachránit od varianty „výzkum pro výzkum“. Je velmi žádoucí, aby si každý, kdo plánuje výzkum s využitím dotazníku uvědomil a zodpověděl několik otázek:

- Na jaké téma bude výzkum zaměřen? Je možné výzkumný problém obsáhnout srozumitelnými otázkami a získat na ně odpovědi?
- Na co se ptát a na co se neptat? Budeme potřebovat i otevřené otázky?
- V případě, že plánujeme dotazník o více sekcích, jak co nejvhodněji určit jejich pořadí?
- Byl již na toto téma nějaký výzkum proveden? Existují zkušenosti s dotazníky na naše téma?

9.3 Určení relevantních otázek

Jak jsem již předeslal, výchozím bodem pro volbu správných otázek je uvědomění si tématu výzkumu. Okruhy otázek a otázky samotné mohou být získány na základě studia odborné literatury, hledání v již proběhnuvších výzkumech na podobné téma nebo prostě vytvořeny na základě předchozích zkušeností za použití vlastního úsudku. To vše platí v případě, že dotazník již předtím nevznikl na straně zadavatele výzkumu. Dotazník také znovu zpravidla nevytváříme při opakovaných šetřeních nebo kontinuálních výzkumech, kdy se snažíme udržovat stejné znění otázek za účelem srovnatelnosti dat. Učebnicový příklad volby a vzniku otázek pro výzkum, který má objasnit faktory vedoucí mladé lidi k volbě studia a získání titulu na vybrané univerzitě uvádí Burgess.

Tabulka 2

Otázka:	Uvažujeme:
Co přivádí kandidáty právě ke studiu pro získání konkrétního titulu?	Kandidát hledá: - změnu pracovního zařazení - lepší pracovní uplatnění - lepší finanční ohodnocení - jiné
Mohla rozhodnutí jít studovat na tento titul ovlivnit minulá pracovní zkušenost?	- kolik let praxe má kandidát za sebou?
Mohou za rozhodnutím stát genderové rozdíly?	- je kandidát muž nebo žena?
Mohla rozhodnutí jít studovat na tento titul ovlivnit již dosažené vzdělání?	- jaké je nejvyšší dosažené vzdělání kandidáta? - v jakém oboru kandidát dosáhl této úrovně vzdělání?
Atd.	Atd.

[Burgess, 2001:8]

Jak jsme viděli, při definování výzkumného problému se objevuje množství otázek. Je nesmírně důležité správně určit, které z nich k rozklíčování souvislostí, o které se ve výzkumu zajímáme vedou a jsou tedy relevantní a těmi ostatními, které jen navyšují počet otázek v dotazníku, dělají ho nepřehledným a prodlužují délku dotazování. To je nejen nepříjemné pro respondenta, ale i neekonomické. Vybíráme tedy otázky, které mají potenciál získat důležité odpovědi a jsou tematicky relevantní. Odfiltrováním irelevantních otázek hned na začátku výzkumu se vyhneme situaci, kdy jsme vynaložili časové a finanční prostředky na sběr dat, ze kterých pro analýzy nebo závěrečnou zprávu použijeme pouze část. V ideálním případě bychom se měli ptát pouze na otázky řešící cíle výzkumu a otázky typu „zeptáme se, protože by mohlo být zajímavé to vědět“ nedotazovat. Ačkoliv byla již dávno prokázána souvislost mezi délkou a response rate dotazníku, lze se v praxi setkat s dotazníky, které přesahují čas dotazování 45 minut. V případě, že se v mnoha případech rozhovor podaří (variantně za použití odměny pro respondenta) úspěšně dokončit, nabízí se otázka, zda v pozdějších sekcích dotazníku dostáváme validní odpovědi nebo odpovědi respondenta, který chce mít již takto zdlouhavou proceduru co nejdříve za sebou a proto dělá vše pro to, aby čas dotazování co nejvíce zkrátil. O problému ovlivnění výsledků nízkou response rate se rozepisují v jedné z dalších částí textu.

9.4 Pilotáž, testování dotazníku

Pilotáž dotazníku je jako výzkumná fáze často opomíjena, i když zvláště u výzkumů prováděných poprvé může mít velký vliv na výslednou kvalitu výzkumného nástroje. Velmi zkráceně - pilotáží ověřujeme, zda dotazník funguje tak, jak jsme si představovali. Pilotáží nebo předvýzkumem je nejčastěji míněna zkrácená forma hlavní studie. Tato forma by měla mít naprosto stejný design jako její následující část. V praxi bychom tak ideálně dotazník v rámci pilotáže předložili respondentům se shodnými charakteristikami jako jsou ti, kteří byli vybráni pro hlavní studii a měli bychom je dotazovat ve srovnatelných podmínkách. Pokud zjistíme jakoukoliv nesrovnalost, máme šanci dotazník opravit. Toto už není možné ve fázi, kdy jsme dotazníky vytiskli a rozeslali tazatelům do terénu. Kontaktování tazatelů s upřesněním dotazového úkolu je nákladné a někdy i bohužel technicky neproveditelné.

Při pilotáži nás zajímá:

- Obsahuje otazník takové otázky, na které se tazatel musel respondenta znovu ptát?
- Jaké komentáře se vyskytly ze strany tazatelů a respondentů ve spojitosti s vyplňováním dotazníku? Je možné získat odpovědi na všechny otázky navrženým způsobem?
- Objevila se otázka, které respondenti častěji nerozuměli? Objevila se otázka, které tazatelé nerozumí nebo jim její čtení dělá potíže?
- Měli respondenti problémy rozumět otázkám? Vykládají si je správně nebo je tazatelé musí usměrňovat?
- Působí dotazník na respondenty pozitivním dojmem, motivuje je k odpovědím?
- Pokud se jedná o otázku sahající do minulosti respondenta – jak dobře si byl schopen vybavit odpověď?
- Jak dlouho trval přibližně jeden rozhovor? Co rozhovor nejvíce protahuje?

Pokud již při pre-testu dotazníku zasílaného poštou struktura testovaného nástroje odpovídá zamýšlené struktuře pro hlavní studii, můžeme získat představu nejen o item nonresponse⁴² slabých položek v dotazníku, ale můžeme také odhadnout response rate. Přesnost takového odhadu samozřejmě záleží na rozsahu testu. Nelson uvádí, že k odhalení špatného znění otázek a nesprávné skladby dotazníku postačí testovat dotazník již na 30-ti respondentech. Pokud chceme generovat relevantní předkódované nabídky odpovědí u semi-otevřených otázek, musíme do testování zahrnout 50-100 respondentů a ještě vyšší počet respondentů vidí Nelson jako potřebný při vývoji škál pro vyjádření postojů [Nelson, 1985: 181-182 in Journal of Official Statistics].

Existuje celá plejáda metod testování dotazníku. Za tradiční metody lze jmenovat testování dotazníku na zmenšeném souboru respondentů s následným sezením s tazateli a diskusí nad jejich postřehy (debriefing), expertní posouzení dotazníku, dodatečné zkoušení pochopení významu otázek respondenty (probing), sledování rozhovorů⁴³ (observation of interviews), následné hodnocení rozhovorů (post interview evaluations) [Nelson, 1985: 185 in Journal of Official Statistics]. Vývojově mladší jsou ty metody testování dotazníku, které bych označil za kvalitativní. Tzv. kognitivní laboratorní techniky a behaviorální testování využívají poznatků sociální psychologie. Pozornost si získaly zejména díky experimentům

⁴² Item nonresponse je mírou toho, nakolik respondenti odpovídají na konkrétní otázku nebo položku v baterii otázek. Unit nonresponse je označení pro celkovou míru atraktivity výzkumu. Unit označuje jednotku výzkumu ať už je jí respondent nebo domácnost a udává, zda se tato jednotka výzkumu po oslovení zúčastnila nebo ne.

⁴³ Příkladně CATI náslechy rozhovorů

prováděných týmy Belsona, Tourangeaua, Oksenberga, DeMaio a dalších [srovnej s Schaeffer, 1995: 83 in Journal of Official Statistics]. Z klasických kvalitativních metod výzkumu se, jak ukáží, mohou uplatnit také skupinové diskuse (focus groups) [Desvousges et al, 1989: 349-363 in Journal of Official Statistics]. Podrobněji si představíme ty nejpoužívanější. Podíváme se, jaké mají možnosti využití a jaké jejich limity musíme brát v úvahu.

9.4.1 Následné hodnocení rozhovorů

Provádějí ho tazatelé po ukončení testu. Tazatelé musí být schopni shrnout své pocity z průběhu testování do formuláře tak, aby jejich zpráva byla srovnatelná se závěry jiných tazatelů. Tazatelé se možná shodnou na tom, že určitý typ respondentů neodpovídá na konkrétní otázky nebo že podstatná část respondentů se doptávala na skutečný význam konkrétní otázky, že požadovala vysvětlení odborného výrazu. Následně mohou být tazatelé přeškoleni a dotazník ve svých problematických bodech upraven. Mohu potvrdit, že následné hodnocení lze využít také pro posouzení míry osvojení dotazníku tazatelem. Před časem nás britská výzkumná společnost požádala o to, zda bychom vyslechli nahrávky rozhovorů, které z Velké Británie uskutečňovali tazatelé české národnosti. Tazatelé telefonovali českým zákazníkům společnosti na výrobu chemikálií, kteří měli vyjádřit míru spokojenosti s touto společností ve velmi konkrétních, technických parametrech. Vedoucí oddělení telefonického dotazování s dlouholetými zkušenostmi práce s tazateli měla posoudit nejen zda tazatelka dotazuje předepsaným způsobem, ale také v jakých úsecích dotazníku respondenti vyjadřují nejistotu nad tím, co se po nich žádá a jak tazatelka zvládá takové situace. Na konci každého dne jsme posílali podrobnou zprávu hodnotící nejen výkony jednotlivých tazatelů jako podklady pro školení, ale také zpětnou vazbu k reakcím respondentů na dotazník. Z pohledu testování výzkumného nástroje je metodologie CATI mezi ostatními technikami dotazování výjimečná v tom ohledu, že umožňuje dotazník v případě nutnosti ladit v průběhu sběru dat.

9.4.2 Behaviorální testování

Pro znázornění možností využití této metody testování dotazníku jsem se rozhodl reprodukovat výsledky případové studie, jejíž závěry v roce 1991 se svými kolegy publikoval Lois Oksenberg [Oksenberg et al, 1991: 349-365 in Journal of Official Statistics]. Dotazník se zabýval problematikou zdraví a obsahoval 60 otázek na toto téma. Pro behaviorální testování bylo určeno 60 rozhovorů provedených 6-ti tazateli, pro dodatečné zkoušení

(probing) určil tým 104 rozhovorů a 9 tazatelů. Pro účely testování dotazníku a práce tazatelů byla navržena sada kódovacích klíčů, které vystihují možné alternativy nabídnutí otázky tazatelem a alternativy následného chování respondenta (viz. tabulka 3). Nahrávky telefonických rozhovorů kódovali zkušení tazatelé tak, že mohli jednomu rozhovoru přiřadit i více kódů.

Tabulka 3

Kódy pro chování tazatele	Popis kódu	Vysvětlení významu kódu	Index chybovosti
E	Přesné	Čtená otázka se neodlišuje od vzoru v dotazníku	-
S !	Změněné	Čtená otázka se mírně odlišuje od vzoru v dotazníku, ale změna nemění význam otázky	12%
M !	Zavádějící	Čtená otázka se natolik odlišuje od vzoru v dotazníku, že to mění její původní význam NEBO Tazatel nečte otázku v celém rozsahu	4%
Kódy pro odpověď respondenta			
Kódy pro odpověď respondenta	Popis kódu	Význam kódu	
1 !	Přerušení odpovědi	Respondent odpovídá dříve, než byla dočtena otázka	4%
2 !	Ujasnění	Respondent žádá upřesnění nebo zopakování otázky	10%
3	Správná odpověď	Odpověď respondenta uspokojuje nároky otázky	-
4 !	Kvalifikovaná odpověď	Odpověď respondenta uspokojuje nároky otázky, ale mohla by být výstižnější	7%
5 !	Nesprávná odpověď ⁴⁴	Odpověď respondenta neuspokojuje nároky otázky	24%
6 !	Respondent na otázku nezná odpověď	-	4%
7 !	Respondent na otázku odmítl odpovědět	-	0%

! indikuje možný problém s otázkou

Zdroj: [Oksenberg in Journal of Official Statistics, 1991:352]

Poté, co bylo všech 60 rozhovorů opatřeno příslušným kódem, bylo možné spočítat jakýsi index chybovosti, který udává četnost výskytu konkrétního druhu chyby napříč všemi 60-ti otázkami. Kódy E a 3 se v indexu nepromítají, protože nesignalizují ohrožení kvality dotazování ani dotazníku. V tabulce 3 jsou průměrné skóre indexu zobrazeny v posledním sloupci tabulky. Jistě by nebylo problémem zjistit parciální hodnoty indexu pouze za kódy

⁴⁴ V originálním znění „inadequate answer“. Kahn a Kannel [1957:217-218] in Moser a Kalton [1971:276] rozlišují celkem pět variant nesprávné odpovědi: 1 – částečně správná odpověď, 2 - non-response (respondent na otázku neodpoví), 3 – irelevantní odpověď (odpověď na něco jiného než na co se ptá otázka), 4 – neakurátní odpověď (odpověď je zkomolená nebo zkreslená), 5 – sdělení problému (respondent vysvětlí tazateli, proč na otázku nemůže odpovědět. Důvodů může být několik, ale nejčastěji proto, že otázce nerozumí, nemá dostatek informací k zodpovězení otázky, považuje ji za irelevantní)

pro chování za jednotlivé tazatele a takové kritérium použít pro hodnocení jejich práce a podklad pro další výcvik. Kumulované hodnoty indexu v části odpovědí respondenta nám signalizují, zda se v dotazníku vyskytují otázky vyvolávající konkrétní reakce. Stejně jako v předchozím případě s tazateli nás pro odhalení problémů v dotazníku bude zajímat % rozložení kódů u jednotlivých otázek. Oksenberg si jich pro tento účel vybral osm a jejich kvalitu nechal zobrazit v tabulce 4.

Tabulka 4

Kódy pro chování tazatele/respondenta	Popis kódu	Číslo otázky							
		1	2	3	4	5	6	7	8
S	Změněné	2	8	30	8	3	7	5	0
M	Zavádějící	3	19	17	0	2	0	8	2
1	Přerušeni odpovědí	0	35	23	0	0	0	0	0
2	Ujasnění	2	3	10	3	3	27	30	10
4	Kvalifikovaná odpověď	5	3	3	22	27	12	3	0
5	Nesprávná odpověď	5	8	17	13	87	18	30	77
6	Respondent na otázku nezná odpověď	0	0	8	3	12	40	3	5
7	Respondent na otázku odmítl odpovědět	0	0	2	0	0	0	0	0

Zdroj: [Oksenberg in Journal of Official Statistics, 1991:354]

Jak vidíme, pouze první z osmi náhodně vybraných otázek nepůsobila tazatelům a respondentům problémy. Při čtení druhé⁴⁵ i třetí⁴⁶ otázky byli tazatelé často přerušováni. Oksenberg má za to, že respondenti si v polovině otázky mysleli, že již nemá pokračování nebo už si mysleli, že ji dokáží zodpovědět bez dalšího upřesnění. Čtvrtou otázku⁴⁷ sice respondenti nedokázali zodpovědět jen ve třech procentech případů, ale každý pátý respondent vyjádřil obavy o tom, zda je jeho odpověď dostatečně podložená. K šesté otázce „Jaké jsou podle Vás průvodní znaky rakoviny?“ respondenti jednak požadovali často další ujasnění nebo na ni téměř v polovině případů nevedli odpověď. Hypotéz má Oksenberg několik. Je možné, že respondenti nemají představu tom, co jsou to průvodní znaky⁴⁸. Dále je možné, že se respondenti s takovou informací již setkali, ale nemohou si ji vybavit. Je možné otázku více specifikovat a omezit tak případy, kdy respondenti budou potřebovat ujasnění otázky. Pokud se ale respondenti s podobným druhem informace ještě nesetkali

⁴⁵ „Kdy naposledy jste byl(a) na prohlídce u svého lékaře – minulý měsíc, před jedním až šesti měsíci, před šesti až dvanácti měsíci nebo před více než rokem?“

⁴⁶ „Kolik jste lékaři při Vaší poslední návštěvě zaplatil(a)? Nezapočítávejte sem prosím poplatky, které proplatila nebo proplatí pojišťovna. Pokud si nevzpomínáte na přesnou částku, prosím Vás alespoň o odhad.“

⁴⁷ „Řekl(a) byste, že na svůj věk jste fyzicky více aktivní, méně aktivní nebo stejně aktivní jako ostatní lidé ve Vašem věku?“

⁴⁸ v originále „warning signs“

nebo se s ní setkávají tak vzácně, že si ji nedokáží vybavit, není v silách dotazníku od nich tuto informaci získat. Sedmá otázka zjišťovala kdy naposledy se nechal respondent ošetřit a vztahovala se k tématu péče o zuby, které bylo rozebíráno těsně předtím - v otázce na počet návštěv zubního lékaře v minulém roce. Vysoký počet případů žádostí o dovysvětlení otázky lze přičíst tomu, že respondenti možná rychle ztratili souvislost s minulou otázkou a bylo by jim bývalo třeba zdůraznit, že se stále bavíme o ošetření u zubaře. Stejně vysoký podíl nesprávných odpovědí Oksenberg dává za vinu vágní formulaci otázky, které nenabízí žádné možnosti a tím nevyomezuje prostor respondentova uvažování. Jak ale ukazuje otázka osmá, ani kompletní nabídka možností nemusí být zárukou pro to, aby je respondenti používali. Otázka se dotazuje na možné příčiny srdečních onemocnění. Respondentovi je postupně nabízeno 7 možných zdrojů onemocnění a pro každý z nich se má vyjádřit, zda riziko onemocnění rozhodně nebo spíše zvyšují, případně snižují. Respondenti ve více než polovině případů (viz tabulka 4, sloupec hodnocení osmé otázky) odpověděli buď rozhodně nebo spíše, ale už nedodali zda zvyšuje nebo snižuje. Oksenberg tento příklad uvádí jako chybu v nabídce odpovědí, já ji považuji jako dobrý příklad nebezpečnosti otázek pracujícími se sety výroků k ohodnocení. Pokud má respondent odpovídat na tyto typy otázek, je pochopitelné, že (i když jsou tazatelé instruováni k opakování kompletní nabídky odpovědí po každé čtvrté až páté položce) si odpovědi začne dříve či později zjednodušovat a je povinností tazatele v takových případech doptat se na pravý význam jeho odpovědi. To samozřejmě prodlužuje délku dotazování.

Pokud se u otázky často objevují kódy chyb typu 1 nebo 4, můžeme problém vyřešit poměrně nenáročnou úpravou znění otázky. Po shlédnutí výsledků testování otázek se výzkumný tým rozhodl změnil znění například u otázek č. 2 a 3 takto:

„Byl(a) na prohlídce u svého lékaře minulý měsíc, před jedním až šesti měsíci, před šesti až dvanácti měsíci nebo před více než rokem?“

„Následující otázka se týká toho, kolik jste lékaři zaplatil(a) při poslední návštěvě. Když nepočítáte poplatky, které proplatila nebo proplatí pojišťovna, kolik jste lékaři zaplatil při poslední návštěvě?“

Dále bych chtěl ukázat, jak důležité je testovat otázky i poté, co již byly upraveny. Experiment provedl tým výzkumníků pod vedením Theresy DeMaio [1998: 345-346 in de Vauss] v letech 1994-1995. během první vlny ověřování funkčnosti otázek metodou behaviorálního kódování se jako problematické jevíly dvě otázky v tomto znění:

„Kolik Vaše domácnost utratila v minulém týdnu za potraviny v supermarketu, kdybyste nepočítal(a) útratu za jiné zboží než potraviny, jako jsou čisticí prostředky, kancelářské potřeby, krmivo pro domácí zvířata nebo cigarety?“

„Kolik Vaše domácnost utratila minulý týden za potraviny při nákupech v obchodech s potravinami nebo večerkách?“

Podle dat z kódování chování respondentů na první ze dvou otázek dokázalo poskytnout smysluplnou odpověď pouze 70% z nich. Ukázalo se, že respondenti se často soustředili více na upřesnění otázky místo vlastního jádra otázky a odpovídali tazatelům v tom smyslu, že nekouří a podobně. Použité upřesnění otázky také mnohým respondentům zakrylo časový údaj v otázce, takže tazatele museli žádat o upřesnění, jaké době se útrata v obchodě vztahuje. Druhá otázka v kombinaci s první se v tomto znění ukázala jako nepoužitelná. Pro některé respondenty je supermarket a obchod s potravinami to samé. Ve druhém kole testování byly respondentům předloženy otázky v takto upraveném znění.

„Kolik Vaše domácnost utratila v minulém týdnu při nákupech v supermarketech nebo obchodech s potravinami?“

Kolik z této sumy (doplňte číselný údaj z první odpovědi) jste utratili za jiné zboží než potraviny, například za čisticí prostředky nebo kancelářské potřeby?

I po úpravách reflektujících výsledky první vlny testování otázky působily problémy asi čtvrtině respondentů. Odpovědi se nevztahovaly k útratám z minulého týdne, ale často k průměrné hodnotě útraty za týden a do této částky byly navíc zahrnovány položky za útratu jinde než v supermarketech a obchodech s potravinami!

9.4.3 Dodatečné zkoušení pochopení významu (probing)

Co když respondent otázce nerozumí? Jak naznačují závěry stejného Oksenbergova výzkumu, je možné postupovat cestou probing, neboli dalšího „zkoušení“ respondenta. Můžeme se ptát jak si respondent vykládá význam otázky - kam si myslí, že otázka směřuje. Další cestou je požádat respondenta o další rozvedení jeho odpovědi. Je také možné chtít od respondenta co nejsobitější popsání problému, který s otázkou má. Tímto je možné včas odhalit dezinterpretace významu otázky. Oksenberg touto metodou testoval osmnáct otázek a zjistil, že přibližně 15% respondentů si vykládá špatně význam dvanácti z nich a přitom tyto lidé nevyjadřovali jakoukoliv pochybnost o chápání významu otázky! Podívejme se na následující otázku:

„Za posledních 12 měsíců, tzn. od ledna 1987, kolik dní jste kvůli nemoci nebo zranění zůstal(a) více než polovinu dne na lůžku?“

Tazatelé se po dokončení rozhovoru respondentů zeptali, jak interpretují časový údaj „polovina dne“ a setkali se s poměrně pestrým výčtem odpovědí, které zahrnují možnosti jako od rána do oběda, 2-4 hodiny nebo 12 hodin. Dále tazatele zajímalo, zda respondenti počítají do odpovědi „zůstat na lůžku kvůli nemoci“ také stav, kdy respondent zůstává na lůžku, protože cítí, že by mohl onemocnět. Zjistilo se, že takto otázku interpretují pouze dvě třetiny respondentů. V jiném testu dodatečného ověření interpretace otázky se tazatelé ptali těch respondentů, kteří na otázku

„Věnoval(a) jste se v posledních čtyřech týdnech některé fyzické aktivitě jako je cvičení, sport nebo záliby vyžadující pohyb?“

odpověděli „ne“ zda měli ve zmiňovaném období vůbec nějaký pohyb. Třetina respondentů, kteří původně odpověděli ne, připustili, že chodili na procházky. Jak se zjistilo dodatečným zkoumáním odpovědí⁴⁹, stejný podíl respondentů se kvalifikoval do kategorie „ano“ právě na základě chození na procházky. Tak můžeme dopadnout v případě, že nedostatečně jasně definujeme předmět otázky. V tomto případě jsme špatně definovali koncept fyzické aktivity a získáváme nevalidní odpovědi.

Jiný výzkumný tým provedl jinou studii v domácnostech v roce 1981 [Sykes et al, 1987: 191-207 in Journal of Official Statistics]. Dotazované byly hlavy domácnosti, z toho 60 z nich bylo vybráno pro experiment s využitím metody dodatečného zkoušení s tím rozdílem oproti Oksenbergerově výzkumu, že za respondenty chodili jiní tazatelé než ti, kteří s nimi vedli rozhovor a to s jednodenním zpožděním, vybaveni dotazníkem vyplněným respondentem dané domácnosti. Respondent byl detailně dotazován na chápání jedné se sedmi vybraných experimentálních otázek a jen zběžně bylo ověřeno, že správně pochopil význam zbylých šesti otázek. Jako velice problémová se ukázala následující otázka, na kterou respondenti odpovídali s pomocí karty:

⁴⁹ Tazatel se na respondenta obrátil s otázkou „Řekl(a) jste, že jste v uplynulých 4 týdnech cvičil(a), sportoval(a) nebo provozoval(a) záliby vyžadující pohyb. Mohl(a) byste mi o tom říci více?“

„Co si myslíte o množství hluku pocházejícího z osobních aut, nákladních vozů a jiných dopravních prostředků v místě Vašeho bydliště?“

Jsem rozhodně spokojen(a)	1	(23%)
	2	(12%)
	3	(12%)
Nejsem spokojen(a) ani nespokojen(a)	4	(25%)
	5	(11%)
	6	(3%)
Jsem rozhodně nespokojen(a)	7	(13%)
Nevím	8	(1%)

Zdroj: [Sykes in Journal of Official Statistics, 1987:198]

Čísla v závorkách u kódů odpovědí udávají rozložení odpovědí na otázku po pořízení a vyhodnocení odpovědí z celkem 544 rozhovorů⁵⁰. Při pohledu na rozložení si všimneme, že u kódů se slovními popisky se soustředí vyšší počet odpovědí. V padesáti procentech opakovaných testů se zjistilo, že respondenti nepochopili účel kódů bez popisky a proto je nezvolili. Protože původní rozhovory byly nahrávány na pásek, bylo možné zpětně dohledat způsob prezentování otázky tazatelem. Porovnáním jejich přepisů a originálního znění otázky se zjistilo, že různí tazatelé si otázku různě modifikují, což bylo zdrojem signifikantní odchylky mezi hodnotami dat otázky za různé tazatele. Tvořivost tazatelů dosáhla vrcholu v tomto příkladě, kdy si tazatel vyložil otázku svérázným způsobem.

„Na této kartě vidíte škálu od jedné do sedmi. Kdybyste bydlel(a) na venkově a nebyl(a) byste obtěžován(a) hlukem, pravděpodobně byste zvolil(a) možnost odpovědi 1. Pokud byste ale bydlel(a) na rušné hlavní třídě, pravděpodobně byste zvolil(a) možnost odpovědi 7. Kdybych se Vás zeptal(a), kam byste se na této škále zařadil(a), co byste řekl(a)? Jaké číslo?“

Při experimentálním, opakovaném čtením otázky požadovalo 10% respondentů další vysvětlení jejího významu, 17% odpovědí nebylo zařaditelných do stávajících možností odpovědí. To všechno jsou informace, které mají zásadní vliv na interpretovatelnost dat a které bychom nezjistili jinak než například dodatečným ověřováním pochopení významu otázek.

9.4.5 Kognitivní rozhovory

Metoda testování otázek během kognitivních rozhovorů spočívá v podněcování respondenta k tomu, aby o plánovaných odpovědích mluvil nahlas. Podle DeMaio [1998: 340 in de Vauss] je možné metodu kombinovat s výše popisovanou metodou probing s tím, že ze má být opravdu dáno nejvíce prostoru respondentovi samotnému a dodatečné otázky pokládá

⁵⁰ druhá, experimentální návštěva provedena pouze v 60-ti případech

výzkumník pouze v případě nejasné odpovědi. Dodatečné otázky je možné pokládat v okamžiku, kdy k nim byl dán podnět nebo po ukončení rozhovoru. Je výhodné se respondenta doptat v okamžiku, kdy si ještě dobře pamatuje důvody konkrétní odpovědi, DeMaio si ale myslí, že taková interakce s tazatelem může vést k ovlivnění způsobu jeho dalšího odpovídání.

DeMaio a Wellens měli šanci v roce 1997 testovat otázky pro výzkum Census 2000 Dress Rehearsal. Dvacetí kognitivními rozhovory byly například testovány dvě varianty téže otázky zjišťující, zda se v domácnosti nachází někdo s fyzickým nebo duševním postižením a jak je toto postižení vážné. Dvě varianty se lišily především nabízeným formátem.

Varianta otázky A

„Pro každou z následujících aktivit zvolte tu variantu, která nejlépe popisuje nakolik je postižená osoba ve Vaší domácnosti omezena v jejím vykonávání:“

		Bez omezení	Určité omezení	Velké omezení
A	Výkon duševních činností jako učení, pamatování si, soustředění se...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	Oblékání se, koupání se, pohyb po bytě/domě bez pomoci druhé osoby...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	Návštěva lékaře nebo provedení nákupu bez pomoci druhé osoby...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

„Trpí postižená osoba některými z následujících dlouhodobých omezení?“

		Ano	Ne
A	Slepota nebo závažné onemocnění zraku	<input type="checkbox"/>	<input type="checkbox"/>
B	Hluchota nebo závažné onemocnění sluchu	<input type="checkbox"/>	<input type="checkbox"/>
C	Omezením fyzické výkonnosti, která ji omezuje například při chůzi, zdolávání schodů, zvedání předmětů, přenášení břemen atd.	<input type="checkbox"/>	<input type="checkbox"/>

Varianta otázky B

„Vlivem fyzických nebo duševních potíží trvajících 6 a více měsíců, má postižená osoba ve Vaší domácnosti potíže s vykonáváním následujících činností?“

		Ano	Ne
A	Učení se, pamatování si, soustředění se	<input type="checkbox"/>	<input type="checkbox"/>
B	Mluvení, zrak (s brýlemi), sluch	<input type="checkbox"/>	<input type="checkbox"/>
C	Chůze nebo přenášení nákupu	<input type="checkbox"/>	<input type="checkbox"/>
D	Práce nebo údržba domácnosti	<input type="checkbox"/>	<input type="checkbox"/>
E	Návštěva lékaře nebo provedení nákupu bez pomoci druhé osoby...	<input type="checkbox"/>	<input type="checkbox"/>
F	Oblékání se, koupání se, pohyb po bytě/domě bez pomoci druhé osoby...	<input type="checkbox"/>	<input type="checkbox"/>

Protože DeMaio mohla sledovat respondenty při vyplňování dotazníku a mohla sledovat jejich myšlenkové pochody, odhalila, že první verze otázky přináší do výzkumu vinou špatného formátování chybu měření. Respondenti si po vyplnění prvního bloku otázek mysleli, že i nadále, zaznamenáním odpovědi v levém sloupci deklarují zdravotní neomezenost postižené osoby v domácnosti v hodnoceném ohledu.

Druhá varianta otázky sice netrpí špatným formátováním, respondenti ale mají problémy s její interpretací. Potíže působilo především slovní spojení „zrak (s brýlemi)⁵¹“, které si respondenti vykládali různě. Jedna skupina respondentů měla za to, že tento impuls indikuje potíže vidění i s nasazenými brýlemi. Jiní ho interpretovali jako potíže s viděním a odtud potřeba nosit brýle. Respondenti, kteří nosí brýle nesprávně zvolili možnost odpovědi „ano“. Čili toto je chyba měření projevující se nadhodnocením hodnoty proměnné. Ve stejné otázce se ale vyskytl zrcadlově obrácená ukázka chyby měření, když respondenti s evidentní vadou řeči neindikovali takový problém ve své odpovědi.

9.4.6 Využití skupinových diskusí při hodnocení kvality otázek

Poslední metodou testování otázek, které se chci věnovat, jsou skupinové diskuse. Skupinové diskuse jsou vedle individuálních hloubkových rozhovorů⁵² nejpoužívanější výzkumnou technikou z kategorie kvalitativních výzkumů. Jak ale ukazuje výzkum Desvousgese a Freye [Desvousges a Frey, 1989: 349-363 in Journal of Official Statistics], mohou být s úspěchem použity také při testování otázek pro kvantitativní typy výzkumů. Abych byl přesný, ve výzkumu zaměřením na vnímání nebezpečí nebezpečného odpadu metodou modelování hypotetických situací skupinové diskuse měly pomoci v těchto úkolech:

1. vyvinout vizuální karty pro demonstrování modelových situací
2. vyhodnotit vhodnost použití alternativ designů studie
3. modelovat vhodné modelové situace
4. určit nejlepší pořadí otázek ve struktuře dotazníku
5. navrhnout efektivní způsob oslovení respondentů

Prvním úkolem bylo respondentovi na modelových situacích dobře vysvětlit, jaká rizika představuje nebezpečný odpad a přimět ho k tomu, aby na základě pochopení takové situace vyjádřil jím vnímanou úroveň rizika hrozícího jemu a jeho blízkým. I když kvalitativní

⁵¹ v originále „see (with glasses)“

⁵² in-depth interviews

výzkum sledoval více cílů, my se podíváme jakým vývojem prošly během cyklu šesti skupinových diskusí koncepty karet znázorňující jednak nebezpečnost odpadů a také karty nastiňující modelové situace.

Na schématu 4 příloh si můžeme prohlédnout, jak vypadala jedna z karet k výzkumu před a po testování. Pro autentičnost jsou schémata prezentována v původní podobě. Tato karta chtěla komunikovat riziko smrti při vystavení nebezpečnému odpadu, ale ve své původní verzi byla nepoužitelná. Diskutující k ní měli hned několik výtek počínaje tím, že vyšrafovaní části kruhové výseče není nejpřehlednější cestou znázornění pravděpodobnosti. Dále nebylo účastníkům diskuse zcela zřejmé, jakým způsobem byla odvozena kombinovaná pravděpodobnost. V neposlední řadě se moderátor diskuse dozvěděl, že udávané příklady rizik se jen málo vztahují k rizikům vystavení nebezpečnému odpadu. Diskuse nad kartami s modelovými situacemi naznačily nepoužitelnost karet s obecným líčením představitelných alternativ a jako dobré řešení se neukázaly ani karty zabíhající do přílišných detailů, protože respondenti na nich ulpívají a neodpovídají na jádro otázky. Jako přínosné se ukázalo používat karty popisující výhradně situace reálné v daném regionu dotazování a zasadit typ nebezpečného odpadu do reálií respondentova bydliště. První a finální verze karty je v původním znění k nahlédnutí ve schématu 5 příloh.

Myslím si, že ze všech uvedených příkladů uvedených v sekci testování dotazníku by měla být zřejmá důležitost pilotních výzkumů. Několikrát jsme viděli, jak samotné odpovězení na otázku neznámá, že respondent otázce rozumí. Když pomyslíme, jak malá část ostrých dotazníků prochází fází testování na skutečných respondentech, musíme se ptát jaké výsledky zadavatelům prezentujeme a bojím se přemýšlet nad odpovědí. Na druhou stranu musím zdůraznit dvě vlastnosti všech metod testování otázek bez ohledu na to, zda se jedná o debriefing nebo focus groups. Testování dotazníku se neděje ve vzduchoprázdnu, ale v reálném čase, který musí někdo zaplatit. Čas, prostředky a zkušenosti výzkumného týmu jsou veličiny, které se podílejí na kvalitě výzkumu. Dovolím si formulovat lehce kontroverzní myšlenku. Kvalita výzkumu měřeno kvalitou výzkumného nástroje je otázkou kompromisu. Zadavatel si musí uvědomit, jaké jsou jeho priority. Potřebuje znát požadované informace rychle a levně nebo je ochoten si za jejich kvalitu připlatit a nechat výzkumný tým déle pracovat na zhodnocení výzkumu? Pokud je pověřená instituce tlačena k první variantě, je pouze na etice výzkumného týmu, zda na základě expertního odhadu dokáže posoudit dotazník jako funkční a dobře měřící nebo zda „zavře oči“ a nechá ho projít terénem. Samozřejmě nic není černobílé a již jednou otestovaný dotazník může v opakovaném výzkumu získat požadované informace rychleji a levněji. Z vlastní zkušenosti mohu ale říci, že netestované dotazníky se používají také v kontinuálních studiích.

9.5 Design dotazníku

Cannel a Kahn [in Moser a Kalton, 1971:271] vidí nejméně tři nutné podmínky pro úspěšný rozhovor. Dostupnost informace je první z nich. Pokud respondent odpoví na otázku nemá, můžeme se ptát z jakého důvodu. Informaci mohl znát, ale zapomněl ji; nebo to byl druh informace vyvolávající emoční stres a proto její vyslovení potlačil; nebo ji možná i zná, ale otázka nebyla položena tak, aby mu ji připomněla. Nutnou podmínkou je i rozpoznání toho, o jaký druh informace tazatel usiluje. Z otázky, na kterou má odpovědět, musí být zřetelné jak podrobné odpovědi a v jakých souvislostech se má tazateli dostat. Tento moment jsme dostatečně rozebrali v kapitole věnující se pilotáži dotazníku. Poslední ale neméně důležitá podmínka je motivovanost respondenta. Tazatel musí dokázat přesvědčit respondenta, aby svolil k rozhovoru, ale ochotu respondenta pokračovat v rozhovoru musí udržovat po celou dobu jeho trvání.

Ještě před zahájením vlastního dotazování by proto mělo dojít k navázání důvěry mezi respondentem a tazatelem. Děje se tak většinou představením výzkumné organizace, která výzkum provádí a seznámením respondenta se základními etickými pravidly každého výzkumu. Samozřejmostí je poučení respondenta o tom, že účast na výzkumu je zcela dobrovolná a že účast může odmítnout, dále pak poučení o způsobu zacházení s daty a osobními údaji respondenta. Tazatel by měl pokračovat informováním respondenta o předmětu výzkumu. Za slušnost se považuje uvedení pravdivého údaje o přibližné délce rozhovoru. Tento údaj vznikl již při pilotáži dotazníku.

Doporučuje se začít dotazník otázkami, které získají pozornost respondenta, které zaujmou. Pokud působí úvodní otázky nepřesvědčivě, budí v respondentovi pocit, že je zkoušen nebo v něm dokonce otázky vyvolávají negativní pocity (zlost, strach, ...), šance na úspěšné provedení rozhovoru se zmenšují. Stejně tak jako v utváření dojmu o lidech hrají největší roli první okamžiky, měl by i dotazník respondenta hned zpočátku vtáhnout do výzkumné problematiky. Neméně důležité je ho v tomto stavu co nejdéle udržet. Těmito technikami se ještě budeme dále zabývat. Zde uvedu jen některé z nich. Otázky musí být jednoznačné. Otázky musí být jednoduché pro pochopení. Otázky musí být srozumitelné a výstižné. Musí být stručné. To vše snižuje nebezpečí nedorozumění mezi tazatelem a respondentem. K nežádoucím chybám dochází i na straně tazatelů. Cestou ke zpřehlednění dotazníku může být zvýraznění a odlišení otázek od pokynů pro tazatele, průběžné opakování způsobů zaznamenávání odpovědí, uvedení instrukcí pro práci s kartami k výzkumu, upozornění na problematická místa dotazníku aj.

Klíčové otázky výzkumu by měly být obsaženy v první polovině dotazníku. Takové opatření oceníme v případě, že respondent se z nějakého důvodu rozhodne předčasně ukončit rozhovor. Dotazové archy sice nebudou vyplněny kompletně, ale o důležité informace nepřijdeme.

Řekl jsem, že respondentovu pozornost je nejen třeba získat, ale i udržet. Jedním ze způsobů jak udržet respondentovu pozornost je vytvoření několika různých tematických okruhů. S tím souvisí řazení otázek do kontextu ve správné logické posloupnosti. Další možností může být střídání formátu otázek, vyhýbání se tzv. „setům“, což jsou velké bloky tematicky podobných otázek nebo položek v rámci jedné otázky.

10. Chyby při tvorbě dotazníku

Když jsem měl před několika lety navrhnout dotazník pro svůj první výzkum - univerzitní výzkum postojů studentů k práci a zaměstnání, stál jsem před problémem výběru „těch správných“ otázek. Mohl jsem se podívat například do datového archivu Sociologického ústavu AVČR, zda byl již výzkum na podobné téma již někdy proveden a zda by nebylo možné hledat inspiraci pro výběr otázek právě zde. Udělal jsem to a ukázalo se to jako dobré rozhodnutí. Nicméně jsem zamýšlel u studentů zjišťovat takové informace, které dříve podle výsledků mého hledání nikdo nezjišťoval a proto bylo potřeba vymyslet způsob, jak se na tyto údaje správně zeptat. V té době jsem poprvé narazil na problém, se kterým se potýkáme při každém pokusu o zachycení nového výzkumného problému relevantními a validními otázkami. Jak dále uvidíme, existují případové studie, které referují o úspěšnosti testovaných variant otázek. My sice vidíme, jakých výsledků která otázka v daném případě dosáhla, ale máme problém tyto závěry zobecnit tak, aby mohly být používány dále. Výsledky testu otázky jsou platné pro daný čas a danou populaci. Neexistuje registr dobrých otázek stejně tak jako neexistuje soupis těch špatných. Jediné co máme, je několik málo obecně platných pokynů. Už jsem bohužel nedohledal, kdo vyslovil výrok, který naši situaci pěkně ilustruje a který by zněl asi takto. Kdybychom pověřili 10 velmi zkušených odborníků, aby vytvořili sadu otázek měřící zcela konkrétní koncept, na konci stěží najdeme dvě stejné.

Dostáváme se do fáze, kdy bych chtěl některé zkušenosti s přípravou dotazníků demonstrovat na konkrétních příkladech. Nejprve se zaměřím na některá závažná témata z okruhu problematiky správného znění otázek. Jako názornou pomůcku použiji závěry výzkumů, o kterém se podrobně rozepisuje Labaw [1980 bez odkazu na stranu textu] in Gendall a Hoek [2002 bez uvedení stránkování – jedná se internetovou prezentaci výsledků]. Jednalo se o sérii šesti výzkumných šetření, které v rámci programu zahraniční

politiky USA proběhly v období let 1976-1978 a zabývaly se problematikou správy Panamského Kanálu. I když výzkumný tým v každé ze šesti vln měnil znění otázek (nikoliv význam), výsledky těchto vln nevykazují signifikantní odchylky v získaných datech. V tabulce 5 příloh se můžeme podívat na výsledky tohoto dvouletého experimentu. Labaw odtud usuzuje, že znění otázek má pouze minimální vliv na výsledky. Toto platí do té doby, dokud otázky vyvolávají stejné nebo podobné emoce. Šest verzí otázky naměřilo v různých vlnách výzkumu srovnatelné výsledky. Pak ale tým otázku změnil vnesením zcela nového prvku a výsledné odpovědi jsou signifikantně odlišné (viz zvýraznění ve spodní části tabulky 5). Je to bezpochyby dobrá zpráva pro ty, kteří dotazníky navrhují. Jejich práce by se stala extrémně obtížnou, pokud by i jen sebemenší změna ve znění otázky vedla k jiným výsledkům. Nicméně, nebylo by správné utvářet si generalizující představy na základě jednoho výzkumu.

Na několika dalších příkladech bych chtěl ukázat, že jediný způsob, jak vytvořit ucelený a spolehlivý soubor pravidel stavby otázek je jejich testování v různých obměnách, v různém pořadí, v různých kontextech a to až do doby, kdy budeme schopni s jistotou říci, která pravidla mají obecnou platnost a která jsou specifická pro různé situace. Především se zaměřím varianci výsledků plynoucích z volby mezi otázkami

1. uzavřenými a otevřenými,
2. směřujícími do minulosti respondenta
3. senzitivními, např. ve výzkumech zjišťující příjmy jednotlivců a domácností nebo výzkumy se zdravotní/sexuální tematikou
4. pozitivně a negativně orientovanými,
5. vlivem jedné otázky na jinou otázku

a dalšími rozdílnostmi výsledků způsobenými změnami ve znění otázek, především

6. zavedením nového konceptu do otázky a
7. obrácením výroku.

Zmíním se také o vlivu pořadí otázek (vliv jedné otázky na druhou), vlivu neutrálního bodu mezi možnostmi odpovědi a téma bych uzavřel výčtem nejčastějších chyb dotazníků, tzn. kladení otázek dvojitých, negativních, zavádějících (navádějících).

10.1 Uzavřené a otevřené otázky

Uzavřené otázky těší velké oblibě jak v komerčních průzkumech trhu tak ve výzkumech společensko-vědních a v nejrůznějších dotaznících nad otázkami otevřenými početně převažují. Existuje pro to hned několik důvodů. Uzavřené otázky v dotazníku mají respondenti „raději“ než otázky otevřené, protože nad možnostmi odpovědí nemusí přemýšlet tak dlouho jako v případě otázky otevřené. Tato skutečnost je spíše než rozumným argumentem psychologickým momentem. Respondent má pocit, že mu zodpovězení uzavřené otázky zabere méně času než v případě otázky otevřené. Ve skutečnosti může respondentovi složitě kódovaná otázka působit v široké nabídce odpovědí orientační potíže. Uzavřené otázky jsou méně náročné na zpracování ve fázi pořizování dat. V případě, že dotazník je obsáhlý, uzavřené otázky ještě dále neprodlužují čas dotazování a méně výmluvní respondenti nejsou znevýhodněni.

Naopak nelze zamlčet některá nezpochybnitelná negativa spjatá s jejich využíváním. Jak jsem již naznačil, tyto nedostatky můžeme částečně zmírnit zařazením poslední možnosti odpovědi „jiné-upřesněte“. Nekompletní nabídkou spektra odpovědí riskujeme zkreslení výsledků výzkumu. Ztrácíme spontánnost a další vysvětlení respondentových odpovědí. Pokud nabízíme příliš mnoho odpovědí, riskujeme variantu odpovědi vybrané náhodně jako důsledek rezignace na rozumnou volbu ze strany respondenta. Tato metoda je také vystavena nezamýšlené chybě zaznamenání odpovědi tazatele (zaznamenání jiného kódu odpovědi). Nemáme jistotu, že všichni respondenti rozumí nabízeným možnostem odpovědí stejně. Nemáme jistotu, že respondenti odpovědi neuvidí (u otázek bez použití karet nebo nabídnutí odpovědí). Navíc víme, že tazatelé mají tendenci napovídat. U nabízených možností odpovědí mají již dopředu některé z nabízených variant řekl bych pozitivní hendikep, který spočívá v jejich „výhodném“ umístění v nabídce odpovědí. Mosteler [1949: 70 in Moser a Kalton, 1971:345] dokázal, že varianty odpovědí umístěné v čele a na konci výčtu odpovědí jsou respondenty voleny častěji. Moser s Kaltonem, tentokrát samostatně, pozorují, že u otázek, kde respondenti volí pouze ze dvou variant odpovědi, pozitivních a negativních, si respondenti raději vyberou tu první. Tento efekt se někdy označuje jako „acquiescence response set“, což lze chápat jako set souhlasných odpovědí. [Moser a Kalton, 1971:345]. Tomuto efektu se můžeme bránit trochu dvojsečnou zbraní negativních výroků, kterými prokládáme sety pozitivně orientovaných výroků a spoléháme na to, že tím respondenta zachráníme před upadnutím do monotonního odpovídání jedním způsobem (ano nebo ne/ souhlasím nebo nesouhlasím). Nebezpečí negativních výroků spočívá v tom, že si respondent negaci otázky nemusí uvědomit. Na nepoužitost otázek s dvojitými

negacemi snad netřeba upozorňovat. Negativně orientovaným otázkám a opačným výroklům se ještě budeme věnovat.

Podobné srovnání kladů a záporů je třeba provést také pro otázky otevřené. Otevřenou otázku zařazujeme zpravidla tehdy, pokud se chceme dozvědět spontánní znalost respondenta. Proto ji vždy v kombinaci s otázkou uzavřenou (předkódovanou) zařadíme jako první. K výhodám otevřených otázek můžeme s jistotou zařadit širší pole pro svobodné vyjádření respondentových odpovědí. Respondenti mohou vyzdvihnout témata, která výzkumný tým předtím nebral v úvahu. V odpovědích na otevřenou otázku se mohou objevit cenná emocionální zabarvení, která pomohou pochopit zjišťované příčiny a souvislosti. Otevřené otázky nevedou ke zkreslení reality „náповědou“ předem vybraného znění odpovědí. Obecně jsou otevřené otázky vhodnější zařadit v případě, že se ve zkoumané problematice nedostatečně orientujeme.

Na druhou stranu i otevřené otázky jsou pronásledovány některými negativy. Odpovědi z otevřených otázek jsou jednoznačně složitější zpracovatelné než odpovědi na uzavřené otázky. Výzkumník provádějící její kódování pracuje s velmi nesourodým souborem výroklů, které zařazuje do kategorií na základě subjektivního chápání jejich významů. Samotné zodpovězení otázky závisí na míře ochoty respondenta spolupracovat. Ve výsledku odpovědi na otázky leží na jiných úrovních obecnosti. U tištěných dotazníků mohou být některé odpovědi respondentů nečitelné. O tyto odpovědi samozřejmě přicházíme. Tlak odpovědět na otevřenou otázku je menší než u otázky uzavřené. Svědomité vyplnění otevřené otázky je obvykle časově náročnější než zodpovězení otázky uzavřené. Míra pečlivosti zápisu je různá u různých tazatelů. Jak kontrolovat, že tazatelé zapisují všechny informace poskytnuté respondentem? Zapisují je v původním znění nebo si je agregují podle potřeby? Je jasné, že zde je jeden zdroj možného zkreslení výsledků ze strany tazatele. Tomu je možné se částečně bránit instruováním tazatelů k zápisu otevřených otázek v původním znění. Takové výpisy si někdy zadavatelé nechávají posílat pro kontrolu práce nejen tazatelů, ale také toho, kdo otevřené otázky kódoval. Je nutné přiznat, že takový postup je zvláště u výzkumů s papírovými dotazníky časově náročný. Všechny otevřené otázky se musí znovu elektronicky pořizovat. Neméně časově náročná (ve fázi vyhodnocování) je alternativní technika zkvalitňování výsledků otevřených otázek, kterou popisují Moser a Kalton [1971: 281]. Spočívá v pořízení audio záznamu rozhovoru, což je samozřejmě exaktnější způsob kontroly pořízení věrného znění odpovědi než spoléhání se na to, že tazatelé respektují pokyn k doslovnému zápisu. Pokyny tazatelům ale na rozdíl od přítomnosti nahrávacího zařízení nemají negativní vliv projevující se nižší response rate. Navíc máme podle Mosera s Kaltonem důvod domnívat se, že nahrávání rozhovoru má vliv

na pravdivost odpovědí poskytovaných respondentem. Citují závěry výzkumu provedého Belsonem, který nahrával odpovědi na otevřené otázky ve výzkumu se čtenáři denního tisku a zjistil, v tomto případě měli tendenci zakrývat skutečný stav věcí respondenti z vyšších sociálních skupin. Opačný efekt pozoroval na druhém konci společenského žebříčku.

Otevřené otázky mají svou nezastupitelnou úlohu při pilotních studiích. Mohou nám pomoci definovat škálu odpovědí na otázku, kterou plánujeme zařadit ale nechceme dopředu respondenty svazovat nevyhovujícím výčtem odpovědí. Není od věci rozhodnout se pro otevřenou variantu otázky v případě, že dotazujeme takové téma výzkumu, které není pro respondenta zcela běžné, ve kterém se třeba ne úplně orientuje a nabídkou odpovědí bychom ho sváděli k výběru jakékoliv varianty, jen aby odpověděl a nevypadal před tazatelem špatně.

Díky již provedeným studiím máme možnost porovnat, jakého rozložení odpovědí se při položení otázky stejného významu dočkáme pokud respondentovi varianty odpovědí nabídneme a s jakou frekvencí se budou tytéž odpovědi objevovat jako spontánní výroky. V prvním příkladě se jedná o novozélandský omnibusový výzkum z roku 1989, realizovaný studenty Massey University, fakulty Marketingu. Cílovou populací byly domácnosti v regionu Palmerston North. Více podrobností o výzkumu podávají Gendall a Hoek [2002 bez uvedení stránkování – jedná se internetovou prezentaci výsledků]. Zvláštností tohoto výzkumu bylo rozdělení výběrového vzorku na 2 skupiny se stejnými demografickými charakteristikami. Každá z těchto dvou skupin odpovídala na jinou verzi dotazníku. V komentáři k metodologii výzkumu se lze dočíst, že i když oba soubory byly jen přibližně stejně zastoupené (N1=347, N2=311), byly převažovány podle proměnných pohlaví a věk. V první verzi dotazníku respondenti odpovídali na otevřenou otázku tohoto znění:

„Jakému nejzávažnějšímu problému podle Vás v současné době čelí Nový Zéland?“

Ve druhé verzi dotazníku byla otázka změněna na uzavřenou s předkódovanými možnostmi odpovědí:

„Podívejte se na následující kartu. Který z těchto problémů je podle Vás nejzávažnější a Nový Zéland mu v současné době čelí?“

Respondenti vybírali z těchto možností odpovědí: AIDS , inflace, vysoký kurz měny, zákonnost, výše úrokové míry, nezaměstnanost, ekonomika obecně, rasové problémy. Výsledky tohoto experimentu je možné srovnat v tabulce 6.

Tabulka 6

	Verze dotazníku 1 – uzavřená otázka (v %)	Verze dotazníku 2 – otevřená otázka (v %)
Nezaměstnanost	56	50
Ekonomika obecně	22	11
Zákonnost	9	5
Rasové problémy	7	10
AIDS	6	-
Neví	1	-
Nestabilní vláda	-	5
Upadající morálka obyv.	-	3
Jiné*	-	16
Celkem	100	100

*Zahrnuje: školství, drogy, nedostatek křesťanské víry, lenost, zhroucení rodiny

Výsledky experimentu naznačují, že ačkoliv existují rozdíly v zastoupení odpovědí v řádu procentních bodů, celkové spektrum odpovědí na uzavřenou a otevřenou otázku stejného smyslu by nemuselo být výrazně jiné. Vidíme, že největší vážnost je přisuzována v obou případech problému nezaměstnanosti. V jiných problémech nenajdeme významnějšího rozdílu. Celkem logicky se ve výčtu odpovědí na druhou (otevřenou) verzi otázky setkáváme s jinými než navrženými možnostmi odpovědí. To nám připomíná další triviální vlastnost uzavřených otázek – odpovědi jsou již předem ovlivněny omezeností nabídky odpovědí. V praxi se tento problém řeší nabídnutím možnosti odpovědi „jiné - upřesněte“, což je vlastně vnořená otevřená otázka. Používá se v situaci, kdy dotazujeme tematiku, která ještě není dokonale zmapovaná nebo když chceme zachytit jedinečnost odpovědí různých skupin respondentů.

S předchozím výsledkem novozélandského omnibusu se zásadně rozchází závěry studie Kaltona a Schuman [in de Vauss, 2002:220], kteří se svých respondentů zeptali na důležitost atributů jejich zaměstnání (podle stejného scénáře jako Gendall a Hoeková). Otevřená otázka byla tohoto znění:

„Následující otázka se bude týkat Vaší práce. Lidé pracují z různých důvodů. Čeho si na práci nejvíce ceníte Vy osobně?“

Podoba otázky s předkódovanými možnostmi odpovědí:

„Následující otázka se bude týkat Vaší práce. Podívejte se prosím na tuto kartu a řekněte mi, co z této nabídky nejvíce vystihuje to, čeho si na práci Vy osobně nejvíce ceníte?“

Tabulka 7

	Verze dotazníku 1 – uzavřená otázka (v %)	Verze dotazníku 2 – otevřená otázka (v %)
Má práce je pro mě důležitá ⁵³	59,1%	Práce mi dává podněty ⁵⁴ 21,3%
Šance na povýšení	17,2%	Příležitost k postupu 1,8%
Vysoký příjem	12,4%	-
Jistota zaměstnání	7,2%	-
Méně práce a více volného času	3%	-
Pracovní podmínky	-	14,9%
Příjemná práce	-	15,4%
Práce, kterou dělám, mě těší	-	17%
Ostatní	1%	29,6%*
Celkem	100	100

Zdroj: [Kalton, Schuman in de Vauss, 2002:220]

*Kalton a Schuman do kategorie „Ostatní“ řadí také procenta odpovědí u položek „Práce mi dává podněty“ a „Příležitost k postupu“. Protože jim ale sami našli ekvivalenty („Práce, kterou dělám, je pro mě důležitá“ a „Šance na povýšení“), považují za logické je do „Ostatních“ nezahrnovat.

Odpovědi na otevřenou otázku daly vzniknout dalším třem silně zastoupeným kódům odpovědí. Naopak pokud respondentovi nebyly předsunuty první dva důvody, skórují tyto výrazně slaběji. V dalším pokračování experimentu byla zařazena otázka z předkódovanými možnostmi odpovědí z otevřené otázky (bez nabídky respondentovi; pouze pro účely snadnějšího zaznamenání odpovědi pro tazatele) a překvapivě, silné kódy definované v prvním kole dohromady vysvětlily nyní pouze 58% procenta variability odpovědí!

U případových studií v sociálních vědách není výjimkou, že se jejich závěry rozcházejí. Nemůžeme se divit. Lidé se často chovají jinak, než říkají. Tento jev lze rozdělit na dvě části. První, která je nasnadě, je to, že prostě neříkají pravdu. Respondenti často neříkají to, co si myslí, protože jsou zvyklí upravovat skutečnost ve svém každodenním životě a jsou přesvědčeni o tom, že drobná lež je ku prospěchu věci. Nechtějí vyvolávat konflikty. Například mnoho lidí není zvyklých říkat upřímně to, že jsou s něčím nespokojeni. To pak platí dvojnásob, pokud by tuto nepříjemnou pravdu měli říci někomu přímo do očí. Právě v případě, že agentura provádí průzkum pro dodavatele, pro respondenty je to téměř totéž, jako říci tuto nepříjemnost dodavateli přímo.

Druhá část jevu, že se lidé chovají jinak než říkají, spočívá v tom, že si konkrétní důvody svého chování neuvědomují. Typicky to je například vliv cen, který je obvykle nadhodnocován na úkor jiných, často daleko významnějších vlastností výrobku nebo služby. Na rozdíl od problému zakrývání skutečnosti zde máme naopak mnoho možností, jak

⁵³ V originále „The work is important and gives feeling of accomplishment“

⁵⁴ V originále „Stimulating work“

odpovědi pro interpretaci zpracovávat. Tyto možnosti jsou empirické, problém lze však řešit také s využitím statistických metod. Respondenti nejsou laboratorní myšky. Když nám v rozhovoru řeknou, že určitým způsobem jednají nebo přemýšlí, nemusí to nutně znamenat, že to takto funguje v opravdovém životě. Netvrdím, že nám tvrdí nepravdu záměrně. Můžeto tak být proto, že v daný moment dané odpovědi věří nebo věřit chtějí. Také my, kteří se jich na informace dotazujeme, nepracujeme v laboratorních podmínkách. Když se kteréhokoliv občana z regionu Prahy zeptáme na nejzávažnější problémy v místě jeho bydliště, s velkou pravděpodobností nebude na prvním místě jmenovat nezaměstnanost. Nezaměstnanost je v hlavním městě dlouhodobě nejnižší v porovnání s ostatními regiony České republiky. Možná bude jmenovat problémy s dopravou nebo korupci. Opakuji se, ale chci tím znovu připomenout, že výsledky případových studií z oboru společenských věd, sociologické studie nevyjímaje, mají velmi nízkou schopnost být univerzálně platné.

10.2 Otázky směřující do minulosti

Nevyhňeme se tomu, abychom se respondentů museli ptát na informace z jejich minulosti. Odpovědi, které získáme, jsou ovlivněny minimálně dvěma faktory. Důležitá je doba, která uplynula mezi tím, kdy došlo k události (věk při koupi prvního osobního vozu, poslední návštěva lékaře, počet zahraničních dovolených v posledních pěti letech) a okamžikem dotazování. Důležitost, kterou respondent přisuzuje události, je druhým faktorem [Moser a Kalton, 1971:331]. Je celkem logické, že respondent události, kterým respondent přisuzuje menší význam, jsou zapomínány dříve. I pro události, které jsou pro nás z nějakého důvodu významné, funguje naše paměť selektivně, takže si některé podrobnosti pamatujeme více než ostatní složky téže události. Tazatelům potom jako respondenti prezentujeme neúplné záznamy.

Moser na příkladu výzkumu čtenosti, poslechovosti a sledovanosti médií, popisuje, jak lze dosáhnout přesnějších odpovědí s použitím karet. Poté co se tazatel zeptá respondenta, které noviny četl předchozí den, předloží mu kompletní seznam deníků vydávaných v daném místě a zaznamená všechny odpovědi respondenta. Obdobně postupuje v případě měření poslechovosti rozhlasu a sledovanosti televize. Kromě problému nápovědy a rizika o kterém Moser nemluví a které spočívá v tom, že respondent vybere ze seznamu cokoli, aby se předvedl v lepším světle jsou tu minimálně dva problémy spojené s omezeností paměti respondenta. Prvním je zapomenutí informace. Druhý, označovaný jako telescoping, nastává když respondent hlásí události, které nastaly před měřením

obdobím. Případě otázky na čtenost včerejších deníků by nahlásil titul, který četl již někdy dříve.

S problémem telescopingu se lze vypořádat několika způsoby. Při opakovaných rozhovorech v panelových studiích může tazatel přijít za respondentem vybaven souhrnem odpovědí z minulé vlny. Pokud respondent nahlásí konkrétní událost znovu, může tazatel připomenout respondentovi, jaké události hlásil při poslední návštěvě a společně ověří, zda se nejedná o duplicitu. Neter s Waksbergem [in Moser a Kalton, 1971:340] změřili vliv telescopingu právě metodou porovnání s již existujícími záznamy u amerických domácností v roce 1965. Výzkum zjišťoval kolik a jak často domácnosti vydají za opravy domů. Po měsíci od prvního dotazování u oprav převyšujících hodnotu zakázky 100 dolarů naměřili incidenci o 55% vyšší u domácností dotazovaných bez korekce na duplicitu respondentem. V intervalu tří měsíců od prvního dotazování byla incidence už „jen“ o 26% vyšší.

Jiným způsobem ošetření problému s omezenou pamětí respondentů je využití deníčkové metody, při které domácnost samostatně zaznamenává relevantní události hned při jejich výskytu. Pokud se domácnost podaří přemluvit ke spolupráci, je nejtěžší na celé metodě zajistit, aby domácnost zaznamenávala události předepsaným způsobem a v čas výskytu události, ne retrospektivně. Moserem zmiňovaná nevýhoda minimálně dvojí návštěvy respondenta v domácnosti (jednou při rekrutaci do panelu, podruhé při odevzdání deníku) lze eliminovat změnou designu ve smyslu zaslání vyplněných deníků poštou. Větší problém s metodou záznamu do deníku je psychologický vliv přítomnosti deníku. V kapitole Metody výzkumu jsem popisoval situaci, ve které si hospodyně uvědomovala, že je sledována tazatelem a pravděpodobně uzpůsobila své nákupní chování. Deníček může mít podobně negativní vliv. Některé domácnosti si možná nevedou statistiku měsíčních výdajů a proto při zjištění, kolik vlastně za některé produkty a služby vydávají, své jednání změní.

Pokud by schopnost odpovídat na otázky sledující historické události byla stejně dobrá nebo stejně špatná u všech skupin respondentů, byla by to lepší zpráva než ta, kterou na základě výzkumu mezi populací seniorů sepsal tým Knäuperové [Knäuper et al, 1997:181-199 in Journal of Official Statistics]. Ve svém výzkumu vyšli z předpokladu, že zodpovězení jakékoliv otázky respondentem, je závislé na úrovni jeho kognitivních schopností. Mezi kognitivní schopnosti se řadí schopnost provádět procesy zdůvodňování, chápání a řešení problému, ale také pamatování si. Kromě toho, že starší respondenti mají tendenci odmítat odpovědi na určité druhy otázek [Colsher a Wallace, 1989 in Knäuper et al, 1997:182] a mají tendenci odpovídat na škálové otázky stereotypně [Andrews a Herzog, 1986 in Knäuper et al, 1997:182], bylo hned několika výzkumy předvedeno, že kognitivní schopnosti se s rostoucím věkem horší [Gergen a Beck, 1988 in Knäuper et al, 1997:182;

Herzog a Dielman, 1988 in dtto a další]. Závěry výzkumu Knäuperové to jen potvrzují. Proč je to problém? Pokud budou respondenti s menšími kognitivními schopnostmi méně častěji odpovídat na složitější otázky nebo otázky směřujícími do minulosti budeme mít potom tyto skupiny respondentů v datech podhodnoceny. Zvláště ve výzkumech se zdravotní tematikou, která se týká sice všech skupin populace, ale starších lidí přednostně.

Je zřejmé, že v otázkách na zdravotní historii, ale třeba i profesní nebo vzdělanostní dráhu nám karty ani deníčkové metody pomoci nemohou. Východiskem ze situace by snad mohla být metoda časové přímky. Princip metody je možná časově náročný, ale velmi jednoduchý pro pochopení i provedení respondentem. Tazatel požádá respondenta, aby na připravenou kartu zaznamenal všechny významné události ve sledovaném období. V průběhu rozhovoru a při odpovídání na otázky z minulosti se respondente na časovou přímku může kdykoliv podívat. Předpokládá se, že umístěním událostí různého charakteru pomůže respondentovi osvěžit již zapomenuté souvislosti a snáze si vybaví jejich správnou chronologii [Jobe a Mingay, 1989 in Van den Vaart, 2004: 303 in Journal of Official Statistics]. Van der Vaart provedl sekundární analýzu holandských dat na vzorku 1257 respondentů z dlouhodobého sociologického výzkumu z let 1987 a 1991. Kopii časové karty pořízené respondentem a upravené pro účely publikace je možné vidět na schématu 6.

Schéma 6

1984						1985						Age = 24	
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July		
x-----science---X													
X-----sociology-----X													
										X-car dealer--O			
accident			wedding			party			Spain				
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July		
1985						1986						Age = 25	
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July		
X-sociology--						X---didactics-----X							
O- car dealer---X												Xunem pl.oyed X---teacher-----O	
			L.A.			Dory			Mexico				
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July		

zdroj: [Van den Vaart, 2004: 315 in Journal of Official Statistics]

Výzkum se týkal vzdělávání a dotazník obsahoval otázky zjišťující počet navštěvovaných vzdělávacích kurzů, rok zahájení kurzu a typ vzdělávání. Možnost vytvořit časovou přímku byla nabídnuta jen části respondentů, aby mohl být odhadnut přínos techniky. U vzorku respondentů, kteří kartu použili, bylo možné sledovat signifikantně lepší výsledky u otázek zaměřených na počet a typ vzdělávacích kurzů. Stejně jako Van der Vaart neumím odhadnout, zda za zlepšením stojí samotná možnost zaznamenat si relevantní události ve spojitosti s jinými mezníky v životě respondenta, možnost přizpůsobit si časovou kartu vlastním požadavkům respondenta a udělat ji libovolně podrobnou nebo jen psychologický vliv aktivnější role. Určitě se ale ukázalo, že použití časové přímky u konkrétního typu otázky je přínosné z hlediska úplnosti získávané informace.

10.3 Senzitivní otázky

Pokud senzitivní otázky nejsou nezbytné pro zjištění vhodnosti respondenta pro dotazování, zbytečně riskujeme odmítnutí spolupráce. Někteří lidé neradi hovoří na osobní témata jakými jsou příjem (ať už domácnosti, hlavy rodiny nebo respondenta samotného) nebo úroveň dosaženého vzdělání. Citlivé otázky zařazujeme do poslední části dotazníku. Obvykle respondenta informujeme o tom, že tyto otázky, stejně tak jako jeho ostatní odpovědi, budou zpracovány hromadně a jeho jméno nebude s odpověďmi spojováno. V případě, že se tyto bezpochyby osobní údaje podaří získat, mohou později posloužit jako popisná nebo třídící proměnná v datech, vysvětlující proměnná (pohlaví, věk,...) nebo proměnná k určení reprezentativity vzorku, pokud citlivá témata nejsou předmětem výzkumu.

Už jsem několikrát zmínil, že otázky by měly být co nejpřímější, aby pokud možno nenechávaly respondentovi žádný prostor pro interpretaci a chybné pochopení významu. Dotazování citlivých témat jako zdravotní stav, příjmy, sexuální život atd. vyžaduje jiný přístup. Víme, že v případě takových témat, které zasahují do respondentova soukromí, získáme pravdivou informaci spíše za použití nepřímých otázek nebo projektivních metod. A nebo, jak se dočteme v kapitole *Developing the questionnaire* u Blankenshipa [Blankenship et al, 1992: 214-215] se můžeme pokusit zeptat se přímou otázkou poté co jsme respondenta navedli k problematice nepřímou. V jednom ze svých výzkumů⁵⁵ měl Blankenship zjistit míru konzumace piva v jednotlivých demografických skupinách. Obával se, že by minimálně částí konzumenti nechtěli uvádět skutečně konzumované objemy. Proto nejdříve

⁵⁵ Podrobnosti o výzkumu: A.B.Blankenship. 1961. *Creativity in consumer research*. Journal of Marketing 25, s.34-38.

respondentům (muži a ženy ve věku 18+ let) položil zdánlivě nesouvisející set otázek a až poté se jich zeptal na množství piva, které týdně spotřebují.

Q1. Jaká je Vaše oblíbená značka mýdla? Jaká je Vaše oblíbená značka zubní pasty? Jaká je Vaše oblíbená značka piva?

Q2. Nyní budeme mluvit o množství piva, které konzumujete. Do jaké z těchto kategorií byste se zařadil(a)? Řekněte mi pouze písmeno vyhovující kategorie.

A, Velmi silný konzument (více než 60 lahví nebo plechovek piva týdně)

B, Silný konzument (49-60 lahví nebo plechovek piva týdně)

C, Nadprůměrně silný konzument (37-48 lahví nebo plechovek piva týdně)

D, Průměrný konzument (25-36 lahví nebo plechovek piva týdně)

E, Podprůměrně silný konzument (13-24 lahví nebo plechovek piva týdně)

F, Slabý konzument (7-12 lahví nebo plechovek piva týdně)

G, Velmi slabý konzument piva (6 nebo méně lahví nebo plechovek piva týdně)

Blankenhipa samozřejmě nezajímalo, čím se respondenti myjí nebo jakou značku zubní pasty používají. Spoléhal na to, že respondenti nepostřehnou přechod od obvyklého spotřebního zboží k pivu a dále že touto otázkou bude moci rozlišit konzumenty a ty, kteří pivo nepijí. Druhá otázka je pouze pro konzumenty a nabízená karta je navržena tak, aby si respondent při pohledu na realistické a průměrně konzumované množství piva (kategorie G až max. C) uvědomil, že je stále ještě „nejhůře“ nadprůměrně silný konzument. Slovní popisky u rozmezí konzumovaného piva mají tedy pouze psychologický efekt. Navíc při hlášení odpovědi jako písmene značící kategorii respondent nemusí udávat konkrétní čísla. Protože výsledky extrapolované na populaci v regionu výzkumu přibližně odpovídaly množství prodaného piva v dané době (porovnání se záznamy prodejců), lze říci, že taková metoda dotazování je úspěšná.

Kategorie výzkumů zabývajících se sexuálním chováním tvoří ještě specifitější kategorii mezi senzitivními tématy výzkumů. Sběr dat v podobných výzkumech patří k těm nejtěžším. Kromě toho, že se jedná velmi osobní témata, spjatá s postavou a do jisté míry image respondenta, jsou otázky sexuality velmi ostře sledovány okolní společností. Například americká společnost je zatížena určitým tabu homosexuality. Připouštění potratů nebo prostituce je považováno za nepřipustné [Smith, 1992:309 in Journal of Official Statistics] a není proto divu, že navázat vztah důvěry je pro tazatele velmi obtížné.

Aby se respondentům ulehčilo v odpovídání na takto náročné otázky, můžeme použít již zmíněné projektivní metody, při kterých se respondenta neptáme přímo na jeho postoje, ale na to, co by řekl někdo v jeho postavení nebo co si myslí jiní lidé. Moser a Kalton [1971:327] popisují jiné tři techniky. Respondentovi se předloží kreslený příběh, včetně textu, které jednoduché postavy říkají. Je úkolem respondenta doplnit chybějící text u postav, které zatím mlčí. Dále se může použít metoda dokončování výroků v určitém

časovém intervalu, protože nám jde o spontánní respondentovu odpověď a nechceme mu dát šanci, aby měl čas své spontánní myšlenky zakrýt.

Nejsofistikovanějším nástrojem dotazování senzitivních otázek je pravděpodobně metoda náhodného dotazování položek⁵⁶ vyvinutá Warnerem [Warner, 1965 in Moser a Kalton, 1971:328] a rozvíjená Greenbergem [Greenberg et al, 1969 in Moser a Kalton, 1971:328], ve které počítač se známou pravděpodobností vybírá ze série výroků, mezi kterými jsou i senzitivní výroky a předkládá je respondentovi k posouzení. Všechny výroky směřují pouze ke dvěma variantám odpovědi - souhlasím nebo nesouhlasím. Tazatel zapisuje odpovědi zatímco neví, na jakou otázku vlastně respondent odpovídá. Poměrně jednoduchou cestou [srovnej s Moser a Kalton, 1971:328] se znalostí pravděpodobnosti dosažení každého z výroků je možné odvodit výsledek pro skutečně sledované výroky. Předpokládá se, že když respondent nemá pochyb o tom, že když tazatel nezná výsledek losování výroků, poskytne pravdivější odpověď.

Šanci na dosažení velmi dobrých výsledků z hlediska response rate a pravdivosti odpovědí v senzitivních otázkách mají samovyplňovací dotazníky, jak ukázal Smith [1992: 309-325 in Journal of Official Statistics]. V ročníku 1990 výzkumu sexuálního chování byly mimo jiné dotazovány otázky zjišťující počet a pohlaví sexuálních partnerů respondenta od 18-ti let jeho věku a frekvenci pohlavních styků. Obě tyto otázky nebyly zařazeny v hlavním dotazníku, ale byly předloženy respondentovi na zvláštní kartě, kterou po vyplnění zalepil do obálky a předal zpět tazateli⁵⁷. Bohužel ve výzkumu nebyla kontrolní skupina respondentů, kteří by na obě citlivé otázky odpovídali jako na součást hlavního dotazníku, takže nemáme možnost srovnání, ale response rate 85,5% považuji za velmi solidní výsledek. Z výsledků analýzy non-response obou otázek známe charakteristiky těch, kteří nejčastěji neodpovídali.

⁵⁶ Randomized response technique

⁵⁷ Malé odbočení od Smithova výzkumu NORC 1990. V březnu roku 2004 proběhl na Slovensku F2F výzkum s názvem Sexuálne správanie sa obyvateľstva SR. Oproti designu měl některá vylepšení, ačkoliv si to v té době vedoucí výzkumu možná neuvědomovali. Zprvce byli respondenti oslovováni velmi specifickým způsobem: „Takéto projekty a prieskumy sa už v minulosti uskutočnili vo viacerých krajinách sveta, na Slovensku je to prvá štúdia tohto typu. Prieskumy tohoto typu slúžia jednak ako zdroj informácií v oblasti zdravotníctva, prevencie ako aj ďalšieho výskumu. Na základe tohto prieskumu bude vypracovaná odborná vedecká publikácia, o ktorej sa určite dozviete z tlače a bude možné si ju zakúpiť. Rovnako o výsledkoch tohoto prieskumu budete informovaní prostredníctvom médií“. Takové oslovení má vyšší šanci na upoutání respondentovy pozornosti než standardní způsob, protože poukazuje na další využití výsledků pro lékařské nebo vědecké účely. Výraznější výhodou byla ale nabídka respondentům, aby samovyplňovací část dotazníku neodevzdávali v případě nedůvěry tazateli, ale aby ho poslali poštou na adresu instituce provádějící výzkum.“

1. Jsou to stejní respondenti, jejichž frekvence neodpovídání nebo zadávání odpovědi „nevím“ byla vysoká také v otázkách hlavního dotazníku měřícího postoje.
2. Response rate u citlivých otázek v samovyplňovacím modulu je signifikantně vyšší u respondentů s vyšším vzděláním.
3. Statistické testování nepotvrdilo vazbu mezi sexuálními postoji a response rate.
4. Naopak byla statisticky prokázána souvislost mezi response rate a mírou sexuální aktivity.
5. Nižší response rate zaznamenaly také ženy a starší lidé (provázané s nižší mírou sexuální aktivity ve skupině starších lidí).

Kompletní přehled proměnných a jejich vazeb na non-response je ve schématu 7 příloh. Zájemcům o jiné metodologické výzkumy sledující podobné cíle doporučuji výživný text Rogera Tourangeaua „Sources of Error in a Survey on Sexual Behaviour“⁵⁸.

10.4 Pozitivně a negativně orientované otázky, opačné výroky

Nejen, že negativní otázky působí na respondenty hůře než otázky pozitivně orientované, ale vystavujeme se riziku, že v otázce vznikne dvojitá negace. Ta je vlastně v konečném výsledku pozitivním výrokem, takže si nejen přiděláváme práci, ale zároveň, a to je závažnější, mateme respondenta. Zkušenosti z různých výzkumů naznačují, že se respondenti neradi prezentují jako negativně uvažující a proto tíhnou spíše k volbě pozitivnějších/souhlasných odpovědí. Tuto skutečnost musíme mít na paměti především při navrhování konkrétních škál pro odpovědi, ale také při formulaci otázek. Další z experimentálních otázek, zařazených do Palmerstonského omnibusu [Gendall a Hoek, 2002 bez uvedení stránkování] hypotézu o větší oblibě pozitivních odpovědí příliš nepotvrzují. Obě verze dotazníku obsahovaly následující otázku:

„Nyní Vám přečtu několik výroků o králičím mase. Pro každý z výroků mi, prosím řekněte, zda s ním souhlasíte či nesouhlasíte. Pokud Vaše stanovisko nedokážete jednoznačně určit, nezvolíme žádnou odpověď. Souhlasíte nebo nesouhlasíte s tím, že...?“

⁵⁸ [Tourangeau et al, 1997:341-365 in Journal of Official Statistics]

Následně tazatelé doplňovali postupně výroky z přiložené baterie výroků a pro každý z nich zaznamenávali respondentovu odpověď. Výroky se lišily pro obě verze dotazníku. Pracovní hypotéza tohoto experimentu byla, že se podaří naměřit vyšší procento souhlasných odpovědí pro výroky v první verzi dotazníku v porovnání s procentem nesouhlasu s negativními výroky dotazníku verze 2. V ideálním případě, podle logiky otázek, by se údaje ve druhém a třetím sloupci tabulky 8 sobě měly rovnat.

Tabulka 8

	Verze dotazníku 1 – pozitivní výroky (% souhlasu)	Verze dotazníku 2 – negativní výroky (% nesouhlasu)	Rozdíl odpovědí verze1 – verze2
Králičí maso . . .			
Je/není jídlo pro každý den	10	10	0
Je/není vždy dostupné v supermarketu	10	12	+2
Je zdravé/je nezdravé	87	89	+2
Je nízkokalorické/ vysokokalorické	98	96	-2
Je snadné/obtížné na přípravu	87	81	-6
Jeho kvalita odpovídá/neodpovídá ceně	54	61	+7
Hodí/nehodí se pro hosty	51	59	+8
Má příjemnou/nepříjemnou vůni při přípravě	56	64	+8
Je levné/je drahé	61	53	-8
Je z něj málo/hodně odpadu	90	75	-15*

*Rozdíl signifikantní na 95% hladině významnosti

Vidíme, že procentuální rozdíly odpovědí se pohybují v kladných i záporných číslech. I kdyby byly všechny odchylky pozitivní nebo negativní, pouze u jediného z 10-ti výroků je rozdíl signifikantní. Takový výsledek nedovoluje zobecňující úsudek. Takto nejednoznačný výsledek může být vysvětlován tím, že respondenti výroky nepovažovali za opačné. Daleko lepší výsledek z hlediska vysledování trendu nabízí tabulka 9, která je opět porovnáním odpovědí na tentokrát opravdu zrcadlově opačně orientované výroky týkající se domácí politiky Nového Zélandu.

Tabulka 9

	Verze dotazníku 1 – pozitivní výroky (% souhlasu)	Verze dotazníku 2 – negativní výroky (% nesouhlasu)	Rozdíl odpovědí verze1 – verze2
Zákon by měl povolit ⁵⁹ /zakázat ⁶⁰ veřejná slyšení propagující rasismus	13	31	18*
Plavidlům vybavený jadernými zbraněmi by mělo být umožněno /zamezeno návštěvám Nového Zélandu	32	28	-4
Policie by měla/neměla být ozbrojena	31	42	11*
Homosexualita by měla/neměla být považována za zločin	20	29	9*
Potraty by měly být za určitých podmínek legální/ by neměly být legální za žádných podmínek	81	75	-6
Většina příslušníků národnosti Maori má/nemá dobré podmínky pro vstup do naší země	77	72	-5
Lidé by měli/něměli mít právo zvolit, zda budeme nebo nebudeme patřit k měnové unii	89	78	-11*

*Rozdíl signifikantní na 95% hladině významnosti

Rozdíly naměřené v odpovědích na tyto zrcadlové výroky jsou sice výraznější než rozdíly v tabulce 8, nicméně stále nedovolují jednoznačný závěr. Většina z nich nebyla interpretovaná jako logické protiklady.

Protikladným anglickým slově „forbid“ – zakázat a „allow“ – dovolit ve výrocích zjišťující mínění veřejnosti se již dříve věnovali např. Converse & Presser, 1986; Hippler & Schwarz, 1986; Kalton, Collins & Brook, 1978; Kalton & Schuman, 1982; Schuman & Presser, 1981. Zjistili, že více lidí spíše něco „nedovolí“ spíše než bude souhlasit s tím, aby to bylo „zakázáno“. Tuto skutečnost potvrdily i výsledky novozélandského omnibusu když 83% respondentů nesouhlasilo s výrokem, že by zákon měl povolit veřejná slyšení propagující rasismus, ale jen 66% respondentů souhlasilo s tím, že by taková slyšení měl zakázat.

V tomto výzkumu bychom našli ještě další dvojice antonym. Je zřejmé, že odlišně formulované výroky mohou a nemusí přinášet odlišné výsledky a pokud různý výsledek vytvoří, pak ho jen stěží umíme dopředu předpovědět. Proto, pokud provádíme opakované

⁵⁹ v originálním znění „allow“

⁶⁰ v originálním znění „forbid“

šetření na dané téma, snažíme se znění otázek v zájmu srovnatelnosti dat neměnit. Samozřejmě za předpokladu, že znění otázek není v rozporu s pravidly jejich tvorby.

10.5 Vliv jedné otázky na jinou otázku

Jedním z úkolů v již zmiňovaném omnibusovém výzkumu bylo ověřit předpoklad, že na výsledky může mít vliv pořadí, v jakém pokládáme otázky. Obě verze dotazníku obsahovaly otázku zjišťující postoje k osobám nakaženým chudobou AIDS. V první verzi dotazníku jsme mohli najít otázku na míru souhlasu s výrokem, že

„Lidé nakažení AIDS si za své onemocnění mohou především sami“

Druhá verze předkládala k posouzení opačný výrok. Bez závislosti na odpovědi respondenta na tuto otázku následoval pro obě verze identický výrok v tomto znění:

„Společnost s lidmi nakaženými AIDS sympatizuje daleko méně než by měla.“

Opět se zjišťovala míra souhlasu. Pracovní hypotéza je nasnadě. Výzkumníci předpokládali, že v odpovědích se projeví vliv tzv. principu smíření (tzv. phenomenon of acquiescence). Pokud respondent projevil nesouhlas s prvním výrokem (vlastní vina nakaženého) pak by podle tohoto principu bylo pravděpodobné, že bude souhlasit s výrokem komentujícím míru sympatií k těmto lidem ze strany společnosti a naopak. Výsledky testu jsou zobrazeny v tabulce 10.

Tabulka 10	Souhlasím %	Nesouhlasím %	Neví %
Verze 1 „Lidé nakažení AIDS si za své onemocnění mohou především sami.“	38	54	8
Verze 2 „Lidé nakažení AIDS by neměli ze svého onemocnění obviňovat především sami sebe.“	33	44	23
Verze 1 a 2 „Společnost s lidmi nakaženými AIDS sympatizuje daleko méně než by měla.“	62 x 61	20 x 21	18 x 18

I když bychom naměřená procenta mohli považovat za učebnicový příklad rozdílnosti výsledků způsobenou pozitivním a negativním zněním otázky, výsledky testování hypotézy o ovlivnění odpovědí otázkou předchozí jsou negativní. Autoři své závěry srovnávají se závěry studií uskutečněných týmy Converse & Presser, 1986; Crespi & Morris, 1984; Kalton & Schuman, 1982; Schuman & Presser, 1981, které rozdílnost odpovědí zaznamenali. Takto znovu dostáváme nejednoznačný závěr – v určitých případech může dojít k ovlivnění respondenta pořadím otázek, bohužel ale nejsme schopni s jistotou predikovat s jakou pravděpodobností a v jakém směru.

10.6 Zavedení nového konceptu do otázky

Stále čerpáme z omnibusového šetření na Novém Zélandu. Spíše jen pro zajímavost zde uvedu další dva příklady toho, jak změna ve znění otázky může ovlivnit dosažené výsledky výzkumu. První z nich popisuje situaci, kdy ve stejném výzkumu byla ve dvou mutacích dotazníku zařazena otázka na správnost rovného ohodnocení pracovního výkonu mužů a žen.

Tabulka 11	Souhlasím %	Nesouhlasím %	Neví %
Ženy by za stejnou práci měly být ohodnoceny stejně jako muži.	93	6	1
Ženy by za stejnou práci měly být ohodnoceny stejně jako muži i kdyby toto mělo vést k růstu nezaměstnanosti.	78	14	7

Výsledky takového experimentu nás zřejmě nepřekvapí. Pokud zařadíme do otázky takový prvek, který dělá výrok méně jednoznačný a nutí respondenta k posouzení hned dvou informací, přinejmenším se nám část respondentů zařadí do kategorie „Nevím“, v horším případě najdeme odpověď nevyplněnou.

10.7 Obrácení pořadí výroku v otázce

Dva výroky z tabulky 12 se liší „pouze“ pořadím slov „žena“ a „muž“.

Tabulka 12	Souhlasím %	Nesouhlasím %	Neví %
Pro politiku je emočně lépe připravena většina mužů než žen.	17	73	11
Pro politiku je emočně lépe připravena většina žen než mužů.	20	46	34

Za povšimnutí stojí více než ztrojnásobení frekvencí nerozhodných záznamů.

10.8 Některé další zásady bezpečné práce s dotazníkem

Poslední část textu je velice stručným soupisem několika dalších poznatků, které se uplatní při tvorbě každého dotazníku a jejichž nedodržením ohrožujeme reliabilitu výsledků výzkumu.

Již při pilotáži jsme se ujistili o tom, že respondenti otázce rozumí a nevykládají si ji různě. Určitě by se v otázkách neměla objevit cizí slova. Ne všichni respondenti jim rozumí a jejich výklad je mnohdy nejednoznačný. Používání odborných termínů, slangu, výrazů specifických pro určitou kulturní oblast je na zvážení výzkumného týmu. Rozhodnutí závisí na povaze předmětu výzkumu a povaze cílové populace. V každém případě by měl mít tazatel připraveno vysvětlení všech takových termínů v pokynech pro výzkum nebo příručce pro tazatele k výzkumu. Vedle odborných termínů, žargonu a zkratk se ke slovům, která je lepší do otázek nezařazovat, řadí také vybraná přídavná jména a příslovce. Za všechny bych jmenoval např. slova jako obvykle, často, někdy, příležitostně, zřídka kdy, běžný (běžně). Ačkoliv se jedná o jinak často používaná slova, jejich význam může být subjektivní. Smysl otázky musí být respondentovi jasný již po prvním přečtení otázky tazatelem. Většina lidí věnuje větší pozornost závěrečné části otázky. Na to musíme pamatovat zvláště tehdy, když otázka otevírá novou sekci dotazníku a respondent se ještě nestihl zorientovat. Proto se někdy věty, které otevírají nové kapitoly v dotazníku uvádějí odděleně od otázek. Předejdeme tak splynutí otázky s doprovodným textem a nutnosti zopakovat otázku pro respondenty, kterým smysl otázky unikl mezi řečí.

Názvy je třeba používat konzistentně v celém dotazníku, tak, aby respondent nebyl v žádném okamžiku zmaten použitím jiného názvu (pro tentýž předmět, pro tutéž společnost, službu ...). Pokud se nemůžeme vyhnout názvům, pojmům, zkratkám... , které nejsou pro respondenty/tazatele zcela běžné, musíme je velmi dobře vysvětlit.

Vyskytují se v otázce slova nebo slovní spojení, která mají potenciál navést respondenta na určitou odpověď s vyšší pravděpodobností než k jiným možnostem odpovědí? Na to bychom si u každé otázky mělu umět spolehlivě odpovědět. Tyto otázky se v dotazníku nesmí objevit. Otázka typu „*Souhlasíte, jako většina lidí s tím, že naše zdravotnictví se nachází v krizi?*“ mezi takové otázky typově patří.

Otázku také pokud to není nezbytně nutné nezačínáme formulací „*Mohl(a) byste mi, prosím, říci...*“ nebo „*Dokázal(a) byste si vzpomenout...*“. Tyto otázky navozují možnost odpovědi odpovědi ano/ne a již jsme řekli, že odpovědi ano/ne se také snažíme odbourávat.

S dichotomickými otázkami pracujeme velmi často. V zásadě existují dva způsoby dotazování dichotomické proměnné – proměnné nabízející pouze dva kódy odpovědí (např.

ano/ne). V prvním případě respondentovi nabídneme seznam dotazovaných impulsů a zaznamenáváme pouze jeden druh odpovědi. Ve druhém případě pořizujeme odpověď ano/ne u každého impulsu. Ukázalo se, že respondenti ve druhé variantě dotazování častěji odpovídají pozitivně i u impulsů, které by v první variantě nezmínili. Působí zde princip, který jsem popsal již dříve. Respondenti neradi odpovídají negativně.

Obecně se otázky nabízející pouze odpověď ano/ne příliš nedoporučují. Namísto nich je lepší zařadit otázku na intenzitu souhlasu/nesouhlasu. Získáme tak lepší informaci než v případě jednoduché odpovědi ano/ne a respondentovi zodpovězení otázky netrvá o moc déle.

Respondent v žádné fázi dotazování nesmí získat pocit, že jeho odpověď tušíme a že si ji pouze ověřujeme. Tazatelé často z různých (ale většinou časových) důvodů „doříkávají“ odpovědi za respondenta. Takový postup je samozřejmě chybný.

Napovídání odpovědí je taktéž závažná chyba tazatelů, jeden ze zdrojů zkreslení výsledků. Pokud dotazujeme složitější otázky a respondent si není jistý jak odpovědět ne z důvodů nerozhodnosti, ale proto, že se v problematice neorientuje, můžeme mu problém přiblížit pouze obecně. Například v otázce zaměřené na spontánní znalost modelů lehkých užitkových vozů můžeme vysvětlit respondentovi pojem lehkého užitkového vozu, nicméně uvedením příkladu jakékoliv značky nebo modelu se dopouštíme chyby.

Je třeba rozlišovat mezi filtry vstupními a filtry uvnitř dotazníku. Vstupní filtry ověřují, že jsme našli vhodného kandidáta pro rozhovor. Filtrovou otázkou může být například otázka zjišťující věk respondenta, zaměstnání v určitém oboru, dobu uplynulou od jeho poslední účasti na výzkumu trhu, ale také otázky hlídající kvóty určené pro výzkum. Hlavním úkolem vnitřních filtrů je hlídat, aby respondent odpovídal na otázky jemu určené. Je jasné, že nemá smysl ptát se nezaměstnaného člověka na spokojenost se zaměstnavatelem v různých ohledech. Odovědi takového respondenta můžeme samozřejmě ve fázi čištění dat druhotně odfiltrovat což představuje práci navíc při čištění dat. Především ale dotazování irelevantních otázek může vyvolat nepříjemnosti při samotném rozhovoru. V kombinaci s filtry proto používáme pokyny pro tazatele s popisem logických přechodů pro snadnější a rychlejší orientaci v dotazníku.

Úloha škálování je téma samo o sobě a nebudu se jím zde zabývat kromě jednoho momentu. Když přemýšlíme o možnostech odpovědí na určitou otázku a nabízíme jak verbální tak jen numerickou škálu, narážíme na problém středního bodu. Neexistuje závazné pravidlo, které by říkalo že správné je použít škálu se sudým nebo naopak lichým počtem bodů mezi nimiž se objeví jeden ve středu možností odpovědí.

Umístěním neutrálního bodu (např. "neumím posoudit...", "nedokáži rozhodnout...") vždy změní strukturu odpovědí. Výsledky různých výzkumů prokázaly, že také záleží na charakteru středního bodu – jinak na respondenty působí popiska "nevím" a jinak "ani dobrý ani špatný". Vliv má také to, zda neutrální bod fyzicky umístíme do středu nebo až na konec škály, kde má šanci zlákat ještě více "příznivců".

V praxi se ukazuje, že když má respondent možnost "uniknout" k neutrálnímu bodu, často to udělá. Proto některé výzkumné týmy ve svých dotaznících možnost "neví" ani neutrální bod nenabízejí, i když tazatel pole pro záznam vytvořené má. Předpokládá se, že respondent do dotazníku nevidí (což je například dobře ošetřeno ve výzkumech metodologií CAPI). Ani zde ovšem nepanuje shoda, když někteří výzkumníci naopak tvrdí, že regulérní je obě možnosti transparentně od začátku nabízet. Určitým kompromisem by mohlo být nabízet tyto možnosti ve znalostních otázkách a nepoužívat je v otázkách postojevých.

11. Chyby na straně tazatelů

Variabilita výsledků způsobených tazatelem může mít několik příčin a závažnost variability, jak ukáží, je u různých příčin nestejná. Pokud tazatel špatně zaznamená odpověď, čelíme typické chybě měření. Jak si všimli Moser a Kalton [Moser a Kalton, 1971: 379 in De Vauss], tento typ variability pravděpodobně neznamena systematickou chybu měření. Je totiž nepravděpodobné, že všichni tazatelé zapojení ve výzkumu budou zaznamenávat své *odpovědi podle stejného chybného vzorce*. Pokud by stejného 30-ti letého respondenta dotazovalo 10 tazatelů otázkou zjišťující věk a měli v nabídce odpovědí 5 kategorií – do 17 let, 18-29, 30-44, 45-64, 65 a více let, je výběr chybných kategorií 18-29 a 45-64 stejně pravděpodobný, pokud respondent odpoví všem stejně pravdivě. Nemohu si odpustit krásný příklad podobné chyby záznamu odpovědi, který je velmi aktuální. V projektu EuroBarometr se respondentů ve Slovenské republice ptáme, kterých všech států jsou občany. Znění otázky a podoba záznamového archu je součástí příloh, dotazník 2. Pravidelně se ve výsledcích dozvídáme, že na Slovensku žije přibližně 1%-ní menšina občanů Slovenské národnosti. Takový výsledek nás přirozeně poprvé překvapil. Obyvatel Slovinska, kteří by byli schopni vést plynulý rozhovor ve slovenštině jistě nežije na Slovensku takový počet. Daleko pravděpodobněji se jedná o chybu záznamu odpovědi tazatelem. Slovinsko je v tabulce odpovědí nabízeno hned za Slovenskem. Tyto chyby můžeme logicky odfiltrovat a proto nezpůsobí systematické vychýlení výsledků.

Daleko závažnější chyby tazatelů jsou ty, které naopak způsobují bias, systematické vychýlení výsledků, které neklesá s počtem provedených měření. Musíme si uvědomit, že respondent s výjimkou samovyplňovacích dotazníků neslyší otázku tak, jak jsme ji navrhli, ale jak ji prezentuje tazatel. V roce 1914 Stuart Rice v datech ze studie na populaci 2000 sociálně slabých mužů zjistil, že jejich odpovědi se liší v závislosti na tom, který tazatel rozhovor provedl. Deming bohužel ve své citaci výsledků neuvádí, jak byly měřeny charakteristiky tazatelů, ale píše, tazatelé se sklony k prohibicionismu měli tendenci připisovat neutěšenou situaci respondentů konzumaci alkoholu a tazatelé se silným sociálním cítěním měli snahu klást vinu za jejich postavení špatným poměrům ve společnosti a kořistnickému chování průmyslníků [Deming, 1944: 340 in De Vauss]. Pokud respektujeme Demingovo zjištění, že vliv tazatelů je způsoben jejich odlišným politickým, náboženským nebo sociálním cítěním, ekonomickým postavením, věkem nebo vzděláním a naše pozorování, že tazatelé F2F výzkumů se rekrutují z přibližně stejných sociálních podmínek a vykazují podobné sociodemografické charakteristiky (nejčastěji ženy a nižšími nebo průměrnými příjmy, ve věku 35-50 let), potom je pravděpodobné, že dílčí variance

jednotlivých tazatelů se vzájemně nevyruší a spíše povedou k jednosměrnému biasu. Vlivem sociální vzdálenosti tazatele a respondenta se dříve zabývali Katz⁶¹ a Hyman⁶². Případovou studii zkoumající vliv příliš velké nebo naopak příliš malé sociální vzdálenosti nabídla Dohrenwend [Dohrenwend et al, 1968: 142-153 in De Vauss]. Hyman také do problematiky chyby tazatelů přináší koncept zaujetí tazatele výzkumem. Navrhuje rozlišovat mezi totálním zaujetím, které dokáže respondenta negativně ovlivnit k výpovědím ve shodě s vnímaným přesvědčením tazatele a zaujetím pro úkol⁶³, které pomůže respondentovi pochopit jeho úlohu ve výzkumu a zvýšit validitu odpovědí [Beatty, 1995: 45 in De Vauss].

Kromě osobních charakteristik tazatelů a jejich vlastních neskrývaných názorů mají na kvalitu dat vliv také tazatelova očekávání. Moser a Kalton rozlišují tři typy očekávání u tazatelů.

1. Postojově strukturní očekávání

Dřívější odpovědi respondenta mohou tazatele ovlivnit tím způsobem, že další odpovědi vnímá a interpretuje ve specifickém kontextu. To se negativně projeví při záznamu otevřených otázek nebo když respondent odpovídá vlastními slovy na předkódovanou, ale spontánně míněnou otázku.

2. Rolová očekávání

Tazatel si na základě indicií poskytnutých respondentem o něm samém vytváří chtě nechtě v hlavě obraz respondentovy osobnosti. Pokud se mu kdykoli během rozhovoru dostane nejednoznačné odpovědi, interpretuje ji právě ve světle obrazu, který si o respondentovi utvořil. V ideálním případě by nejednoznačné odpovědi samozřejmě interpretovat neměl a měl by zjistit, co respondent svou odpovědí myslí. Extrémním případem, který uvádí Moser a který si dokáží snadno představit, je vyplňování některých otázek v dotazníku za respondenta ať už z důvodu ušetření času nebo pokud si tazatel myslí, že dokáže respondentovu odpověď předvídat.

3. Pravděpodobnostní úvahy

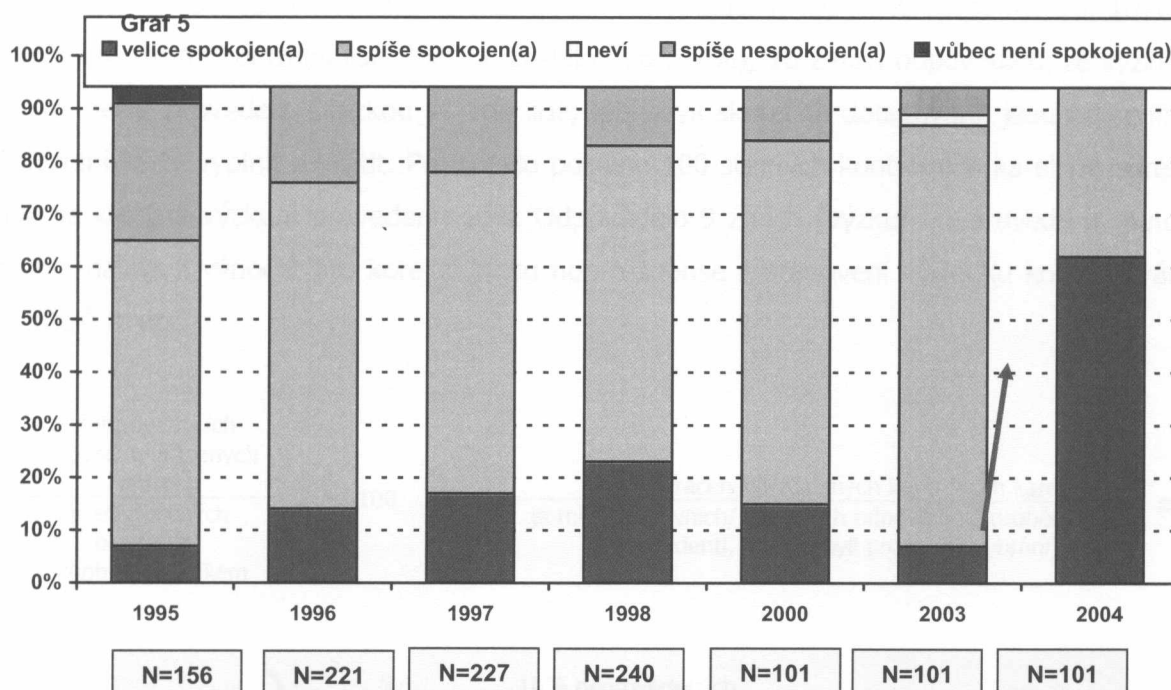
Tazatel, který zná prostředí, ve kterém dotazuje, může odhadovat pravděpodobnost výskytu zastoupení variant sledovaného jevu, např. politických preferencí. Pokud dotazuje v malém městě, kde v nedávných volbách přesvědčivě zvítězili socialisté, bude odpovědi respondenta interpretovat možná právě v politickém kontextu [Moser a Kalton, 1971: 386-387].

⁶¹ „Do Interviewers Bias Polls?“ *Public opinion quarterly*, Vol. 6, 1942, s. 248-268

⁶² Hyman et al. *Interviewing in Social Research*. Chicago, Chicago university press, 1954. s. 159-167

⁶³ v originále „task involvment“

Je těžké určit, který druh očekávání tazatele, pokud vůbec, se projevil v měření spokojenosti obyvatel jednoho z českých regionů v roce 2004, kteří již několikrát rok po sobě odpovídali na otázku do jaké míry jsou celkově spokojeni s životem v jejich obci (region zahrnuje 7 různě velkých obcí⁶⁴). Graf 5 zobrazuje nepravidelnou časovou řadu výsledků dotazování za region od roku 1995. Jedná se o poměrně malý region, kde agentura disponovala a disponuje omezeným počtem tazatelů. Spokojenost obyvatel regionu v krajní kategorii „velice spokojen (a)“ se meziročně zvýšila o 55 procentních bodů, což je s ohledem na trend nastíněný v minulých letech neočekávaný výsledek. Když jsme poprvé viděli výsledky, měli jsme za to, že se jedná o chybu zpracování dat. Až později se ukázalo, že asi čtvrtinu ze 100 provedených rozhovorů v regionu provedla jedna tazatelka. V roce 2004 byla dotazováním ve dvou obcích v regionu pověřena nová tazatelka, která měla k dispozici stejný dotazník jako její předchůdci ve starších ročnících studie. Výsledky kontroly byly v tomto případě provedeny raději ve 100% provedených rozhovorů s pozitivním výsledkem a nemáme proto důvod domnívat se, že tazatelka si odpovědi vymyslela. Bohužel jsme za stejné dvě obce neměli možnost srovnání s daty za jiného tazatele, takže neúmyslné ovlivnění výsledků tazatelem je pouze jedno z možných vysvětlení dosaženého výsledku, který byl s rozpaky prezentován zadavateli výzkumu. Poučení z takového scénáře je jednoduché. Vliv tazatelů můžeme teoreticky zmenšit zadáním rozhovorů většímu počtu tazatelů.



⁶⁴ výpočet kvóty rozhovorů se provádí podle počtu obyvatel obce

Poctivost tazatelů je problém, o kterém se příliš nemluví. Přesto dokáže významně ovlivnit celý průzkum včetně výsledků. Z ekonomických důvodů jsou jako tazatelé pro metodu D2D využíváni zaměstnanci na částečný úvazek. Ti mají k agentuře slabou vazbu a chovají se racionálně. Pokud svému úkolu rozumí tak, že agentura po nich chce vyplněný dotazník, je pro ně jistě jednodušší dotazník zfalšovat, než složitě shánět vyhovujícího respondenta. V praxi se ujal přístup, že pro práci mimo kancelář nejsou využíváni studenti, ale spíše starší lidé, u nichž je předpoklad větší zodpovědnosti.

11.1 Minimalizace a měření velikosti chyby tazatele

Základním způsobem řešení problému nepoctivosti tazatelů jsou kontroly a sankce v případě zjištění podvodu. Obvykle jsou respondenti agenturou zpětně kontaktováni s otázkou, jestli s nimi tazatel rozhovor vůbec provedl – tento postup byl využit také ve výše popisovaném příkladu výzkumu spokojenosti obyvatel regionu. Tento kontakt se děje poštou – opět z ekonomických důvodů – a jen vyjíměčně telefonicky. Agentura dostane zásilku zpět jako nedoručitelnou anebo dostane vyplněný korespondenční lístek s kladnou či zápornou odpovědí. Jak ale interpretovat kontrolní lístky, které do domácnosti dojdou zůstanou nezodpovězeny – tj. spadnou do kategorie non-response?

Studie uskutečněná Institutem pro sociální vědy při organizaci UNESCO nabízí jeden možný způsob řešení takové otázky. Studie se zúčastnilo 3500 respondentů, 1070-ti z nich byl zaslán kontrolní list. Zpět ho zaslalo 396 (37%), 10 z nich odpovědělo, že výzkum s nimi nebyl proveden. Otázkou je, zda lidé, kteří byli skutečně dotazováni, jsou ochotnější kontrolní karty vyplnit a vrátit. Proto bylo posláno 100 stejných kontrolních karet na adresy lidí, se kterými výzkum proveden nebyl. Odpovědělo 9 z nich (výzkum neproveden). Autoři studie navrhují výpočet pro korekci biasu non-response a stanovení výsledku kontrol práce tazatelů takto:

$$\left(\frac{\text{počet negativních odpovědí u vrácených kontrol}}{\text{počet vrácených odpovědí v kontrolách celkem}} \times 100 \right) \times \frac{\text{poměr vrácených/zaslaných kontrolních karet}}{\text{poměr negativních/ zaslaných odpovědí ve druhém kole (respondenti, kteří nebyli pro studii vybráni)}} =$$

$$\left(\frac{10}{396} \times 100 \right) \times \frac{37}{9} = 10\% \text{ neprovedených rozhovorů}$$

V praxi se posílá takový počet kontaktů, aby bylo zkontrolováno alespoň 10% provedených rozhovorů. Těmito kontrolami se však nedají zachytit případy, kdy rozhovor je dlouhý a tazatel jej zkrátí tak, že za respondenta vyplní filtr, který znamená přeskočení celé pasáže dotazníku. Je třeba přiznat, že tento problém je v podstatě neřešitelný a nakládá se s ním porovnáváním výsledků podle tazatelů, zkoumáním délky rozhovoru a podobně. V CAPI výzkumech můžeme důslednost práce tazatele kontrolovat navíc sledováním mezičasů, ve kterých dosáhl konkrétních otázek v dotazníku. Je pouze na našem rozhodnutí a práci programátora, s jak velkým odstupem mezičasy kontrolujeme. V extrémním případě lze kontrolovat délku zodpovídání každé z otázek.

O mnoho lepší je situace při studiových typech průzkumů, kam patří telefonické rozhovory a rozhovory označované jako Central Location Test (CLT). Těm se běžně říká studiové testy. Jak již bylo popsáno v předcházejících kapitolách, telefoničtí tazatelé jsou kontrolováni odposlechem v reálném čase. Práci na studiovém testu můžeme rozdělit na rekrutaci, která se děje obvykle na ulici, a na vlastní dotazování ve studiu. Studio je pod dohledem stálého zaměstnance. Rekrutátoři mohou mít tendenci si práci usnadňovat tak, že vkládají potenciálním respondentům do úst taková rekrutační kritéria, která se jim hodí. Například, je-li někomu 37, rekrutují jej přestože vědí, že potřebují jen respondenty do 35 let. Obvykle se to řeší tak, že stejné otázky jsou respondentovi položeny jiným člověkem na začátku studiového rozhovoru. Takové řešení funguje v případě, že respondent nepotřebuje cokoliv skrývat. Třeba z toho důvodu, že se jedná o tzv. profesionálního respondenta. Tito lidé se výzkumů účastní výhradně za účelem osobního obohacení a vyhledávají ty výzkumy, ve kterých respondentům čas strávený odpovídáním jakkoliv kompenzujeme. Pro to, aby se výzkumu mohli účastnit, jsou ochotni upravit osobní parametry takřka libovolně. Mezi jiné klasické kontroly práce tazatelů patří opakované dotazování stejných respondentů speciálně školenými a vysoce důvěryhodnými tazateli. Rozhovory je také možné nechat tazateli nahrávat a po ukončení výzkumu je analyzovat. Námi naměřená data je možné v případě tvrdých dat porovnat z externími zdroji, kupříkladu oficiálními statistikami.

Sofistikované metody měření variance tazatelů využívají technik vícenásobného vzorkování⁶⁵, které poprvé odzkoušel Mahalonobis (1946). Tato metoda, kdy více tazatelů dotazuje naprosto stejně složené skupiny respondentů, je velice náročná na proces výběru a pokud studie probíhá na geograficky rozsáhlejších území, klade extrémní nároky na mobilitu tazatelů. V celonárodní studii s kvótami na kraje České republiky by tazatelé vybraní pro test museli provést alespoň jeden rozhovor v každém kraji. Z úsporných důvodů se proto využívá metoda omezeného vícenásobného vzorkování (srovnej s Moser a Kalton, 1971: 401).

⁶⁵ replicated/interpenetrating sampling [Moser a Kalton, 1971: 400]

12. Chyby non-response

Non-response (dále jen nr) je vyjádřením stavu, kdy jsme se pokusili neúspěšně získat informaci od vybraného respondenta. Za chybu ve výzkumu a problém je považována proto, že snižuje velikost plánovaného výběrového souboru a je zdrojem biasu ve výsledcích.

V zásadě existují dva druhy nr. Prvním je *Unit nr*, o kterém jsem se již dříve v textu krátce zmínil. De Leeuw [de Leeuw et al, 2003: 155 in Journal of Official Statistics] ji označuje jako nr první úrovně a mluvíme o ní tehdy, když se respondent výzkumu neúčastní, protože spolupráci odmítl naprosto nebo jsme nebyli úspěšní při jeho kontaktování. Nr druhé úrovně je *Item nr*, která označuje nezodpovězení konkrétní položky v dotazníku (respondent se rozhodl spolupracovat, ale na některé otázky nezná nebo nechce dát odpověď). De Leeuw upozorňuje na to, že ještě před nedávnem byla item nr řešena nezahrnováním do analýzy těch případů (cases), které v analyzované proměnné nr vykazují – tzn. mají zde chybějící nebo smysl nedávající hodnotu. To je špatný postup hned z několika důvodů. Zaprvé oslabuje sílu statistických testů. Hlavně je ale nebezpečný tehdy, když je nízká item response funkcí odlišných charakteristik respondentů. Lidé, kteří odpověděli, se mohou nějak lišit od charakteristik typických pro celou cílovou skupinu. Proto je před volbou metody ošetření efektu nr v datech třeba znát charakter chybějících hodnot. Dále vycházím z terminologie de Leeuwové, kterou používá v textu *Prevention and treatment of item of Item Nonresponse*⁶⁶, ale nevylučuji, že v jiné literatuře existuje jiná typologie chybějících hodnot.

*1. Zcela náhodně chybějící hodnoty*⁶⁷

Jsou to ty hodnoty, které jinak nesouvisí s dalšími hodnotami proměnných v datech. Vznikají například přehlédnutím otázky respondentem nebo tazatelem a z definice svého vzniku nemohou mít systematický vliv na výsledky v datech.

*2. Náhodně chybějící hodnoty*⁶⁸

Vznikají, resp. nevznikají vinou například neznalosti odpovědi na danou otázku. Pokud si starší respondent nedokáže vzpomenout, kolik mu bylo let, když ukončil svou povinnou školní docházku, je toto příklad náhodně chybějící hodnoty.

⁶⁶ De Leeuw et al. Prevention and treatment of item nonresponse. Journal of Official Statistics, Vol. 19, No. 2, 2003, s. 153-176

⁶⁷ MCAR – Missing Completely At Random

⁶⁸ MAR – Missing At Random

3. Pokud je ale chybějící hodnota ve vztahu k potenciální hodnotě proměnné, tzn. že respondent z jakéhokoliv důvodu z vlastního rozhodnutí informaci neposkytne i když ji zná, máme co do činění s *nenáhodně chybějícími hodnotami*⁶⁹. Takové chybějící hodnoty jsou vážným zdrojem narušení validity výsledků.

Kromě těchto tří případů výskytu nr zmíním spíše pro úplnost další cesty způsobení efektu nr. Některé otázky jsou jednoduše určeny pouze pro některé respondenty. Například pro muže jsou vyfiltrované otázky týkající se problematiky kojení dětí. Nr může také nastat pokud respondent přeruší spolupráci v rozběhnutém dotazování. Typicky u metodologie CATI mají respondenti menší psychologické zábrany ukončit ze své iniciativy předčasně hovor, pokud dojdou k závěru, že nechtějí pokračovat. Otázky od tohoto momentu zůstanou nevyplněny, mají 100% item nr a je jen na rozhodnutí výzkumného týmu, zda takový rozhovor použije nebo nahradí jiným. Odpověď můžeme ztratit také při pořizování – ve fázi převodu dat do elektronické podoby z papírových dotazníků.

12.1 Minimalizace a napravování efektu nr

De Leeuw a Collins metaanalýzou dat ukázali [de Leeuw et al, 1997 in de Leeuw, 2003: 161 in Journal of Official Statistics], že výzkumy s podporou tazatele (F2F a CATI) dosahují vyšší response rate než výzkumy poštou. Nicholls a Baker [Nicholls et al, 1997 in de Leeuw, 2003: 161 in Journal of Official Statistics] dokázali, že data z dotazování s využitím programovaných skriptů v počítačích přinášejí lepší výsledky co do míry nr než výzkumy P&P. Kvalita dotazníku má na nr vliv zásadní, jak jsme si ukázali v kapitolách věnovaných jeho testování a konstrukci. Na nr se podepisují také socio-ekonomicko-demografické charakteristiky respondenta. Starší a méně vzdělaní respondenti mají ve svých datech obvykle vyší míru nr.

Je-li průzkum dobře připraven a přesto by odpovědní míra byla příliš nízká, lze ji zvýšit předáním odměny. Odměna se předává obvykle na konci rozhovoru a nejčastěji se nepředávají peníze, ale dárky. Odměny mají své světlé a stinné stránky. Peníze se podle zákona mohou v těchto případech předávat pouze jako mzda, což s sebou nese identifikaci příjemce hotovosti včetně rodného čísla a jeho písemné potvrzení, že peníze převzal. Jistě není třeba dodávat, že takovéto obtěžování respondenta by bylo nepřipustné. Daleko závažnějším problémem z hlediska kvality výzkumu je nepříjemná zkušenost, že mnozí respondenti poskytují pod vlivem dárku pozitivnější odpovědi, než jejich skutečné.

⁶⁹ NMAR – Not Missing At Random

Uvědomují si totiž, že odměnu jim vlastně posílá zadavatel výzkumu. Na tomto místě si dovolíme malé odbočení. Podobně jako jsou reakce veřejnosti na reklamu z dlouhodobého hlediska stále vlažnější, v posledních dekádách se s tímto jevem setkáváme také u průzkumu trhu. Přestože situace není tak špatná jako v USA anebo mnoha zemích západní Evropy, i u nás již začíná být jasné, že jde-li nám o typického zástupce cílové skupiny, bude třeba dopřát sluchu racionálním argumentům respondentů. Pokud chceme, aby rozhovorem strávili určitý čas, měli bychom jim poskytnout odpovídající kompenzaci. Odpověď na výše popsané protiargumenty (ovlivnění odpovědí respondenta) lze možná hledat v oblasti interpretace a zvláště benchmarkingu (porovnávání výsledků se standardem). Budeme-li mít včas dostatek výsledků z průzkumů, kde respondent dostal odměnu, budeme schopni tyto výsledky správně interpretovat a snáze si poradíme i s mírně pozitivnějšími odpověďmi.

Na úroveň nr má dále vliv kvalita tazatelů. Při kvalitativním výzkumu tazatelů v roce 1996 pořádaném Holandským statistickým ústavem Ger Snijkers a Joop Hox [Snijkers et al, 1999: 185-198 in Journal of Official Statistics] dokázali definovat nejužitečnější charakteristiky tazatelů použitelné při rekrutaci respondentů. Ve skupinových diskusích s 22 zkušenými tazatelkami (průměrná délka praxe tazatele 11,5 roku), které v tehdy proběhnuvší studii pracovních sil dosahovaly response rate v rozmezí velmi slušných 50%-77% se tazatelek zeptali, jaké strategie považují za nejučinnější pro získání respondenta k výzkumu F2F. Na škále 4 až -1, kde 4 znamená „tato strategie funguje téměř v každém případě“ a -1 znamená „tato strategie způsobuje negativní reakci respondenta“ bylo postupně hodnoceno více než 40 různých strategií. Za nejosvědčenější označily zkušené tazatelky

- zvládnutí prvního dojmu na respondenta (skóre na škále 3,5/ směr. odchylka 0,51)
- otevřeně projevovat čestné úmysly (3,35/0,88)
- zmínit zadavatele výzkumu (3,3/0,92)
- flexibilně reagovat na respondentovy připomínky (3,25/0,72)
- představení sebe sama/samého (3,2/1,11)

To jsou strategie, které jsou prevencí ke vzniku nr. Jak zacházet s nr v momentě, kdy ji máme v datech? Pokud můžeme považovat chybějící data za chybějící zcela náhodně nebo jen náhodně, můžeme odstranit celou proměnnou, pokud je nr zasažena z podstatné části. Stejně tak můžeme z datové matice vyřadit data za jednotlivé respondenty, kteří poskytli neúplné rozhovory. Sofistikovanějšími metodami zpracování nr jsou přímé odhady a imputace chybějících hodnot [De Leeuw, 2003: 168-169 in Journal of Official Statistics].

Přímé odhady využívají zákonností teorie pravděpodobnosti. Imputace je nahrazení chybějící hodnoty hodnotou uměle vygenerovanou. Výhodou teoreticky stoprocentního zaplnění datové matice je možnost provádět analýzy bez obav z nedostatečného množství případů v různých kombinacích kategorií proměnných. Negativem imputací je generování biasu ve výsledcích, protože chybějící hodnoty nejsou často ničím jiným než průměrem proměnné počítané přes existující validní hodnoty ostatních respondentů. V případě, že chybějící hodnoty jsou nenáhodně chybějící, dosazujeme těm respondentům, kteří měli zvláštní důvod neodpovědět, údaje za typologicky odlišné skupiny respondentů. Navíc imputací uměle zvyšujeme počet validních hodnot proměnných, což vede k posílení signifikance vychýlených testů hypotéz. Tento problém řešil Rubin [Rubin, 1987 in De Leeuw, 2003: 169 in Journal of Official Statistics] vícenásobnou imputací. Při vícenásobné imputaci program vygeneruje více umělých hodnot proměnné a vytvoří tomu odpovídající počet setů dat. Výhodnost přístupu spočívá v technice, s jakou jsou různé varianty umělé proměnné počítány. Uvažuje a zahrnuje se totiž chyba odhadu průměrné hodnoty proměnné s určitým rozložením. Analýza chybějících dat pomocí metody vícenásobné imputace probíhá ve všech setech dat a konečný odhad doplňované hodnoty je potom spíše očištěn od biasu.

13. Závěr

V mém metodologickém příspěvku k teorii průzkumu trhu a sociologického výzkumu jsem rozebíral výsledky vybraných empirických šetření experimentálních studií. Pro své závěry a doporučení jsem čerpal jednak ze sekundárních zdrojů – publikovaných výzkumných zpráv a také z vlastních zkušeností, které jsem doposud načerpal v každodenní práci na straně výzkumné instituce. Porovnával jsem zjištění různých výzkumných týmů především v tématice measurement error – tedy chyb vzniklých ve fázi měření. Pokud mi to zkušenost dovolila, interpretoval jsem jejich závěry s výsledky studií, u kterých jsem mohl osobně být. Vysvětlil jsem, proč jsem se v této práci nezabýval teorií smplování. Jak jsem se zmínil, chyby non-observation dokážeme dnes spolehlivě měřit, dokážeme jim předcházet a dokonce je i pomocí metod například vážení odstraňovat. Naproti tomu se ukázalo, že se profesionálové v odvětví výzkumu trhu a sociologického výzkumu dodnes nedokázali shodnout na tom, jak spolehlivě měřit chyby nevýběrového typu, protože tyto chyby se na kvalitě dat podepisují s nestejnou intenzitou a jejich vliv nedokážeme dopředu odhadnout. I když již byly navrženy modely k měření velikosti chyby, neumíme určit, jak jeden typ chyby ovlivňuje druhý; neexistuje shoda o závažnosti chyb měření a závěry uskutečněných případových studií se jen těžko zobecňují. Respondenti, se kterými pracujeme, jsou velmi nestálý „materiál“ a s tím musíme počítat. Respondenti se mohou jinak chovat a jinak své

chování deklarovat nebo zpětně hodnotit. Některé informace, o kterou usilujeme, si lidé neuvědomují. Jak jsem ukázal, i pokud zjišťujeme zcela konkrétní informace, potýkáme se s úskalím senzitivity otázek nebo omezeností paměti respondenta. Výzkumy probíhají v různých kulturních prostředích. Dotazujeme skupiny lidí s jinou historickou zkušeností, životními preferencemi, návyky. Výsledky případových studií z oboru společenských věd, sociologické studie nevyjímaje, mají velmi nízkou schopnost být univerzálně platné. Proto není možné definovat obecně závazné pokyny zaručující kvalitu výsledků jakéhokoliv šetření. K projektům je třeba přistupovat se znalostí prostředí, ve kterém výzkum provádíme.

Na začátku textu jsem si položil otázku co je to kvalita výzkumu. V textu jsem několikrát zdůraznil, že prevence a odhad způsobu myšlení a jednání respondentů a tazatelů je pro kvalitu dat získaných výzkumným šetřením spíše přínosná než dodatečné ošetřování získaných výsledků. I když tomu okolní podmínky mnohdy ve skutečnosti nejsou nakloněny z úsporných nebo časových důvodů, musíme v maximální možné míře využívat fungující opatření vedoucí k zajištění vyšší validity a reliability výzkumů, i když představují zvláštní náklady k výzkumu. Kvalita odhadu populační charakteristiky je rozhodně závislá na způsobu, jakým charakteristiku měříme, jak se na ni ptáme. Proto jsem se zaměřil především na vývoj výzkumného nástroje, jeho testování a práci s ním. Od správnosti sestavení otázek se odvíjí chybovost práce tazatelů, vychýlenost odpovědí respondentů s ochota respondentů účastnit se výzkumů měřeno response rate. Tvorba dotazníku je proces, na jehož konci musí být výzkumný nástroj, který je nejen dobře postavený (ve smyslu splnění metodologický požadavků), ale také atraktivní pro respondenta. Pokud se nám toto nepodaří, budou výsledky studie zkreslené, nejednoznačně interpretovatelné nebo dokonce zavádějící.

V rozsáhlé části textu jsem se proto věnoval činnostem spojeným s ověřováním funkčnosti výzkumného nástroje. Viděli jsme, že různí respondenti položené dotazy interpretují jinak, že ne všichni respondenti jsou stejně ochotní odpovídat na konkrétní otázky, že různě položené otázky stejného obsahu přinášejí odlišné výsledky atd. Ze všech uvedených příkladů uvedených v sekci testování dotazníku by měla být zřejmá důležitost pilotních výzkumů. Viděli jsme, že výsledky testovaných variant otázek jsou platné pro daný čas a danou populaci. Neexistuje registr dobrých otázek stejně tak jako neexistuje soupis těch špatných. Ukázal jsem, že jediný způsob, jak vytvořit ucelený a spolehlivý soubor pravidel stavby otázek je jejich testování v různých obměnách, v různém pořadí, v různých kontextech a to až do doby, kdy budeme schopni s jistotou říci, která pravidla mají obecnou platnost a která jsou specifická pro různé situace.

14. Resume

The text being offered is methodological micro-tribute to the theory of market research, sociological surveys and opinion polling. It analyses results of already evaluated empirical examinations about the nature of interviewing and measurement error streaming from this source. The results are discussed and completed by authors empirical experience gained in full service market research agency on both domestic and international projects. We are able to measure sampling error or, as some researchers say, errors of non-observation, but we have serious doubts about measurement and prevention of those arising in the data-collection phase. Those errors negatively affect the quality of survey results. Whether it is respondents reactions, interviewer tactics or non-response phenomena, we are not able to foresee its volume and direction in advance. Partially, because those effects are interacting. Results of sociological survey are uneasy to be generalised because of impossibility to compare data collected in different cultural, historical and political surroundings. The text aims to define survey quality rather in a way what it is not. The conclusions of the analyses show that it is safer to rely on prevention of the errors rather than on subsequent data processing. Also, validity and reliability schemes must be applied in a as much as possible extend in order to help increase survey data. Piloting the questionnaire, known but infrequently used technique is most recommended precaution. Different pretesting techniques are presented and specific pros and cons discussed. It shows up, that even if we finally discover the interviewing formula for one case study, this result is location dependable. There is not any database of correct and false questions. We only have several general rules to become familiar with and with help of those, test as many question alternatives as possible to obtain those of most valid and reliable.

15. Přílohy

15.1 Použité zdroje

- BLANKENSHIP, A.B., BREEN, E.B. (1992) *State-of-the-art marketing research*. Chicago: NTC Publishing Group. ISBN: 0-8442-3457-5
- DE VAUSS, D. (2002). *Social surveys: Sage benchmarks in social research methods*. England, Cambridge: Cambridge university press. ISBN 0-7619-7338-9 (4 svazky).
- DISMAN, M. (2002). *Jak se vyrábí sociologická znalost*. Praha: Univerzita Karlova. 3. vydání. ISBN: 80-246-01397.
- GLENDALL, P., HOEK, J. (2002). *A question wording*.
- HENDL, J. (2004). *Přehled statistických metod zpracování dat. Analýza a metaanalýza*. Praha: Portál. 1. vydání. ISBN: 80-7178-820-1.
- LABAW, P.J. (1980). *Advanced Questionnaire Design*. Cambridge, Massachusetts: Art Books.
- MCDONALD, C., VANGELDER P. (1998). *The ESOMAR handbook of market and opinion research*. Amsterdam. 4. vydání. ISBN: 9-2831-1271-7
- MOSER, C.A., KALTON, G. (1971). *Survey methods in social investigation*. London: Heinemann educational books limited (second edition).
- SCHUMAN, H., PRESSER, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York: Academic press.

Články v archivu:

Journal of Official Statistics, Statistics Sweden

- BAILAR, B.A. (1987). *Nonsampling errors*. Vol.3, no.4, str. 323-325.
- DE LEEUW, E.D., HOX, J., HUISMAN, M. (2003). *Prevention and treatment of item nonresponse*. Vol. 19, no.2, Str. 153-176.
- DE LEEUW, E.D. (2005). *To mix or not to mix data collection modes in surveys*. Vol.21, no.2, str. 233-355.
- DESVOUSGES, W.H., FREY, J.H. (1989). *Integrating focus groups and surveys: examples from environmental risk studies*. Vol.5, no.4, str. 349-363.
- JABINE, T.B. (1987). *Non-Sampling errors: some reflections*. Vol.3, no.4, str. 335-338.
- FORSMAN, G. (1989). *Early survey models and their use in survey quality work*. Vol.5, no.1, str. 41-55.
- KNÄUPER, B., BELLI, R.F., HILL, D.H., HERZOG, A.G. (1997). *Question difficulty and respondents cognitive ability: The effect on data quality*. Vol.13, no.2, str. 181-199

- NELSON, D.D. (1985). *Informal testing as a means of questionnaire development*. Vol.1, no.2, str. 181-182.
- OKSENBERG, L., CANNEL, CH., KALTON, G. (1991). *New strategies for pretesting survey questions*. Vol.7, no.3, str. 349-365.
- SCHAEFFER, N.C. (1995). *A decade of questions*. Vol. 11, no.1, str.79-92.
- SMITH, T.W. (1992). *A methodological analysis of the sexual behaviour questions on the general social surveys*. Vol. 8, no.3, str. 309-325.
- SNIJKERS, G., HOX, J., DE LEEUW, E.D. (1999). *Interviewers tactics for fighting survey nonresponse*. Vol. 15, no 2, str. 185-198.
- SYKES, W., WILLIAMS, J.M. (1987). *Evaluating survey questions*. Vol. 3, no.2, str. 191-207.
- TORTORA, R.D. (1987). *Quantifying nonsampling errors and bias*. Vol.3, no. 4. str. 339-342.
- TOURANGEAU R., RASINSKI K., JOBE B.J., SMITH W.T., PRATT W.F. (1997). *Sources of error in survey on sexual behavior*. Vol. 13, no.4, str. 341-365.
- VAN DEN VAART, W. (2004). *The time-line as a device to enhance recall in standardized research interviews: A split ballot study*. Vol.20, no.2, str. 301-317.

Články ve sborníku:

DE VAUSS, D. (2002). *Social surveys: Sage benchmarks in social research methods*. England, Cambridge: Cambridge university press. ISBN 0-7619-7338-9 (Sborník 4 svazky).

- BEATTY, P. (1995). *Understanding the standardized/nonstandardized interviewing controversy*. Str. 42-53.
- DE MAIO, T., ROTHGEB, J., HESS, J. (1998). *Improving survey quality through pretesting*. Str. 345-346.
- DEMING, W.E. (1944). *On errors in surveys*. Str. 335-350.
- DOHRENWEND B.S., COLOMBOTOS, J., DOHRENWEND, B.P. (1968). *Social distance and interviewer effect*. Str. 142-153.
- GROVES, R.M (1987), *Research on survey data quality*. Str.350-365.
- KALTON, G., SCHUMAN, H. (1982). *The effect of the question on survey responses: a review*. Str. 209-256.

Články v jiných časopisech:

- BLANKENSHIP, A.B. (1961). *Creativity in consumer research*. Journal of Marketing 25, s.34-38.

Internet:

<http://www.leeds.ac.uk/iss/documentation/top/top2.pdf>

BURGESS, T.F. (2001). *Guide to design of questionnaires. General introduction to the design of questionnaires for survey research*. University of Leeds.

<http://www.TFT.net>

POLLAND, R.J. (1998). *Essentials of survey research and analysis. A workbook for community researchers*. Duval County Health Department.

<http://www.ESOMAR.org>

Esomar World Research Report. Industry study on 2004

http://www.cmg.wkac.ac.uk/courses/fyp/FYP_QUIK.DOC

<http://www.copn.cz/clan/>

[http://www.czso.cz/csu/edicniplan.nsf/t/7E002FD06F/\\$File/tab10.pdf](http://www.czso.cz/csu/edicniplan.nsf/t/7E002FD06F/$File/tab10.pdf)

<http://www.exodusresearch.com/resources.htm>

<http://www.iso.cz/hledat.asp>

<http://www.iso.org/iso/en/iso9000-14000/index.html>

<http://www.simar.cz/default.aspx>

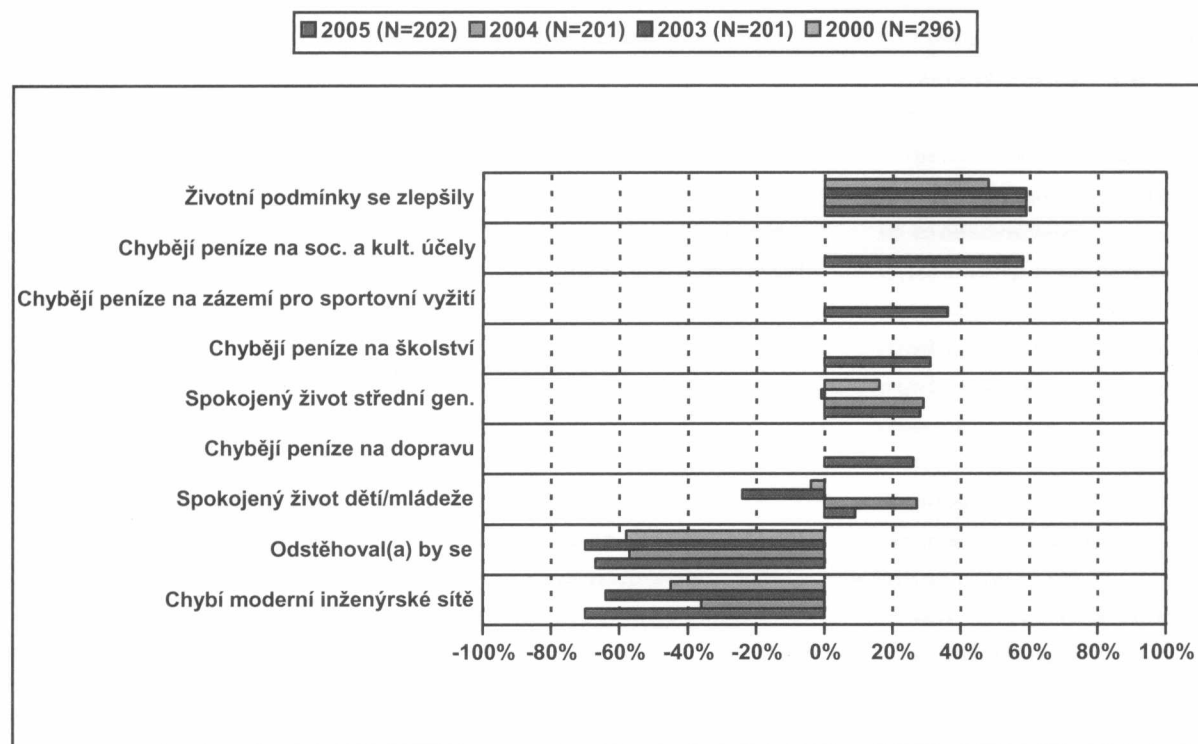
<http://www.surveysystems.com/online.html>

http://www.telecom.cz/infocentrum/tiskove_centrum

<http://www.tns-aisa.cz/omnibus.asp?lg=cz>

15.2 Grafy, schémata, tabulky a dotazníky

Graf 1 Rozdíl souhlasných a nesouhlasných hodnocení výroků vztahujících se ke kvalitě života v obci; údaje za ročníky výzkumu 2000,2003,2004,2005.



Graf 2 Odpovědi obyvatel regionu na otevřenou otázku: Co vám ze strany společnosti [jméno zadavatele výzkumu] chybí (spontánně)?

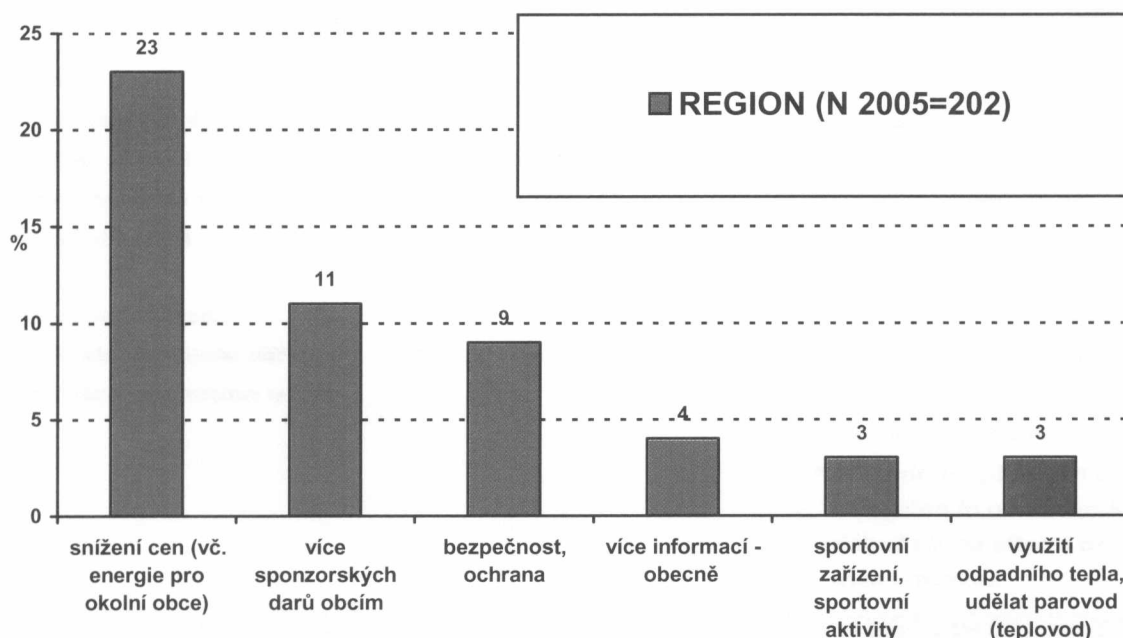
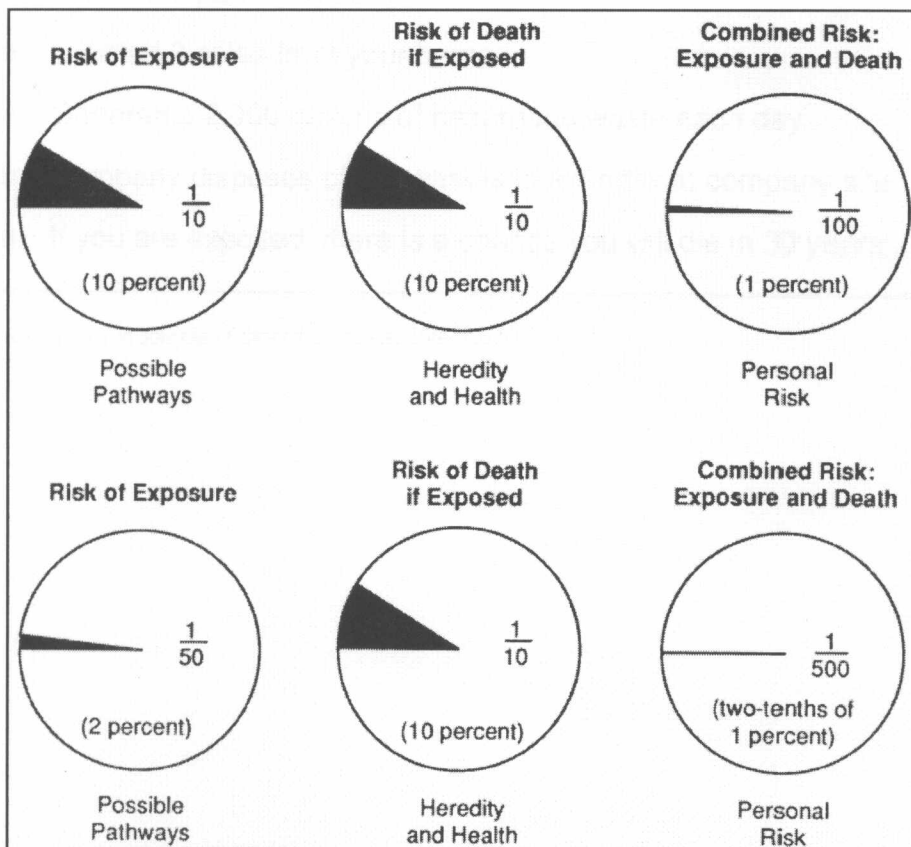
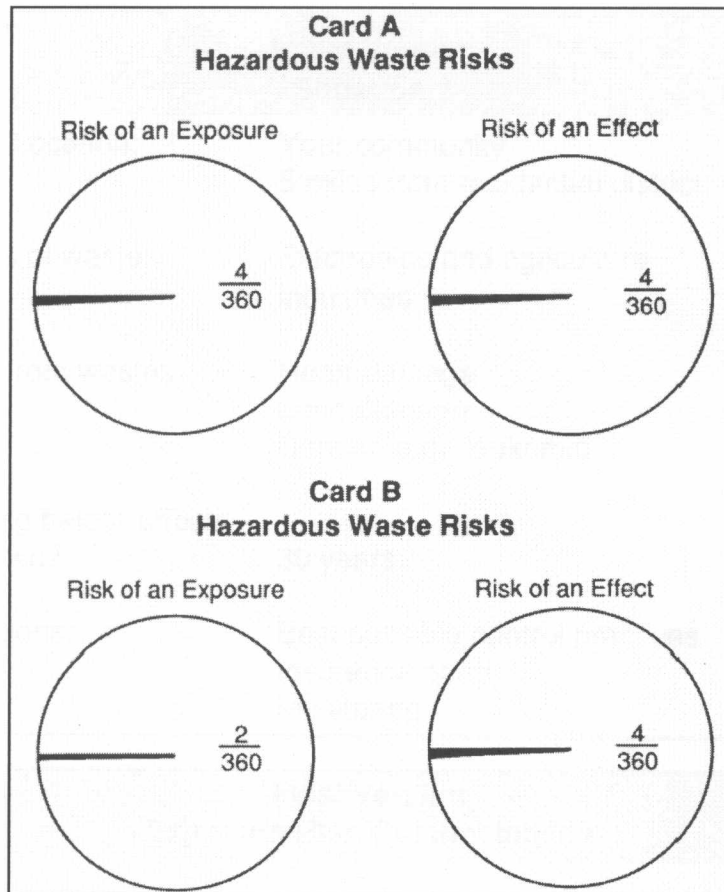


Schéma 3

<i>Mixed-mode survey system</i>	<i>Rationale for implementation</i>	<i>Effect on survey quality</i>
<i>Contact phase</i>		
Advance notification in mode different from data collection mode	<ul style="list-style-type: none"> ▪ Correct sampling frame ▪ Raise response rate ▪ Enhance credibility/trust 	<ul style="list-style-type: none"> ▪ Reduce coverage and nonresponse error ▪ No threats to measurement (if data collection is uni-mode)
Recruitment / Screening / Selection in mode different from data collection mode	<ul style="list-style-type: none"> ▪ Reduce cost ▪ Enhance efficiency ▪ Update / expand contact information for main mode 	<ul style="list-style-type: none"> ▪ Improved timeliness ▪ If pure screening no threats to measurement ▪ If screening plus first part data collection in other mode, then risk of potential mode effects on measurement
<i>Response phase</i>		
One sample, one time period, one questionnaire, but different sample persons, different modes	<ul style="list-style-type: none"> ▪ Reduce costs ▪ Improve coverage ▪ Improve response 	<ul style="list-style-type: none"> ▪ Reduce coverage and nonresponse error ▪ Mode effects on measurement (confounded with subgroups)
One sample, one time point, but different modes for different parts of questionnaire (for same person)	<ul style="list-style-type: none"> ▪ Improve privacy of measurement ▪ Reduce social desirability bias 	<ul style="list-style-type: none"> ▪ Improved data quality, especially with very sensitive questions
One sample, multiple time points, but same persons measured with different modes at different time points	<ul style="list-style-type: none"> ▪ Reduce costs 	<ul style="list-style-type: none"> ▪ Measurement differences causing confounding of time effects and mode effects
Different samples, different modes, sometimes even different times and questionnaires	<ul style="list-style-type: none"> ▪ Comparative research ▪ Different research traditions ▪ Different coverage ▪ Different cost structures 	<ul style="list-style-type: none"> ▪ Coverage error ▪ Nonresponse error ▪ Measurement error ▪ Incomparability
<i>Follow-up phase</i>		
Reminder in mode different from data collection mode	<ul style="list-style-type: none"> ▪ Raise response rate 	<ul style="list-style-type: none"> ▪ Reduce nonresponse error ▪ If pure reminder no threats to measurement ▪ If reminder plus part of data collection in other mode then risk of mode effects on measurement
Partly based on:	Dillman (2000)	Balden (2004)

zdroj: [de Leeuw in Journal of Official Statistics, 2005:238]

Schéma 4



zdroj: [Desvousges a Frey in Journal of Official Statistics, 1989: 353 a 355]

Schéma 5

Initial Version Situation 1	
Landfill Location:	Your community 5 miles from residential district
Sources of waste:	Electronics and agriculture industries
Effects from wastes:	Heart damage Lung damage Cancer e.g., leukemia
How long before effects are known?	30 years
Regulations:	Best possible control practices Insurance bond Monitoring

Final Version Exposure Risk Circumstances	
<ul style="list-style-type: none"> ● Electronic parts company ● Located 3 miles from your home ● Generates 2,000 gallons of hazardous waste each day ● Company disposes of the wastes in a landfill at company site ● If you are exposed, there is a chance you will die in 30 years 	

zdroj: [Desvousges a Frey in Journal of Official Statistics, 1989: 357]

Schéma 7

	% Answering both items	Prob.
A. Cooperation		
Interviewer ratings		
Rated as cooperative	85.9	.001
Rated as uncooperative	77.8	
DKs to attitude items		
No DKs	86.1	.000
1+ DKs	80.1	
B. Cognitive ability		
Education		
Less than high school	74.6	.000
High school	85.0	
Some college+	91.3	
Vocabulary score		
0-3 words correct	77.9	.001
4-7 words correct	82.4	
8-10 words correct	87.8	
Comprehension rating		
Good	86.9	.000
Fair	74.5	
Poor	63.6	
C. Sexual beliefs		
Approve of premarital sex		
Always wrong	84.0	.781
Almost always wrong	85.5	
Sometimes wrong	86.6	
Not wrong at all	88.3	
Approve homosexual sex		
Always wrong	83.2	.670
Not always wrong	87.9	
Approve extra-marital sex		
Always wrong	83.3	.047
Not always wrong	88.5	
D. Sexual behaviors		
Number of partners		
None	80.2	.000
1+	85.8	
Marital fidelity		
Unfaithful last year	89.7	.436
Faithful last year	85.1	
Sexual orientation		
Bisexual/homosexual	94.1	.991
Heterosexual	86.5	
E. Other demographics		
Gender		
Men	86.8	.000
Women	82.5	

Pokračování schématu 7 na další straně!

Pokračování schématu 7

Age		
18-48	89.0	.000
49-69	79.0	
70+	70.8	
Marital status		
Married	84.7	.440
Not married	84.0	
Religious orientation		
Fundamentalist	81.1	.480
Moderate	85.0	
Liberal	87.6	
Community type		
Large central city	76.8	.412
Smaller central city	82.8	
Suburb of large central city	83.9	
Suburb of smaller central city	85.5	
Other urban	84.7	
Rural	78.8	
Race		
White	86.1	.000
Non-white	73.5	

zdroj: [Smith, 1992:309-325 in Journal of Official Statistics]

Tabulka 5

„Souhlasíte s tvrzením, že vláda by měla eventuelně vrátit kontrolu nad Panamským Kanálem do rukou Panamské vlády?“ (Květen 1976)	
Souhlasím	24%
Nesouhlasím	52%
Nevím	24%
"Myslíte si, že přišel čas pro změnu znění smlouvy o Panamském Kanálu nebo bychom měli stále trvat na původním znění smlouvy?" (Leden 1977)	
Pro změnu	24%
Proti změně	53%
Nevím	23%
"Jste pro navrácení Panamského Kanálu pokud by si vláda Spojených Států podržela právo na jeho ochranu?" (Březen 1977)	
Pro navrácení Kanálu	29%
Pro držení Kanálu	53%
Nevím	18%
„Myslíte si, že by měly Spojené státy jednat s Panamou o smlouvě, která by definovala podmínky, za jakých by Panama eventuelně mohla opět vlastnit a spravovat Panamský kanál?“ (Květen 1977)	
Měli bychom jednat	27%
Neměli bychom jednat	51%
Nevím	22%
„Senát nyní jedná o smlouvách podepsaných prezidentem Carterem zaručujících Panamě převzetí kontroly nad panamským kanálem v roce 2000. Schvalujete nebo neschvalujete tyto smlouvy?“ (Leden 1978)	
Schvaluji	29%
Neschvaluji	51%
Nevím	20%
„Měl podle Vašeho názoru senát schválit smlouvy zaručujících Panamě převzetí kontroly nad panamským kanálem v roce 2000 nebo je měl odmítnout?“ (Červen 1978)	
Měl je schválit	30%
Neměl je schválit	52%
Nevím	18%

"Byl(a) byste pro nebo proti smlouvě o Panamském Kanálu pokud by k ní byl přidán dodatek umožňující Spojeným Státům intervenovat pokud by měl být kanál vojensky ohrožen?" (Leden 1978)	
Pro takto změněnou smlouvu	65%
Proti takto změněné smlouvě	25%
Nevím	10%

zdroj: [Labaw,1980 in Gendall a Hoek, 2002]

Dotazník1 - otázky pro omnibus a CATI

Q1. Máte v současnosti půjčku do 40 tisíc korun z jedné z následujících kategorií?**NABÍDNĚTE.** Zaznamenejte všechny odpovědi.

		Má	Nemá
A	Půjčka od přátel nebo rodiny	1	0
B	Půjčka od zaměstnavatele	1	0
C	Krátkodobá hotovostní půjčka od nebankovní instituce	1	0
D	Půjčka od banky	1	0
E	Půjčka ze zastavárny	1	0
F	Půjčka od úvěrového družstva	1	0
G	Koupě zboží na splátky	1	0
H	Kontokorentní úvěr	1	0
I	Kreditní karta od banky	1	0
J	Kreditní karta od společnosti nabízející nákup na splátky	1	0

Q2. Jak je pravděpodobné, že byste si v případě potřeby v budoucnu pořídili úvěr od následujících společností. Použijte prosím následující škálu:**NABÍDNĚTE JMÉNA SPOLEČNOSTÍ.** Předložte kartu Q2. Zaznamenejte odpověď u všech společností.

- 5- velmi pravděpodobné
 4- spíše pravděpodobné
 3- ani pravděpodobné ani nepravděpodobné
 2- spíše nepravděpodobné
 1- velmi nepravděpodobné
 99- neví, bezodpovědi [NENABÍZEJTE]

A	Česká spořitelna	1	2	3	4	5	99
B	Komerční banka	1	2	3	4	5	99
C	Provident Financial [čti: provident fajnenšl]	1	2	3	4	5	99
D	GE Capital Bank [čti: dží í kepitl benk]	1	2	3	4	5	99
E	Multiservis	1	2	3	4	5	99
F	Home Credit [čti: houm kredit]	1	2	3	4	5	99

Dotazník2 – EuroBarometr Slovensko, 2005

Q1. Jaké jste státní příslušnosti? Prosím uveďte název státu nebo států, jejichž jste občanem.*Respondent může uvést více odpovědí!*

		Je občanem
A	Belgie	1
B	Dánsko	1
C	Německo	1
D	Řecko	1
E	Španělsko	1
F	Francie	1
G	Irsko	1
H	Itálie	1
I	Lucembursko	1
J	Nizozemsko	1
K	Portugalsko	1
L	Velká Británie	1
M	Rakousko	1
N	Švédsko	1
O	Finsko	1
P	Kypr (jižní část)	1
Q	Česká republika	1
R	Estonsko	1
S	Maďarsko	1
T	Lotyšsko	1
U	Litva	1
V	Malta	1
W	Polsko	1
X	Slovensko	1
Y	Slovinsko	1
Z	Jiná země (VYPIŠTE)...	1
AA	Neví, bez odpovědi [NENABÍZEJTE]	1

15.3 Projekt diplomové práce

Pavel Kohout
Sociologie
4.ročník, 2004/2005
6.6.2004, Praha

Název práce: „Vybrané problémy průzkumu trhu a sociologických výzkumů“

Námět práce:

Jako námět diplomové práce jsem si zvolil fenomény, se kterými se lze setkat při plánování, provádění a vyhodnocování marketingových průzkumů, dále průzkumů volebních a spotřebitelských preferencí, životního stylu nebo veřejného mínění. Konkrétně se plánuji zaměřit na problémové situace, které se při realizaci různých výzkumů mohou vyskytnout. Cílem práce je popsat postupy vedoucí k úspěšnému řešení takových situací a eliminující chyby výzkumu, ať již se jedná o chyby zaviněné či nezaviněné.

Předpokládané metody zpracování:

Při zpracovávání této diplomové práce budu vycházet především z dostupné odborné tematické literatury. V odborné literatuře, především zahraničního původu, lze najít závěry četných metodologických projektů, které se zabývaly vysvětlováním příčin poškození validity výsledků výzkumů. Předběžně odhaduji, že práce by mohla pojednávat o následujících typech výzkumných chyb:

- Chyby dotazníku
- Chyby na straně tazatelů
- Chyby na straně respondentů
- Non-response
- Chyby výběru
- Chyby při zpracování dat
- Chyby interpretace výsledků

Mým cílem je provést smysluplnou syntézu závěrů již publikovaných výzkumných zpráv.

Konzultant:

Mgr. Jiří Remr

Orientační seznam literatury:

Babbie, E (1992). *The Practice of Social Research*, Wadsworth, Belmont, California.

Bárta, V. *Výzkum trhu*.

Bouma, G. (1996). *The Research Process*, Oxford University Press, Melbourne.

Converse, J. M. & Presser, S. (1986). *Asking questions: Thehandcrafted questionnaire*. Thousand Oaks, CA: Sage.

Davidson, C. a Tolich, M. (1999). *Social Science Research in New Zealand*, Addison Wesley Longman, Auckland.

De Vaus, D.A. (1995). *Surveys in Social Research* (4th Edition), Allen and Unwin, Sydney.

De Vaus, D.A. (2002). *Social surveys, Sage benchmarks in social research methods*, volumes 1-4. Sage publications Ltd., London.

Foddy, W. (1993) *Constructing questions for interviews and questionnaires – theory and practice in social research*, Cambridge University press.

Fowler, F. J. (1995). *Improving survey questions*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Franklin, B.J., *Research Methods*.

Glazer, M. (1972). *The researcher adventure: promise and problems in field work*, Random house, NY.

Goode a Hatt. *Methods of social research*.

Gray, G. a Guppy, N. *Successfull surveys*.

Krejčí, J. *Výzkum stranických preferencí*.

Lebeda a Rendlová. *Výzkumy veřejného mínění*.

Oppenheim. *Questionnaire design attitude measurement*.

Payne, S. L. (1951). *The Art of Asking Questions*, Princeton University Press, Princeton.

Příbylová, M. *Marketingový výzkum v praxi*.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). *Handbook of survey research*. Orlando, FL: Academic Press.

Sarantakos, S. (1993). *Social Research*, McMillan Education Australia, Melbourne.

Sheatsley, P. B. (1983). *Questionnaire construction and item writing*.

Sudman, S., & Bradburn, N. (1982). *Asking questions*. San Francisco: Jossey-Bass.

Walizer a Weinir. *Research methods and analysis: searching for relationships*.