

Franciszek Sztabiński

**OCENA
JAKOŚCI DANYCH
W BADANIACH
SURVEYOWYCH**

<http://rcin.org.pl/iis>

Franciszek Sztabiński

**OCENA
JAKOŚCI DANYCH
W BADANIACH
SURVEYOWYCH**

Franciszek Sztabiński

**OCENA
JAKOŚCI DANYCH
W BADANIACH
SURVEYOWYCH**

Wydawnictwo IFiS PAN

Warszawa 2011

<http://rcin.org.pl/ifis>

Projekt okładki
Joanna Moraszczyk

Redaktor
Elżbieta Morawska

Copyright © by Franciszek Sztabiński and Wydawnictwo IFiS PAN, 2011

ISBN 978-83-7683-037-7

Wszelkie prawa zastrzeżone.

Żadna część niniejszej publikacji nie może być reprodukowana, przechowywana jako źródło danych i przekazywana w jakiegokolwiek formie zapisu bez pisemnej zgody posiadacza praw.

Wydawnictwo Instytutu Filozofii i Socjologii PAN
00-330 Warszawa, ul. Nowy Świat 72, tel. (22) 65 72 897
e-mail: publish@ifispan.waw.pl
Wydanie I

Spis treści

Od autora	7
Wstęp	9

Rozdział 1. Ocena jakości danych: perspektywa wewnętrzna i zewnętrzna 19

1.1. Perspektywa wewnętrzna.....	19
1.1.1. Ujęcie dynamiczne: analiza przebiegu procesu badawczego.....	23
1.1.2. Ujęcie statyczne: analiza elementów procesu badawczego	35
1.1.3. Błędy pomiaru	45
1.1.3.1. Błędy niezwiązane z obserwacją	49
1.1.3.2. Błędy związane z obserwacją.....	53
1.1.3.3. Błędy opracowania	60
1.2. Ocena wartości danych: perspektywa zewnętrzna.....	61
1.2.1. Trafność i rzetelność pomiaru.....	61
1.2.1.1. Trafność jako kryterium jakości pomiaru.....	62
1.2.1.2. Rzetelność jako kryterium jakości pomiaru	71
1.2.2. Weryfikacja i walidacja danych	77
1.2.2.1. Weryfikacja danych	78
1.2.2.2. Walidacja danych.....	85
1.3. Implikacje praktyczne	90

Rozdział 2. Ocena jakości danych: podejście pośrednie i bezpośrednie 97

2.1. Ocena danych: metody nieterenowe i terenowe	100
2.1.1. Nieterenowe metody oceny	101
2.1.2. Terenowe metody oceny	105

2. 2. Ocena jakości danych: podejście pośrednie.....	114
2. 2. 1. Nieprawidłowości związane z realizacją próby.....	116
2. 2. 2. Nieprawidłowości związane z procedurą badania ...	119
2. 3. Ocena jakości danych: podejście bezpośrednie	130
2. 4. Implikacje praktyczne	134
Rozdział 3. Bezpośrednia ocena jakości danych: analiza	
struktury pytania	139
3. 1. Dane	143
3. 2. Powtórzony pomiar: struktura pytania	148
3. 2. 1. Pytania otwarte	150
3. 2. 2. Pytania zamknięte.....	151
3. 2. 3. Pytania prekategoryzowane	155
3. 3. Implikacje praktyczne	156
Rozdział 4. Bezpośrednia ocena jakości danych: analiza	
przedmiotu pytania	159
4. 1. Powtórzony pomiar: przedmiot pytania	159
4. 1. 1. Pytania o opinie	162
4. 1. 2. Pytania o fakty	167
4. 1. 3. Pytania o zachowania	178
4. 1. 4. Pytania o wiedzę	188
4. 1. 5. Pytania o cechy	190
4. 2. Przedmiot a struktura pytania	203
4. 3. Trzykrotny pomiar	209
4. 3. 1. Deklaracje respondentów	209
4. 3. 2. Deklaracje respondentów a dane dokumentalne	211
4. 4. Wpływ rozbieżności informacji na wynik badania.....	212
4. 5. Implikacje praktyczne	221
Zakończenie	225
Bibliografia	231
Aneksy	257

Od autora

Przygotowując tę pracę, spotkałem się z pomocą i życzliwością bardzo wielu osób i instytutów badawczych. Wszystkim im należą się podziękowania. Przede wszystkim profesorowi Henrykowi Domańskiemu, profesor Antoninie Ostrowskiej i mojemu bratu – Pawłowi Sztabińskiemu – za wyrozumiałość i wsparcie. Dziękuję też Kierownictwu instytutów badania rynku i opinii publicznej, na rzecz których realizowane były badania kontrolne, oraz Kierownictwu firmy EPI B.A., która była bezpośrednim wykonawcą większości badań kontrolnych.

Dziękuję także Wydawnictwu IFiS PAN, w szczególności Elżbiecie Morawskiej, za nadanie ostatecznego kształtu niniejszej pracy.

Chciałbym też podziękować wszystkim, którzy brali udział w analizowanych przeze mnie badaniach zasadniczych, a więc Respondentom. W szczególności chciałbym podziękować wszystkim, którzy zgodzili się na powtórna rozmowę dotyczącą ich udziału w badaniu i opinii na ten temat.

Na koniec chciałbym podkreślić, że niniejsza praca wyrasta z tak zwanej łódzkiej szkoły metodologicznej lat 60., 70. i 80., kształtowanej przez Zygmunta Gostkowskiego, Krystynę Lutyńską i Jana Lutyńskiego. W tych latach tam właśnie studiowałem, tam się uczyłem i rozpoczy-
nałem pracę zawodową. Nie sposób więc oderwać się od korzeni, co jednak nie znaczy, że nie można mieć krytycznego spojrzenia na pewne elementy owego dziedzictwa.

Franciszek Sztabiński

Wstęp

Jednym z podstawowych dogmatów socjologii ilościowej jest zachowanie „obiektywizmu” badania. „Obiektywizm” ów można jednak rozumieć dwojako. Po pierwsze jako rozdzielenie przedmiotu badania od jego podmiotu. Oznacza to, iż badacz winien reprezentować podejście „zewnętrzne”, wolne od wartościowania. Po drugie postulat „obiektywizmu” badań odnosi się także do ich rezultatów. W tym przypadku chodzi o to, aby wynik badania był wolny od „zanieczyszczeń” czy – inaczej mówiąc – błędów. W obu rozumieniach „obiektywizm” jest oczywiście wyłącznie postulatem. Naiwnością byłoby bowiem sądzić, iż w naukach społecznych możliwy jest pomiar w pełni obiektywny. Pisał o tym ponad pół wieku temu Herbert H. Hyman, podkreślając, iż świadomość błędu pomiaru świadczy o poziomie rozwoju nauki (Hyman 1954: 4).

Realizacja każdego badania praktycznie zawsze jest związana z tworzeniem pewnych sztucznych warunków umożliwiających jego prowadzenie (niezależnie od tego, każda z technik rodzi specyficzne zniekształcenia). W przypadku wywiadu kwestionariuszowego konsekwencją owej sztuczności są przede wszystkim błędy związane z osobą ankietera oraz respondenta i interakcją między uczestnikami procesu badawczego. Ich źródłem mogą być obserwowalne cechy ankietera i respondenta (w tym charakterystyki społeczno-demograficzne, takie jak: płeć, wiek, wykształcenie itp., oraz „pozaspoleczne”: wrażenie ogólne, sposób zachowania), a także ich cechy nieobserwowalne – „psychologiczne” (system wartości, opinie, wyobrażenia, umiejętności, wiedza, motywacje itp.). W przypadku respondenta może to być również skłonność do udzielania społecznie akceptowanych odpowiedzi, „zgadzania się”, ale także wyobrażenia o badaniu, jego celach czy realizującym je ośrodku itp. (Sztabiński P. 1997: 14–35). Jednocześnie, już w 1968 roku Krystyna Lutyńska

w swoim artykule *Ankieta do ankietera jako źródło wiedzy o wywiadzie kwestionariuszowym* pisała: „we wszystkich podręcznikach technik badań socjologicznych można znaleźć o wiele więcej zaleceń metodologicznych dotyczących zbierania materiałów przy pomocy wywiadów kwestionariuszowych niż uwag na temat oceny wartości uzyskanych materiałów. W praktyce terenowej obserwuje się również, iż sam badacz poświęca na ogół o wiele więcej czasu problemom metodologicznym przed rozpoczęciem przeprowadzania wywiadów niż po zakończeniu tego etapu badań” (1968: 184).

Niespełna dwadzieścia lat później ta sama autorka zauważyła, że „W związku z rozwojem empirycznych badań socjologicznych prowadzonych zwłaszcza techniką standardowego wywiadu kwestionariuszowego, (...) badacze z różnych krajów, w tym także i polscy, przejawiają zaniepokojenie wywołane troską o wartość otrzymywanych tą drogą informacji. (...) Zdarza się jednak, że już po przeprowadzeniu wywiadów, wartość uzyskanych tą drogą danych wzbudza szereg wątpliwości. W wielu przypadkach jest to związane z pośpiechem, z jakim przeprowadza się badania, z niewłaściwym przygotowaniem samego narzędzia, a także, i może przede wszystkim, z brakiem wiedzy metodycznej niektórych badaczy i niechęcią do refleksji metodologicznej związanej z własnymi badaniami i ich wynikami” (Lutyńska 1984: 7–8).

Skąd więc bierze się owa niechęć badaczy do refleksji metodologicznej? Kazimierz M. Słomczyński wskazuje, że „w Polsce, w ostatnim trzystoletniu, nagromadzono bardzo dużo cennego materiału na temat sytuacji wywiadu i wiarygodności wypowiedzi w wywiadzie (...). Jednocześnie w polskich pracach metodologicznych stosunkowo mało miejsca poświęcono refleksji, jak błędy wprowadzone na różnych etapach procesu badawczego można wykorzystać w analizie merytorycznych wyników. (...) Ogólnie, istnieje rozdźwięk między nagromadzoną w Polsce wiedzą empiryczną o badaniach kwestionariuszowych a jej praktycznym wykorzystaniem” (Słomczyński 2004: 7–9)¹.

¹ Jest to o tyle dziwne, że przedstawiciele innych nauk społecznych, choćby historycy, bardzo dobrze rozwinęli metodologię oceny materiałów będących przedmiotem ich analizy (Topolski 1984). Bez wątpienia warsztat metodologiczny socjologa jest różny od warsztatu historyka, niemniej odmienną jest ta nie „usprawiedliwia” braku zainteresowania problemem oceny materiałów przez tych pierwszych. O możliwościach uwzględnienia błędów pomiaru pisał między innymi Zbigniew Sawiński (1988: 45–62).

Problem krytycznej oceny wartości danych jest tak ważny, ponieważ socjologowie bazują głównie na materiałach „wywołanych”, a nie „zastanych”. Ocena uzyskanych materiałów jest koniecznym elementem fazy terenowej badań. Jej szczególna rola wynika stąd, że w chwili rozpoczęcia badacz praktycznie traci kontrolę nad ich realizacją: nie wie, w jakich warunkach przebiegały rozmowy, czy były prowadzone przez właściwych (odpowiednio przeszkolonych) ankierów, z właściwymi, to znaczy wylosowanymi, respondentami itd. Badacz staje się więc w istocie klientem firmy badawczej czy instytutu marketingowego i jest od nich całkowicie uzależniony. Czy tak być musi? Chyba nie. Każdy badacz – klient – po zakończeniu badania powinien dokonać, niezależnie od wykonawcy, oceny zebranego materiału. Jest to klasyczne zalecenie, które w podręcznikach badań ankietowych uważane jest za standard (Hyman 1954, 32; Lininger 1975; Mann 1968; Noelle 1963; Riesman 1958: 267, 282; Selltitz i in. 1959: 240–242, oraz: Nowak 1965: 86–87)².

W Polsce problem jakości danych z badań stał się widoczny po roku 1990, kiedy to, na skutek wprowadzenia zasad gospodarki rynkowej, nastąpił gwałtowny wzrost zapotrzebowania na informacje, których źródłem są wyniki badań (Domański, Sztabiński, Sztabiński 1995; Sułek 2001: 119–142; Sułek 2002: 46–80). We wcześniejszym okresie, kiedy badania socjologiczne prowadziły głównie ośrodki akademickie, ogólne zasady pracy naukowej niejako same wymuszały ich jakość. Ocenie społeczności naukowej poddawana była nie tylko wartość merytoryczna wyników, ale również, przynajmniej w pewnym zakresie, przestrzeganie zasad warsztatowych i etycznych gromadzenia danych. Obecnie, gdy liczba ośrodków prowadzących badania, wliczając również instytuty i agencje o zasięgu regionalnym, przekroczyła pięćdziesiąt, praktycznie

² O konieczności kontroli jakości realizowanych badań i jakości uzyskiwanych materiałów mówi większość kodeksów ogólnych i etycznych branży badawczej. Wymienić tutaj należy choćby kodeks ICC/ESOMAR, International Statistical Institute Declaration on Professional Ethic, ISO Quality Standard, ale także brytyjskie IQCS (Interviewer Quality Control Scheme), MRQSA (Market Research Quality Standards Association), UK Market Research National Standard BS7911 czy polski PKJPA (Program Kontroli Pracy Ankieterów). Wszystkie one mówią, między innymi, o kontroli pracy ankieterów i sytuacji wywiadu, kontroli procesu przetwarzania informacji (kodowania, przenoszenia na inne nośniki itp.) oraz kontroli zbioru.

nikt nie jest w stanie rzetelnie ocenić wartości ich pracy³. Problem zaś jest o tyle ważny, że jakość pracy ośrodków kształtuje „image” naszej dyscypliny i to w dwojakim sensie: poprzez stopień trafności przewidywań i diagnoz oraz przestrzeganie zasad etycznych w kontaktach z badanymi.

Nasuwa się zatem zasadnicze pytanie, w jaki sposób badacz ma określić jakość uzyskanych danych? Na to właśnie pytanie próbuję odpowiedzieć w niniejszej pracy, co sprawia, że nawiązuje ona do nurtu rozważań dotyczących wartości danych uzyskiwanych w badaniach surveyowych. Wbrew pozorom jest to szeroki nurt rozważań, a wymienienie choćby wszystkich stosowanych kryteriów oceny jest właściwie niemożliwe. Najczęściej wymienia się: poprawność sformułowania koncepcji i przygotowania schematu badania; poprawność wyboru sposobu uzyskiwania danych (techniki badania) i przygotowania narzędzi (w tym: długość kwestionariusza, porządek pytań, ich długość i sposób sformułowania); poprawność definicji badanej zbiorowości i schematu doboru próby (w tym: wielkość błędu losowania i pokrycia); szkolenie i monitoring pracy ankierów; reprezentatywność badania, która definiowana jest poprzez wielkość błędu „braku odpowiedzi” na poziomie całego wywiadu (*unit non-response*) i związaną z tym możliwość generalizacji wyników badania (*generalizability*); prawidłowość administrowania badania (w tym przestrzeganie zasad i procedur zbierania danych w terenie); kompletność danych, określaną wielkością błędu „braku odpowiedzi” na poziomie poszczególnych pytań (*item non-response*); poprawność procesu wprowadzania i przetwarzania danych; poprawność analizy danych i prezentacji wyników badania; dokładność pomiaru (*accuracy*) i jego precyzję (*precision*) oraz oczywiście – wywodzącą się z teorii testów – trafność i rzetelność, w których zawiera się większość wymienionych powyżej aspektów⁴.

Niektóre z tych kryteriów są bardzo ogólne, stąd też autorzy na różne sposoby operacjonalizują je, uszczegóławiając lub wprowadzając dodatkowe. Charakterystyczną ich cechą jest jednak to, iż niektóre dotyczą

³ Ostatni katalog PTBRiO (PTBRiO, 2010/11) wykazuje 50 firm badawczych, z których 26 posiada Certyfikat PKJPA, w zakresie minimum jednego typu badań.

⁴ Inną w miarę pełną i spójną enumerację kryteriów oceny jakości badań surveyowych proponuje de Leeuw (2004). Oparta jest ona na koncepcji „przydatności użytkowej” („fitness for use”) i zawiera wiele z wymienionych wyżej kryteriów.

danych całościowych, a więc wyniku całego badania (na przykład, gdy kwestionariusz został błędnie przygotowany lub próba błędnie dobrana), natomiast inne – danych całościowych lub/i jednostkowych (na przykład błąd braku odpowiedzi – „unit” i „item nonresponse”). Dodatkowo, przy zastosowaniu niektórych z wymienionych kryteriów, sformułowanie oceny jest bardzo proste, podczas gdy zastosowanie innych wymaga dodatkowych nakładów, nie tylko finansowych. Stąd też konieczne jest doprecyzowanie przedmiotu moich zainteresowań.

Po pierwsze rozważania dotyczące możliwości oceny wartości danych uzyskiwanych w badaniach surveyowych ograniczam do pewnej grupy technik. Technik surveyowych, zwanych ankietowymi, sondażowymi lub ilościowymi jest bowiem wiele. Obok wywiadu kwestionariuszowego PAPI (*Paper and Pencil Interviewing*) należy do niej zaliczyć wywiady osobiste wspomagane komputerowo CAPI (*Computer Assisted Personal Interviewing*), wywiady telefoniczne wspomagane komputerowo CATI (*Computer Assisted Telephone Interviewing*), ale także wszelkiego rodzaju techniki ankietowe, w których respondent samodzielnie wypełnia kwestionariusz (w tym także w formie elektronicznej, jak ma to miejsce w przypadku technik CSAQ (*Computerized Self-Administered Questionnaires*)⁵. Stąd też konieczne wydaje się dalsze uściślenie przedmiotu moich badań: zasadniczo interesuje mnie wyłącznie wywiad kwestionariuszowy, przede wszystkim w wersji PAPI, ale także CAPI. Pomijam tutaj techniki ankietowe, ponieważ status danych uzyskiwanych za ich pomocą jest inny niż dane z wywiadu. Wiąże się to z odmiennością procesu ich uzyskiwania danych, w efekcie czego materiał badawczy z wywiadu (zapis w kwestionariuszu) nigdy nie jest tożsamy ze źródłem informacji (wypowiedzią respondenta), podczas gdy w ankietach zapis w kwestionariuszu zawsze jest tożsamy ze źródłem informacji (Daniłowicz 1981; Sztabiński F. 1997: 23–47)⁶. Mówiąc o wywiadzie kwestio-

⁵ Zdaniem niektórych metodologów (de Leeuw, Nicholls 1996) wysoka jakość danych jest niejako „przypisana” technikom wspomaganim komputerowo (CADAC). Wynika to przede wszystkim z ograniczenia błędów w regułach przejść, błędów zakresu i spójności odpowiedzi, błędów wprowadzania danych, ale także ograniczenia efektu porządku i skłonności ankietowanych do niepożądanego zachowań.

⁶ W wywiadzie sytuacja taka występuje ze względu na pośrednictwo ankietera, który nie jest neutralnym „przekaznikiem”. Przekaz, który dociera do respondenta (pytanie kwestionariuszowe zadane respondentowi przez ankietera), nigdy nie jest

nariuszowym, mam na myśli proces uzyskiwania informacji w drodze interogacji (kierowanego komunikowania się), charakteryzującej się bezpośrednim (naocznym) kontaktem badacza (ankietera) z respondentem, w której dzięki osobistemu kontaktowi możliwa jest bezpośrednia wymiana zestandaryzowanych przekazów ustnych (Sztabiński F. 1997: 23–47, por. też Lutyński 1968, 1972b).

Po drugie w części empirycznej ograniczam analizę do danych pierwotnych, a więc zapisanych w kwestionariuszu, ściśle mówiąc – danych zapisanych w zbiorze z badania zasadniczego (przed tak zwanym czyszczeniem zbioru) oraz danych z badania kontrolnego.

Trzecie ograniczenie ma w pewnym sensie charakter merytoryczny. Otóż większość analizowanych przeze mnie badań to badania rynku, w których analizowane pytania mają specyficzny charakter, biorąc pod uwagę ich przedmiot. Są to pytania o konsumpcję różnych produktów, ich znajomość itp. Podobnych badań realizuje się najwięcej, a co za tym idzie, najwięcej jest też badań kontrolnych, a ich baza materiałowa jest największa⁷.

Zanim przejdę do omówienia własnej propozycji oceny jakości danych, podejmę próbę „uporządkowania” tradycyjnych sposobów oceny jakości danych (rozdział 1). Wyróżniam dwa ujęcia: wewnętrzne i zewnętrzne. Pierwsze z nich, nazwane przeze mnie „perspektywą wewnętrzną”, w ocenie jakości danych odwołuje się więc wyłącznie do kryteriów związanych z procesem badawczym oraz prawidłowością jego przebiegu. Drugie, nazwane „perspektywą zewnętrzną”, odwołuje się natomiast do kryteriów niezwiązanych z procesem badawczym, a więc na przykład do danych uzyskanych z innych źródeł. W przypadku tego ujęcia badacze skupiają się nie na procesie badawczym, lecz na jego ostatecznym rezultacie, a więc uzyskanym wyniku pomiaru. Co więcej, w tym przypadku sam proces badawczy, w szczególności etap zbierania informacji, znajduje się zasadniczo poza ich polem zainteresowania.

bowiem tożsamy z przekazem badacza (pytanie w kwestionariuszu przeznaczone do odczytania respondentowi przez ankietera), a ten, który dociera do badacza (zapis ankietera odpowiedzi respondenta), nigdy nie jest tożsamy z przekazem respondenta (odpowieź respondenta).

⁷ Nie oznacza to, że przeciwstawiam badania akademickie marketingowym. Zarówno jedno, jak i drugie mają określone cele i aczkolwiek różne, nie należy ich wartościować. W badaniach akademickich można też korzystać z doświadczeń badań marketingowych i *vice versa* (zob. Sztabiński, Sztabiński, Sawiński red., 2004a).

Przyjmuje się, że jeżeli uzyskany (nieważne, w jaki sposób) wynik badania jest „prawdziwy”, a więc zgodny ze stanem faktycznym, wtedy jest on wartościowy i wiarygodny.

Przedstawione „tradycyjne” możliwości oceny jakości danych (perspektywa wewnętrzna i zewnętrzna) mają swoje ograniczenia i zalety, biorąc pod uwagę realne możliwości ich zastosowania, efekty oraz pracę i kosztochłonność (por. Groves 1989). Stąd propozycja wskazania innego podejścia: pośredniego i bezpośredniego w ocenie jakości danych, które są komplementarne, a jednocześnie łączą pewne elementy scharakteryzowanych wcześniej perspektyw (rozdział 2). Nawiązując do teorii agencji (*firm/agency theory*) Jensena (1994; Jansen, Meckling 1976), proces badawczy można rozpatrywać w kategoriach „konfliktu interesów”. Konflikt taki występuje przede wszystkim między zleceniodawcą a instytucją badawczą. Wykonawcy (instytutowi) zależy przede wszystkim na tym, aby dobrze wykonać zlecenie (zrealizować badanie), zleceniodawcy (badaczowi) natomiast – aby uzyskane materiały (dane) były zgodne z rzeczywistością. Innymi słowy, instytut jest zainteresowany procesem zbierania danych, a badacz jego efektem, końcowym rezultatem. Można nawet zaryzykować stwierdzenie, iż badacz nie interesuje jak pracowali ankieterzy, czy realizowali wywiady czy nie, byle ostateczny rezultat był „prawdziwy”, wiarygodny. Ów „konflikt interesów” wyznacza dwa podejścia w ocenie uzyskanych materiałów: pośrednie i bezpośrednie⁸.

W przypadku pierwszego z ujęć (podejście pośrednie) oceny wartości uzyskanych materiałów dokonuje się na podstawie oceny pracy ankieterów. Przyjmuje się założenie, że jeśli ankieterzy pracowali właściwie (poprawnie dobierali respondentów, realizowali z nimi wywiad, w sposób zgodny z instrukcją, wywiad był realizowany w sprzyjających warunkach itd.), to wartość uzyskanych danych jest wysoka. To podejście zgodne jest z „interesami” instytutu badawczego.

Podejście drugie, bezpośrednie, bazuje na ocenie rezultatu pracy ankietera (uzyskanego wyniku), a nie ocenie jego pracy. Podejście to polega

⁸ „Konflikt” ten jest szczególnie widoczny w badaniach rynkowych: o ile klienta interesuje wynik badania, na przykład jaki odsetek badanych jest użytkownikami produktu X, o tyle ośrodek jest przede wszystkim zainteresowany tym, czy ankieterzy pracowali uczciwie czy też nie. Inspiracją do rozpatrywania problemu oceny jakości danych w kategoriach „konfliktu interesów” stanowiła dla mnie praca licencjacka Małgorzaty Kulej (2004).

na ocenie odpowiedzi respondenta na wybrane pytania poprzez powtórne ich zadanie. Jak wspomniano, badania są typem pomiaru, opartym na subiektywnych ocenach ankietera bądź raportowaniu stanu rzeczy przez respondenta, i stąd mogą występować w nim błędy. Ponieważ kwestionariusz wywiadu rzadko pozwala na oszacowanie tych błędów poprzez powtórzony pomiar (na przykład w kwestionariuszu zawarte są te same pytania, zadawane w różnych fazach wywiadu), funkcję tę pełnić może powtórne zadanie części pytań z kwestionariusza zasadniczego w badaniu kontrolnym. Badanie kontrolne pozwala wówczas ocenić rzetelność zebranych w badaniu zasadniczym materiałów nie pod kątem pracy ankieterów, ale ze względu na ich wartość – rzetelność odpowiedzi udzielanych przez respondentów (ich zgodność)⁹. To podejście zgodne jest z kolei z „interesami” zleceniodawcy – klienta instytutu badawczego.

W kolejnych częściach (rozdziały 3 i 4) przedstawiam wyniki analiz empirycznych. Bazuję na badaniach kontrolnych realizowanych w latach 1992–2005. Przedmiotem analizy jest replikowalność odpowiedzi respondentów na pytania ze względu na ich strukturę (pytania otwarte, dychotomiczne, kafeteryjne, pytania ze skalą i prekategoryzowane) oraz ich przedmiot (a więc pytania o opinie, fakty, zachowania, wiedzę i cechy). Jest to ważne, ponieważ, jak wiadomo, pytanie na dany temat sformułowane może być na różne sposoby, może mieć inną strukturę. W rozdziale 4 podejmuję również problem „spłaszczania” kategorii odpowiedzi, a więc różnicy między odpowiedziami szczegółowymi i ogólnymi (na przykład odpowiedzi na skali 5- i 3-punktowej). Oczywiście jest, iż zagadnienia te można rozstrzygnąć intuicyjnie. Nikt nie zaprzeczy, że zgodność odpowiedzi na pytania o fakty będzie wyższa niż w przypadku pytań o opinie, że odsetek odpowiedzi zgodnych będzie wyższy w pytaniach dychotomicznych niż otwartych. O ile wiem, nikt tego nie pokazał dokładnie. Ponieważ bazuję na materiałach z 34 wyselekcjonowanych badań kontrolnych (23 685 przypadków), postaram się rozstrzygnąć tę kwestię.

Zgromadzone materiały upoważniają mnie również do wyznaczenia benchmarkingu, a więc swoistego „punktu odniesienia” umożliwiającego ocenę zebranych badań. Jeśli na przykład w 25 badaniach realizowanych przez różne instytuty badawcze średni odsetek odpowiedzi zgod-

⁹ Jest to zatem nic innego jak „powtórzony pomiar”, a więc ocena jego rzetelności rozumianej jako „stabilność”. Tego typu ujęcie wydaje się w pełni uzasadnione, jeśli przyjmiemy, że rzetelność pomiaru jest koniecznym warunkiem jego trafności.

nych na pytanie o liczbę osób w gospodarstwie domowym wynosi 84%, to można uznać to za „benchmark”. Jeśli natomiast w jakimś badaniu odsetek ten wynosi tylko 75%, oznacza to, iż zebrane informacje są mało rzetelne, a ich wartość dyskusyjna. W rozdziale podejmę również problem konsekwencji posługiwania się danymi nie w pełni wiarygodnymi dla wyników badań, a więc zagadnienie wpływu błędów na zależności między zmiennymi. Wiadomo, iż pomiar praktycznie każdej cechy obarczony jest błędem, ale nie wiadomo, jakie są tego konsekwencje. Gdyby bowiem okazało się, iż korelacje między odpowiedziami respondentów na pytanie X a ich odpowiedziami na pytanie Y oraz analogicznymi uzyskanymi w powtórny pomiarze (tzn. X' i Y') różnią się między sobą, to wówczas można by zakwestionować wartość wniosków opartych na korelacji między odpowiedziami na pytanie X i Y.

Rozważania zawarte w niniejszej pracy mają między innymi cele praktyczne. Po pierwsze mają zachęcić do krytycznego spojrzenia na jakość danych uzyskiwanych w badaniach surveyowych, po drugie – wskazywać łatwy sposób ich oceny. Łącząc bowiem w badaniach kontrolnych elementy oceny pośredniej i bezpośredniej uzyskujemy informacje nie tylko na temat pracy ankietatorów, ale też rzetelności odpowiedzi respondentów (przynajmniej na wybrane pytania). Jest to prosty i tani sposób, ponieważ obowiązkiem firmy, zgodnie ze wspomnianymi kodeksami branży badawczej – jest kontrola każdego realizowanego badania. Powtórne zadanie respondentom części pytań z kwestionariusza zasadniczego nie podwyższa zaś, w sposób znaczący, kosztów badania kontrolnego. I wreszcie, praca ta daje podstawę do oceny jakości uzyskiwanych danych, wskazując wspomniane benchmarki.

Ocenę, na ile praca ta wypełnia, choćby w niewielkim stopniu, wspomnianą przez Kazimierza M. Słomczyńskiego „lukę” między nagromadzoną wiedzą empiryczną o badaniach kwestionariuszowych a jej praktycznym wykorzystaniem, pozostawiam Czytelnikowi.

ROZDZIAŁ 1.

Ocena jakości danych: perspektywa wewnętrzna i zewnętrzna

1.1. Perspektywa wewnętrzna

W literaturze poświęconej metodologii badań socjologicznych można znaleźć wiele pozycji na temat jakości danych uzyskiwanych w badaniach ilościowych. Wydaje się to oczywiste, biorąc pod uwagę, że tego typu badania są w najszerszym zakresie stosowanym – praktycznie we wszystkich krajach – sposobem zbierania informacji dotyczących wielu różnych zagadnień społecznych i rynkowych. Każdy, kto prowadził tego typu badania, bez wątpienia zastanawiał się nad wartością uzyskanych danych, nad tym, na ile ich wyniki są „prawdziwe”, zgodne ze „stanem faktycznym”. Badania takie opierają się bowiem na raportowaniu przez badanych informacji o swoich zachowaniach, opiniach, poglądach itd. Jednocześnie tak zwana socjologia ilościowa wywodzi się w prostej linii z pozytywizmu i opiera na wzorach nauk przyrodniczych. Tym samym badanie ilościowe, ze względu na swoje konstytutywne cechy (przede wszystkim standaryzację, masowy charakter oraz losowy dobór przypadków do badania) jest pomiarem. Jego wynik prezentuje się w formie liczbowej, a co więcej – uogólnia na całą populację. Jednocześnie pomiar ten może być, co oczywiste, obciążony błędem. W naukach przyrodniczych błąd kontrolowany jest poprzez wielokrotne powtórzenie pomiaru. Niestety socjologia jest dziedziną, w której wielokrotny pomiar jest praktycznie niemożliwy. Stąd też przeszło pół wieku temu, jak już wspominałem, Herbert Hyman napisał, że „...wykazanie błędu wskazuje na

zaawansowany stan nauki. Każde badanie jest obarczone błędem i jest daleko lepiej mieć jego świadomość (...), niż być ignorantem w zakresie błędów obecnych w danych. Ignorowanie błędu nie oznacza jego braku” (Hyman 1954: 4).

Analizując literaturę poświęconą problemom oceny wartości danych pochodzących z socjologicznych badań ilościowych, można jednak zauważyć, iż większość autorów dokonuje jej, przyjmując różne kryteria. Przykładowo mówi się o: błędach pomiaru, przyjmując zresztą różne jego definicje (na przykład Groves 1989; Sztabiński P. 1997); trafności pomiaru (na przykład Babbie 2004: 166–169; Frey i Oishi 1995: 25–33); jego rzetelności (na przykład Babbie 2004: 163–165; Frey i Oishi 1995: 25–33); możliwości generalizacji (*generalizability*), która praktycznie jest równoznaczna z trafnością zewnętrzną, a więc możliwością uogólniania wyników badania na całą populację (na przykład Frey i Oishi 1995: 25–33); reprezentatywności wyników, która wiąże się z możliwością ich uogólniania (*representativeness*), której wskaźnikiem jest odsetek realizacji próby założonej (*response rate*) lub brak błędu braku odpowiedzi na poziomie jednostki losowania (*unit nonresponse error*) [na przykład de Leeuw i van der Zouwen 1991: 10–11, 17–18; Frey i Oishi 1995: 30–32; Lejeune i Bied-Charreton 1991: 70–77; Stoop 2005: 23–32; Sztabiński F. 1997: 56–64]; kompletności informacji (*completeness*), której wskaźnikiem jest brak błędu braku odpowiedzi (*item nonresponse error*) oraz liczba informacji w odpowiedziach na pytania otwarte lub wielowyborowe (na przykład de Leeuw i van der Zouwen 1991: 10–11, 17–18; Lejeune i Bied-Charreton 1991: 70–77; Frey i Oishi 1995: 28–29; Sztabiński F. 1997: 56–64), czy wreszcie o precyzji (*precision*) i dokładności (*accuracy*), która jest określana przez zgodność z informacjami zewnętrznymi (*response validity*), brak błędu społecznie akceptowanych odpowiedzi (*social desirability*) oraz podobieństwie rozkładów między odpowiedziami uzyskanymi przy zastosowaniu różnych technik (na przykład de Leeuw i van der Zouwen 1991: 10–11, 17–18; Babbie 2004: 162–163; Lavrakas 2008: 598–599; Sztabiński F. 1997: 56–64).

Wydaje się, iż generalnie wszystkie te odmienne stanowiska można sprowadzić do dwóch niezależnych ujęć. Jedno z nich można nazwać wewnętrznym, drugie zewnętrznym¹. Zgodnie z pierwszym z nich (per-

¹ W literaturze metodologicznej, nie tylko z zakresu socjologii, można wskazać podobne przykłady tego typu dwóch ujęć. W naukach historycznych mówi się, jak

spektywa wewnętrzna) przedmiotem zainteresowania jest analiza procesu badawczego, kolejnych jego etapów, faz i elementów oraz potencjalnych błędów. Analizując poszczególne etapy, fazy i elementy badania, ocenia się prawidłowość jego przebiegu (wskazuje jego ewentualne błędy – odstępstwa od zakładanego), a tym samym określa jakość zebranych danych. W tym przypadku przyjmuje się, że jeśli proces badawczy przebiegał zgodnie z założeniami (prawidłowo, a więc nie stwierdzono błędów), to zebrane materiały są wartościowe, wiarygodne. Ujęcie to nazywam „wewnętrznym”, ponieważ w ocenie nie wykracza ono poza sam proces badawczy, nie sięga się do dodatkowych informacji, uzyskanych poza danym aktem badawczym, a więc „zewnętrznych”. Ocenie podlegają wyłącznie poszczególne etapy i fazy danego badania lub jego elementy oraz potencjalne błędy. Przyjmując perspektywę wewnętrzną, jakość uzyskanych danych można jednak oceniać w dwojaki sposób: na podstawie analizy samego procesu badawczego bądź też poprzez analizę błędów pomiaru w procesie badawczym.

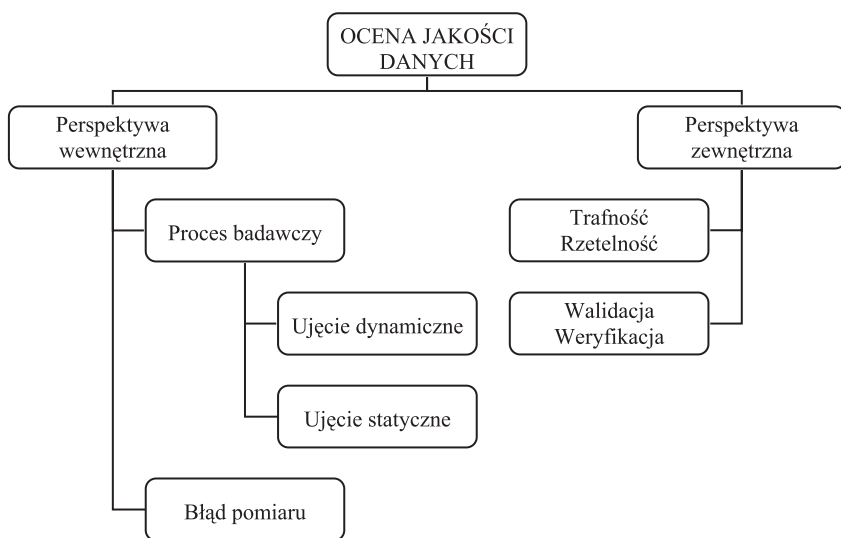
W pierwszym przypadku (analiza procesu badawczego) wskazać można dwa odmienne ujęcia. Jedno z nich określam mianem dynamicznego, drugie – statycznego. W ujęciu dynamicznym przedmiotem oceny jest przebieg procesu badawczego, a więc poszczególne jego etapy: przygotowanie badania, jego realizacja w terenie oraz selekcja i opracowanie zebranego materiału. W ujęciu statycznym przedmiotem zainteresowania nie są poszczególne etapy i fazy badania, lecz jego elementy, przede wszystkim uczestnicy: badacz, ankieter i respondent.

W drugim przypadku, w ramach perspektywy wewnętrznej, jakość danych określa się na podstawie analizy błędów w badaniu. Chodzi tutaj o trzy podstawowe kategorie błędów: błędy obserwacji (*observational errors*), które związane są z wyborem techniki, przygotowaniem narzędzia oraz, co oczywiste, z ankieterem i respondentem (a więc terenową realizacją wywiadów); błędy niezwiązane z obserwacją (*errors of nonobservation*), będące pochodną operatu losowania, procesu doboru próby i realizacji próby w terenie oraz błędy opracowania zebranego materiału, a więc błędy kodowania, wprowadzania danych itd. W przypadku perspektywy zewnętrznej badacze skupiają się nie na procesie badawczym, przebiegu i elementach, lecz na jego ostatecznym rezultacie,

wspomniałem, o dwóch typach oceny: tak zwanej krytyce źródeł wewnętrznej i zewnętrznej (Topolski 1984: 357–376).

a więc uzyskanym wyniku pomiaru. Co więcej, w tym przypadku sam proces badawczy, w szczególności etap zbierania informacji, znajduje się zasadniczo poza ich polem zainteresowania. Można powiedzieć, że w tym ujęciu badacze w ogóle nie interesuje, w jaki sposób zostały zebrane dane (czy ankieterzy w ogóle rozmawiali z respondentami, czy też nie, czy respondenci udzielali odpowiedzi zgodnych z własnymi przekonaniami, czy próba została prawidłowo dobrana itd.). Przyjmuje się, że jeżeli uzyskany (nieważne, w jaki sposób) wynik badania jest „prawdziwy”, a więc zgodny ze stanem faktycznym, wtedy jest on wartościowy i wiarygodny. Tutaj także wskazać można dwie odmienne perspektywy. Jedna z nich odwołuje się do koncepcji trafności i rzetelności, druga zaś do procedury walidacji i weryfikacji.

Reasumując, prezentowane w literaturze zróżnicowane perspektywy oceny jakości danych schematycznie przedstawia rysunek 1.



Rysunek 1. Główne podejścia w ocenie jakości danych pochodzących z socjologicznych badań ilościowych

Poniżej omówię kilka wybranych stanowisk prezentujących powyższe dwie perspektywy.

1. 1. 1. Ujęcie dynamiczne: analiza przebiegu procesu badawczego

Pierwsze z ujęć w ramach perspektywy wewnętrznej koncentruje się, jak wspomniałem, na analizie przebiegu procesu badawczego, poszczególnych jego etapach i fazach następujących po sobie w określonej sekwencji czasowej. Stąd też ujęcie to nazwałem dynamicznym. Badanie traktuje się tutaj jako proces zbierania informacji, a nie jednorazowy akt. Każde badanie, niezależnie od problematyki, celu i zakresu obejmuje bowiem trzy podstawowe etapy (por.: Alreck i Settle 1985; Antoine 1990: 147–199; Babbie 2004: 127–134, 541–545; Laniel i Lavallée 1991: 265–284; Lutyński 1968: 11–12)²:

A. Etap przygotowania badania, na który składa się sformułowanie problematyki i koncepcji badawczej oraz przygotowanie narzędzia badawczego.

B. Etap zbierania materiałów, który obejmuje: szkolenie ankierów, realizację terenową oraz bieżącą kontrolę realizacji badania.

C. Etap opracowania danych, a więc ocenę i selekcję zebranego materiału, właściwe jego opracowanie: analizę i interpretację oraz przygotowanie raportu.

Realizacja każdego z etapów zakłada wykonanie określonych czynności badawczych. Koniecznym warunkiem „przejścia” do następnego etapu badania jest „spełnienie” warunków określonych w fazie wcześniejszej. Rozstrzygnięcia podjęte wcześniej wyznaczają więc w pewnym zakresie następne, a te ostatnie są uzależnione od uprzednich. Dotyczy to zarówno poszczególnych etapów, jak i zadań badawczych. Można więc powiedzieć, że badanie jest złożonym systemem wzajemnie warunkujących się czynności.

Biorąc pod uwagę złożoność procesu badawczego, prawdopodobieństwo „odchylenia” faktycznego przebiegu procesu badawczego od

² Niektórzy z przywołanych autorów wskazują więcej niż trzy etapy badania. Przykładowo, Alreck i Settle (1985) oraz Laniel i Lavallée (1991) wskazują sześć podstawowych etapów badania, przy czym odmiennie je określają. Laniel i Lavallée mówią o: planowaniu badania, przygotowaniu badania, realizacji badania, walidacji wyników badania, upowszechnieniu wyników badania, porównawczej analizie uzyskanych rezultatów i wyników innych badań. Wyróżniając trzy etapy, chodziło mi o wyraźne oddzielenie zadań badawczych, które muszą poprzedzać realizację terenową badania, oraz tych, które po niej następują.

zakładanego – modelowego – a tym samym destandaryzacji pomiaru i w efekcie obniżenia wartości gromadzonych danych jest wysokie. Innymi słowy, jeśli którykolwiek etap badania lub którakolwiek jego faza ulegnie destandaryzacji, uzyskany wynik jest obciążony „błędem”. O ogólnej wartości zebranego materiału (*data quality*) stanowi błąd, który jest sumą błędów na różnych etapach i w różnych fazach procesu badawczego. W ujęciu tym przyjmuje się, że o wartości danych decydują trzy elementy: ścisłość, czy precyzja informacji (*response accuracy*), ich kompletność (*completeness*) oraz pokrycie próby (*sample coverage*). Pierwsza z tych kategorii definiowana jest poprzez trafność i rzetelność pomiaru. Wskaźnikiem kompletności jest niewystępowanie błędu braku odpowiedzi. Trzecia z kategorii wiąże się z możliwością generalizacji, a więc tego, do jakiego stopnia wnioski z badania są prawdziwe dla całej populacji (praktycznie jest ona równoznaczna z trafnością zewnętrzną). Jej wskaźnikiem jest odsetek realizacji próby założonej (Frey i Oishi 1995: 25–38).

Jak zapobiec lub zminimalizować prawdopodobieństwo powstania tak rozumianego błędu, aby uzyskać w pełni wartościowe dane? Skoro każdy etap i faza procesu badawczego może ulec destandaryzacji, praktycznie w odniesieniu do każdego etapu i fazy prowadzić należy kontrolę jakości (Quality Control – QC). QC to system działań (w tym także zarządzanie realizacją badania) zmierzających do podniesienia jakości uzyskiwanych danych lub, inaczej, zastosowanie procedur optymalizujących rezultaty badania, a więc pozwalających na identyfikację i eliminację potencjalnych błędów (Frey i Oishi 1995: 7; Laniel i Lavallée 1991: 263–265; Lavrakas 2008: 649–651)³. Zakłada się, że jeśli QC badania jest prowadzona właściwie, zapewnia wysoką jakość uzyskanych danych.

Jakie konkretne działania kontrolne można przypisać poszczególnym fazom i zadaniom badawczym?

³ Laniel i Lavallée (1991) wprowadzają termin *assurance de la qualité* (AQ). Oznacza on zasadniczo to samo, co *quality control* (QC). AQ można bowiem określić jako zespół działań zapewniających, że ostateczny produkt (jako całość) będzie odpowiedniej jakości i zgodny z oczekiwaniami (tamże: 264). Z kolei Biemer i Caspar (1994) mówią o *continuous quality improvement* (CQI), podkreślając, iż działania zmierzające do zapewnienia wysokiej jakości danych określa się także mianem *total quality management, strategic quality management, total quality control* itp. (zob. także Lyberg, Biemer 2008).

A. Etap przygotowania badania to przede wszystkim, jak wspomniałem, opracowanie koncepcji badawczej oraz narzędzia badawczego.

Przygotowując koncepcję, musimy określić nie tylko problematykę badania (a więc to, „co badać?”), próbę i zakres badania („kogo i ile badać?”), metody i techniki badania („jak badać?”), ale także „jak analizować?” (zob. Sztabiński F. 1995: 49–60; Sztabiński F. 2003: 148–149). Innymi słowy, musimy dokładnie zaplanować nasze badanie „na papierze”. W tej fazie procesu badawczego najistotniejszym elementem QC są konsultacje merytoryczne i metodologiczne przygotowanej koncepcji badawczej. Konsultacje pozwalają odpowiedzieć na pytanie, na ile przyjęte przez nas założenia i rozwiązania są słuszne. Chodzi tutaj przede wszystkim o zasadność (por. Antonie 1990: 150–151; Babbie 2004: 140–162; Lutyński 1994: 77–80; Nowak 1970: 214–251) dotyczących rozstrzygnięć:

- problematyki badania. Na tym etapie ocenie podlega nie tylko konkretyzacja i hierarchizacja problemów badawczych, ale także kompletność naszego zapotrzebowania informacyjnego;

- jednoznacznego określenia – w zależności od problematyki – badanej zbiorowości i schematu doboru jednostek do badania. Jednym z istotnych elementów jest ocena realnych możliwości dotarcia do tak określonej zbiorowości, a więc praco- i kosztochłonności doboru respondentów;

- wyboru właściwej procedury badania (techniki podstawowej oraz technik uzupełniających) z punktu widzenia jej adekwatności metodologicznej (tzn. jej informatywności i wiarygodności) oraz psychologiczno-społecznej, a więc jej przydatności ze względu na cechy emocjonalne, intelektualne oraz zakres doświadczeń badanych osób (Sztabiński F. 2003: 150–154);

- przygotowania ramowego schematu analizy uzyskanego materiału (szczególne znaczenie ma określenie zestawu informacji metryczkowych, będących pochodną przewidywanego planu analizy zależności). Plan analizy jest jednocześnie narzędziem oceny poprawności sformułowania i konceptualizacji problematyki badania.

Istotnym elementem przygotowania koncepcji badawczej są również określone decyzje organizacyjne. Wśród nich wymienić należy przede wszystkim harmonogram realizacji badania w terenie oraz wybór instytutu badawczego (Antoine 1990: 149–150; ESS Round 5: *Specification for participating countries*, 2011).

Drugą zasadniczą fazą etapu przygotowania badania jest opracowanie narzędzia badawczego. Faza ta nie ogranicza się wyłącznie do ułożenia pytań. Elementami narzędzia badawczego, oprócz kwestionariusza wywiadu, są bowiem: instrukcja ogólna do badania (tak zwana instrukcja dla ankietera), materiały pomocnicze (itd. karty respondenta, fotografie, plansze itd.) oraz list zapowiedni (informujący respondenta o badaniu). W tej fazie procesu badawczego celem QC jest zapewnienie, że nasze zapotrzebowanie informacyjne zostało adekwatnie „przełożone” na język pytań w kwestionariuszu, a sama realizacja wywiadów w terenie będzie przebiegała bezproblemowo. Najistotniejszymi elementami QC w tej fazie badania (por. Alreck i Settle 1985: 92–94; Antonie 1990: 150–151, 165–170; Babbie 2004: 281–282; Boeswillwald 1991: 296–298; Lutyński 1968: 11–12; 1975: 24–26) są:

- badanie rekonesansowe (zwiad badawczy), przy zastosowaniu technik jakościowych, przede wszystkim FGI lub IDI. Badanie to służyć może nie tylko doprecyzowaniu koncepcji badawczej (wcześniejsza faza), ale stanowić może pomoc przy konstrukcji kwestionariusza, kafe-terii do pytań itd. W tego rodzaju badaniach chodzi bowiem o poznanie sposobu postrzegania przez badanych wymiarów podjętej problematyki, kategorii pojęciowych służących do ich wyodrębnienia i opisu, używany przez nich język itd. (Boeswillwald 1991: 296–298; Nowak 1970: 232–236; Sztabiński i Sztabiński 2005b);

- konsultacja metodologiczna narzędzia badawczego, mająca na celu ocenę poprawności sformułowania poszczególnych pytań kwestionariusza, poprawności jego struktury statycznej (tematycznej, logicznej i psychologicznej) i dynamicznej, poprawności instrukcji oraz przygotowania materiałów pomocniczych do badania;

- pilotaż narzędzia badawczego, służący ocenie przygotowanego narzędzia badawczego i mający rozstrzygnąć wątpliwości badawcze dotyczące jego funkcjonowania w terenie oraz zastosowanej procedury. Badanie to ma wyłącznie cele metodologiczne: jego celem jest przede wszystkim sprawdzenie, czy respondenci udzielają odpowiedzi na pytania kwestionariusza bez większych trudności, ustalenie rodzajów odpowiedzi na poszczególne pytania (tzn. stwierdzenie, czy odpowiedzi respondentów są zgodne z zapotrzebowaniem badawczym), ustalenie, czy uzyskiwane informacje są prawdziwe, ocena funkcjonowania pozostałych elementów narzędzia, ale także sprawdzenie funkcjonowania założonej procedury badania poprzez analizę sytuacji wywiadu (Antonie

1990: 165–170; Boeswillwald 1991: 296–298; Lutyńska 1975; Sztabiński i Sztabiński 2005b)⁴. Oznacza to, iż w badaniu pilotażowym narzędzie badawcze testowane jest nie tylko „na respondentach” przez ankieterów, ale także „na kwestionariuszach” przez koderów (Frey i Oishi 1995: 108);

– badania próbne, które w przeciwieństwie do badania pilotażowego ma cel substancywny. Służy ono sprawdzeniu jak „pracują” wskaźniki, więc jego celem jest ocena przyjętego sposobu operacjonalizacji koncepcji badawczej. Dlatego też zebrany materiał (zbioru danych z badania) analizuje się analogicznie do planowanego w badaniu zasadniczym. Uzyskane rozkłady odpowiedzi respondentów i wyniki innych analiz pozwalają ocenić, czy przyjęte wskaźniki rzeczywiście mierzą to, co z założenia mają mierzyć. Celem badania próbnego jest nie tylko „przetestowanie” narzędzia w szerszym zakresie, ale także organizacji badania i wstępnie – hipotez badawczych (uzyskane dane opracowuje się bowiem zgodnie z przewidzianym schematem analizy badania zasadniczego). Odmiennie wyniki, niż można było oczekiwać na gruncie przyjętej teorii, stanowią wskazówkę do ponownego przemyślenia sposobu operacjonalizacji przyjętych wskaźników lub nawet sposobu konceptualizacji badania (Alreck i Settle 1985: 92–94; Sztabiński i Sztabiński 2005b)⁵. Badania próbne przeprowadza się na większej próbie niż

⁴ Istotnym elementem jest tutaj analiza czasu trwania wywiadu i jego wpływu na wartość uzyskiwanych informacji. Przykładowo, w jednym z prowadzonych przeze mnie badań pilotażowych respondenci odmówili kontynuowania wywiadu (przerwali wywiad) w 10 na 12 przypadków. Najdłuższy wywiad, do momentu przerwania, trwał 200 min, najkrótszy zaś – 50 min. Tylko dwa wywiady udało się zrealizować w całości: zakończyły się one „sukcesem” wyłącznie dlatego, że ankieter dopuścił możliwość, aby w trakcie rozmowy respondentki zajmowały się drobnymi pracami domowymi (sprzątanie/przygotowywanie posiłku). Z pewnością jednak sytuacja ta przesądza o niskiej wartości uzyskanych odpowiedzi.

⁵ Warto podkreślić, iż badanie próbne od pilotażowego różni się nie tylko celami, ale także liczebnością i charakterem próby, na której jest realizowane. Badanie próbne prowadzi się na próbach liczących zazwyczaj 200–500 osób i stanowiących „miniaturę” próby do badań zasadniczych (próba dobierana jest w analogiczny sposób jak planowana próba w badaniu zasadniczym). Z kolei badania pilotażowe prowadzi się z reguły na niewielkiej, liczącej 20–50 osób próbie (jej niewielka liczebność związana jest z jakościowym, nie zaś ilościowym charakterem opracowania wyników badania). Co więcej, próba nie musi odzwierciedlać struktury próby w badaniach zasadniczych, choć jest wskazane, aby wszystkie jej najważniejsze segmenty były reprezentowane. Próba połączenia funkcji tych dwóch typów badań została

pilotaż. Jest to próba analogiczna do tej, na której realizowane będzie badanie zasadnicze.

Wśród wskazanych przez Lutyńską sposobów zapewnienia właściwej jakości danych, szczególne miejsce zajmuje wspomniane powyżej konsultacja metodologiczna oraz pilotaż⁶. Przy czym, co podkreśla Lutyńska, wpływ badacza na jakość uzyskanych informacji ma jednak charakter pośredni: nie tylko poprzez narzędzie badawcze, ale także poprzez ankietorów, którzy zadają pytania i zapisują odpowiedzi (por. de Maio 2009).

B. Drugim zasadniczym etapem procesu badawczego jest zbieranie materiałów. Etap ten obejmuje trzy zasadnicze fazy: szkolenie ankietorów, realizację terenową oraz bieżącą kontrolę realizacji badania. Procedury QC na tym etapie procesu badawczego stanowią specyficzną kategorię w dwojakim sensie. Po pierwsze, ponieważ odnoszą się do tego etapu badań, który jest zasadniczo niepowtarzalny. O ile inne elementy tego procesu (przygotowanie schematu losowania próby, konstrukcja kwestionariusza czy wprowadzanie danych i ich analiza) podlegać mogą ewentualnej korekcie, to kontakt ankietera z respondentem ma charakter jednorazowy. Ewentualne zniekształcenia uzyskanych rezultatów badań nie są zazwyczaj możliwe do skorygowania. Po drugie specyfika proce-

podjęta w związku z realizacją badań wstępnych na potrzeby projektu ESS. Metodologia tego badania polega na połączeniu analizy ilościowej z jakościową i pozwala przezwyciężyć ograniczenia badania pilotażowego i próbnego (tamże).

⁶ Mówiąc o pilotażu, Lutyńska (1975) ma na myśli tak zwany pilotaż „pogłębiony”. W tego typu pilotażu, w trakcie wywiadu zadaje się respondentowi pytania dodatkowe (tzw. pilotażowe), w wyniku czego tylko w niewielkim stopniu przypomina on „normalny” – zasadniczy – wywiad (w skrajnym przypadku tego rodzaju wywiad pilotażowy może przekształcić się we wspomniany wywiad o wywiadzie). Realizując pilotaż „pogłębiony”, uzyskuje się wiele informacji dotyczących poszczególnych pytań, niewiele jednak dowiadujemy się na temat funkcjonowania narzędzia jako całości. Można jednak wskazać inny rodzaj pilotażu, a mianowicie pilotaż „symulacyjny”. Imituje on standardowy wywiad kwestionariuszowy w sferze interogacyjnej, co oznacza, iż ankieter pilot nie zadaje respondentowi żadnych dodatkowych pytań, dzięki czemu wywiad nie zostaje zakłócony. Tym samym przynosi on wiele informacji na temat sytuacji wywiadu i funkcjonowania całego narzędzia, jednak na jego podstawie w ograniczonym zakresie możemy wypowiadać się na temat rozumienia pytania i wiarygodności odpowiedzi (Sztabiński, Sztabiński 2005). W obu typach pilotażu niezwykle istotne jest prowadzenie przez ankietera obserwacji zachowań (werbalnych i niewerbalnych) respondenta oraz sytuacji wywiadu i jego warunków zewnętrznych. Wyniki obserwacji są mianowicie podstawą formułowanych wniosków (por. także de Maio, July 2009).

dur QC na tym etapie polega na tym, że ich przedmiotem są ankieterzy, a więc osoby spoza zespołu badawczego (w tym sensie „zewnętrzne”), a co więcej, osoby nieposiadające specjalistycznego przygotowania zawodowego, lecz „przyuczone”. Jednocześnie musimy pamiętać, że etap zbierania materiałów w terenie jest najdroższym elementem procesu badawczego.

Na tym etapie procesu badawczego podstawowe działania z zakresu QC (por.: Babbie 2004: 290–295; Boeswillwald 1991: 298–301; Frey i Oishi 1995: 109–146) obejmują:

- selekcję ankieterów do badania, przykładowo ze względu na cechy społeczno-demograficzne: wiek, płeć, sytuację zawodową, ale także doświadczenie, umiejętności lub predyspozycje. Jest to istotny element, ponieważ charakterystyki ankieterów mogą wpływać na uzyskiwane odpowiedzi (Sztabiński P. 1997);

- centralne szkolenie koordynatorów do badania (w przypadku sieci zdecentralizowanej – „dwustopniowej”). Szkolenie takie stanowi jedyną możliwość bezpośredniego przekazania wykonawcom badania informacji na jego temat, a więc jego założeń teoretycznych, specyfiki ujęcia poruszanej problematyki itd. Stąd też istotnym elementem QC jest osobiste uczestnictwo badacza w szkoleniu. Pełni ono funkcje kontrolne (mamy pewność, iż szkolenie zostało przeprowadzone właściwie) oraz motywacyjne (osobiste uczestnictwo „podnosi” rangę projektu). W jednym i drugim przypadku nie pozostaje to bez znaczenia dla jakości realizacji badania. Mówiąc o szkoleniu jako elemencie QC, nie można oczywiście zapominać o innych jego elementach, takich jak konieczność przekazania uwag ogólnych do próby i narzędzia badawczego, szczegółowej analizie kwestionariusza („pytanie po pytaniu”), a także sprawach organizacyjnych (Antonie 1990: 191–192; ESS: *Field Procedures in the ESS R3. Enhancing Response Rate*, 2006; Sztabiński F. 2005: 119–127);

- regionalne szkolenie ankieterów do badania (także w przypadku sieci zdecentralizowanej). Szkolenie w terenie winno być, ze względu na wymóg standaryzacji „repliką” badania centralnego (pomijając oczywiście osobiste uczestnictwo badacza w szkoleniu)⁷;

⁷ Należy tutaj jednak wspomnieć, iż w ostatnio realizowanej fali badania panelowego „Struktura społeczna 1988–2008” (projekt „Pol-Pan”, kierowany od 1988 roku przez Kazimierza M. Słomczyńskiego), uczestnikami regionalnych szkoleń ankierskich byli także członkowie zespołu badawczego.

– kontrolę szkoleń regionalnych przez zewnętrzną – niezależną od ośrodka badawczego – instytucję. W tym przypadku mamy pewność, iż szkolenie w regionie zostało przeprowadzone, a co więcej, mamy możliwość oceny, na ile zostało ono przeprowadzone zgodnie w wymogami, co oczywiście nie pozostaje bez wpływu na jakość realizacji badania w terenie przez ankierów (Sztabiński F. 2005b: 120);

– obserwację realizacji wywiadów próbnych, której wyniki traktować można jako sprawdzenie „efektywności” szkolenia ankierów. Ten element QC przez większość autorów nie jest wymieniany, ze względu na jego kosztowność⁸;

– bieżąca kontrola realizacji badania (monitoring), która obejmuje nadzór koordynatora-centrali (w zależności od struktury sieci ankierskiej) nad realizacją badania w terenie przez ankierów. Jego przedmiotem może być: odsetek realizacji wywiadów (w tym wywiadów skonwertowanych), poprawność wypełnienia kwestionariuszy, ale także koszty realizacji wywiadów (Frey i Oishi 1995: 13–14, 142–145). W tej fazie procesu badawczego, narzędziem QC są bieżące sprawozdania z realizacji badania przez poszczególnych ankierów w każdym z regionów. Mogą to być sprawozdania obejmujące kilka dni lub nieco dłuższy okres itd. Procedura ta umożliwia nadzór i wczesną identyfikację problemów realizacyjnych w przypadku poszczególnych ankierów, regionów i całego badania. Dzięki temu można podjąć działania zaradcze (zob. Babbie 2004: 295; Sztabiński F. 2004: 143–166; ESS: *Fortnightly progress reports from survey organizations*, 2002).

C. Trzecim wreszcie zasadniczym etapem procesu badawczego jest opracowanie zebranego materiału. Etap ten obejmuje następujące fazy: ocenę i selekcję materiałów, właściwe ich opracowanie (kodowanie i wprowadzanie) oraz przygotowanie raportu (analiza i interpretacja danych).

Pierwsza z nich (ocena i selekcja materiałów) ma bezpośredni związek z wcześniejszym etapem badania (etap zbierania materiałów), ponie-

⁸ Element ten jest jednak standardową procedurą stosowaną przez Centre d'Etude des Supports de Publicité (Boeswillwald 1991: 299–301; Chavenon, Dupont, Haering, Libs 1999: 18–19). Sytuację tę należy jednak odróżnić od techniki wywiadu obserwowanego, która traktowana jest jako jedna z metod kontroli pracy ankiera (w tym ostatnim przypadku kontrolowane w ten sposób wywiady eliminowane są z opracowania).

waż odnosi się do pracy ankieterów. W tej fazie realizacji badania chodzi ogólnie o ocenę autentyczności materiałów (Topolski 1984: 357–376). Ze względu na to, że materiały są zbierane przez ankietera, jest to zatem ocena rzetelności i uczciwości ankietera. Chodzi tutaj nie tylko o sam fakt realizacji wywiadu, ale realizację wywiadu z właściwą osobą, to znaczy wskazaną, poprawnie wylosowaną lub poprawnie dobraną (w zależności od rodzaju próby), a co więcej – realizację wywiadu zgodnie z przewidzianą procedurą badania⁹. Musimy bowiem pamiętać, że w momencie rozpoczęcia realizacji terenowej badacz praktycznie traci kontrolę nad badaniem: nie wie, w jakich warunkach przebiegały rozmowy, czy były prowadzone przez właściwych (odpowiednio przeszkolonych) ankieterów, z właściwymi, to znaczy wylosowanymi, respondentami itd. Dlatego też każdy badacz po ich zakończeniu dokonuje, a przynajmniej powinien dokonać oceny zebranego materiału. Jeśli w stosunku do danego wywiadu można wysunąć uzasadnione wątpliwości, winien on zostać usunięty z dalszego opracowania. Jest to klasyczne zalecenie, które zamieszcza się w podręcznikach badań jako standard (Hyman 1954: 32; Lininger 1975; Mann 1968; Noelle 1963; Riesman 1958: 267, 282; Sellitz i in. 1959: 240–242 oraz Gostkowski 1966: 48–49, 59)¹⁰.

W tej fazie procesu badawczego działania z zakresu QC mają zazwyczaj charakter ekstensywny (Antonie 1990: 151–152, 184–194; Boeswillwald 1991: 298–303; Frey i Oishi 1995: 144–146; Sztabiński F. 1995; ESS R. 5. *Specification for participating countries*, 2011) i obejmują:

– ocenę terenową zebranego materiału. Sprowadza się ona do sprawdzenia – w trakcie kontaktu niezależnego inspektora z respondentem – realizacji przez ankietera wywiadu i poszczególnych elementów procedury badania. W tej fazie procesu badawczego jest ona najistotniejszym

⁹ Earl Babbie (2004: 165–166) mówi wprost o rzetelności osób zatrudnionych przy realizacji badań.

¹⁰ Warto podkreślić, iż w ciągu ostatnich lat stosunek do zagadnień kontroli uległ w Polsce istotnej zmianie. O ile wcześniej zdarzało się, iż kontrola traktowana była jako wyraz nieufności i próba podważenia rzetelności oraz wiarygodności nie tylko ankieterów, ale i instytutu badawczego, o tyle dziś systematyczna kontrola fazy terenowej każdego badania daje możliwość pozytywnego „odróżnienia się” i to zarówno jeśli chodzi o ośrodek, jak i ankieterów. Świadectwem owej „przemiany” są między innymi zapisy zawarte w Programie Kontroli Jakości Pracy Ankieterów (zob. Program Kontroli Jakości Pracy Ankieterów, www.ofbor.pl).

elementem QC, ponieważ pozwala ocenić, oprócz autentyczności zebranych informacji, także i inne aspekty czy warunki realizacji badania. Przedmiotem tej oceny mogą być różne elementy wywiadu w zależności od charakteru badania. Zazwyczaj są to: tematyka rozmowy (chodzi o ocenę kompletności zadawanych pytań), wykorzystanie materiałów pomocniczych, termin wizyty i czas trwania wywiadu, jego atmosfera, ocena ankietera itd. Kontrola terenowa może jednak skupiać się nie tylko na ocenie pracy ankieterów (poprawności realizacji wywiadu czy doboru respondenta), ale także na ocenie uzyskanych informacji (określa się ją na podstawie „powtórzonego pomiaru”, a więc powtórnym zadaniu respondentowi wybranych pytań z kwestionariusza)¹¹;

– ocenę nieterenową zebranego materiału. Chodzi tutaj przede wszystkim o ocenę wewnętrznej spójności materiałów zebranych przez ankietera, ale także informacji przez niego dostarczonych (przygotowanych). W pierwszym przypadku to spójność informacji zawartych w kwestionariuszu, ich zgodność logiczna, ale także ich spójność z informacjami uzyskanymi z innych wywiadów, a więc zgodności ich profilu z profilem całości uzyskanych informacji. W drugim przypadku chodzi o spójność informacji zawartych w kwestionariuszu oraz danych zawartych w materiałach realizacyjnych (np. formularzach doboru, kartach realizacji badania, ankiecie do ankietera po wywiadzie itd.)¹².

W fazie opracowania materiałów, a więc kodowania danych, wprowadzania ich na nośnik komputerowy oraz przygotowania zbioru, QC sprowadza się zasadniczo do następujących elementów (Babbie 2004: 424–435):

– ocena poprawności przygotowania kluczy kodowych dla odpowiedzi na pytania otwarte. Rzeczą oczywistą jest, że przygotowując klucze kodowe, należy kierować się koncepcją badawczą i sformułowanym zapotrzebowaniem informacyjnym. Poprawność przygotowania kluczy

¹¹ Tego typu ocena terenowa materiału (w szerszym lub węższym zakresie) winna być prowadzona przynajmniej na dwóch poziomach: przez koordynatora lub/i ośrodek badawczy oraz przez zewnętrzną (od ośrodka) instytucję. Ten ostatni typ oceny terenowej – ocena przez zewnętrzną instytucję – wydaje się optymalny. Instytucja zewnętrzna nie jest bowiem, w przeciwieństwie do koordynatorów lub/i ośrodka badawczego, w żaden sposób zainteresowana wynikiem oceny.

¹² Rzeczą oczywistą jest, iż przedmiot oceny terenowej i nieterenowej nie musi być rozłączny. Najprostszym przykładem jest wspomniany termin wizyty i czas trwania wywiadu. Do problemów tych powrócę w kolejnym rozdziale.

można ocenić, zlecając to zadanie dwóm niezależnym osobom, a następnie porównując rezultaty ich pracy;

– ocena poprawności kodowania danych. Podobnie jak w przypadku przygotowania kluczy, aby ocenić poprawność kodowania, czynność tę można powierzyć dwóm niezależnym zespołom koderskim, porównując następnie rezultaty ich pracy;

– ocena poprawności wprowadzania danych, w tym kontrola zakresu wartości zmiennych. Ocena taką umożliwi między innymi dwukrotne wprowadzenie danych i porównanie uzyskanych rezultatów¹³;

– ocena logicznej spójności przygotowanego zbioru danych, która może być dokonana poprzez analizę danych z pytań wzajemnie kontrolujących się lub logicznie powiązanych¹⁴.

Ostatnia faza tego etapu, a jednocześnie ostatnia faza całego procesu badawczego to analizy statystyczne, interpretacja danych i przygotowanie raportu. QC sprowadza się tutaj zasadniczo do oceny poprawności zastosowania określonych mierników i modeli statystycznych. Nie chodzi wyłącznie o oczywistą rzecz, jaką jest trafność stosowanych wskaźników, ale przede wszystkim stopień odpowiedniości materiału empirycznego i założeń modeli statystycznych, dotyczących między innymi rozkładów cech w badanej populacji (Babbie 2004: 438–458, 478–510; Laniel i Lavallée 1991: 265–284; Sztabiński F. 2003: 154–172).

¹³Tego typu rozwiązania w zakresie oceny poprawności kodowania danych oraz oceny poprawności wprowadzania danych przyjęte zostały w projekcie Programie Międzynarodowej Oceny Umiejętności Uczniów (The Programme for International Student Assessment, PISA). W projekcie tym przed rozpoczęciem kodowania materiału z badania zasadniczego koduje się próbnie kilkanaście zestawów każdego rodzaju kwestionariuszy (jest to zatem pewnego typu „pilotaż”). Robią to każdorazowo dwie niezależne osoby. W efekcie możliwe jest sprawdzenie nie tylko kluczy kodowych, ale także określenie, jakiego rodzaju błędy popełniają koderzy. Informacje te są wykorzystywane podczas szkolenia wszystkich koderów, którego elementem jest próbne kodowanie materiału pod nadzorem kontrolera. Zgodnie z przyjętą procedurą, po zakodowaniu całego zebranego materiału dodatkowo kodowano jeszcze trzykrotnie materiały z pytań otwartych. Porównanie rezultatów umożliwia ocenę wielkości błędów w fazie kodowania. Zakodowane dane także zostają wprowadzone do komputera dwukrotnie, w celu sprawdzenia zgodności klucza kodowego z programem do przenoszenia danych na elektroniczne nośniki informacji (zob. Program PISA: przygotowanie i przebieg badań, www.ifispan.waw.pl/ifis/badania).

¹⁴Tego typu kontrolę prowadzi się automatycznie, wykorzystując specjalnie przygotowywane oprogramowanie komputerowe.

Charakteryzując dynamiczne ujęcie problemu oceny jakości danych, którego egzemplifikacją może być podejście *quality control*, należy podkreślić, iż wbrew nazwie, nie jest to kontrola w sensie potocznym, a więc ocena następująca po zakończeniu danego etapu czy fazy badania. *Quality control* jest bowiem, jak wspomniałem, systemem działań zmierzających do podniesienia jakości realizowanych zadań lub inaczej, zastosowaniem procedur optymalizujących rezultaty badania i eliminujących potencjalne błędy. Z punktu widzenia centralnej fazy procesu badawczego, a więc etapu realizacji terenowej badania, działania QC można podzielić na trzy grupy:

- działania „ex ante”, obejmujące konsultację koncepcji, narzędzia badawczego, pilotaż narzędzia, szkolenia ankierów itd.,
- działania „on going”, podejmowane w trakcie realizacji terenowej, oraz
- działania „ex post”, po zakończeniu etapu terenowego, a więc kontrola ankierów i selekcja materiałów, kontrola kodowania, kontrola wprowadzania danych itd.

Bardziej trafna wydaje się jednak charakterystyka QC jako procesu nie tyle oceniania, ile wspomagania w optymalizacji poszczególnych faz badania. Ów proces „wspomagania” nie przebiega jednak „obok”, ale „równolegle” z zasadniczym procesem badawczym.

Druga uwaga dotyczy poszczególnych elementów QC, odpowiadających kolejnym fazom badania. Rzeczą oczywistą jest, iż wskazana ich „lista” nie jest wyczerpująca i ostateczna. Winna być każdorazowo modyfikowana: uzupełniana lub ograniczana, w zależności od charakteru i specyfiki badania. Na przykład w badaniach międzykrajowych element konsultacji narzędzia badawczego winien być zastąpiony procesem tłumaczenia narzędzia badawczego z języka źródłowego na języki krajowe. Proces ten winien także mimo wszystko podlegać określonym procedurom z zakresu QC. Jedną z nich jest tak zwana procedura TRAPD, która jest akronimem terminów: *Translation, Reviewing, Adjudication, Pre-testing, Documentation* (Harkness 2003: 35–56, 2007). Podobnie, w fazie opracowania materiałów rezygnujemy niekiedy z przygotowania kluczy kodowych do pytań otwartych, wykorzystując już istniejące. Jest tak na przykład w pytaniach otwartych o zawód: do ich kodowania wykorzystuje się klasyfikację International Standard Classification of Occupations (ISCO 2008) lub Społeczną Klasyfikację Zawodów (SKZ–2007) [Domański, Sawiński, Słomczyński 2007]. Inną sprawą także związa-

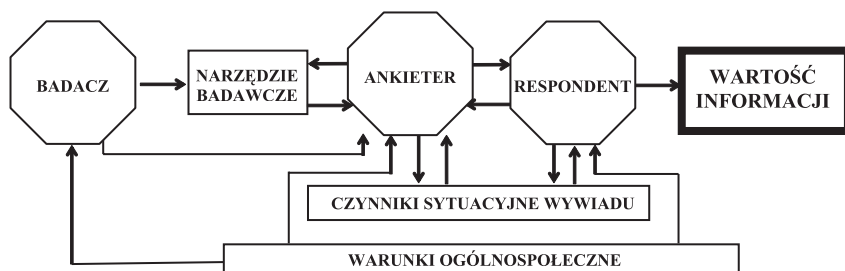
na z problemem „listy” elementów QC, pomijając merytorycznie uzasadnione jej modyfikacje, jest konieczne niekiedy jej „urealnienie” ze względu na dwa czynniki „zewnętrzne”: koszty badania oraz długość okresu realizacji badania (Groves 1989). Autorzy omawianych przeze mnie prac abstrahują od tych dwóch istotnych ograniczeń, prezentując modelowy przebieg procesu badawczego. Jest to z pewnością słuszne, jednak realizując konkretne badanie, zawsze musimy brać pod uwagę jego budżet oraz okres realizacji i prowadząc „rachunek zysków i strat”, modyfikować listę QC.

1. 1. 2. Ujęcie statyczne: analiza elementów procesu badawczego

Z pewnego punktu widzenia badanie można traktować jako określonego rodzaju sytuację komunikacyjną, a ściślej biorąc „sieć komunikacyjną” (Lutyńska 1984, 1993; Lutyński 1968, 1972; Sztabiński F. 1990, 1997). Proces komunikowania przebiega bowiem nie tylko w obrębie i między całym zespołem badawczym: badaczem, konsultantami oraz respondentami i zespołem pomocniczym (ankieterzy, koderzy, osoby wprowadzające dane, analitycy itp.), ale także czytelnikami, którzy są odbiorcami rezultatów badań. Tak rozumiany proces badawczy można analizować, biorąc pod uwagę uczestników procesu komunikowania oraz wymieniane między nimi przekazy. Koncentrujemy się wówczas przede wszystkim na poszczególnych elementach procesu komunikowania, a więc jego uczestnikach (badacz, ankieter, respondent itd.), oraz wymienianymi między nimi przekazami, nie zaś następstwie czasowym poszczególnych faz badania. Stąd też ujęcie to nazywam statycznym, w przeciwieństwie do dynamicznego, które scharakteryzowałem w poprzedniej części.

Uzemplifikacją statycznego ujęcia mogą być rozważania Krystyny Lutyńskiej (1984: 34 i n.), która analizując różne czynniki wpływające na wartość uzyskiwanych informacji, przedstawia proces badawczy (por. Brzeziński 1978: 4) w sposób następujący:

Rysunek 2. Główne kierunki oddziaływania różnych grup czynników na odpowiedzi respondenta*



* Strzałki oznaczają najważniejsze kierunki oddziaływania poszczególnych grup czynników.

Zgodnie z powyższym schematem wywiad kwestionariuszowy jest pewnego typu łańcuchem komunikacyjnym, którego głównymi uczestnikami są: badacz, ankieter i respondent. Osoby te wzajemnie na siebie wpływają, a ich interakcje przebiegają w określonych warunkach i w konkretnej sytuacji społeczno-kulturowo-historycznej. Celem interakcji (procesu komunikacji) jest uzyskanie określonego rodzaju informacji. Tym samym wywiad traktowany jest jak relacja między badaczem, respondentem i ankieterem w określonej sytuacji społeczno-kulturowej.

Zarówno uczestnicy owej relacji, jak i czynniki na nich oddziałujące mogą wpływać na wartość otrzymywanych informacji. Według Lutyńskiej (1984: 27–31, 1993: 5–6) o wartości danych decyduje ich kompletność, istotność oraz trafność lub też, zgodnie z ujęciem „socjologiczno-metodycznym” – pełność i wiarygodność¹⁵. Uzyskanie infor-

¹⁵ Kategorie powyższe autorka odnosi do odpowiedzi na dane pytanie danego respondenta, nie zaś ostatecznego wyniku w odniesieniu do danego pytania. Mówiąc więc o narzędziu badawczym, ma na myśli pojedyncze pytanie. Pierwsza z kategorii – pełność – wydaje się odpowiadać temu, co inni autorzy nazywają „kompletnością” (De Leeuw i van der Zouwen 1991: 10–11, 17–18; Lejeune i Bied-Charreton 1991: 70–77; Frey i Oishi 1995: 28–29). Należy także podkreślić, iż kategorie istotności i trafności autorka odnosi do tak zwanej rozszerzonej informacyjnej koncepcji pytania (Lutyński 1978, 1979).

macji, a więc jej kompletność, stanowi warunek konieczny dalszej analizy. Odpowiedź respondenta na dane pytanie można oceniać z punktu widzenia jej istotności (Lutyńska 1975: 57–64, 1984; Lutyński 1972: 68–86). Odpowiedź istotna to taka, która przynosi poszukiwaną przez badacza jednostkową informację, natomiast odpowiedź nieistotna takiej informacji nie przynosi. To, czy dana odpowiedź zaspokaja zapotrzebowanie informacyjne badacza, czy też nie, rozstrzyga się na jej podstawie analizy istotności. Jak wspomniałem, analizę tę prowadzi się indywidualnie dla każdego pytania. Polega ona na ocenie zgodności zapisanej w kwestionariuszu odpowiedzi respondenta z zadaniem postawionym w pytaniu. Daną odpowiedź można uznać za istotną, gdy możliwe jest zaliczenie jej do którejś z kategorii badanych zjawisk. Ogólnie można powiedzieć, iż o istotności danej odpowiedzi decyduje stopień jej „użyteczności” dla rozwiązania danego problemu. (Lutyński 1972). Sens określenia „istotność odpowiedzi” zależy jednak zawsze od zapotrzebowania badawczego¹⁶.

Nie każda jednak istotna informacja (odpowiedź respondenta) może zostać uznana za w pełni wartościową – trafną, to znaczy prawdziwą (Lutyński 1972: 62–86). Przykładowo odpowiedzi udzielane przez respondenta „na chybił-trafił” mogą być trafne jedynie przypadkowo. Innymi słowy, odpowiedzi nietrafne lub budzące zastrzeżenia co do swojej trafności to takie, które niezależnie od tego, że mogą być istotne – z takich lub innych względów przynoszą informacje nieprawdziwe. O sukcesie procesu interwencji świadczą bowiem jedynie te odpowiedzi, które przynoszą prawdziwe informacje, a więc takie, które pozwalają na zaliczenie odpowiedzi respondenta do tej kategorii badanych zjawisk, do której faktycznie należą. Można przyjąć, iż informacje takie uzyskuje się w rezultacie zakładanego procesu formułowania odpowiedzi, a więc „prawidłowego” (z punktu widzenia badacza) przebiegu procesów psychicznych u respondentów.

¹⁶ Jeżeli jedną i tę samą odpowiedź respondenta będziemy klasyfikować z różnych punktów widzenia, to z jednej strony ze względu na jedną poszukiwaną informację może ona być nieistotna, z drugiej zaś – ze względu na inną poszukiwaną informację – istotna. Przykładowo odpowiedź „nie mam zdania” na pytanie o „stosunek do aborcji” jest odpowiedzią nieistotną, gdy interesuje nas opinia respondenta na ten temat. Ta sama odpowiedź stanie się jednak istotna, gdy interesuje nas nie opinia, lecz sam fakt posiadania (lub nie) przez niego opinii na dany temat.

O tym, czy dana odpowiedź jest „prawdziwa” czy też nie, rozstrzyga się na podstawie analizy trafności¹⁷. Polega ona na ocenie procesów psychicznych prowadzących do udzielenia przez respondenta odpowiedzi na dane pytanie. Jeśli w trakcie odpowiadania na pytanie zaszły w respondencie pożądane przez badacza procesy psychiczne (tzw. procesy konieczne), to uzyskaną informację uznaje się za trafną. Informacje o procesach koniecznych uzyskać można, między innymi, na podstawie obserwacji zachowań werbalnych i niewerbalnych respondenta oraz całej sytuacji badania, a także analizy odpowiedzi respondenta na pytania mające ustalić przebieg tych procesów¹⁸. Analiza trafności odpowiedzi pozwala zatem na pełniejszą ocenę uzyskanej informacji niż analiza istotności.

Jaki jest zatem wpływ wspomnianych czynników (uczestników i warunków badania), na tak rozumianą wartość informacji? Zdaniem Lutyńskiej (1984: 10, 29–33, 1993: 5–6) źródłem nietrafności czy braku wiarygodności uzyskiwanych odpowiedzi, a więc ogólnie niskiej jakości informacji może być¹⁹:

- badacz i narzędzie badawcze, w tym przede wszystkim błędy pytań;
- ankieter, jeśli wywiad realizowany jest w niewłaściwy sposób;
- respondent, który ze względu na cechy osobowościowe, stan psychofizyczny (na przykład zmęczenie), ale także na przykład chęć popisania się lub przedstawienia się w dobrym świetle (chodzi o udzielanie odpowiedzi społecznie pożądanych), udziela odpowiedzi nietrafnych, ale także:

¹⁷Oczywiście prezentowane przez Lutyńskiego rozumienie analizy trafności jest ogólniejszym, a dokładniej, krzyżującym się pojęciem, niż analiza trafności wskaźników.

¹⁸Chodzi tutaj o tzw. introspekcję pytaniami w wywiadzie o wywiadzie (Lutyński 1972; 1975; 1975a; Przybyłowska 1975: 457–488). Jej zastosowanie w przypadku masowych badań wydaje się jednak problematyczne. Do sprawy tej powrócę pod koniec niniejszej części. Należy podkreślić, iż odpowiedzi trafne, o których mówi Jan Lutyński, nie mają nic wspólnego z trafnością, tak jak jest ona rozumiana na gruncie psychologii. Autor ten odpowiedzi „trafne” nazywał wcześniej „pożądanymi”

¹⁹Należy podkreślić, że termin „błąd” Lutyńska używa w innym rozumieniu niż ja w kolejnej części poświęconej błędom pomiaru. Błąd rozumiany jest tutaj jako odstępstwo od określonych zasad, dyrektyw lub reguł, a więc różnica między modelowym (zakładanym), a faktycznym przebiegiem poszczególnych faz badania (por. Kistelski 1981).

- czynniki mikrospołeczne związane z sytuacją wywiadu, w tym warunki zewnętrzne (na przykład miejsce wywiadu, osoby trzecie, hałas itp.), środowiskowe oraz wewnętrzne (tzn. relacje między ankieterem i respondentem);
- warunki kulturowe, ogólnospołeczne (makrospołeczne).

Oddziaływanie tych czynników na wartość uzyskiwanych informacji (istotność i trafność odpowiedzi respondenta) jest wielorakie i różnokierunkowe i może być, co oczywiste, pozytywne i negatywne. Pierwsza ich grupa to narzędzie badawcze i badacz, który ma bezpośredni wpływ na jego ostateczny kształt²⁰. Lutyńska podkreśla, że właśnie narzędzie badawcze może w sposób decydujący wpływać na wartość uzyskiwanych informacji – im lepiej więc zostanie ono przygotowane, tym większa szansa na uzyskanie odpowiedzi istotnych i trafnych. Narzędzie winno być tak przygotowane, aby ani dla respondenta, ani dla ankietera nie było zbyt trudne: należy bowiem brać pod uwagę ich możliwości intelektualne, psychiczne, językowe i fizyczne (Lutyńska 1984: 34–55)²¹. Oznacza to, iż przygotowując kwestionariusz, należy mieć świadomość, iż staje się on narzędziem badawczym we właściwym sensie dopiero w terenie i musi uwzględniać nie tylko zapotrzebowanie badawcze, ale także, a może przede wszystkim, respondenta i ankietera (chodzi tutaj nie tylko o takie sprawy jak sformułowanie pytań, ale także długość wywiadu, „przeładowanie” kwestionariusza, jego „niespójność” językową i tematyczną itp.).

Druga grupa czynników wpływających na wartość uzyskiwanych informacji związana jest z osobą ankietera. Z pewnego punktu widzenia ankieter jest postacią „centralną” w procesie badawczym, ponieważ to

²⁰ Lutyńska utożsamia tutaj, jak się wydaje, wpływ narzędzia badawczego i badacza. Jest to słuszne, o ile mówiąc o badaczu, ma się na myśli wyłącznie twórcę narzędzia. Nie można jednak pomijać „instytucjonalnej przynależności” badacza. W literaturze metodologicznej można znaleźć wiele przykładów wskazujących na występowanie tak zwanego efektu sponsora (zob. m.in.: Dillman 1978: 6; Hyman 1954; Sułek 2002: 46–80; Śmiłowski 1990).

²¹ W tym przypadku Lutyńska mówi o narzędziu badawczym w szerokim sensie, a więc nie tylko kwestionariuszu do wywiadu, ale również instrukcji dla ankieterów oraz tzw. materiałach pomocniczych (Lutyńska 1984: 9; Lutyński 1968; 1972). Równocześnie autorka wspomina o narzędziu badawczym także w wąskim sensie, mając na myśli pojedyncze pytanie (por. Lutyńska 1975; 1984; Lutyński 1964; 1978). O tych dwóch podejściach pisze między innymi Paweł B. Sztabiński (1997: 19–21). Por. przypis 15.

jemu powierza się zbieranie informacji, a więc realizację niepowtarzalnej fazy badania. Jednocześnie przyjmuje się, że ankieter jest jednym z najbardziej znaczących czynników odkształcania danych uzyskiwanych w badaniach kwestionariuszowych (Lutyńska 1993: 95). Wiąże się to z tym, iż w trakcie realizacji wywiadu ankieter wchodzi w bezpośrednią interakcję z respondentem i to on określa jej charakter i przebieg. Jej przebieg, mimo licznych zabiegów standaryzujących, jest różny w przypadku różnych ankieterów, co rzecz jasna nie pozostaje bez wpływu na odpowiedź respondenta, a tym samym na wartość uzyskiwanych informacji i wynik badania. Zjawisko to określa się mianem wpływu ankieterskiego²². Ankieter wywiera wpływ na respondenta nie tylko w sposób bezpośredni, ale także poprzez przekazy badacza, a które także oddziałuje. Tym samym podstawowe drogi wpływu ankieterskiego stanowią: zadawanie pytań, zapisywanie odpowiedzi oraz wzajemna interakcja ankietera i respondenta (wykraczająca poza zadawanie pytań i zapis odpowiedzi).

Wpływ ankietera na odpowiedź respondenta może być zarówno pozytywny, jak i negatywny. Lutyńska (1984: 39–40) przyjmuje, że pozytywny wpływ ankieterski polega na zwiększeniu liczby odpowiedzi istotnych (przynoszących poszukiwane informacje) i odpowiedzi trafnych (przynoszących informacje prawdziwe), negatywny zaś, na zwiększeniu liczby tego typu odpowiedzi. Głównymi źródłami wpływu ankieterskiego są: cechy ankieterów (społeczno-demograficzne, osobowościowe i kulturowe), ich poglądy; oczekiwania co do odpowiedzi respondentów, ale także ich stosunek do narzędzia badawczego i instytutu badawczego. W szczególności istotny jest tu wpływ poglądów ankieterów, ich oczekiwań oraz wykształcenia i pewnych cech osobowych (Lutyńska 1978, 1993: 102–112; Sztabiński P. 1997). Właściwe przygotowanie narzędzia badawczego, szkolenie ankieterów, realizacja wywiadów próbnych, odpowiednia organizacja badania, systematyczna kontrola pracy ankieterów itp. pozwalają jednak w pewnym zakresie ograniczyć liczbę odpowiedzi nieistotnych i nietrafnych (Kistelski 1981; Lutyńska 1984: 41–43).

Dla oceny zebranego materiału niezwykle istotne są analizy, których celem jest określenie rozmiarów negatywnego wpływu ankieterskiego. Zdaniem Lutyńskiej (1993: 99) określenie wpływu ankieterskiego jest

²² Mówiąc o wpływie ankieterskim, Lutyńska (1978) pomija problem oszustw ankieterskich, które są świadomym naruszeniem zasad pracy ankietera.

jednak bardzo trudne, ponieważ zakłada przeprowadzenie weryfikacji uzyskanych danych na poziomie jednostkowym, przy pomocy sposobu pewnego, co nie zawsze jest możliwe (dopiero gdy znany jest wynik „prawdziwy”, można stwierdzić kierunek wpływu ankietarskiego). Znacznie łatwiej określić tak zwany efekt ankietarski, który jest mniejszy lub co najwyżej równy wpływowi ankietarskiemu i jest różnicą między rezultatami oddziaływania różnych ankietarów, przy spełnieniu warunku losowego przydziału wylosowanych osób do ankietarów (Lutyńska 1978, 1993: 98–99; Sztabiński P. 1997). W tym przypadku porównuje się rozkłady odpowiedzi uzyskane przez różnych ankietarów i stwierdza, czy są między nimi jakieś różnice²³.

Kolejne grupy czynników wpływających na wartość uzyskiwanych informacji to mikrospołeczne i makrospołeczne warunki wywiadu i badania (Lutyńska 1984: 44–54, 1985, 1993: 5–6). Pierwsze z nich (warunki mikrospołeczne) kształtowane są przez psychospołeczne czynniki sytuacyjne, w których odbywa się wywiad. Są one specyficzne dla każdego wywiadu. Można wśród nich wyróżnić czynniki związane z warunkami zewnętrznymi i wewnętrznymi wywiadu. Warunki zewnętrzne wywiadu to przede wszystkim:

– miejsce prowadzenia wywiadu (mieszkanie, zakład pracy, ulica itp.). Czynnikiem ten może mieć bezpośredni wpływ na charakter i treść odpowiedzi respondenta, ponieważ miejsce wywiadu w znacznej mierze określa rolę społeczną, w której występuje respondent: jeśli wywiad prowadzony jest przykładowo w miejscu pracy, to najprawdopodobniej przyjmie on rolę zawodową. Podobnie, jeśli wywiad realizowany jest w domu, to będzie on występował w roli głowy gospodarstwa/gospodyni domowej lub też ojca/matki (Lutyńska 1993: 58–76).

Czynnikiem drugim jest obecność i aktywność osób trzecich – on również może mieć bezpośredni wpływ na odpowiedzi respondenta przy czym, istotne jest: kim są te osoby (osoba obca vs członek gospodarstwa domowego: współmałżonek, rodzice, dzieci itp.), oraz czego dotyczy pytanie (chodzi o przedmiot pytania). Obecność osób obcych

²³ Innym możliwym sposobem takiej oceny jest porównanie wyników otrzymanych przez najlepszych i najgorszych ankietarów. Różnica wyników wskazuje przybliżoną wartość wpływu w przypadku najsłabszej grupy ankietarów. Brak różnicy świadczy o tym, że różnice między ankietarami nie wywarły istotnego wpływu (zob. Borucki 1978; Zurn 1976).

w trakcie wywiadu może potęgować u respondenta poczucie zagrożenia, skrępowania i naruszenia prywatności. Z kolei obecność osób z rodziny może osłabiać to zagrożenie lub/i wprowadzać element „samokontroli” respondenta, ale także „korekty” w wypowiedzi respondenta (Lutyńska 1970: 215, 1983, 1993: 38–57)²⁴.

Czynnik trzeci to pora roku, miesiąc, pora i rodzaj dnia, kiedy przeprowadzany jest wywiad. Te elementy warunków mikrospołecznych mogą mieć wpływ nie tylko na charakter udzielanych przez respondenta odpowiedzi, ale także na sam fakt wyrażenia zgody na wywiad. W niektórych przypadkach „ekstremalne” warunki atmosferyczne wywiadu (na przykład zima, niska temperatura) mogą pozytywnie wpływać na przebieg interakcji między ankierem i respondentem (zob. Sztabiński, Sztabiński, Przybysz 2007).

Jak wskazano wyżej, wpływ zewnętrznych czynników sytuacyjnych może mieć zarówno pozytywny, jak i negatywny wpływ na jakość uzyskiwanych danych, a więc ich kompletność i wiarygodność. Jednak, ze względu na brak możliwości wskazania ogólnego kierunku ich oddziaływania, zawsze należy dążyć do ich standaryzacji. Możliwe jest to poprzez przestrzeganie zaleceń zawartych w instrukcji dla ankiera, a ich wpływ można kontrolować dzięki informacjom zawartym w ankiecie dla ankiera i późniejsze analizy sprawdzające (Lutyńska 1968)²⁵.

²⁴ Przykładem „korekty” deklaracji respondenta pod wpływem obecności osób trzecich może być przypadek, w którym w trakcie realizacji wywiadu respondent stwierdził, że „zasadniczo nie pali papierosów”. Obecny w trakcie wywiadu wnuczek respondenta w tym momencie wtrącił się i powiedział: „A dziadziuś to tylko tak mówi. Wiele razy obiecywał, że rzuci palenie, ale dalej pali”. Po tej wypowiedzi respondent stwierdził: „No tak, chcę rzucić palenie, ale to nie tak łatwo mi idzie i jeszcze palę”. Innym przykładem pozytywnego wpływu obecności osób trzecich na wartość uzyskiwanych informacji mogą być wywiady retrospektywne, dotyczące faktów i zdarzeń przeszłych. W tym przypadku osoby trzecie mogą „wspomagać” pamięć respondenta.

²⁵ Z pewnością należy się tutaj zgodzić z autorką, iż czynniki zewnętrzne mają wpływ na wartość uzyskiwanych informacji. Problematiczna jednak wydaje się możliwość kontroli tego wpływu na podstawie opisu sytuacji wywiadu pochodnego od ankiera, a więc danych z tak zwanej ankiety do ankiera. Na podstawie badań własnych (1996) uważam, że jest to niemożliwe ze względu na bardzo niską wiarygodność informacji przekazywanych przez ankierów w ankiecie po wywiadzie. Oznacza to, iż wpływ czynników zewnętrznych istnieje, jednak praktycznie nie sposób określić jego rozmiaru.

Na wewnętrzne warunki sytuacyjne składają się natomiast psychospołeczne relacje między respondentem i ankieterem. Chodzi tutaj o charakter aranżacji wstępnej, atmosferę wywiadu, rolę, w której występuje ankieter, styl prowadzenia wywiadu itd. Zakłada się, że czynniki te mogą mieć wpływ na uzyskanie odpowiedzi istotnych i trafnych (Lutyńska 1984: 44–54)²⁶. Należy podkreślić, iż tego typu warunki sytuacyjne są kształtowane przede wszystkim przez ankietera, ale oddziałują nie tylko na respondenta, lecz także na niego samego.

Z kolei makrospołeczne warunki wywiadu i badania kształtują aktualne wydarzenia związane z sytuacją społeczno-polityczną, historyczną, kulturową, ekonomiczną kraju, społeczeństwa czy regionu. Jest to ogólny kontekst nie tylko całego badania, ale także realizowanych wywiadów²⁷. Oznacza to, iż warunki ogólnospołeczne wpływają nie tylko na badacza, ale także na ankietera i respondenta. Badacz funkcjonuje bowiem w określonych strukturach instytucjonalnych, poprzez które owe czynniki makrospołeczne determinują nie tylko wybór problematyki i tematu badań, często procedur badawczych, terminu badań itp., ale również jego własne preferencje (Grawitz 1972: 140–146; Lutyńska 1984: 33–38). W przypadku ankietera ich oddziaływanie jest nieco słabsze, aczkolwiek nie pozostają one bez wpływu na jego zaangażowanie i motywację oraz poglądy, które z kolei są istotnym źródłem wpływu ankieterskiego, o którym wspominałem powyżej. W przypadku respondentów warunki makrospołeczne mogą z kolei wpływać na postrzeganie i stosunek do badań i tym samym na zgodę na uczestnictwo w nich, a pośrednio na wartość udzielanych odpowiedzi – ich istotność i trafność (Lutyńska 1993: 131–155, 171–183; Sztabiński F. 2004; Sztabiński i Sztabiński 1994; Sztabiński P. 1995, 2006)²⁸.

²⁶ Na gruncie polskim rozważania dotyczące koncepcji ankietera i jego roli, i co się z tym wiąże, stylem prowadzenia wywiadu podejmowali między innymi Kistelki (1985:16), Krak (1978) i Sitek (1976).

²⁷ O znaczeniu wpływu czynników ogólnych na wartość uzyskiwanych informacji oraz ich wadze przy interpretacji wyników świadczy także to, że w ramach międzynarodowego projektu Europejski Sondaż Społeczny poszczególne kraje mają obowiązek zbierać tak zwane dane kontekstowe, opisujące najważniejsze wydarzenie o charakterze ogólnospołecznym w okresie poprzedzającym realizację badania oraz w trakcie jego realizacji (zob. Stoop 2008; Sztabiński P. 2004).

²⁸ Praktycznie nie jest jednak możliwe do przewidzenia, w jakim kierunku i w jakich sytuacjach czynniki makrospołeczne będą oddziaływały na różne katego-

Oddziaływanie czynników sytuacyjnych i makrospołecznych jest wielokierunkowe, wielostopniowe i nieraz bardzo skomplikowane. Nie występują one bowiem w izolacji, lecz zawsze w powiązaniu, wywołując różnorodne zjawiska i procesy psychiczne u respondenta i ankietera. Respondent i ankieter nie są jednak biernymi przedmiotami tych oddziaływań, ich „efektywność” uzależniona jest bowiem od wielu cech ich osobowości, dyspozycji psychicznej, stanu psychofizycznego, podatności na wpływy zewnętrzne, poziomu pobudliwości emocjonalnej, postaw i systemów wartości, cech społeczno-demograficznych itp. Wszystko to powoduje, iż wpływ wspomnianych czynników jest praktycznie niemożliwy do przewidzenia przez badacza, a co więcej, zazwyczaj mu nieznanym.

Scharakteryzowane powyżej statyczne ujęcie koncentruje się przede wszystkim na poszczególnych elementach procesu badawczego, które mają wpływ na wartość zbieranych danych. Jak twierdzi Krystyna Lutyńska wartość uzyskiwanych odpowiedzi zależy od: badacza (i narzędzia badawczego), ankietera, respondenta oraz warunków sytuacyjnych i ogólnych, w których oni funkcjonują. Autorka, analizując wpływ wymienionych czynników, skupia się przede wszystkim na narzędziu badawczym, wpływie ankietarskim oraz mikro- i makrospołecznych warunkach wywiadu. W swoich rozważaniach pomija wpływ respondenta na wynik badania, aczkolwiek jest on (respondent) uwzględniony jako „element” schematu. Osoba respondenta „obecna” jest w jej rozważaniach, ale tylko o tyle, o ile stanowi on „przedmiot” oddziaływania ankietera oraz warunków badania. Tymczasem jego roli w procesie uzyskiwania informacji nie sposób przecenić. Co więcej, to właśnie od niego zależy między innymi ich wartość (Auriat 1993, 1996; Rosenthal i Rosnow 1969).

Bez wątpienia jednak propozycja Lutyńskiej stwarza pewne możliwości oceny „ex ante”, ale tylko i wyłącznie narzędzia badawczego jako

rie respondentów. Można jedynie przypuszczać, iż pewne istotne wydarzenia społeczno-polityczne mogły mieć generalnie pozytywny wpływ (na przykład „odwilż” 1956 roku, wybór Karola Wojtyły na papieża w 1978 roku, powstanie „Solidarności” w 1980), inne – negatywny (na przykład wprowadzenie stanu wojennego czy Ustawy o ochronie danych osobowych). Jednak zawsze, niezależnie od warunków ogólnych, napotykamy problemy związane z „opiniami oficjalnymi i prywatnymi”, „odpowiedziami społecznie pożądanymi” czy z sytuacją „kłamstwa obronnego” (Gostkowski 1961, 1966; Lutyńska 1993; Rostocki 1992; Słomczyński 1983; Staszewska 1985; Sztabiński F. 2004).

całości (kwestionariusza, materiałów pomocniczych, instrukcji) i tym samym możliwości podniesienia jakości zbieranych danych. Należy jednak pamiętać, iż badanie pilotażowe ma ograniczony zakres ze względu na niewielką, w specjalny sposób dobraną próbę (realizuje się je na próbie nie więcej niż pięćdziesiąt osób dobranych w sposób celowy). Tym samym nie mamy pewności, czy w trakcie badania ujawniły się wszystkie problemy związane z zastosowanym narzędziem, nie wiemy też, jaki jest zakres występowania ujawnionych problemów.

1. 1. 3. Błędy pomiaru

Drugie podejście w ramach perspektywy wewnętrznej nie skupia się na przebiegu procesu badawczego, jego elementach czy fazach, lecz na badaniu jako pomiarze. Podstawowym elementem pomiaru jest standaryzacja²⁹. Dzięki niej, a także losowości przy doborze próby, możliwa jest kwantyfikacja badanych zjawisk, ich ściśle oszacowania liczbowe, a także uogólnianie wniosków na całą badaną populację. Należy jednak pamiętać, iż badanie ilościowe jest „specyficznym” pomiarem, opartym na raportowaniu ankietowemu przez respondenta faktów, opinii i poglądów. W związku z tym pomiar ten może być obciążony błędem – można nawet powiedzieć, że niemożliwy jest pomiar w pełni „obiektywny”, to znaczy wolny od zanieczyszczeń³⁰. Praktycznie wynik każdego badania (pomiaru) obciążony jest błędem. Stąd też Norman Bradburn i Seymour Sudman piszą, iż „użytkownik danych surveyowych powinien rozumieć źródła błędów i ograniczenia, jakie te błędy mogą powodować w interpretacji danych” (1988: 179).

W literaturze wskazuje się na dwie koncepcje błędów pomiaru związane z podstawowym przedmiotem zainteresowania statystyki oraz psycholo-

²⁹ O standaryzacji można mówić, mając na myśli sposób zbierania danych (technikę badawczą), narzędzie badawcze (kwestionariusz, materiały towarzyszące itd.), sposób doboru próby (ujednoczenie sposobu doboru jednostek), zachowania ankieterskie (zasady prowadzenia wywiadu, wyjaśnienia), warunki zbierania informacji (miejsce wywiadu: mieszkanie badanego, miejsce pracy, miejsce publiczne, ulica; obecność osób trzecich itd.), ale także sposób opracowania danych (w tym przede wszystkim sposób kodowania). Elementy te stanowią różne wymiary standaryzacji (zob. Sztabiński P. 2005; Maison 2001: 15 i n.)

³⁰ Należy podkreślić, iż w tym ujęciu „błąd” rozumiany jest inaczej niż poprzednio (por. przypis 18).

gii (Groves 1989; 1991; Sztabiński P. 1997; Sawiński 1988: 49–51)³¹. W ujęciu statystycznym przyjmuje się założenie o istnieniu wartości prawdziwej. Tym samym przez błąd pomiaru rozumie się dowolne odchylenie (różnicę) wyników badania od wartości „prawdziwej”, która jest przedmiotem badania. Jego źródłem jest przede wszystkim sposób doboru i realizacji próby. Z kolei w ujęciu psychologicznym przyjmuje się, że wartość prawdziwa nie istnieje (istnieje co najwyżej wartość „ukryta”), a błąd jest „zanieczyszczeniem” wyniku pomiaru, spowodowanym wpływem jakiegoś czynnika. Przykładowo, jeśli stosując dwa różne pytania na ten sam temat (poprawnie sformułowane), uzyskuje się różne wyniki, to można podejrzewać, że jedno z nich „lepiej” mierzy badane zjawisko. Dlatego często mówi się nie o „błędzie”, lecz „efekcie” sformułowania pytania.

Moim celem nie jest szczegółowa analiza różnych koncepcji błędów, trzeba wszakże wspomnieć, iż niezależnie od ujęcia wyróżnia się dwa ich rodzaje: błąd zmienny (losowy) oraz stały (systematyczny). Pierwszy z nich nie rzutuje na wielkości parametrów, takich jak średnia, czy mierniki zależności między zmiennymi. Średnia tego typu błędów jest równa zeru. Drugi rodzaj błędów określa się też mianem wypaczenia. Jest on efektem jednokierunkowego wpływu jakiegoś czynnika na cechę mierzoną lub operację pomiaru. Wpływa on na wielkości parametrów, a średnia tego typu błędów nie jest równa zeru³².

³¹ Paweł Sztabiński, za Robertem Grovesem, przyjmuje statystyczne i psychologiczne rozumienie błędu. Z kolei Zbigniew Sawiński mówi o dualizmie teorii testów i wymienia ujęcie tradycyjne i współczesne. W ujęciu klasycznym zakłada się, że każdy z badanych obiektów posiada „wartość prawdziwą” danej cechy. W ujęciu współczesnym wartości prawdziwe mierzonej cechy nie istnieją poza pomiarem. Stąd też nieuzasadnione jest pytanie o zgodność wyników pomiaru z wartościami prawdziwymi.

³² Zarówno pojęcie błędu zmiennego (Moser i Kalton 1971), jak i wypaczenia związane są w tym ujęciu z traktowaniem każdego konkretnego badania jako jednego z nieskończonej ilości możliwych jego powtórzeń, w których stosuje się dokładnie to samo narzędzie w odniesieniu do tej samej populacji. Konceptualnie zakłada się, że są one powtarzane w tym samym czasie i że odpowiedź udzielona w jednym z nich nie wpływa na uzyskaną w następnych. Respondent w kolejnych badaniach może udzielać różnych odpowiedzi, choć mogą być one również identyczne. Ich wariancja stanowi indywidualny błąd zmienny danego respondenta. Jeśli dla jego odpowiedzi obliczymy średnią, to może ona, ale nie musi, być zgodna z prawdziwą wartością. Gdy nie jest zgodna, mówimy o wypaczeniu.

Przedmiotem zainteresowania socjologii ilościowej są zarówno charakterystyki faktualne społeczeństwa, jak i jego opinie: przedstawia ona wyniki w postaci uzyskanych rozkładów i zależności. Jej ambicją jest także unikanie błędów związanych zarówno z próbą, jak i z procedurą pomiaru. Stąd też w niektórych podręcznikach metod badań surveyowych dokonywana jest swoista kompilacja konceptualna: o błędach związanych z próbą mówi się, używając języka statystyki, błędy pomiaru zaś analizuje się językiem zbliżonym do psychologicznego (Bradburn i Sudman 1988). Przykładem może być podejście Kahna i Cannella (1957), którzy nawiązując do ujęcia psychologicznego, używają wspomnianego określenia „wartość prawdziwa”. Uznają oni, że w zależności od tego, jak zdefiniuje się cechę będącą przedmiotem badania, to samo narzędzie spowoduje mniejszy lub większy błąd pomiaru, a tym samym jego wynik będzie mniej lub bardziej odbiegał od „wartości prawdziwej”. Na przykład wykształcenie można zdefiniować jako szkołę najwyższego szczebla, którą respondent ukończył, lub też jako liczbę lat skolaryzacji. W przypadku zarobków należy rozstrzygnąć, czy uwzględniać czy też nie, premie, nagrody, gratyfikacje niepieniężne, zarobki dodatkowe itd. Poważne problemy zaczynają się jednak, gdy cechą mierzoną są złożone, nieobserwowalne konstrukty, na przykład postawy czy stereotypy. Kahn i Cannell piszą, iż „w pewnych sytuacjach prawdziwa wartość atrybutu może być abstrakcją, którą można jedynie określić w przybliżeniu, nie zaś dokładnie ustalić”. I dalej: „koncepcja prawdziwej wartości, przynajmniej jeśli odnosi się do postaw, musi być traktowana jako relatywna. Zatem z definicji koncepcja błędu także staje się relatywna. Błąd może być oszacowany tylko w kategoriach związku między aktualnie otrzymanym pomiarem i wartością, która reprezentuje nasz cel. (...) To, czy nasza technika lub sformułowanie powoduje wypaczenie, czy też nie, zależy od celu, jaki chcemy osiągnąć, i wypaczenie to może być oceniane tylko w relacji do tych celów” (1957: 172 i 175). W rozumieniu Kahna i Cannella w przypadku postaw prawdziwa wartość jest pewną intuicją badacza, która nie przyjmuje konkretnych wartości liczbowych i stanowi pochodną przyjętej definicji cechy³³. W przypadku badań ilościowych

³³ W przypadku tych cech o szacowaniu wielkości błędu pomiaru w sensie statystycznym (a więc w stosunku do zewnętrznej wartości prawdziwej) możemy mówić tylko w wyjątkowych przypadkach. Oprócz sytuacji eksperymentalnych występują one na przykład w badaniach przedwyborczych. Jeśli na podstawie badań prowa-

można więc mówić o pewnej „dwoistości” ujęć. Otóż w każdym badaniu zadaje się respondentom pytania o pewne ich cechy, tak zwane metryczkowe (płeć, wiek, wykształcenie itd. – cechy pierwotne), oraz pytania dotyczące ich opinii, poglądów, przekonań (tzw. cechy wtórne), które zgodnie z pozytywistycznymi założeniami socjologii są zależne od tych pierwszych lub nimi uwarunkowane. Ponieważ w przypadku cech społeczno-demograficznych istnieje „wartość prawdziwa”, należy w odniesieniu do nich mówić o statystycznej koncepcji błędu. Z kolei w przypadku pytań o cechy wtórne (opinie, przekonania) – o psychologicznym ujęciu błędu. Nie można więc powiedzieć, iż określając błąd pomiaru w danym badaniu, przyjmujemy psychologiczne lub statystyczne jego rozumienie. Należy zatem mówić raczej o „psychologicznym” lub „statystycznym” podejściu do błędów pomiaru, w zależności od rodzaju pytań: ich przedmiotu czy poszukiwanej informacji.

Jak wspominałem, moim celem nie jest szczegółowa analiza różnych koncepcji błędów, lecz wskazanie możliwości wykorzystania tych ujęć do oceny danych uzyskiwanych w wywiadzie kwestionariuszowym. Należy obecnie wskazać źródła potencjalnych błędów pomiaru. Robert Groves (1989, 1991), za Leslie Kishem (1965) przyjmuje, że obydwie typy błędów (a więc błąd zmienny i wypaczenie) mogą mieć swoje źródło w:

- błędach niezwiązanych z obserwacją, które są wynikiem tego, że pomiary pochodzą jedynie od części badanej populacji (jest to błąd związany z próbą);
- błędach związanych z obserwacją oraz
- błędach opracowania (*processing errors*), które nie są bezpośrednio związane z procesem pomiaru³⁴.

dzonych na tydzień przed elekcją chcemy wypowiadać się o tym, jakie będą rzeczywiste wyniki wyborów, a nie jakie są aktualne preferencje, wówczas istnieje wartość prawdziwa (wynik wyborów) i można w relacji do niej określić wielkość błędu. Podobna sytuacja występuje w badaniach marketingowych, w których pytamy o zamiar kupna danego produktu. Uzyskane od respondentów odpowiedzi można odnieść do rzeczywistej wielkości sprzedaży.

³⁴ Jest to oczywiście jedno z wielu możliwych ujęć źródeł błędu. Niekiedy przyjmuje się, że można wskazać pięć następujących rodzajów błędów: (i) błąd związany z reprezentacyjnym charakterem badania, a więc faktem, że badanie realizowane jest na próbie, nie zaś całej populacji (*sampling error*); (ii) błąd związany z niepełną realizacją próby oraz brakiem odpowiedzi (*nonresponse error*); (iii) błąd pokrycia

Stąd też, obecnie mówi się o całkowitym błędzie badania (*total survey error*, TSE) na oznaczenie czterech podstawowych błędów: losowania, pokrycia, pomiaru i braku odpowiedzi (de Leeuw, Hox, Dillman 2008: 6–13; Groves i in., 2004: 39–63; Weisberg 2005).

1. 1. 3. 1. Błędy niezwiązane z obserwacją

Wśród błędów niezwiązanych z obserwacją wyróżnia się następujące.

A. Błąd losowania próby (błąd statystyczny lub standardowy błąd estymacji). Jego wielkość zależy od stosunku liczebności próby do populacji (jest ona odwrotnie proporcjonalna do pierwiastka z liczebności próby, a dokładniej, pierwiastka z ilorazu odchylenia standardowego zmiennej i wielkości próby).

Oznacza to, iż przykładowo, aby zmniejszyć błąd dwukrotnie, trzeba liczebność próby zwiększyć czterokrotnie. Tym samym, im próba jest większa, tym wynik bardziej dokładny, obarczony mniejszym błędem. Nie jest to jednak zależność prostoliniowa, powyżej pewnej granicy zwiększanie liczebności próby powoduje stosunkowo niewielki wzrost dokładności pomiaru.

B. Błąd „pokrycia”. Błąd ten związany jest z niedoskonałością operatu losowania, to znaczy jego niekompletnością lub błędami operatu. Źródłem błędu pokrycia są różnice między operatem doboru próby a populacją. Wynikają one z:

- braku niektórych jednostek w operacie (na przykład osób bezdomnych brak w imiennych operatach doboru osób);

- nieprawidłowego ich zaklasyfikowania (na przykład w operacie firm danej branży część firm może mieć nieprawidłowo określoną branżę;) oraz

- duplikacji (na przykład w operacie abonentów telefonów stacjonarnych to samo gospodarstwo może być uwzględnione dwukrotnie). Podobnie, może się zdarzyć, że do próby wejdą przypadki, które nie należą

(*coverage error*) związany z tym, że w próbie „z góry” brakuje pewnych kategorii badanych lub występują kilkakrotnie; (iv) błąd pomiaru (*measurement error*), mający swoje źródła w kwestionariuszu (*questionnaire effects*), technice i sposobie zbierania danych (*data collection mode effects*), wpływie ankietera (*interviewer effects*) oraz osobie respondenta (*respondent effects*); (v) błąd przetwarzania danych (*processing error*) [Kasprzyk 2001]. Wszystkie rodzaje wspomnianych błędów odnaleźć można w propozycji Grovesa (2004) [por. także Sztabiński P. 1997 i in.].

do badanej populacji (na przykład operat adresowy uwzględnia cudzoziemców przebywających czasowo w Polsce). Oznacza to, że w operacji pewne kategorie osób nie są uwzględniane (brakuje ich zupełnie i w związku z tym nie mają szansy zostać wylosowane bądź są niedoreprezentowane) lub też są nadreprezentowane (podstawowym i najczęściej występującym błędem pokrycia populacji jest jednak nieuwzględnienie części populacji w badaniu). W efekcie ich udział w próbie jest systematycznie niższy niż w populacji bądź systematycznie wyższy (*Sampling for the European Social Survey, Round 3: Principles and requirements*, 2006). W każdym przypadku próba będzie zniekształcona, a tym samym uzyskane wyniki badania będą wypaczone, gdyż zostaną nimi objęte (lub nie zostaną) pewne kategorie osób (na przykład bezdomni, gospodarstwa nieposiadające telefonu stacjonarnego), a więc ludzie o określonym statusie. Wydaje się, iż błędy pokrycia populacji są najrzadziej uwzględnianym rodzajem błędów zawartych w wynikach badania.

C. Błąd „braku odpowiedzi”. O błędzie tym mówi się wtedy, gdy jednostka należy do badanej populacji, została dobrana w skład próby i z różnych przyczyn nie zrealizowano z nią wywiadu. Błąd tego rodzaju może dotyczyć zarówno całego wywiadu (chodzi o niedostępność części respondentów w badaniu ze względu na ich odmowę udziału w badaniu lub niemożność zastania itp.), jak i jego fragmentów, a więc poszczególnych pytań (*unit i item nonresponse error*) [zob. Grzeszkiewicz-Radulska 2009]. Podstawową miarą wielkości błędu tego rodzaju jest tak zwany współczynnik realizowalności (*response rate*), który określa, z jaką częścią wylosowanych osób udało się zrealizować badanie (jest to proporcja liczby zbadanych osób do liczby osób wylosowanych, należących do badanej populacji)³⁵. Można zaobserwować ogólnie stały spadek odsetka realizacji próby. Przykładowo, w ciągu ostatnich szesnastu lat odsetek realizacji próby adresowej (w której jednostką losowania jest gospodarstwo domowe) w projekcie Polski Generalny Sondaż Społeczny (PGSS) zmniejszył się o 30 punktów procentowych: z 82% w 1992 do 52% w 2008. Oznacza to, że w 2008 roku nie zrealizowano 48% wy-

³⁵ O różnych sposobach liczenia współczynnika realizowalności piszę w odniesieniu do badań realizowanych techniką ankiety pocztowej (Danilowicz, Sztabiński 1992). W Europejskim Sondażu Społecznym przyjęte są dwie miary odsetka realizacji: tak zwany prawdziwy i terenowy (*ESS Round 5., Specification for participating countries*, 2009, zob. także: Lavrakas 2008: 758–760).

losowanej próby (Cichomski, Jerzyński i Zieliński 2008). Tendencję tę potwierdzają również dane z badań CBOS: w roku 2000 średni odsetek realizacji na próbach adresowych wynosił 67,7, w 2001 – 65,4, w 2002 – 62,2, a w 2003 – 61,9. Te same dane pokazują, iż analogiczna tendencja występuje w przypadku prób imiennych (losowanych z systemu PESEL, gdzie jednostką losowania jest konkretna osoba): w roku 2003 średni odsetek realizacji wynosił 62,3, w 2004 – 58,8, a w 2005 – 57,0 (Dąbała 2005)³⁶. Tendencję taką można zaobserwować nie tylko w Polsce: w Holandii odsetek realizacji w badaniach czytelnictwa spadł poniżej 50 w 1998 r., a w 2000 roku nie osiągnął 40 (Bassler, Tchaoussoglou 2001; Stoop 2005). Podkreślę jednak, że istnieją pewne możliwości minimalizacji błędu „braku odpowiedzi”, a więc podniesienia odsetka realizacji próby (*Field Procedures in the European Social Survey: Enhancing Response Rates*, 2006; Sztabiński F. 2004). Minimalizacja błędu „braku odpowiedzi” jest bardzo ważna, ponieważ wysoki odsetek realizacji próby zapewnia reprezentatywność badania³⁷. Błąd ten nie ma charakteru losowego, lecz jest błędem systematycznym. Jednostki niedostępne różnią się w sposób istotny od pozostałych, co powoduje ryzyko tendencyjności próby ze względu na cechy społeczno-demograficzne, psychologiczne oraz opinie, a w rezultacie – wypaczenie wniosków z badania. Na przykład w badaniach ogólnopolskich kategoriami szczególnie trudno dostępnymi są obecnie mieszkańcy wielkich miast, ludzie młodzi – do 24 lat oraz mężczyźni (Dąbała 2005)³⁸. Oczywiście tego typu wypaczenia

³⁶ Paweł Sztabiński (1997: 29) pisze, iż w połowie lat 90. odsetek realizacji wynosił około 80 dla prób adresowych. W przypadku prób imiennych był niższy i wynosił, jeśli uwzględnić tylko próbę zasadniczą (bez rezerwowej), około 70–75.

³⁷ Współczynnik realizowalności jest, jak wspominałem, uznawany za jedną z najważniejszych miar jakości realizacji badania (de Leeuw, van der Zouwen 1991: 10–11, 17–18; Frey i Oishi 1995: 30–32; Lejeune i Bied-Charreton 1991: 70–77; Stoop 2005; Sztabiński F. 1997: 56–64). Oznacza to, że błędy niezrealizowania części próby mają najpoważniejsze konsekwencje dla wyników badania.

³⁸ Tendencje wskazane przez Wiesławę Dąbałę w odniesieniu do prób imiennych potwierdzają dane zebrane w kolejnych rundach Europejskiego Sondażu Społecznego (www.europeansocialsurvey.org/data/CF). Należy jednak podkreślić, iż odsetek jednostek niedostępnych można ograniczyć między innymi poprzez wielokrotnie podejmowane próby kontaktu z wylosowaną osobą. W Europejskim Sondażu Społecznym Runda 1, przy ogólnym odsetku realizacji 70,9, jedynie 39,7% wywiadów udało się zrealizować podczas pierwszej wizyty ankietera, a pozostałe 31,2% podczas drugiej i kolejnych wizyt. Musi tutaj jednak zostać spełniony wa-

mogą być „poprawiane” poprzez różne procedury ważenia. Lista cech, w przypadku których znamy prawdziwy ich rozkład w populacji (na podstawie danych ze spisu powszechnego), obejmuje jednak tylko kilka charakterystyk społeczno-demograficznych. Nie obejmuje natomiast całego szeregu charakterystyk psychologicznych, a tym samym ważenie wyniku rozwiązuje problem wypaczenia jedynie częściowo. Rzeczą oczywistą jest bowiem, iż osoby wylosowane, które nie wzięły udziału w badaniu (tzw. nierespondenci), różnią się od respondentów także cechami niedemograficznymi. Wyniki badań w krajach zachodnich wskazują bowiem, że w badaniach chętniej biorą udział osoby o wyższym statusie zawodowym, o większej potrzebie akceptacji, wyższym poziomie inteligencji oraz mniej autorytarne (Rosenthal, Rosnow 1969). Z kolei z badań prowadzonych w Polsce wynika, że nierespondenci, w porównaniu z respondentami, są relatywnie bardziej nieufni wobec ludzi, bardziej niezadowoleni z demokracji oraz gorzej oceniają swoje samopoczucie, ale jednocześnie są relatywnie bardziej zainteresowani polityką i mają mniej tradycyjne poglądy dotyczące roli kobiety w rodzinie. Tym samym nierespondenci stanowią specyficzną kategorię, różniącą się od respondentów (Sztabiński, Sztabiński, Przybysz 2007a). Oznacza to, że błąd „braku odpowiedzi” może być błędem systematycznym i powodować wypaczenie wyniku badania. W przypadku niektórych opinii dokonywanie uogólnień na populację na podstawie przebadanej próby może nie być więc uprawnione. Jednocześnie, ponieważ niektóre zależności między cechami społeczno-demograficznymi a opiniami są różne wśród nierespondentów i respondentów, ważenie próby nie zawsze poprawia estymacje³⁹. Błąd „braku odpowiedzi” może wystąpić także, o czym wspominałem, na poziomie pytania. Chodzi o odmowę udzielenia odpowiedzi na poszczególne pytania (*item nonresponse error*), na przykład pytanie o dochody, posiadane dobra materialne. Tego typu błędy nie mają jednak charakteru systematycznego w skali całego badania, a co

runek wydłużonego okresu realizacji badania (Sztabiński, Sztabiński i Przybysz 2007b, 2009).

³⁹ Problemowi doskonalenie reprezentatywności i estymacji błędu będącego konsekwencją niedostępności części jednostek poświęcony był jeden z projektów Połączonych Działań Badawczych 2 (Joint Research Activities – JRA2), realizowanych w ramach Europejskiego Sondażu Społecznego (ESS). Prowadził go Jaak Billiet (Katholieke Universiteit Leuven, Belgia). Zob. *Improving Representativeness and Response*, 2009.

najwyżej są wypaczeniem na poziomie ankietera (jest to zatem raczej efekt ankieterki niż błąd „braku odpowiedzi”)⁴⁰.

1. 1. 3. 2. Błędy związane z obserwacją

Z kolei wśród błędów związanych z obserwacją wyróżnia się następujące.

A. Błąd związany ze sposobem uzyskiwania danych (technika). Błąd ten wynika stąd, że każde badanie stwarza sztuczną sytuację, umożliwiającą jego prowadzenie. Dodatkowo, każda technika badawcza ma właściwe sobie ograniczenia, wprowadza specyficzne zniekształcenia (może to być na przykład sama fizyczna obecność ankietera, forma komunikowania się). Wszystko to wpływa na uzyskany wynik. W rezultacie dane uzyskiwane za pomocą jednej techniki mogą być nieporównywalne z danymi uzyskiwanymi za pomocą innej⁴¹. Przykładowo, różnice rozkładów odpowiedzi na pytanie o „spontaniczną” znajomość nazw firm sektora „X” (pytanie otwarte) uzyskane w CATI i wywiadzie kwestionariuszowym PAPI przekraczają 9 punktów procentowych (na „niekorzyść” CATI). W tym samym badaniu różnice rozkładów odpowiedzi na pytanie o wspomaganą znajomość nazw firm (pytanie zamknięte)

⁴⁰ Problemowi braku odpowiedzi poświęcona jest książka Ineke Stoop (2005), *The Hunt for the Last Respondent*. Problem ten podejmują również autorzy *Improving Survey Response. Lesson learned from the European Social Survey* (Stoop, Billet, Koch, Fitzgerald 2010). Zob. także Stoop (2007) Sztabiński, Sztabiński, Przybysz (2007, 2009).

⁴¹ W literaturze metodologicznej wpływ techniki na uzyskiwane rezultaty określa się mianem „mode effects”. Jak wspomniałem, jest on źródłem błędu. Niemniej obecnie próbuje się określić, na ile zastosowanie różnych technik zbierania informacji (*mixed mode data collection*) w ramach tego samego badania może być przydatne w minimalizowaniu błędu braku odpowiedzi na poziomie wywiadu (*non-response*). Wymaga to oczywiście przygotowania wersji pytań, które w niewielkim stopniu byłyby podatne na efekt techniki, a następnie dokonanie oceny skutków zastosowania *mixed mode*, to znaczy wypracowania technik integrowania danych i korygowania błędów pomiaru związanego z efektem techniki (Roberts 2007; Roberts, Jakle, Lynn 2007; Sztabiński, Sztabiński 2007). Próby te podejmowane są w ramach projektu Połączonych Działań Badawczych 1 (Joint Research Activities, JRA1), prowadzonego przez Caroline Roberts i Gillian Eva z City University (Londyn), pod kierunkiem Rogera Jowella (City University), we współpracy z Peterem Lynnem (University of Essex) i Willemem Sarisem (ESADE, Barcelona). Zob. *Modes of Data Collection 2006*.

sięgają 30 punktów procentowych (na „korzyść” CATI). Wynika to między innymi z faktu, że w wywiadzie CATI respondenci mają skłonność do udzielania „szybkich” odpowiedzi, „zgadzania się” („potakiwania”), przerywania odczytywanych kafeterii, a w przypadku pytań otwartych – udzielania odpowiedzi skrótowych, schematycznych i niepogłębionych (Sztabiński P. 1999). Z kolei immanentną cechą techniki ankiety pocztowej jest tak zwana konsultacja wewnątrzdomowa (Moser i Kalton 1971; Sztabiński F. 1997), w efekcie czego uzyskujemy opinie „zakotwiczone” społecznie, „uzgodnione”, nie zaś indywidualne. Ze względu na to, że respondenci mają możliwość zastanowienia się, przemyślenia problemów będących przedmiotem pytania, ich odpowiedzi są mniej spontaniczne (bardziej wyważone) niż w wywiadzie. W rezultacie w ankiecie i wywiadzie nie występują różnice w odpowiedziach na pytania o fakty; natomiast w przypadku pytań o opinie w ankiecie uzyskujemy odpowiedzi mniej skrajne (różnice dochodzą do 18 punktów procentowych). Dodatkowo w ankiecie pocztowej ze względu na brak pośrednictwa ankietera zapis odpowiedzi należy traktować jako pełny, a zebrany materiał badawczy jako tożsamy ze źródłem informacji. W wywiadzie natomiast ankieter jest do pewnego stopnia „współautorem” uzyskanych materiałów i stąd nigdy nie jest on tożsamy ze źródłem informacji, a więc wypowiedzią respondenta (Sztabiński F. 1997)⁴². Ponieważ każda technika stwarza odmienną sytuację badawczą i w rezultacie przynosi inne dane, ważnym problemem jest dobór odpowiedniej techniki ze względu na badany problem (co badać?) i badaną zbiorowość (kogo badać?). Jest to zagadnienie adekwatności metodologicznej, psychologicznej i społecznej techniki badawczej (Eva, Widdop 2007; Gostkowski 1965, 1966; Roberts 2007; Słomczyński 1968; Sztabiński F. 2003).

B. Błąd związany z pytaniem (określany także błędem narzędzia). Pytanie może być źródłem błędu w trojakim sensie: błąd może być pochodną treści pytania (chodzi o trafność doboru wskaźników), sposobu sformułowania pytania i wreszcie jego miejsca w kwestionariuszu (chodzi o strukturę kwestionariusza i tak zwany efekt kontekstu, w szczegól-

⁴² W powyższych przypadkach z pewnością możemy mówić o efekcie techniki, a nie o błędzie. Błędem techniki byłoby natomiast na przykład zastosowanie ankiety audytoryjnej wśród dzieci początkowych klas szkoły podstawowej (chodzi oczywiście o zróżnicowany poziom umiejętności sprawnego czytania i pisanie).

ności efekt kontrastu i asymilacji)⁴³. Najczęstsze są dwa ostatnie rodzaje błędów. Mówiąc o błędach strukturalnych (błędach sformułowania pytania), mam przede wszystkim na myśli następujące:

- niedostosowanie pytania ze względu na sytuację respondenta (chodzi o to, czy pytanie „ma zastosowanie”, odnosi się do sytuacji wszystkich respondentów?);

- trudność pytania, która może być spowodowana jego nadmierną długością, zawartym w jednym pytaniu żądaniem dwóch lub więcej informacji, zbyt wysokim poziomem ogólności pytania, założeniem o posiadaniu przez respondenta wiedzy, której faktycznie nie posiada, oraz użyciem trudnych terminów czy też sformułowaniem pytania w formie przeczącej (zwłaszcza przy ocenie stwierdzeń na skali słownej);

- niejasność pytania wynikająca z: wieloznaczności użytych terminów i niejasności zapotrzebowania informacyjnego, użycia różnych terminów na określenie tego samego zjawiska (terminy te mają zazwyczaj różny zakres lub różną waloryzację), niezręczności sformułowania („zamykający” językowy);

- multiplikacja, a więc stosowanie zbitek pojęciowych (na przykład stosunek do religii i Kościoła), co może powodować, iż respondenci będą oceniać jeden z elementów lub je rozdzielać;

- drażliwość pytania, co może skłaniać do udzielania społecznie akceptowanych odpowiedzi, zgodnych z normami (chodzi o sytuacje, w przypadku których normy społeczne określają, co jest pożądane, a co nie), wzbudzać niepokój respondenta (chodzi o sprawy, które zgodnie z normami społecznymi należą do sfery tabu, na przykład: życie seksualne, przemoc w rodzinie, posiadane oszczędności) lub też powodować irytację respondenta i stwarzając ryzyko odmowy udzielenia odpowiedzi na cały blok pytań lub przzerwania wywiadu (może to być na przykład efekt nagromadzenia pytań trudnych);

- wi sugestia, której źródłem może być kafeteria „zamykająca” mentalnie respondentów, ograniczając się oni do wyboru spośród możliwych odpowiedzi (przykładem może być fałszywa dychotomia, brak logicznego punktu „środkowego”, niewyczerpująca, nierozłączna lub niehomogeniczna kafeteria, niezbalansowanie pytania, co oznacza, że istnieje nie-

⁴³ Problemom efektu kontekstu poświęcona jest między innymi praca Anety Krzewińskiej (2006), która empirycznie testuje niektóre hipotezy Norberta Schwarza.

równa szansa wyboru prezentowanych możliwych odpowiedzi). Źródłem sugestii może być także użycie terminów wartościujących (stwarza to ryzyko, że respondenci uznają je za wskazówkę, które odpowiedzi są „właściwe”) oraz ukryte w pytaniu założenie odzwierciedlające system wartości i przekonania badacza (respondenci w zdecydowanej większości zaakceptują to założenie i zgodnie z nim udzielą odpowiedzi i stąd jest to jeden z najbardziej niebezpiecznych błędów).

Mówiąc o błędach strukturalnych, trudno nie wspomnieć o efekcie sformułowania pytań. Pytanie jest bardzo wrażliwym instrumentem pomiaru i nawet niewielkie różnice jego sformułowania (przy zachowaniu ekwiwalentności logicznej) mogą powodować znaczące różnice w rozkładach odpowiedzi⁴⁴. Przykładowo może to być efektem rodzaju zastosowanego pytania (pytanie otwarte vs zamknięte, pytanie z odczytywaną kafeterią vs nieodczytywana), porządku prezentacji możliwych odpowiedzi, zamieszczenia punktu środkowego skali, który może w istotny sposób zmodyfikować uzyskane rozkłady odpowiedzi i wyniki analiz (Sztabiński P. 2002), długości skali (jej czułości), określenia skali częstości zachowań, która może dookreślać sens pytania (Schwarz, Hippler 1987). Również w przypadku złożonych pytań tabelarycznych należy pamiętać o efekcie sposobu zadawania pytania: w zależności od tego, czy pytanie zadajemy „poziomo” (wg boczku tabeli) czy też „pionowo” (wg jej główki), uzyskane wyniki mogą się różnić. I wreszcie, w przypadku pytań otwartych możemy uzyskiwać różne informacje w zależności od sposobu sondowania oraz instrukcji zapisu. Musimy mieć świadomość, iż praktycznie w przypadku każdego poprawnego pytania mamy do czynienia z efektem sformułowania. I nie ma możliwości eliminacji tego efektu: chcąc uzyskać jakąkolwiek informację, musimy zadać jakieś pytanie. A to, jak je sformułujemy, może mieć wpływ na uzyskany wynik.

Drugi najpowszechniejszy rodzaj błędów narzędzia jest związany z miejscem pytania w kwestionariuszu (chodzi o strukturę kwestionariusza i tak zwany efekt kontekstu). Mówiąc o strukturze kwestionariusza, mam na myśli przede wszystkim strukturę tematyczną oraz psychologiczną (Sztabiński F. 2005). Struktura tematyczna określana jest przez

⁴⁴ Schuman i Presser (1981) podają przykład pytania zadawanego w dwóch sformułowaniach: „zabronić – nie zabraniać” oraz „nie pozwolić – pozwolić”. Różnice odsetka wskazań sięgały 21 punktów procentowych. Z kolei Greymy (1995) opisuje przykład, w którym różnice sięgają 10 punktów.

związki pomiędzy poszczególnymi częściami kwestionariusza. Jest oczywiste, iż szczególne znaczenie ma tutaj kolejność bloków pytań, ich baterii (w ramach bloków) oraz poszczególnych pytań (w ramach baterii), zarówno w części merytorycznej, jak i metryczkowej⁴⁵. Odpowiedzi respondentów udzielane na poszczególne pytania są ze sobą powiązane. Kontekst, w jakim pojawia się każde z nich, lub jego miejsce w sekwencji mogą wpływać na udzielane odpowiedzi. Dlatego zmiana kolejności zadawanych pytań spowoduje udzielanie różnych odpowiedzi (efekt kontekstu). Zamieszczone wcześniej pytania mogą zatem ukierunkowywać uwagę respondenta na pewne informacje, wpływać na jego tok myślenia i treść udzielanych odpowiedzi. Jednym z aspektów porządku pytań są relacje między tak zwanymi pytaniami ogólnymi i szczegółowymi (odpowiedź na pytanie ogólne może być „średnią odpowiedzi” na poprzedzające je pytania szczegółowe lub też odpowiedzi na pytania szczegółowe mogą zbyt nie odbiegać, mieścić się „w granicach” odpowiedzi na poprzedzające je pytanie ogólne). Innym aspektem porządku pytań jest „efekt asymilacji” (jeśli informacja wcześniejsza jest użyta do formułowania odpowiedzi na następne pytanie) oraz „efekt kontrastu” (jeśli informacja jest celowo pomijana przy formułowaniu odpowiedzi) [Schwarz, Strack, Mai 1991]⁴⁶. Schuman i Presser (1981: 24) uważają, że efekt porządku pytań jest prawdopodobnie najczęściej przedstawianym wyjaśnieniem nieoczekiwanych lub niezgodnych z innymi wyników badania (pomijając błędy próby). Nie mniej istotną sprawą jest właściwa struktura psychologiczna kwestionariusza. Jest ona określana przez psychologiczną funkcję pytań i musi uwzględniać dynamikę wywiadu. Wiadomo, że na początku wywiadu respondent nie orientuje się do końca, na czym polega jego rola: stąd kwestionariusz winien zaczynać się od pytań łatwych, tak zwanych pytań uczących roli. Dopiero potem, gdy respondent „wejdzie” w swoją rolę, następują pytania trudniejsze, przy

⁴⁵ Wpływ porządku bloków tematycznych w badaniach czytelnictwa tygodników i miesięczników na wyniki badania pokazuje Dalbard-Martin (1996). Jeśli pytania o tygodniki zamieszczone są po pytaniach o miesięczniki, ze względu na większą częstość ich ukazywania się respondenci mają skłonność do częstszego ich wskazywania. Jeśli natomiast pytania o miesięczniki zamieszczone są po pytaniach o tygodniki, są wskazywane rzadziej.

⁴⁶ Efekt kontekstu może także polegać na aktywizacji pewnych norm na przykład zasady wzajemności. Jeśli dopuszcza się pewne zachowania jednych grup, nie można odmówić analogicznego prawa innym (Kubicki 2002; Schuman, Presser 1981).

czym należy pamiętać, że w poszczególnych częściach kwestionariusza winna być zachowana ciągłość tematyczna. Skakanie po tematach bardzo utrudnia respondentowi udzielanie odpowiedzi, powoduje zmęczenie i znużenie wywiadem, co oczywiście nie pozostaje bez wpływu na jakość uzyskiwanych informacji. Z kolei przed nowymi blokami pytań, zmieniając temat rozmowy, należy zamieścić wprowadzenie lub zastosować tak zwane pytania-łączniki (aczkolwiek ich funkcja jest częściowo odmienna, ponieważ w jakiś sposób odwołują się do poprzednich pytań). I wreszcie, tak zwane pytania-wata, które pozwalają respondentowi „odpocząć i zrelaksować się” po szczególnie uciążliwych blokach. Są to zazwyczaj pytania łatwe i interesujące dla niego⁴⁷.

C. Błąd związany z osobą ankietera. Jego źródłem mogą być charakterystyki ankietera i może on wystąpić w związku z pytaniami do nich nawiązującymi. Mogą to być cechy obserwowalne, a więc charakterystyki demograficzne (takie jak płeć, wiek) i do pewnego stopnia społeczne (na przykład wykształcenie, status) oraz pozaspołeczne (takie jak jego ogólny wygląd i sposób zachowania), ale także cechy nieobserwowalne, a więc „psychologiczne”. Chodzi tutaj przede wszystkim o jego cechy stałe (osobowościowe, system wartości), poglądy, wyobrażenia o opiniach różnych grup i kategorii społecznych, ale także wyobrażenia o pracy ankietera, instytucie i narzędziu badawczym, o umiejętności, wiedzy, motywacje, opinie w sprawach poruszanych w wywiadzie itp. (Sztabiński P. 1997: 46–72)⁴⁸.

D. Błąd związany z osobą respondenta. Jego źródłem mogą być także obserwowalne cechy społeczno-demograficzne i pozaspołeczne (na przykład sposób zachowania się) oraz cechy psychologiczne. Wśród tych ostatnich poza cechami osobowościowymi (także takimi jak na przykład skłonność do zgadzania się czy potakiwania, tendencję do udzielania

⁴⁷ Na błędy związane z psychologiczną strukturą kwestionariusza zwraca uwagę Paweł Sztabiński (1997), omawiając błąd związany z dynamiką wywiadu, który powstaje w wyniku interakcji między uczestnikami procesu badawczego, a więc ankieterem i respondentem.

⁴⁸ Należy wyjaśnić, iż Paweł Sztabiński (1997) przyjmuje, że błąd ankietera oznacza jego wpływ na uzyskaną i zanotowaną w kwestionariuszu informację. Błąd odnosi on zatem do wpływu szeroko rozumianych cech społecznych i psychologicznych ankieterów na wartość pomiaru, a nie konkretnych dróg oddziaływania ankieterów na uzyskiwane odpowiedzi (na przykład poprzez stosowanie sugerującego sondażu odpowiedzi).

społecznie akceptowanych odpowiedzi, skłonność do „prezentowania się” w pozytywnym świetle, to znaczy nieprzyznawania się do pewnych zachowań, które są negatywnie oceniane w społeczeństwie), wymienić należy jego wyobrażenia o potrzebie i sensowności badań, o ich celach, o instytucie badawczym itp. (tamże). Niezależnym od wskazanych, dodatkowym źródłem błędów respondenta jest ich pamięć⁴⁹. W wielu kwestionariuszach znaleźć można tak zwane pytania retrospektywne, wymagające od respondenta odwołania się do pamięci. Jednocześnie wiadomo, że odpowiadając na tego typu pytania, respondenci popełniają błędy: zapominają, że pewne wydarzenia w ogóle miały miejsce, a niektórzy z nich przypominają sobie wydarzenia, które w ogóle nie miały miejsca (ta ostatnia sytuacja jest znacznie rzadsza). Są to tak zwane błędy pominięcia i popełnienia. Relatywnie najczęstsze są jednak sytuacje, w których respondenci pamiętają, iż dane wydarzenie miało miejsce, ale w sposób błędny umiejscawiają je w czasie (tzw. efekt teleskopowy). Polega to na wydłużaniu dystansu czasowego, a więc raportowaniu, iż miało ono miejsce wcześniej niż faktycznie (*backward telescoping*), lub też skracaniu dystansu czasowego (kompresja), a więc raportowaniu, iż miało ono miejsce później niż faktycznie (*forward telescoping*) [zob. Auriat 1996; Gaskel, Wright, O’Muirheartaigh 2000; Sztabiński F. 2001].

Aczkolwiek wskazane powyżej źródła błędów obserwacji są z sobą powiązane, to najbardziej bezpośredni związek występuje jednak między dwoma ostatnimi rodzajami błędów: błędami ankietera i respondenta. Są to bowiem błędy powstające w wyniku interakcji między uczestnikami procesu badawczego. W większości przypadków błąd, którego źródłem jest ankieter, wystąpi tylko w związku z określoną reakcją respondenta. Z kolei wiele błędów respondenta jest pochodną określonych zachowań ankieterskich. Jeśli ankieter sugeruje respondentom podczas wywiadu

⁴⁹ Paweł Sztabiński (1997) pomija błędy związane z zawodnością pamięci respondenta. Na gruncie przyjętej przezeń psychologicznej koncepcji błędu, w której odpowiedzią prawdziwą jest ta, jakiej respondent udzieliłby sam sobie, analizowanie ich nie jest zasadne. Jeśli respondent nie pamięta jakiegoś wydarzenia i takiej odpowiedzi udziela w wywiadzie, to trudno uznać ją za błędną. Byłaby ona błędna przy przyjęciu koncepcji statystycznej, gdyż nie dostarczałaby informacji o stanie faktycznym. Jednakże, jak wspominałem, nie można mówić o ujęciu psychologicznym czy statystycznym, abstrahując od przedmiotu pytania. Stąd też wskazanie na błędy pamięci respondenta wydaje się celowe.

swoje poglądy, to aby wystąpił błąd, sugestii tej musi się poddać. Podobnie skłonność respondenta do udzielania odpowiedzi społecznie akceptowanych może zależeć od statusu ankietera (Sztabiński P. 1997: 28).

1. 1. 3. 3. Błędy opracowania

Mówiąc o błędzie przetwarzania danych mam na myśli sytuację, w której w fazie przetwarzania danych dokonano nieintencjonalnej zamiany zapisu oryginalnych informacji uzyskanych w badaniu na informacje inne co do wartości. Wśród tak rozumianych błędów opracowania wyróżnia się (Biemer, Caspar 1994; Weisberg 2005) następujące.

A. Błąd kodowania. Występuje on w przypadku pytań otwartych. Jego źródłem mogą być: wyobrażenia i oczekiwania koderów dotyczące struktury postaw respondentów oraz subiektywizm podczas porządkowywania odpowiedzi respondentów kategoriom kodowym. Kodowanie pytań otwartych opiera się bowiem na interpretacji wypowiedzi badanych. Poszczególni koderzy mogą różnić się swoją interpretacją i podejmować niejednakowe decyzje klasyfikacyjne. Rzeczą oczywistą jest także to, iż błędy te zależą również od jakości klucza kodowego.

B. Błąd wprowadzania danych. Zarówno podczas ręcznego wprowadzania danych, jak i skanowania optycznego – pewne symbole oryginalnego zapisu mogą zostać mylnie zinterpretowane (na przykład są nieczytelne) i błędnie wprowadzone.

C. Błąd edycji danych. W fazie edycji danych podejmowane są decyzje dotyczące interpretacji zapisu niepełnego lub niespójnego. Przyjęte procedury postępowania mogą prowadzić do błędnych rozstrzygnięć.

D. Błąd uzupełnienia braków danych. W tej fazie korzysta się z software'owych implementacji przyjętych procedur. Implementacje te mogą mieć jednak błędy i ich skutkiem może być przypisanie wartości w sposób niezgodny z przyjętą procedurą.

1. 2. Ocena wartości danych: perspektywa zewnętrzna

Wcześniej wspomniałem, iż różni autorzy, dokonując oceny wartości danych w badaniach ilościowych, przyjmują różne kryteria ich oceny. Wśród nich wymieniłem trafność i rzetelność pomiaru. Traktując trafność i rzetelność jako wskaźniki wartości czy też oceny jakości pomiaru, przyjąłem, iż te dwa kryteria odnoszą się nie tyle do procesu pomiaru, ile do jego rezultatu, a więc uzyskanego wyniku. W tym przypadku ocenie podlega bowiem nie sam proces badawczy czy jego elementy, a jego ostateczny rezultat, wynik. Stąd też podejście to uznałem za takie, które przyjmuje perspektywę zewnętrzną wobec aktu (a właściwie procesu) pomiaru. Taki też „zewnętrzny” charakter mają owe dwa kryteria oceny.

Podobnie „zewnętrzny” charakter mają dwa kolejne kryteria oceny wartości danych, a mianowicie weryfikacja i walidacja.

1. 2. 1. Trafność i rzetelność pomiaru

Koncepcję trafności i rzetelności wypracowano co prawda na gruncie psychologii, ale można ją stosować do wyników badań socjologicznych, mając jednak na uwadze odmienność przedmiotu badania (Styczeń 1968)⁵⁰. Aby ocenić jakość pomiaru, a zwłaszcza jego trafność, niezbędne jest przede wszystkim jasne i precyzyjne zdefiniowanie problematyki

⁵⁰ Styczeń (1968) pisze wprost, iż koncepcja trafności wywodzi się i ma zastosowanie przede wszystkim w psychologii. Jednocześnie może być przydatna w socjologii, o ile bada się nieobserwowalne bezpośrednio zjawiska, a więc pewne konstrukty. Chodzi tutaj o to, co Lutyński (1978, 1979) nazywa koncepcją testową pytania. Jak stwierdza, jest ona przeniesiona do badań socjologicznych z psychologii, a właściwie z psychometrii. W koncepcji tej pytanie traktuje się jako jedną z wielu pozycji w całym teście i nie może ono występować samodzielnie. Celem testu jest bowiem przypisanie respondentowi pewnego wyniku liczbowego, który określi jego miejsce na skali. Każde z pytań dostarcza konkretnej informacji, ale jest to informacja minimalistyczna: znaczenie posiada jedynie ogólny wynik (tak jest na przykład w przypadku skali autorytaryzmu). Renate Mayntz, Kurt Holm i Peter Hubner (1985: 155) piszą wprost, że „Kwestie trafności mniej ważą przy pytaniach o fakty niż przy pytaniach o opinie lub o cechy niedostępne bezpośredniej obserwacji”.

badania (co chcemy zmierzyć) oraz poprawne sformułowanie pytania badawczego⁵¹.

1. 2. 1. 1. Trafność jako kryterium jakości pomiaru

Najogólniej, trafność pomiaru mówi o tym, czy narzędzie mierzy to, co ma mierzyć, lub z jaką dokładnością zmierzylimy to, co chcieliśmy zmierzyć. Odzwierciedla ona relację, jaka zachodzi między koncepcją właściwości psychologicznej, a użytymi wskaźnikami, czyli adekwatność operacjonalizacji (Fink 2003: 49–53; Hornowska, Paluchowski 2004; Litwin 1995: 33–46). Z kolei Arleck i Settle (1985: 423) określają trafność jako stopień, w którym dane z badania lub rezultaty są wolne od błędów systematycznych i losowych, ponieważ pomiaru dokonuje się za pomocą określonego narzędzia. Selltiz, Wrightsman i Cook (1977) definiują ją jako stopień, w jakim różnice wartości uzyskane za pomocą danego instrumentu bardziej przekładają się na rzeczywiste różnice między jednostkami w zakresie charakterystyk, które chcemy badać, niż na błąd systematyczny lub losowy. Nie istnieją więc narzędzia pomiaru całkowicie trafne, to znaczy takie, które mierzą wyłącznie zróżnicowanie charakterystyk, które są przedmiotem badania. Dzieje się tak, dlatego że w pomiarze nie jesteśmy w stanie wyeliminować wszystkich „obcych” czynników go zanieczyszczających, a zatem trafność ma charakter stopniowalny.

Ponieważ w psychologicznej koncepcji błędu nie znamy wartości prawdziwej, nie istnieje bezpośredni sposób określenia trafności pomiaru. Dlatego też trafność możemy ocenić jedynie ze względu na stopień zgodności wyniku z innymi właściwymi faktami. Ze względu na rodzaj faktów, które przyjmuje się jako podstawę oceny trafności, Selltiz, Wrightsman i Cook (1977) wyróżniają⁵²:

⁵¹ Czasami mówi się o trafności narzędzia oraz trafności wyniku uzyskanego przy zastosowaniu danego narzędzia. Można więc mówić o rzetelności narzędzia i rzetelności respondenta (Lavrakas 2008: 938, 714). Niekiedy mówi się także o rzetelności wyniku, a więc na poziomie zbiorczym (Groves, Fowler Jr., Couper, Lepkowski, Singer, Tourangeau 2004: 260).

⁵² Autorzy ci nie wspominają o trafności treściowej (*content validity*), która określa możliwie wyczerpującą liczbę wskaźników i miar wykorzystywanych do pomiaru zjawiska. Dane narzędzie cechuje się tym wyższą trafnością treściową, im obszerniejszy zakres wskaźników obejmuje (Fink 2003: 49–53; Litwin 1995: 33–46; Lavrakas 2008: 938).

(i) Trafność kryterialną (*criterion validity*), która odzwierciedla, czy przewidywane wyniki uzyskane za pomocą danego narzędzia pomiarowego, mogą zostać potwierdzone za pomocą innych wybranych narzędzi pomiarowych lub zmiennych, które traktujemy jako znaczące kryteria. Jej ocena wymaga stosunkowo pewnego i rzetelnego kryterium, które wykorzystujemy jako podstawę porównań wyników pomiaru. Oceny dokonuje się tutaj na podstawie pomiaru określonej właściwości – innej niż ta, która jest bezpośrednio mierzona. Trafność kryterialna jest to zatem stopień zgodności pomiarów otrzymanych za pomocą narzędzia, którego trafność oceniamy, oraz pomiarów otrzymanych w wyniku oceny kryterium.

Kryterium jest właściwością, o której chcemy wnioskować z wyniku pomiaru. Kryterium jest zmienną zewnętrzną w stosunku do pomiaru (Babbie 2004; Hornowska, Paluchowski 2004). Istnieją dwa rodzaje trafności kryterialnej: diagnostyczna (kryterium dla naszego testu jest pomiar zachowania lub cechy aktualnie występującej u badanych) oraz prognostyczna (kryterium naszego testu stanowi pomiar zachowania lub cechy mającej wystąpić w przyszłości u badanych). Różnica między trafnością diagnostyczną i prognostyczną nie polega na czasie zbierania danych kryterialnych, lecz na tym, że są różne przedmioty badania: przy diagnozie interesuje nas stwierdzenie stanu istniejącego, przy prognozie natomiast – prawdopodobieństwo wystąpienia pewnych skutków stanu aktualnego (Brzeziński 1984; Hornowska, Paluchowski 2004). Trafność prognostyczna określa wartość przewidywań dokonywanych na podstawie badania i używana jest do prognozy przyszłych wyników (na przykład test przydatności do zawodu). Aby ją ocenić, musimy posiadać stosunkowo pewne i rzetelne kryterium, które wykorzystujemy jako podstawę porównań wyników pomiaru (wyniki osiągnięte w pracy – najlepiej mierzone w sposób pewny, na przykład wartością sprzedaży). W badaniach socjologicznych może to być przewidywanie wyników wyborów lub sprzedaży na podstawie deklaracji respondentów.

(ii) Trafność konstrukcyjną (*construct validity*), która z kolei określa, czy skonstruowane narzędzie jest właściwe w rozumieniu teorii, do której się odnosi, a więc sprawdza je pod kątem jego powiązania z pojęciami teoretycznymi i założeniami teoretycznego kontekstu badania. W tym przypadku nie odwołujemy się do zewnętrznego kryterium oceny, ale dokonujemy jej na podstawie logicznych powiązań między zmiennymi.

Ponieważ problem trafności jest jednym z podstawowych w psychologii, stąd wielu autorów wyróżnia jeszcze inne, specyficzne rodzaje jej rozumienia. Celowa wydaje się ich prezentacja, aczkolwiek musimy mieć świadomość, iż w niektórych przypadkach ujęcia te nie są ze sobą rozłączne. Poniżej wyliczam je, nie podejmując jednak próby ich systematyzacji. I tak, mówi się jeszcze o (Litwin 1995: 44–45, Fink 2003: 49–52).

(iii) Trafność fasadową lub nominalną (*face validity*), która oparta jest na intuicjach i wyobrażeniach dotyczących badanego zjawiska lub też konwencjach co do znaczenia prostych pojęć (gospodarstwo domowe, praca zarobkowa itp.). Trafność fasadowa dotyczy tego, co „wydaje się mierzyć”, a nie tego, co rzeczywiście się mierzy (Brzeziński 1984). Określając trafność fasadową, musimy odpowiedzieć na pytania: na ile zastosowany instrument pomiaru mierzy zachowania (zjawiska), które chcemy mierzyć i czy daje zadowalającą próbkę tego rodzaju zachowań (zjawisk). Pod pewnymi względami jest to zatem ocena logiczna (Goode, Hatt 1952; Babbie 2004).

(iv) Trafność treściową, zawartości – logicznej lub wewnętrznej (*content validity*), która jest subiektywną, ale systematyczną oceną tego, czy narzędzie pomiarowe pokrywa cały obszar problemowy mierzonego konstrukt (na przykład wszystkie elementy związane z wizerunkiem marki), a więc tego, czy na podstawie części właściwości uwzględniliśmy wszystkie elementy wchodzące w zakres pojęcia (cechy badanej). Innymi słowy, jest to zakres, w którym itemy właściwie reprezentują interesujący na konstrukt (Brzeziński 1984; Hornowska, Paluchowski 2004)⁵³. Należy zatem określić wszystkie możliwe rodzaje elementów, które mogą służyć do pomiaru cechy (chodzi o główne aspekty pojęcia), i dokonać ich wyboru ze względu na użyteczność w badaniu interesującej nas cechy (traktując każdy z nich jako niezależną całość). Trafność treściowa bada więc, czy treść testu jest zbieżna z oczekiwaniami wobec

⁵³ Brzeziński (1984) odróżnia jednak trafność treściową od wewnętrznej, choć przyznaje, że są one zbliżone. W przypadku trafności treściowej ocenia się ją poprzez wykazanie stopnia, w jakim pozycje testu są próbą zbioru wszystkich pozycji mających związek z daną cechą. Można to robić metodą kompetentnych sędziów, którzy dokonują analizy logicznej poszczególnych pozycji w kwestionariuszu. O trafności wewnętrznej mówimy wtedy, gdy kwestionariusz odpowiada koncepcji założonej przez jego autora. Można to określić na podstawie analizy logicznej, której celem jest stwierdzenie, czy kwestionariusz mierzy to, co ma mierzyć.

tę, co ma badać (podobnie jak trafność fasadowa, jednak opiera się na głębszej analizie). Na przykład, test matematyczny musi obejmować nie tylko dodawanie, ale też odejmowanie, mnożenie i dzielenie (Babbie, 2004).

Pewnego typu „narzędziem” oceny trafności treściowej może być rozszerzona informacyjna koncepcja pytania Jana Lutyńskiego (1978, 1979; Kubiak, Przybyłowska 1989). Autor ten zakłada, że przygotowując badanie, należy stworzyć pewien system pytań i odpowiedzi. Jego elementami są:

- pytania problemowe (problematyka badania), a więc ogół pytań, na które badacz chce dostarczyć odpowiedzi, badając dane zjawisko. Wśród pytań problemowych wyróżnić można pytania problemowe naczelne (ogólne) i podporządkowane (szczegółowe). Uzyskanie odpowiedzi na pytanie szczegółowe jest konieczne do uzyskania odpowiedzi na pytanie ogólne;

- poszukiwane jednostkowe informacje (jednostkowe pytania badacza), które są konieczne do rozwiązania pytań problemowych. Dotyczą one poszczególnych respondentów. Jednostkowe pytanie badacza to pytanie o przynależność jednostki do kategorii badanych zjawisk (formułowane w odniesieniu do każdej badanej jednostki). Pytanie to w sposób szczegółowy określa zapotrzebowanie badawcze;

- pytania kwestionariusza stworzone na podstawie pełnej listy jednostkowych poszukiwanych informacji. Często polega to na zmianie formy gramatycznej, jednak relacje między tymi pytaniami bywają znacznie bardziej skomplikowane. Zdarza się bowiem, że nie można zadać respondentowi jednego pytania kwestionariuszowego odpowiadającego jednemu pytaniu badacza (czy na przykład X jest religijny). W takich sytuacjach należy rozbić poszukiwaną informację na pytania bardziej szczegółowe, którym dopiero będą odpowiadać pytania w kwestionariuszu;

- pytania kodera, to znaczy pytania, które należy zadać, aby właściwie zaklasyfikować odpowiedź respondenta (jest ono logicznie wtórne w stosunku do jednostkowego pytania badacza).

Wyróżnione elementy stanowią pewien system pytań i odpowiedzi: jeśli badacz sformułował jednostkowe pytanie badacza, przemyślał zasady przechodzenia od odpowiedzi respondenta na dane pytanie kwestionariuszowe do jednostkowej poszukiwanej informacji oraz w konsekwencji do odpowiedzi na szczegółowe pytania problemowe, będzie mógł po zebraniu i opracowaniu materiałów rozwiązać interesujące go

problemy i sprawdzić swoje hipotezy. Zdaniem Lutyńskiej (1984: 18), wymagania stawiane przez tę koncepcję pytania „dają gwarancję, iż zebrane materiały – dotyczące zwłaszcza opinii – w większym stopniu będą odzwierciedlały rzeczywiste poglądy respondentów niż spekulacje badacza oparte na jego własnej intuicji”.

(v) Trafność zewnętrzną (*external validity*), która odnosi się do możliwości generalizacji wyników, a więc zakresu ich uogólniania (Babbie 2004; Wojciszke 2004). Chodzi o stopień, do którego badacz chce uogólnić otrzymane – na podstawie pomiaru pewnej grupy jednostek – wyniki badania.

(vi) Tak zwany *expert validity* (ocena ekspertów), która polega na tym, że eksperci sprawdzają trafność skali, kierując się swoją intuicją i znajomością przedmiotu. Opiera się ona na tych samych założeniach co ocena trafności fasadowej (Goode, Hatt 1952; Mayntz, Holm, Hubner 1985).

(vii) Tak zwany *known groups validity* (porównanie grup o znanych właściwościach, znajomość własności grup), w przypadku której skala zostaje sprawdzona w grupach badanych, o których wiadomo, że reprezentują wysoką i niską wartość danej cechy, będącej przedmiotem pomiaru. Średnią wyników takich grup porównuje się ze średnią grup docelowych. Istotna różnica tych dwóch średnich dowodzi trafności skali (Goode, Hatt 1952; Mayntz, Holm, Hubner 1985). W przypadku choćby skali postaw wobec Kościoła badamy osoby praktykujące i niepraktykujące: jeśli wyniki się nie różnią, skala jest nietrafna. Metoda ta pociąga za sobą jednak niebezpieczeństwo, że zakłócać mogą inne cechy grup, wpływające na uzyskany wynik.

Ze względu na tak duże zróżnicowanie ujęć trafności, wspomnieć należy o klasyfikacji jej typów zaproponowanej przez Thomasa D. Cooka i Donalda T. Campbella (za: West 2004)⁵⁴. Wyróżniają oni cztery rodzaje rozumienia trafności i jej warunki.

(i) Trafność wnioskowania statystycznego. Zagrożeniem dla niej mogą być: (a) niespełnione założenia testów statystycznych (na przy-

⁵⁴ Mówiąc o zróżnicowaniu koncepcji trafności, mam na myśli nie tylko badania o charakterze ilościowym. W badaniach o charakterze jakościowym mówi się bowiem dodatkowo na przykład o *instrumental validity* (trafność wyboru narzędzia, na ile narzędzie jest dostosowane do problemu) i *consultative validity* (zgodność wniosków z przyjętym, powszechnym przekonaniem na temat którego dotyczą dane) [Goodyear 1998].

kład rozkład normalny lub zależność pomiarów w badaniach longitudinalnych) oraz (b) nieadekwatny plan badania, ze względu na narzędzie pomiaru, błędy realizacji badania oraz próbę (liczebność).

(ii) Trafność wewnętrzna. Mówimy o niej, gdy manipulowanie zmienną niezależną powoduje uzyskanie zamierzonego wyniku. Zagrożenia trafności wewnętrznej wynikają przede wszystkim z: (a) zmian, jakim podlegają sami badani i ich doświadczenia, (b) pomiaru (testowania i narzędzia pomiaru), (c) doboru próby, (d) interakcji wymienionych czynników. Zagrożenie trafności wewnętrznej można zminimalizować poprzez: losowy dobór osób do grup (nie chodzi wyłącznie o randomizację), udoskonalenie koncepcji badawczej, statystyczne dopasowanie grup oraz uwzględnienie wyników uprzednich badań i teorii.

(iii) Trafność teoretyczna. W tym przypadku istnieje problem wskazania i nazwania zmiennej zależnej oraz wskazania związków między zmienną zależną i niezależną na poziomie konceptualnym i operacyjnym. Może się bowiem okazać, iż mierzymy coś dodatkowego lub zupełnie coś innego, niż zakładaliśmy (chodzi o tak zwany konstrukt teoretyczny).

(iv) Trafność zewnętrzna. W tym przypadku chodzi o możliwości uogólniania rozpatrywanych zjawisk na różne populacje (badani – charakter próby), rodzaj otoczenia (sytuację badania: naturalna vs sztuczna – eksperymentalne) oraz na czas (tu i teraz lub w ogóle). Trafność zewnętrzną można zwiększyć między innymi poprzez reprezentatywność próby oraz replikę badania.

Podobny, porządkujący, charakter ma propozycja przedstawiona przez Williama Trochima (2006b), który prezentuje listę typów trafności, przywoływanych, gdy mówi się o jakości pomiaru. Nadrzędną dla niego kategorią jest trafność teoretyczna (*construct validity*), która wskazuje, na ile przyjęta konceptualizacja i operacjonalizacja odzwierciedla badany konstrukt. W jej ramach wyróżnić można dwa szerokie podtypy trafności: translacyjną (*translation validity*) oraz kryterialną (*criterion-related validity*). Pierwsza z nich ogniskuje się na problemie, czy operacjonalizacja jest dobrym odzwierciedleniem konstruktu. Zakłada się, że istnieje dokładna jego definicja i można sprawdzić jej operacjonalizację. W ramach trafności translacyjnej Trochim wyróżnia: (a) trafność faszadową (*face validity*) oraz (b) trafność treściową (*content validity*). Drugi typ zakłada sprawdzenie funkcjonowania operacjonalizacji w przypadku określonych kryteriów. W jej ramach wyróżnia on: (a) trafność pro-

gnostyczną (*predictive validity*), (b) trafność równoczesną (*concurrent validity*), która bada korelację między skonstruowanym testem a innym testem, o którym wiadomo, że mierzy tę samą właściwość, (c) trafność zbieżną (*convergent validity*) oraz (d) trafność różnicową (*discriminant validity*).

Ponieważ, jak wspomniałem, nie jest moim celem rozstrzygnięcie ogólnych problemów trafności pomiaru, nie będę analizował powyższych klasyfikacji. Celowe wszakże wydaje się wskazanie pewnych „priorytetów” wśród typów trafności, jeśli mówimy o ocenie jakości pomiaru. Stephen West (2004) wskazuje, że w badaniach podstawowych (teoretycznych) i stosowanych przyjmuje się różne priorytety. W badaniach podstawowych priorytetem jest trafność wewnętrzna, potem teoretyczna, wnioskowania, a na końcu zewnętrzna, w badaniach stosowanych zaś – trafność wewnętrzna, potem zewnętrzna, trafność wnioskowania, a na końcu teoretyczna. W rzeczywistości jest to, jak podkreśla, wybór między trafnością wewnętrzną a zewnętrzną.

Jeśli chodzi o metody oceny trafności, to w przypadku trafności teoretycznej zasadniczo istnieją dwie ich grupy.

Pierwsza z nich to badanie związków między konstruktem a zachowaniem. Wymienić tutaj należy (Hornowska, Paluchowski 2004):

- metodę różnic międzygrupowych (jest to procedura weryfikowania hipotez dotyczących różnego zachowania się dwóch grup. Hipotezy te wyprowadzane są z teorii mierzonej cechy i zakłada się, iż badane grupy będą różnić się ze względu na ogólny wynik w teście (grupa o niskich wynikach *vs* grupa o wysokich wynikach). Weryfikacji dokonuje się na podstawie wyników testu. Powinny na przykład wystąpić różnice między praktykującymi i niepraktykującymi w teście postaw wobec Kościoła (Brzeziński 1984);

- metodę zmian nieprzypadkowych (analiza ta polega na porównaniu wyników dwukrotnego badania tym samym testem). W przerwie między badaniami wprowadza się oddziaływanie eksperymentalne, wyprowadzone z teorii mierzonej cechy. Wynik porównywania powinien być zgodny z założonymi efektami manipulacji) oraz

- badanie procesu rozwiązywania testu polega na analizie narzędzia z punktu widzenia procesów przetwarzania informacji. Jeśli test zawiera zadania do rozwiązania lub polecenia do wykonania, to obserwacja kolejności wykonywanych czynności, elementów zadania sprawiających najwięcej trudności, elementów pomijanych, popełnianych błędów itd.

dostarcza informacji o mierzonym konstrukcie poprzez wgląd w procesy odpowiadania, w mniejszym stopniu poprzez analizę odpowiedzi)⁵⁵.

Grupa druga to metody oparte na analizie macierzy korelacji testu z innymi testami. W jej ramach wskazać można między innymi:

– metodę badania wewnętrznej struktury testu (zgodność wewnętrzna – homogeniczność), to stopień, w jakim dane narzędzie można uznać za miarę jednego konstruktów (cechy psychologicznej). Jeżeli w świetle teorii można przyjąć, że mierzona cecha jest homogeniczna, to wysoki współczynnik zgodności wewnętrznej (*alfa* Cronbacha) może być traktowany jako dowód trafności teoretycznej narzędzia);

– analizę korelacji – analizę czynnikową (tak zwana eksploracyjna i confirmacyjna trafność czynnikowa). Polega ona na wykrywaniu czynników odpowiedzialnych za współzależność wyników testów korelujących ze sobą. Celem analizy jest odnalezienie nowej grupy zmiennych, mniej licznej niż grupa zmiennych oryginalnych, które wyrażają to, co jest wspólne między oryginalnymi zmiennymi. Jest to technika statystyczna używana do identyfikacji relatywnie małych grup czynników, mogących reprezentować relacje między grupami wielu wzajemnie skorelowanych zmiennych, oraz

– analizę WCWM: Wielu Cech Wielu Metod (MTMM: Multitrait – Multimethod Matrix). Metoda ta została opracowana przez Donalda T. Campbella i Donalda W. Fiske (1959)⁵⁶. Autorzy ci wprowadzili wspomniane dwa pojęcia określające subkategorie trafności teoretycznej: trafność zbieżna lub konwergencyjna (*convergent validity*) oraz trafność różnicowa lub dyskryminacyjna (*discriminant validity*). Pierwsza z nich – trafność konwergencyjna – określa stopień, w jakim pojęcia, które powinny być teoretycznie powiązane, są takie w istocie (tzn. do jakiego stopnia poszczególne skale korelują pozytywnie z innymi pomiarami tej samej cechy, które wydają się silnie związane z badaną cechą). Z kolei trafność dyskryminacyjna określa stopień, w jakim pojęcia, które nie powinny być teoretycznie powiązane, w rzeczywistości takie nie są (nie korelują z innymi konstruktami, w przypadku których zakłada się, że powinny się różnić). Współczynniki korelacji winny być wysokie

⁵⁵ Obie te metody mogą być także traktowane jako metody badania rzetelności.

⁵⁶ Technikę MTMM omawiam za: William Trochim (2006a). Metoda ta, mimo ograniczeń, jest stosowana jako narzędzie oceny wartości pomiaru (zob. Saris 2002; Saris, Gallhofer 2007).

w przypadku trafności zbieżnej i niskie, w przypadku trafności dyskryminacyjnej. Aby określić trafność teoretyczną, należy ocenić trafność zarówno konwergencyjną, jak i dyskryminacyjną. Oceny tej dokonuje się na podstawie wielkości współczynników korelacji z innymi, podobnymi testami i z testami mierzącymi inne cechy.

W rygorystycznej wersji analiza trafności WCWM (MTMM) powinna być poprzedzona dwoma innymi typami analizy: wykazaniem, że różne metody (narzędzia pomiaru) badanej cechy dają podobne wyniki (trafność konwergencyjna) oraz że inne cechy różnicują badaną cechę (trafność dyskryminacyjna) [Campbell i Fiske 1959]⁵⁷. Analiza WCWM (MTMM) jest w istocie testowaniem trafności konceptualizacji i operacjonalizacji badanej cechy. W praktyce wykorzystywana jest jednak raczej koncepcja trafności konwergencyjnej.

Celowe wydaje się bardziej szczegółowe omówienie tej metody analizy. Otóż, jak już wspomniałem, WCWM (MTMM) jest tabelą zawierającą wartości współczynników korelacji między kilkoma różnymi zmiennymi (cechami – *traits*), mierzonymi jedną lub wieloma metodami. Dla przykładu, jeśli badamy 3 różne cechy (C1, C2 i C3) za pomocą dwóch różnych metod (M1 i M2), otrzymujemy 36-półową tabelę. Ponieważ jest to tabela symetryczna, wystarczy analizować jedynie jedną jej część.

Rysunek 3. *WCWM (MTMM) dla pomiaru trzech WCWM (MTMM) cech (C1, C2, C3) dwiema różnymi metodami (M1 i M2)*

Metoda/Cechy		Metoda 1			Metoda 2		
		C1	C2	C3	C1	C2	C3
Metoda 1	C1	r11					
	C2	r21	r22				
	C3	r31	r32	r33			
Metoda 2	C1	r11	r12	r13	r11		
	C2	r21	r22	r23	r21	r22	
	C3	r31	r32	r33	r31	r32	r33

W analizie bierze się pod uwagę wartości współczynników (Trochim 2006a):

⁵⁷ Powyższe metody są nie tylko metodami badania trafności, ale także rzetelności.

(i) rzetelności diagonalnej, a więc pomiar jednej cechy jedną metodą (pola: r_{11} , r_{22} , r_{33} dla metody 1 i r_{11} , r_{22} , r_{33} dla metody 2);

(ii) trafności diagonalnej, a więc pomiar jednej cechy różnymi metodami (pola: r_{11} , r_{22} , r_{33} dla metody 1 i metody 2);

(iii) korelacji w ramach „trójkątów” przy pomiarze różnych cech jedną metodą („trójkąty” utworzone przez pola: r_{21} , r_{31} , r_{32} dla metody 1 oraz r_{21} , r_{31} , r_{32} dla metody 2);

(iv) korelacje w ramach „trójkątów” przy pomiarze różnych cech różnymi metodami („trójkąty” utworzone przez pola: r_{12} , r_{13} , r_{23} oraz r_{21} , r_{31} , r_{32} dla metody 1 i metody 2);

(v) korelacje w ramach „bloków” przy pomiarze jedną metodą („bloki”: metoda 1/metoda 1 oraz metoda 2/metoda 2);

(vi) korelacje w ramach „bloków” przy pomiarze różnymi metodami („blok” metoda 1/metoda 2).

Podstawową zaletą WCWM (MTMM) jest to, że wprowadza metodologię oceny trafności teoretycznej poprzez możliwość równoczesnej analizy trafności zbieżnej i dyskryminacyjnej. Jest to istotne, ponieważ Campbell i Fiske podkreślali konieczność równoczesnej analizy tego, co się mierzy, i tego, jak się mierzy. Mimo wielu zalet, MTMM jest relatywnie rzadko wykorzystywana. Głównym powodem jest rygorystyczna metodologia zakładająca, jak wspomniałem, pomiar każdej cechy każdą z metod. W rzeczywistości w wielu przypadkach jest to niemożliwe. Dodatkowo MTMM bazuje na interpretacji różnych układów współczynników, co powoduje, iż różni badacze, bazując na tej samej macyry, mogą w sposób uprawniony dojść do odmiennych konkluzji. Wreszcie, MTMM nie pozwala na określenie stopnia trafności konstrukcyjnej w badaniu.

1. 2. 1. 2. Rzetelność jako kryterium jakości pomiaru

Najogólniej ujmując, rzetelność pomiaru mówi o tym, z jaką dokładnością zmierzylismy to, co zmierzylismy, lub dokładność, z jaką test mierzy to, co mierzy (Brzeziński 1984)⁵⁸. Arleck i Settle (1985: 418) określają rzetelność jako stopień, w jakim wynik badania jest wolny od błędu losowego.

⁵⁸ Aczkolwiek niekiedy wskazuje się, iż termin „rzetelność” ma przynajmniej cztery znaczenia (Lavrakas 2008: 713).

Ocena rzetelności pomiaru polega na określeniu proporcji wariancji wyniku powstałej wskutek czynników przejściowych, zmieniających się w poszczególnych pomiarach, a więc do jakiego stopnia pomiar jest wolny od błędu losowego. Im słabszy jest wpływ tych czynników na wynik pomiaru, tym narzędzie jest bardziej rzetelne, to znaczy wynik jest bardziej stały. Selltiz, Wrightsman i Cook (1977) wskazują trzy wymiary rzetelności (por. Fink 2003: 46–49).

1. Stabilność w czasie rezultatów zastosowania instrumentu pomiaru, które ocenia się na podstawie stałości wyniku powtórzonego pomiaru dokonanego w kolejnych okresach (test – retest: T – RT). Idea „powtórzonego pomiaru” wywodzi się w prostej linii z nauk przyrodniczych, jedną z cech charakterystycznych których jest replikowalność wyników: wynik uznaje się za pewny, jeśli pomiary przeprowadzone w tych samych warunkach dają te same rezultaty (Heisenberg 1975: 55; Litwin 1995: 30). Z uwagi na pozytywistyczne korzenie badań ilościowych, często przyjmuje się, że replikowalność jest koniecznym ich elementem. Aby zatem ocenić rzetelność pomiaru, musimy go powtórzyć (Bohrstedt 1970; Bryman 1995; Duncan 1975; Heise 1969; Jones 2000). Metoda ta opiera się na szacowaniu zgodności międzyokresowej wyników, gdy badani są poddani pomiarowi za pomocą identycznego testu w dwóch różnych momentach czasowych (najczęściej okres ten wynosi około dwóch tygodni), aczkolwiek w projekcie ESS pytania ze zmienionymi skalami zadaje się pod koniec wywiadu, a więc po około 60 minutach (Saris 2002). Pomiar jest rzetelny, jeśli wyniki uzyskane w kolejnych pomiarach są takie same. Miarą zgodności międzyokresowej może być na przykład współczynnik korelacji Kendalla. Badanie rzetelności metodą T-RT ma jednak pewne ograniczenia. Po pierwsze otrzymany wynik zależy w dużym stopniu od upływu czasu między pomiarami (w tym czasie opinie badanych mogą ulec zmianie); po drugie czynniki zewnętrzne oddziałujące na pomiar danej cechy mogą zmieniać wyniki drugiego pomiaru; i wreszcie, po trzecie, pierwszy pomiar „uczula” badanego na narzędzie pomiarowe, co może wpływać na wynik drugiego pomiaru (pomiar przestają być niezależne). Stąd też, na metodę tę nakłada się następujące warunki (Styczeń 1968): (a) oba pomiary winny być dokonane w takim odstępie czasu, aby mierzone własności nie mogły ulec zmianie; (b) dystans czasowy między tymi dwoma pomiarami winien być na tyle długi, aby na pomiar nie wpływały te same, niekontrolowane czynniki; (c) niezależność pomiaru, to znaczy pierwszy pomiar nie może

wpływać na przebieg i wynik drugiego. Jeśli nie są one spełnione, procedura test – retest prowadzić do błędnych wniosków, ponieważ trudno oddzielić efekt zastosowania procedury od rzetelności narzędzia (Domański, Dukaczewska 1996; Groves, Fowler Jr., Couper, Lepkowski, Singer, Tourangeau 2004: 262).

Następne metody określania rzetelności, a co za tym idzie również koncepcje można określić mianem statystycznych, ponieważ opierają się na jednym pomiarze (w przeciwieństwie do koncepcji rzetelności rozumianej jako stabilności, a więc powtórnego pomiaru) [por. Litwin 1995: 30].

2. Równoważności (alternatywnego pomiaru, *alternate form*). Równoważność rezultatów ocenia się na podstawie zastosowania w tym samym czasie i wobec tych samych osób różnych narzędzi, które z założenia mierzą to samo zjawisko (chodzi o dwa niezależne, ale w pełni ekwiwalentne testy w tym samym badaniu, stąd też metoda ta często nazywana jest metodą testów równoległych lub alternatywnych form skali: tak zwanej skali pierwotnej i wtórnej). W badaniach socjologicznych ma ona ograniczoną przydatność.

3. Homogeniczności (rozczepionych połówek, *internal consistency*). Homogeniczność lub tak zwaną zgodność wewnętrzną rezultatów ocenia się na podstawie korelacji między dwiema połówkami pytań (lub losowo dobranymi pytaniami) tego samego testu, tak zwany test połówkowy (*split-half*). Są one traktowane jako alternatywne wersje tego testu, oczywiście przy założeniu, że każdy item mierzy to samo (Babbie 2004; Carmines i Zeller 1988; Domański, Dukaczewska 1996:72–75). Wykorzystuje się tutaj mierniki zgodności wewnętrznej, którymi są współczynniki zgodności, określające stopień, w jakim odpowiedzi na poszczególne pozycje skali mierzą to samo, co wynik w całej skali. Jeśli jest on rzetelny, wyniki powinny być identyczne. Pamiętać jednak należy, iż jest to metoda pomiaru rzetelności połowy, a nie całej skali. Podziału całej skali na dwie połowy można dokonać w sposób losowy lub wybierając pozycje parzyste i nieparzyste czy grupując pytania o równych parametrach mocy dyskryminacyjnej. Jest to najczęściej stosowany sposób oceny rzetelności w badaniach socjologicznych.

Innym sposobem badania homogeniczności jest interkorelacja między odpowiedziami na wszystkie pytania testu. Przyjmuje się, że różne czynniki „przejściowe” mogą wpływać na poszczególne pytania testu w różnym stopniu. Rzetelność może być więc analizowana ze względu

na wpływ tych czynników na różne pytania, czyli, przenosząc na grunt operacyjny zgodność wyników pomiaru za ich pomocą. Zaletą tej metody jest możliwość zastosowania jednokrotnego pomiaru daną skalą w celu oceny jej rzetelności. Jednakże różne metody podziału całej skali dają różne oceny jej rzetelności. Dlatego próbuje się rozwiązać ten problem poprzez określenie średniej rzetelności dla wszystkich możliwych sposobów podziału skali na dwie połówki. Metoda ta jest podstawą szacowania zgodności wewnętrznej na podstawie wzorów: *alpha* Cronbacha i współczynnik Kuder-Richardsona. Współczynnik *alpha* Cronbacha przyjmuje wartości od: 0 do 1. Na wewnętrzną zgodność (rzetelność) skali wskazują wartości *alpha* większe od 0.6.

Z kolei Jerzy Brzeziński (1984: 181–185) mówi o czterech metodach estymacji rzetelności testów. Są to:

1. Estymacja wiarygodności (dwukrotne badanie tym samym testem raz po razie).

2. Estymacja konsystencji, a więc:

a) dwukrotne badanie równoległymi formami testu grupy osób raz po razie (tzw. równoważność międzytestowa),

b) obliczanie korelacji między równoległymi połówkami testu (tzw. równoważność międzypołówkowa) oraz

c) analiza związków między poszczególnymi zadaniami i całym testem (tzw. zgodność wewnętrzna).

3. Estymacja stabilności, a więc:

a) dwukrotne badanie tym samym testem grupy osób po pewnym odstępie czasowym (tzw. stabilność bezwzględna),

b) dwukrotne badanie dwoma równoległymi formami testu po pewnym odstępie czasowym (tzw. stabilność względna).

4. Estymacja konsensusyjności (porównanie zgodności dwóch sędziów kompetentnych oceniających odpowiedzi testowe).

Zdaniem tego autora analiza związków między poszczególnymi zadaniami i całym testem (tzw. zgodność wewnętrzna: metoda 2.c.) daje szybkie i dokładne oszacowanie rzetelności. Podstawową zaletą tej metody jest jednokrotne testowanie badanej grupy. Podstawowymi technikami analizy zgodności wewnętrznej są: estymacja na podstawie analizy wariancji, estymacja według wzoru Kuder-Richardsona i estymacja na podstawie średniej korelacji pozycji testu z jego ogólnym wynikiem. Z kolei techniki wykorzystujące dwukrotne testowanie (w szczególności metoda 3a) nie spełniają warunków niezależności pomiarów i obciążone

są między innymi wadami: pamięci, uczenia się, zmianą warunków testowania i dowolnością ustalania przedziału czasowego.

Mówiąc o trafności i rzetelności pomiaru, nie można pominąć problemu relacji między tymi pojęciami. Ogólnie biorąc, rzetelność oznacza otrzymywanie spójnych wyników za pomocą tego samego narzędzia, trafność zaś oznacza uzyskiwanie wyników, które są adekwatnym odbiciem mierzonego pojęcia. Trafność testu wiąże się z jego rzetelnością. Odwołując się do porównania z tarczą strzelniczą, można powiedzieć, że rzetelność oznacza „mały rozrzut” niezależnie od miejsca na tarczy (rzetelność jest bowiem funkcją spójności). Trafność to funkcja rozmieszczenia strzałów wokół celu. Brak rzetelności oznacza losowy rozrzut trafień wokół celu. Brak trafności oznacza trafienia systematyczne poza celem. Pomiar nierzetelny i nietrafny jest mało użyteczny (Babbie 2004). Rzeczą naturalną jest dążenie, aby wynik był rzetelny i trafny. W rzeczywistości jednak pomiar rzetelny nie musi być trafny (istnieje możliwość, że jest on obciążony błędem systematycznym). Pomiar trafny jest natomiast rzetelny, gdyż jego obciążenie „zanieczyszczeniami” przejściowymi (a także, z definicji, stałymi, systematycznymi) jest niewielkie. Rzetelność nie daje gwarancji, że pomiar dotyczy tego, co chcemy mierzyć. Tym samym rzetelność pomiaru jest kwestią fundamentalną dla każdego badania. Innymi słowy, jeśli test nie jest rzetelny, to nie można oczekiwać wysokiej jego trafności (w tym sensie rzetelność stanowi granicę trafności). Jednocześnie wysoka rzetelność nie gwarantuje trafności (Brzeziński 1984). Ponieważ jednak na ogół nie znamy z góry trafności zastosowanego narzędzia, a ponadto określamy ją na ogół w związku z konkretnym pomiarem (a zatem nie możemy go odnieść do narzędzia jako takiego), w każdym badaniu powinniśmy ocenić zarówno trafność, jak i rzetelność wyniku⁵⁹.

⁵⁹ Rozważając relacje między trafnością i rzetelnością, warto wspomnieć o propozycji Bogdana Wojciszke (2004), zakładającej określenie rzetelności (powtarzalności) wyniku, sprawdzenie teoretycznej trafności zastosowanych miar, maksymalizację trafności wewnętrznej oraz maksymalizację trafności zewnętrznej i teoretycznej. Chodzi o koncepcję Systematycznie Modyfikowanej Autoreplikacji (SMAR). Jest to strategia polegająca na powtarzaniu tego samego badania wykazującego jakąś prawidłowość przy celowym modyfikowaniu próby, metod pomiaru i zmiennych stanowiących przedmiot badania. Replikacja wyników badania, w szczególności wykonywanych przez niezwiązanych, niezależnych ze sobą badaczy jest niezwykle ważna, ponieważ zastrzeżenia wobec jednorazowo uzyskanego wyniku empirycz-

Mówiąc o trafności i rzetelności, należy też wspomnieć o pewnej szczególnej cesze badań ilościowych i jakościowych. Na problem ten zwraca uwagę Earl Babbie (2004). Przyjmuje on, że techniki ilościowe dostarczają danych bardziej rzetelnych, jakościowe zaś – bardziej trafnych. Innymi słowy, słabą stroną sondaży jest ich trafność, mocną zaś rzetelność. Sztuczność formy sondażu i standaryzacja pytań obciążają trafność wyniku: odpowiedzi na pytania są bowiem tylko przybliżonymi wskaźnikami tego, co badacz chciał zbadać. Z drugiej jednak strony, prezentując wszystkim badanym ten sam bodziec (standaryzowane pytanie), w znacznym stopniu eliminuje się nierzetelność pomiaru powodowaną przez samych badanych (której źródłem może być odmienność sformułowania pytań) oraz nierzetelność, której źródłem jest ankietier (chodzi o interpretację pewnych zachowań, wypowiedzi itp.). Z kolei badania jakościowe dostarczają danych, które charakteryzują się większą trafnością niż dane ilościowe i eksperymenty, są jednak one mniej rzetelne. Dane uzyskiwane w tych badaniach są subiektywne i dwie osoby obserwujące to samo mogą zupełnie inaczej interpretować wyniki⁶⁰.

nego mogą być więcej niż uzasadnione. Wynik ten może być bowiem związany z błędem próby czy zastosowanej metody. Dodatkowym problemem jest to, jaki wynik replikacji świadczy o tym, że można ją uznać za satysfakcjonującą. Zazwyczaj przyjmuje się kryterium istotności statystycznej związku, przy czym winny one uwzględniać wskaźniki siły efektu. Co więcej, nie wiadomo, o czym świadczy udana lub nieudana replikacja, czy jest efektem zmiany próby, zmiany metody czy też zmiany warunków badania itd. O znaczeniu powtarzania eksperymentów w różnych niezależnych laboratoriach i przez różnych, niezależnych badaczy pisał między innymi John Madge (1965). Twierdził on, że ponieważ żaden eksperyment nie przebiega idealnie w tych samych warunkach, aby uniknąć błędu pomiaru i osiągnąć jego trafność, celowe jest jego powtórzenie.

⁶⁰ Rozważając związki badań jakościowych z badaniami ilościowymi można stwierdzić, że badania jakościowe skupiają się na znaczeniu i wadze uzyskiwanych wyników oraz ich trafności, natomiast badania ilościowe koncentrują się raczej na spójności i powtarzalności uzyskiwanych danych. Stąd dążenie i troska badaczy ilościowych o standaryzację warunków wywiadu: bez tego ujednoczenia wyników badania nie można by do siebie dodawać, generalizować na szersze populacje ani mierzyć zmian zjawisk w czasie. Jeszcze inaczej problem rzetelności i trafności przedstawia się w badaniach typu *desk research*, w których metoda i sam fakt badania nie wpływają na badane zjawiska (na przykład analiza danych zastanych, statystycznych, analiza treści) [Babbie 2004]. Istotne są dwa problemy związane z trafnością w przypadku danych zastanych. Jest to logiczne wnioskowanie (ponieważ zastane dane nie zawsze obejmują dokładnie to, co chcemy badać, dlatego mogą

1. 2. 2. Weryfikacja i walidacja danych

Oceniając wartość odpowiedzi czy też jakość z perspektywy zewnętrznej, wskazałem, iż odnosi się ona nie tyle do procesu pomiaru, ile do jego rezultatu, a więc uzyskanego wyniku. W tym przypadku ocenie podlega bowiem nie sam proces badawczy czy jego elementy, lecz ostateczny rezultat, wynik. Jednym z kryteriów oceny uzyskanych informacji jest ich wiarygodność, pewność czy też dokładność (*accuracy*), która jest określana na podstawie porównania ich zgodności z informacjami zewnętrznymi. W literaturze metodologicznej na określenie tego sposobu oceny pojawiają się takie terminy jak: walidacja (*validation*) i weryfikacja (*verification*). Problem polega na tym, że różni autorzy terminów tych używają w różnym znaczeniu, pomijając to, że w języku angielskim *validation* może także oznaczać ustalanie trafności (*validity*) [Stein i Urbang, red. 1983]. Autorzy badania Denver Validity Study, Parry i Crossley (1950) swój artykuł zatytułowali *Validity of Responses to Survey Questions*, aczkolwiek ich rozważania nie mają nic wspólnego z psychologicznym ujęciem trafności, ponieważ wyniki badania odnoszą do danych dokumentalnych. Z kolei John Madge (1965: 301–311) pisze, że w większości przypadków zwalidować (*to valid*) pierwotny rezultat badania można zazwyczaj jedynie poprzez jego replikację (powtórzenie), przy zachowaniu warunku losowości przydziału jednostek do podgrup. Oznacza to, iż walidacja jest dla niego równoznaczna z oceną rzetelności wyniku. Arleck i Settle (1985: 423) twierdzą zaś, że termin walidacja jest powszechnie, acz niewłaściwie, stosowany na określenie weryfikacji odpowiedzi. Ta ostatnia – weryfikacja – to nic innego jak proces sprawdzania, czy z respondentem faktycznie został zrealizowany wywiad i czy wywiad został przeprowadzony w całości i poprawnie, w miejscu i czasie, kiedy powinien zostać zrealizowany. Wspomniany John Madge (1965) mówi jednak także o weryfikacji w innym sensie niż powyżej wspomniany.

nie być trafną reprezentacją tego, o czym chcemy wnioskować) oraz powtarzalność wniosków poprzez stosowanie zamiennych (równoważnych) wskaźników. Jeśli niezależne testy potwierdzają daną hipotezę, wówczas wzmacniają one trafność wnioskowania. Problem rzetelności sprowadza się do zagadnień jakości danych i dokładności opisywania przez nich zjawisk (w tym sposób zbierania danych i prowadzenia rejestrów). A zatem sposób zbierania danych statystycznych i ich zestawiania może mieć wpływ na ich rzetelność (por. Topolski 1984).

Terminu tego używa, mając na myśli porównanie informacji, którymi dysponujemy (tzn. uzyskanymi z wywiadu) z innymi, niekoniecznie dokumentalnymi. W bardzo zbliżonym sensie o weryfikacji mówi Jan Lutyński (1972b, 1975a, 1975b)⁶¹.

Ponieważ moim celem nie jest analiza stosowanej przez poszczególnych autorów terminologii, lecz prezentacja możliwości oceny uzyskiwanych informacji, przyjmuję, że pojęciem o szerszym zakresie jest weryfikacja. Jak wskażę poniżej, przy takim rozumieniu, walidacja jest jedną z możliwych form weryfikacji. Innymi słowy, o walidacji będę mówił w sensie nadanym temu terminowi przez Hugh Parry’ego i Helen Crossley (1950), a więc jako jednej z form weryfikacji – weryfikacji dokumentalnej.

1. 2. 2. 1. Weryfikacja danych

Zdaniem Johna Madge’a (1965: 264) jako aksjomat traktować należy stwierdzenie, że nie można oczekiwać od informatorów, że będą udzielali precyzyjnych i prawdziwych odpowiedzi na stawiane im pytania. Wiadomo, że pewne informacje są szczególnie narażone na błędy: na przykład kobiety mają tendencję do zaniżania swojego wieku, wszyscy zaś mają tendencję do zaokrąglania liczby lat do wielokrotności pięciu (20, 25, 20, 35 itd.). Podobnie jest w przypadku stanu cywilnego. Zadawanie pytań nie zapewnia więc uzyskania obiektywnych faktów. Zdaniem tego autora istnieją cztery sposoby minimalizacji błędów: (i) weryfikacja, (ii) wewnętrzne metody statystyczne, (iii) niestatystyczne oraz *iv* zapewnienie optymalnych warunków dla „prawdy” (tamże: 267–283).

Należy jednak zaznaczyć, iż trzy pierwsze „sposoby” wymienione przez Madge’a są metodami oceny uzyskanych informacji, podczas gdy ostatnia dąży do zapewnienia informacji wiarygodnych. Wśród warunków, które muszą być spełnione, aby uzyskać prawdziwe odpowiedzi od informatora, wskazuje on bowiem między innymi: przedmiot pytania (jeśli pytania dotyczą standardów moralnych, wartości lub bezpośrednio

⁶¹ Paul Lavrakas (2008: 937, 945) pisze z kolei, iż o weryfikacji można mówić, mając na myśli kontrolę terenową realizacji wywiadu lub też tak zwane *record-checks*, a więc sprawdzanie zgodności odpowiedzi respondenta z danymi zewnętrznymi (dokumentalnymi). Z kolei walidacja według Lavrakasa to sprawdzanie poprawności „wypełnienia” kwestionariusza przez ankietera (w przypadku PAPI chodzi przede wszystkim o przestrzeganie reguł przejścia oraz zapisu odpowiedzi na pytania otwarte).

celu wywiadu, spada rzetelność odpowiedzi), jego sformułowanie (pytania powinny być jak najbardziej konkretne), porządek pytań w kwestionariuszu i porządek alternatyw, typ interakcji (czy jest to ankieta czy wywiad), przekonania ankietera dotyczące informatora itd. Wszystko są to zatem warunki, które należy spełnić przed uzyskaniem informacji, a więc *ex ante*. Pozostałe to sposoby oceny *ex post*, a więc już zebranych informacji. Rzeczą oczywistą jest, iż skupię się właśnie na nich.

O weryfikacji Madge mówi w odniesieniu do wszelkiego rodzaju dokumentów utrwalonych (w tym dokumentów osobistych). Ze względu na fakt, iż w przypadku tego typu informacji można mówić o rozmaitych wypaczeniach związanych z osobą informatora (na przykład ich motywacja), wyróżnia on dwa typy oceny ich wartości: weryfikację wewnętrzną (*internal verification*) i zewnętrzną (*external verification*). Pierwsza z nich z nich sprowadza się do „wewnętrznej” analizy zebranych materiałów, druga zaś zakłada odwołanie się do materiałów zewnętrznych (tamże: 84–85, 267–268).

Możliwości zastosowania na szerszą skalę weryfikacji wewnętrznej są ograniczone, co nie oznacza ich braku w ogóle. W ramach weryfikacji wewnętrznej wskazuje on dwie metody oceny materiałów: statystyczne oraz niestatystyczne. Najprostszą statystyczną metodą weryfikacji wewnętrznej jest porównanie odpowiedzi tych samych osób, w tym samym wywiadzie na podobne pytania i ocena ich zgodności. Mogą to być na przykład pytania zależne lub pytania wzajemnie kontrolujące się, takie jak: zarobki respondenta, jego dochody i dochody gospodarstwa domowego. Niestatystyczne metody weryfikacji wewnętrznej polegają z kolei na szacowaniu, na ile coś jest prawdopodobne. Ocenę wiarygodności formułuje się w tym przypadku poprzez ocenę, na ile pytanie jest takie, że informator jest w stanie (ma możliwość) i chce na nie odpowiedzieć zgodnie z prawdą. Możliwość (*ability*) mówienia prawdy zależy między innymi od osobistych kompetencji informatora (w danej dziedzinie, zakresie), jego wiedzy oraz zainteresowania przedmiotem pytania. Z kolei chęć (*willingness*) udzielenia informacji zgodnych z prawdą zależy od poziomu zaangażowania informatora w problem (zawsze chce się przedstawić w dobrym świetle), tendencji do dramatyzowania, stereotypizacji i przestrzegania pewnych konwencji (na przykład tematy tabu). Madge (tamże: 84–85) przywołuje tutaj pracę Louisa Gottschalka (1945: 43–44), który sformułował pięć warunków, które muszą być spełnione, aby można było uznać dokument za wiarygodny. Są one następujące:

– jeśli przedmiot stwierdzenia jest obojętny dla badanego, to wydaje się ono nieobarczone błędem. Jednocześnie brak zaangażowania informatora może negatywnie oddziaływać na postrzeganie i przypominanie sobie pewnych faktów;

– jeśli stwierdzenie stoi w sprzeczności z poglądami informatora lub jego interesami, to wydaje się ono wiarygodne;

– jeśli stwierdzenie dotyczy faktów, spraw powszechnie znanych i opinia badanego jest z nimi zgodna, to wydaje się ono nieobarczone błędem (jest to forma pośredniej konfirmacji);

– jeśli jedna część stwierdzenia jest zupełnie prawdopodobna, a jednocześnie druga absolutnie przypadkowa, to wydaje się ono być wiarygodne (na przykład stwierdzenie: „kawę A&B można kupić w każdym sklepie za cenę 1,50 zł”. To, że można kawę A&B kupić w każdym sklepie jest pewne, natomiast nie jest pewne, iż można ją kupić w każdym sklepie w tej cenie);

– jeśli stwierdzenie badanego nie jest sprzeczne z oczekiwaniami badacza opartymi na wcześniejszej jego wiedzy dotyczącej badanego i jego opinii, to wydaje się ono wiarygodne⁶².

Koncepcja ta jest zbieżna ze stosowaną w naukach historycznych, w której przy ocenie dokumentów używa się tak zwanej krytyki źródeł (Topolski 1984: 357–376). Istnieje krytyka zewnętrzną i wewnętrzną, przy czym podział jest umowny. Głównym celem krytyki (zarówno wewnętrznej, jak i zewnętrznej) jest uzyskanie odpowiedzi na pytanie, czy informacje dostarczane przez źródło są zgodne z rzeczywistością (chodzi o wiarygodność informacji). Pierwszym jej etapem jest jednak badanie autentyczności źródła (krytyka zewnętrzna), a dopiero następnym badanie wiarygodności informatora (krytyka wewnętrzna). Dla określenia wiarygodności informatora konieczne jest zatem określenie autentyczności źródła (przy czym w niektórych przypadkach w celu określenia autentyczności konieczna jest także ocena wiarygodności informatora)⁶³. Określając wiarygodność informatora, bierzemy pod uwagę przede

⁶² W przypadku trzeciego warunku sformułowanego przez Gottschalka (1945) można mieć pewne zastrzeżenia, ponieważ wydaje się, iż nie uwzględnia on faktu, a właściwie jest sprzeczny z tezą o udzielaniu tak zwanych społecznie akceptowanych czy pożądanym odpowiedzi.

⁶³ Dla historyka autentyczność polega na ustaleniu daty i miejsca powstania źródła. Mówiąc o autentyczności, Topolski wyróżnia cztery jej rodzaje, przy czym socjologa interesuje wyłącznie tak zwana autentyczność właściwa, a więc, czy źródło jest

wszystkim to, czy mógł i chciał, świadomie dążył do przekazania prawdziwej informacji (w tym przypadku bierzemy pod uwagę jedynie cele kierujące informatorem). Wiarygodność informatora Giedymin (1964: 105–109) proponuje sprawdzać w dwojaki sposób: na podstawie wartości uzyskiwanych od niego informacji (jeśli większość podawanych przez niego informacji jest prawdziwa, to można uznać, że informator jest wiarygodny) i na podstawie wiedzy o nim samym (przede wszystkim, czy chciał podać prawdziwe informacje).

Drugi sposób oceny wartości dokumentów – weryfikacja zewnętrzna – zakłada odwołanie się do materiałów zewnętrznych. Mówiąc o weryfikacji zewnętrznej, Madge ma na myśli stopień zgodności uzyskanych informacji z informacjami pochodzącymi z innych źródeł. W wielu przypadkach istnieje bowiem więcej niż jedno źródło informacji na dany temat. Jest rzeczą oczywistą, iż należy wybrać to, które jest najlepsze, najbardziej wiarygodne. Przykładowo, informacje na temat zarobków z listy płac są bardziej wiarygodne niż deklaracje respondentów na ten temat. Niestety w większości przypadków możliwość szybkiego uzyskania określonych informacji z innych źródeł niż wywiad jest bardzo trudna. Porównanie informacji uzyskanych w wywiadzie z informacjami pochodzącymi z innych źródeł (raportów, statystyk itp., ale także z obserwacji zachowań), pozwala jednak na ocenę tych pierwszych⁶⁴. Jest to, zdaniem Madge'a, jedyny satysfakcjonujący sposób oceny dokumentów. Co więcej, wyniki weryfikacji zewnętrznej pozwalają nie tylko na ustalenie stopnia „prawdziwości” informacji, ale także wnioskowanie o wiarygodności informatora w innych sprawach niż przedmiot weryfikacji.

W polskiej literaturze termin „weryfikacja” pojawia się w pracach Jana Lutyńskiego (1972b, 1975a, 1975b). Autor ten rozważa problem wiarygodności informacji uzyskiwanych w badaniach kwestionariuszowych w związku z: (i) ustaleniem rodzajów i zakresu błędnych informacji; (ii) oceną wpływu błędów na wartość wniosków; (iii) wykazaniem

sfalszowane czy też nie (w całości autentyczne, w części nieautentyczne i w całości nieautentyczne). Wydaje się, iż chcąc mówić o autentyczności źródła w przypadku materiałów wywołanych (na przykład w badaniach socjologicznych), należy także uwzględnić problem „autentyczności informatora” albo inaczej „właściwego informatora”, a więc problem autorstwa źródeł.

⁶⁴ Porównanie informacji uzyskanych w wywiadzie z informacjami pochodzącymi z innych źródeł, w tym z raportów, statystyk itp., jest tym, co za Parrym i Crossley (1950) proponują nazywać walidacją.

źródeł błędów i oceną ich znaczenia; (iv) sformułowaniem wniosków i zaleceń dla badań oraz wzoru analizy ewaluacyjno-metodologicznej badań.

Podobnie jak John Madge (1965) mówi on bardzo ogólnie, iż weryfikacja to analizy zmierzające do sformułowania jakiegokolwiek oceny wiarygodności („prawdziwości”) wyników liczbowych⁶⁵. Podobnie jak Madge wyróżnia weryfikację wewnętrzną i zewnętrzną. Weryfikacja wewnętrzna opiera się według niego na założeniu, że aby rozstrzygnąć, czy dana odpowiedź jest „prawdziwa”, należy stwierdzić, czy w trakcie odpowiadania na pytanie zaszły w respondencie pożądane przez badacza procesy psychiczne (tzw. procesy konieczne) czy też nie. Informacje o procesach koniecznych można uzyskać, między innymi, na podstawie obserwacji zachowań werbalnych i niewerbalnych respondenta oraz całej sytuacji badania, a także analizy odpowiedzi respondenta na pytania mające na celu, w drodze introspekcji, ustalenie przebiegu tych procesów⁶⁶. Zastosowanie wywiadu o wywiadzie w przypadku badań masowych jest jednak praktycznie niemożliwe. Stąd, chcąc dokonać oceny wiarygodności informacji uzyskiwanych w badaniach surveyowych, musimy się odwołać do procedur weryfikacji zewnętrznej.

⁶⁵ Według Lutyńskiego wiarygodne (prawdziwe) informacje to takie, które pozwalają na zaliczenie odpowiedzi respondenta do kategorii badanych zjawisk, do której faktycznie one należą. Na określenie tego typu odpowiedzi Lutyński (1972a) używa terminu „odpowiedź trafna”.

⁶⁶ Chodzi tutaj o zastosowanie introspekcji pytaniowej w wywiadzie o wywiadzie (Przybyłowska 1975). Wyniki jej zastosowania przedstawili między innymi Pilichowski i Rostocki (1978). Autorzy ci sprawdzali wartość odpowiedzi z wywiadu zasadniczego poprzez wywiad weryfikacyjny, za pomocą którego starano się odtworzyć procesy intelektualne i inne zjawiska psychiczne, składające się na wewnętrzną reakcję respondenta na zadane pytanie kwestionariuszowe. Powtórny wywiad weryfikacyjny łączył cechy wywiadu o wywiadzie i pogłębionego wywiadu weryfikacyjnego. Z jednej strony miał on bowiem dostarczyć informacji o procesach psychicznych prowadzących do udzielenia odpowiedzi (w szczególności rozumienia pytania), z drugiej zaś – starano się uzyskać tę samą informację, o którą chodziło w kwestionariuszu wywiadu zasadniczego, ale przy zastosowaniu różnych zabiegów ułatwiających (wyjaśnianie pytań, rozkładanie na prostsze elementy itp.). W wyniku analizy ponad połowę odpowiedzi wszystkich respondentów na wszystkie pytania uznano za trafną (przynoszące prawdziwe informacje), jedną czwartą – za nietrafną i jedną piątą – za trudną do oceny. Aczkolwiek do tego typu metody można mieć zastrzeżenia, to jak stwierdzają autorzy, umożliwia ona liczbową ocenę trafności odpowiedzi.

W drugim typie analiz weryfikacyjnych, weryfikacji zewnętrznej, oceny uzyskanych danych dokonuje się na podstawie zestawienia ich z danymi zewnętrznymi, zdobytymi w odmienny sposób. Owe dane „zewnętrzne”, stanowiące kryteria oceny, mogą być danymi „obiektywnymi”, tak zwanymi pewnymi (na przykład dane dokumentalne), ale równie dobrze mogą to być dane uzyskane za pomocą innej procedury (na przykład innej techniki badania). W pierwszym przypadku Lutyński mówi o weryfikacji zewnętrznej z zastosowaniem pewnego sposobu otrzymania danych (w przyjętym przeze mnie rozumieniu jest to walidacja danych). Sposób ten odnosi się do poszczególnych respondentów, nie zaś do informacji w całym badaniu. W wyniku weryfikacji zewnętrznej można ustalić zakres błędu brutto, to znaczy całkowitą liczbę błędnych informacji, zakres błędu netto, to jest procentowych różnic między rozkładami zapisanych odpowiedzi przed i po weryfikacji, a także określić, które informacje (zapisane) są fałszywe.

Gdy uzyskane przez nas dane porównujemy z danymi uzyskanymi za pomocą innej procedury (na przykład innej techniki badania) mówimy o weryfikacji zewnętrznej, opartej na porównywaniu wyników niepewnych (Lutyński 1975a)⁶⁷. W tym ostatnim przypadku zakłada się, że porównywane informacje uzyskuje się dwoma różnymi sposobami, o których nie sposób rozstrzygnąć, który jest „lepszy”. Dodatkowo oba sposoby zbierania informacji są niezależne od siebie, a zatem ewentualne błędy pojawiające się w obu przypadkach są innego rodzaju i nie prowadzą do powstania błędu systematycznego. Co więcej, należy przyjąć, iż zbiorowości, z których każda badana jest tylko w jeden określony sposób, są ze sobą porównywalne oraz że oceniane informacje są analogiczne pod względem treści, to znaczy stanowią rozstrzygnięcie tych samych pytań. Przykładowo, procedura polegająca na porównaniu rezultatów uzyskanych od dwóch porównywalnych zbiorowości, dwiema różnymi metodami, przy czym jedną z tych metod uważa się za bardziej pewną, pozwoliłaby na określenie rozmiaru błędów, które popełnia się, wybierając sposób mniej pewny. Wymaga to jednak wskazania, która z dróg jest owym „mniej pewnym” sposobem zbierania informacji (wywiad czy ankieta), a tego, w sposób jednoznaczny, z całą pewnością uczynić nie

⁶⁷ Porównanie wyników (rozkładów odpowiedzi na poszczególne pytania) uzyskanych różnymi technikami jest niekiedy określane jako analiza zgodności (*concurrent analysis*) [Sjoberg i Nett 1964].

można. Wyniki prowadzonej w powyższy sposób analizy mogą być dwójakiego rodzaju: możemy mieć do czynienia z rozbieżnością bądź zgodnością tendencji. Występowanie rozbieżności w odpowiedziach świadczy o tym, iż jeden z wyników jest mniej wiarygodny niż drugi. Jednakże ze względu na to, że trudno rozstrzygnąć, który ze sposobów zbierania informacji jest „lepszy”, nie można także stwierdzić, które z rezultatów są bliższe „prawdy”. W związku z tym można jedynie wskazać kierunki i rozmiary rozbieżności. W drugim przypadku, gdy występuje zbieżność tendencji, należy przyjąć, iż żaden z wyników nie jest obciążony błędem lub oba są obciążone w tym samym stopniu. To z kolei jest podstawą do stwierdzenia, iż wiarygodność informacji uzyskanych tymi dwoma sposobami jest zbliżona (por. Lutyński 1975a)⁶⁸.

Tego typu formę weryfikacji zewnętrznej zastosowała w swoich badaniach Małgorzata Stefanowska (1988). Ich celem było określenie rozmiarów i kierunków błędów, jakie występują w odpowiedziach respondentów w wywiadzie kwestionariuszowym na pytania o konsumpcję kulturalną (tzn. czytanie książek, czasopism, bytność w teatrze, kinie, filharmonii itp.). Aby osiągnąć ten cel, należało uzyskać informacje o konsumpcji kulturalnej z innych źródeł niż wywiad kwestionariuszowy. Autorka założyła, iż informacje te muszą być pewniejsze i dokładniejsze niż relacje respondentów w wywiadzie. W tym celu dokonywano, w oparciu o wypracowaną metodykę, obserwacji aktów konsumpcji kulturalnej respondentów. Obserwację traktuje ona jako pewniejsze źródło informacji niż wywiad kwestionariuszowy⁶⁹. Była to ukryta obserwacja uczestnicząca prowadzona przez członka gospodarstwa domowego badanego. Dane te można uznać za bardziej pewne niż dane z wywiadu. Walidacja polegała zatem na ocenie danych mniej pewnych (z wywiadu)

⁶⁸ Procedura ta pozwala zatem na ocenę rezultatów w skali całej zbiorowości, nie zaś na poziomie informacji jednostkowych – odpowiedzi poszczególnych respondentów.

⁶⁹ Obecnie podstawę walidacji pomiaru konwencjonalnego (tzn. w wywiadzie) mogą stanowić techniki pomiaru elektronicznego. W przypadku badania audytorium telewizji może to być telemetria (Chavenon, Khoury 2002; Harvey, Jarvis, Booth 2002; Olszewski 2004), w przypadku badania audytorium radia może być to radiometria (chodzi o tzw. zegarek Steinmanna) [Steinmann, Dahler, Hackbart 2002], bądź przy pomiarze jakichkolwiek mediów elektronicznych (radio, TV, sygnały z Internetu itp.) – tzw. *portable people meter* (PPM) [Patchen, Webb 2002]. Techniki te także mają swoje ograniczenia.

na podstawie danych bardziej pewnych (z obserwacji). Po miesięcznym okresie obserwacji realizowany był wywiad kwestionariuszowy dotyczący konsumpcji kulturalnej w okresie obserwowanym, a następnie, po upływie kolejnego tygodnia, wywiad dotyczący norm konsumpcji. Obserwatorami byli studenci, członkowie rodzin osób obserwowanych. Dokonując wyboru badanych, kierowano się kryteriami płci oraz wykonywanej pracy (praca fizyczna i umysłowa). Pole obserwacji stanowiło: czytelnictwo książek, czytelnictwo tygodników, uczęszczanie do kina, uczęszczanie do teatru i uczęszczanie do filharmonii. Celem była analiza rozbieżności (ich rozmiar, kierunek i charakter) między wynikami obserwacji a danymi z wywiadu jako zmiennych zależnych od uznawanych norm i stopnia ich realizacji („poczucie „nieadekwatności kulturalnej”), ale także pamięci respondentów. Jak można przypuszczać, wyniki są mało zadowalające dla badacza chcącego ustalić na podstawie wywiadu kwestionariuszowego rzeczywisty stan czytelnictwa książek pod względem zarówno liczbowym, jak i tytułowym. Autorka rozbieżności tłumaczy poczuciem „nieadekwatności kulturalnej”, która powoduje powstanie błędu systematycznego polegającego na zawyżaniu wskaźników uczestnictwa w kulturze. Innym czynnikiem wpływającym na wielkość tego błędu jest z pewnością struktura czasu w ludzkiej pamięci (por. Auriat 1996).

Ponieważ weryfikacja dokumentalna, a więc w moim rozumieniu walidacja, traktowana jest przez Lutyńskiego jako najpewniejszy sposób oceny uzyskanych materiałów. Poniżej nieco szerzej omówię to zagadnienie.

1. 2. 2. 2. Walidacja danych

Idea, aby w ocenie uzyskanych danych odwołać się do istniejących danych dokumentalnych, ma jedno podstawowe ograniczenie: możliwości zastosowania tej procedury zależą od przedmiotu pytań. W przypadku niektórych pytań (na przykład cech metryczkowych) możliwości takie istnieją, ponieważ istnieją dokumenty zawierające tego typu informacje. Przyjmuje się, iż dane takie są w pełni wiarygodne, ponieważ pochodzą z oficjalnych rejestrów, takich instytucji jak urzędy stanu cywilnego, kancelarie kościelne, działy spraw osobowych w zakładach pracy, ośrodki pomocy społecznej, wydziały komunikacji itd. Podobnie w przypadku deklaracji dotyczących niektórych rodzajów zachowań możliwa jest ich

ocena na podstawie istniejących dokumentów. Poniżej przedstawię wyniki kilku badań walidacyjnych.

Jedną z wcześniejszych prób określenia dokładności pomiaru pewnych zmiennych faktualnych było „Denver Validity Survey” – badanie przeprowadzone w 1949 roku przez Opinion Research Center Uniwersytetu Denver. Dotyczyło ono między innymi udziału respondentów w wyborach municypalnych, posiadania przez nich karty bibliotecznej, prawa jazdy oraz telefonu w miejscu zamieszkania (Parry, Crossley 1950: 61–80). Uzyskane deklaracje respondentów były następnie porównywane z danymi dokumentalnymi na poziomie jednostkowym: stąd też mówiono o trafności jednostkowej odpowiedzi (*individual validity*)⁷⁰. W pytaniach dotyczących rejestracji oraz głosowania w sześciu różnych wyborach (prezydenckich, do Kongresu itp.) tylko 33% respondentów udzieliło odpowiedzi zgodnej z danymi dokumentalnymi, 42% zawyżyło swój udział, a 4% zaniżyło. Deklaracje zawyżające udział w poszczególnych wyborach wahały się od 13% do 28% respondentów. Informacje o uiszczaniu składek na cele lokalne okazały się zawyżone w 34%, o posiadaniu karty bibliotecznej w 9%, posiadaniu prawa jazdy w 10%, samochodu w 3%, domu w 3%, telefonu w 1%. W przypadku trzech ostatnich pytań odsetek osób „zaniżających” był zbliżony do „zawyżających”, w pozostałych zaś wyraźnie niższy. Analiza zależności od podstawowych cech społecznych respondentów wykazała, że kobiety nieco częściej udzielały odpowiedzi zgodnych ze stanem faktycznym niż mężczyźni, ludzie młodzi – częściej niż starsi (z wyjątkiem posiadania prawa jazdy i składek na potrzeby lokalne), zaś wpływ statusu nie był jednoznaczny.

Przedmiotem badań walidacyjnych były także zachowania zdrowotne. Chodzi tutaj przede wszystkim o relacjonowane przez badanych pobyty w szpitalu oraz wizyty lekarskich w ciągu ostatniego roku. Laurent, Cannell i Marquis (1972) w swoich badaniach dotyczących tych właśnie problemów porównywali relacje 1491 respondentów dotyczące poby-

⁷⁰ Badanie „Denver Validity Survey” stało się już klasyczne: do niego i uzyskanych rezultatów odwoływano się wielokrotnie. Ewenementem jest fakt, iż jeszcze dwadzieścia lat po przeprowadzeniu badania ukazał się artykuł bazujący na jego wynikach (Cahalan 1968–1969). Wskazany powyżej schemat oceny rezultatów był i jest nadal wielokrotnie wykorzystywany w badaniach socjologicznych i demograficznych (zob. m.in. Auriat 1993; Clausen 1968–1969; Miller, Groves 1985; Poulain, Riandey, Firdion 1991; Weaver 2000; Weiss 1968–1969).

tu w szpitalu z zapisami w szpitalnych kartotekach. Stwierdzono 233 przypadki zaniżeń i tylko 45 przypadków zawyżeń na ogólną sumę 1833 zapisów w kartotekach szpitalnych dotyczących respondentów. Jednocześnie zaobserwowano wyraźną prawidłowość polegającą na tym, że z wydłużaniem się czasu między pobytem w szpitalu a wywiadem liczba zaniżeń wzrasta: od 2% dla okresu pięciu tygodni poprzedzających wywiad, do 22% dla okresu 40 tygodni i 43% dla okresu 52 tygodni. Ci sami autorzy porównywali relacje respondentów dotyczące wizyt u lekarza z rejestracją prowadzoną przez lekarzy. Zaobserwowano podobną prawidłowość jak w przypadku relacji dotyczących pobytu w szpitalu: z upływem czasu liczba zaniżeń wzrasta od 15% dla jednego tygodnia poprzedzającego wywiad, do 30% dla dwóch tygodni.

Z kolei Cannell (1977) relacjonuje wyniki badań, w których uzyskane podczas wywiadu informacje o pobycie w szpitalu i chorobach porównywano z danymi dokumentalnymi. Stwierdzono, że respondenci cierpiący na poważne choroby częściej nie przyznawali się do pobytu w szpitalu niż osoby o schorzeniach lekkich (różnica wynosiła 11 punktów procentowych) i podawali nieprawdziwe diagnozy. Najczęściej nie przyznawano się do zaburzeń psychicznych i zaburzeń osobowości (zaniżenie wśród 67% respondentów), schorzeń systemu nerwowego (45%), a w przypadku kobiet – do schorzeń piersi oraz narządów rodnych (44%). Natomiast dane zawyżone uzyskano dla schorzeń łagodnych i nieokreślonych nowotworów (zawyżenia wśród 51% respondentów) oraz infekcji i chorób pasożytniczych (45%). Autor wyniki te interpretuje, odwołując się do norm grupowych oraz potrzeby przedstawiania pozytywnego obrazu własnej osoby.

W Polsce badania walidacyjne (określane za Lutyńskim jako weryfikacyjne) dotyczące wizyt respondentów w rejonowych poradniach zdrowia prowadziła Krystyna Lutyńska i Neonila Szeszenia (1972). Autorki porównywały relacje 520 respondentów dotyczące wizyt w rejonowych poradniach zdrowia z zapisami w kartotekach poradni. Dla okresu dwóch tygodni stwierdzono 46 przypadków rozbieżności (w większości były to zawyżenia). Zaobserwowano też prawidłowość zbliżoną do odnotowanej przez Laurenta, Cannell i Marquis (1972) w przypadku relacji dotyczących pobytu w szpitalu i wizyt u lekarzy: z upływem czasu rozbieżności wzrastały od 7% dla dwóch tygodni poprzedzających wywiad, do 16,1% dla dwóch tygodni do roku i 29,8% rozbieżności dla jednego roku i powyżej.

Inną informacją, która często jest przedmiotem walidacji, są zarobki i dochody. Krystyna Lutyńska (1975a) porównywała deklaracje robotników o przeciętnej wysokości zarobków w ciągu roku z ich faktyczną przeciętną wysokością, ustaloną na podstawie dokumentów księgowych. W rezultacie porównań okazało się, że 63% respondentów zaniżyło wysokość przeciętnych zarobków, a 16,6% zawyżyło. Za zgodne z dokumentami (w granicach błędu) uznano 20,4% deklaracji. Ogółem 76,7% deklaracji różniło się o 10% od rzeczywistych zarobków, w tym 30,2% o ponad 20%. Również Jerzy Koniarek (1975) porównywał deklaracje robotników o wysokości zarobków netto z faktyczną ich wysokością ustaloną na podstawie dokumentów księgowych. W rezultacie porównań okazało się, że 75 respondentów (spośród 114) zawyżyło wysokość swoich zarobków, a 38 zaniżyło. Tylko w jednym przypadku stwierdzono całkowitą zgodność. W 32 przypadkach rozbieżności (zawyżenia i zaniżenia) były znaczne i przekraczały 20% otrzymywanych faktycznie sum.

Mówiąc o walidacji trudno nie wspomnieć o badaniach prowadzonych przez Pawła Daniłowicza i Pawła B. Sztabińskiego (1986). W badaniach tych, nazwanych weryfikacyjnymi (niezgodnie z tradycją tego typu badań), podjęli oni próbę wykrycia błędów zawartych w odpowiedziach respondentów na tak zwane pytania metryczkowe poprzez porównanie danych pochodzących z wywiadu z danymi dokumentalnymi. Błędne informacje o własnym wykształceniu podała ponad jedna czwarta respondentów. Zdecydowaną ich większość stanowiły zawyżenia (blisko 70% ogółu błędów), których wielkość wynosiła od jednego poziomu (66%) do trzech (5%). Współczynnik korelacji między danymi z wywiadu a danymi dokumentalnymi był jednak wysoki: 0,81. W informacjach o wieku i zawodzie respondenta błędnych odpowiedzi było niewiele lub też były one nieznaczące. Obarczone znacznym błędem były natomiast, zgodnie z oczekiwaniami, informacje o zarobkach respondenta i dochodach jego gospodarstwa domowego. Jeśli przyjmie się założenie, że może się on pomylić o kwotę równą 4% zarobków (dochodów), odsetek błędnych odpowiedzi wyniósł ponad 80. Zdecydowaną większość spośród nich stanowiły zaniżenia deklarowanych kwot (około dwóch trzecich ogółu błędów), a ich wysokość wyraźnie przekraczała wysokość zawyżeń. Pomimo tych błędów struktura zarobków i dochodów ustalonych w wywiadzie pokrywała się z faktyczną: wartości współczynników korelacji wyniosły odpowiednio 0,88 i 0,90. W omawianych badaniach sprawdzeniu poddano również deklaracje respondentów dotyczące powierzch-

ni zajmowanego mieszkania oraz posiadania pojazdu mechanicznego (samochodu, motocykla i motoroweru). Jeżeli chodzi o pierwszą z tych informacji, błędy popełniło blisko 36% respondentów, nieco częściej zaniżając powierzchnię mieszkania, niż ją zawyżając. Zaskakujący jest udział bardzo dużych błędów: aż 40% stanowią pomyłki przekraczające 10 m². W rezultacie współczynnik korelacji między danymi z wywiadu a danymi dokumentalnymi w porównaniu z omawianymi wcześniej charakterystykami jest stosunkowo niski: 0,68. Błąd w informacji o posiadaniu pojazdów wystąpił u 12% respondentów, co przy ówczesnej stopie motoryzacji (badania przeprowadzono w roku 1976) wydaje się odsetkiem znaczącym. Deklaracji o posiadaniu pojazdu przez osoby go nieposiadające było ponaddwukrotnie więcej niż sytuacji odwrotnych, to znaczy zaniżeń stanu posiadania. Ponadto, o ile respondenci z równą częstotliwością nie przyznawali się do posiadania samochodu, motocykla i motoroweru, o tyle „zawyżenia” dotyczyły znacznie częściej dwóch pierwszych pojazdów, które są droższe i tym samym bardziej prestiżowe. W omawianych badaniach analizowano również związek popełnianych przez respondentów błędów z ich cechami społeczno-demograficznymi. Poza wspomnianą zależnością między rodzajem błędu w zarobkach i dochodach a ich wysokością nie wystąpiły silniejsze związki, pozwalające orzekać o podatności poszczególnych kategorii respondentów na udzielenie określonego rodzaju odpowiedzi.

Oceniając wiarygodność deklaracji respondentów na podstawie danych dokumentalnych, przyjmuje się, iż są one „pewne”. Warto może jednak zapytać, czy rzeczywiście tak jest. W bardzo wielu przypadkach dane te są bowiem niczym innym jak tylko deklaracjami respondentów, zapisanymi w danych „urzędowych” (Sułek 2002a).

Mówiąc o walidacji wyników badań, należy także wspomnieć jeszcze o innej możliwości „zewnętrznej” ich oceny, mianowicie walidacji na podstawie rzeczywistego zachowania badanych. Mam tu na myśli wyniki wyborów, które stanowią najlepszą podstawę oceny badań przedwyborczych. Bradburn i Sudman (1979) stwierdzili zawyżanie udziału w wyborach przez 39% respondentów, rejestracji wyborczej zaś przez 15%. W Polsce tego typu ocenę zgodności wyników sondaży z rzeczywistymi wynikami wyborów do Parlamentu Europejskiego 13 czerwca 2004 roku i wyborów parlamentarnych w październiku 2005 roku prowadziło Centrum im. Adama Smitha (2004, 2005). Jako kryterium oceny przyjęto sumę różnic pomiędzy wyrażonymi procentowo wynikami da-

nego sondażu a oficjalnymi wynikami osiągniętymi przez poszczególne komitety wyborcze. Im wskaźnik ten był niższy, tym większa była ocena dokładności danego sondażu. Pod uwagę wzięto wszystkie sondaże, których wyniki opublikowano najpóźniej w przeddzień wyborów.

Należy podkreślić, iż walidacja jako sposób oceny wyników był i jest nadal wielokrotnie wykorzystywany w badaniach socjologicznych i demograficznych (zob. m.in. Auriat 1993; Miller, Groves 1985; Poulain, Riandey, Firdion 1991; Pyy-Martikainen, Rendtel 2009; Weaver 2000; Weiss 1968: 69). Musimy jednak pamiętać nie tylko o ograniczeniach – ze względu na przedmiot – analiz walidacyjnych, ale także o problemach metodologicznych, związanych z łączeniem danych z wywiadu z danymi dokumentalnymi (Miller, Groves 1985).

1. 3. Implikacje praktyczne

Powyżej przedstawiłem różne sposoby oceny jakości danych. Niektóre z nich koncentrowały się na ocenie poszczególnych etapów i faz procesu badawczego, inne na jego elementach lub błędach w nich występujących (mam tutaj na myśli „perspektywę wewnętrzną”). Inne sposoby oceny jakości danych, które nazwałem „perspektywą zewnętrzną”, odwołują się do kryteriów trafności, rzetelności oraz koncepcji weryfikacji i walidacji danych.

Jakie są praktyczne możliwości stosowania każdej z przedstawionych procedur oceny jakości danych? Zacznijmy od praktycznych implikacji wpływających z przyjęcia perspektywy wewnętrznej, a więc analizy przebiegu procesu badawczego (chodzi o perspektywę Quality Control) oraz analizy jego poszczególnych elementów. Na ile podejdzie to stwarza możliwość praktycznego określenia wartości uzyskiwanych danych?

Wydaje się, iż zarówno ujęcie dynamiczne, jak i statyczne w praktyce nie umożliwia całościowej – syntetycznej – oceny wartości zebranych danych. Możliwa jest jedynie ocena „wycinkowa”, na przykład ze względu na zachowania ankietera. Przyjmując perspektywę QC, nie można bowiem określić wielkości potencjalnego błędu pomiaru na przykład w kategoriach rzetelności. Przykładem mogą być kolejne rundy Europejskiego Sondażu Społecznego, w których zastosowano wszystkie elementy QC. Niestety, oceniając rzetelność uzyskanych danych na pod-

stawie powtórnego pomiaru, nie była ona satysfakcjonująca. Oznacza to, iż „sumaryczna” ocena badania na podstawie analizy jego przebiegu nie jest wystarczająca. Jest jednak bardzo ważna z punktu widzenia możliwości podniesienia jakości zbieranych danych i minimalizacji błędów pomiaru. Stąd też winna stać się nieodzownym elementem każdego badania.

Czy pod tym względem lepsze jest ujęcie drugie (statyczne), w którym przedmiotem oceny są poszczególne elementy procesu badawczego? Raczej nie, ponieważ również ono budzi zastrzeżenia dotyczące możliwości praktycznego jego wykorzystania do oceny wartości odpowiedzi pod kątem jednej z podstawowych kategorii, mianowicie „trafności” (w rozumieniu Lutyńskiego). W tym przypadku należy bowiem dokonać oceny prawidłowości procesów psychicznych, które doprowadziły do sformułowania takiej czy innej odpowiedzi (o ile jest ona istotna). Problem ten w połowie lat 70. stanowił przedmiot zainteresowań teoretycznych i empirycznych zespołu badawczego kierowanego przez Jana Lutyńskiego na Uniwersytecie Łódzkim. Próbowano przykładowo opisać model wewnętrznej reakcji respondenta na pytanie (Tuchańska 1983), ale także docierać, na podstawie wywiadu o wywiadzie, do różnych empirycznych wzorów reakcji wewnętrznej respondenta na pytanie kwestionariuszowe (Przybyłowska 1975: 457–488). Jednakże, jak pisze Lutyńska (1975: 72), nie wypracowano sposobów, za pomocą których można by z pewnością stwierdzić, czy każda dana odpowiedź przynosi informację prawdziwą czy nieprawdziwą. Co najwyżej można sformułować zastrzeżenia co do jej „trafności”. Przyczyną braku „trafności” może być: wpływ osób trzecich, emocje respondenta, niezrozumienie pytania, stan psychofizyczny respondenta oraz inne czynniki (tamże: 61–64). Tym samym introspekcja pytaniowa nie przynosi oczekiwanych rezultatów. W przypadku drugiego kryterium oceny wartości odpowiedzi – oceny istotności – sprawa jest relatywnie prosta, ponieważ, jak pisze Lutyńska (1975: 58–60), odróżnienie odpowiedzi istotnych od nieistotnych wymaga jedynie analizy samego tekstu odpowiedzi w zestawieniu z poszukiwaną informacją.

Kolejna sprawa związana z możliwością praktycznego określenia wartości uzyskiwanych danych w ujęciu statycznym dotyczy poziomu analizy. W swoich rozważaniach zarówno Lutyńska, jak i Lutyński przyjmują perspektywę wywiadu traktowanego jako jednostkowy akt badawczy (pojedynczy wywiad), nie zaś jako proces badawczy, a więc

„zwielokrotnione” akty badawcze (tzn. wszystkie zrealizowane wywiady). Jednocześnie, kategorie istotności i trafności są kategoriami jednostkowymi, to znaczy odnoszą się do odpowiedzi danego respondenta na dane pytanie. Tym samym ocena wartości dotyczy nie wywiadu (jako całości, a tym bardziej nie wszystkich wywiadów), lecz jednego pytania, a więc elementu czy składnika owego jednostkowego aktu badawczego. Innymi słowy, przedmiotem oceny jest pojedyncze pytanie w jednym wywiadzie, a nie zbiorczy wynik w odniesieniu do danego pytania (chodzi o odpowiedzi wszystkich respondentów we wszystkich wywiadach), nie mówiąc już o rezultacie całego badania⁷¹. Z pewnością utrudnia to, aczkolwiek nie uniemożliwia, ilościową ocenę zebranego materiału. Lutyńska (1975: 74) pisze bowiem, że w standardowych badaniach można również obliczać liczbę odpowiedzi nieistotnych (tzn. ile odpowiedzi na każde pytanie jest istotnych, a ile nieistotnych). Jest to jednak możliwe tylko przy założeniu, że wszystkie odpowiedzi typu „nie wiem” oznaczają odpowiedź nieistotną⁷². Jednocześnie w tego typu badaniach przy aktualnej ich organizacji, poziomie i rodzaju wyszkolenia ankierów sprawdzenie wszystkich odpowiedzi respondentów na wszystkie pytania kwestionariusza pod kątem tego, czy budzą one zastrzeżenia co do trafności czy też nie – nie jest możliwe. Wydaje się więc, iż abstrahując nawet od czasochłonności, pracochłonności i w efekcie kosztów takiej procedury oceny, należy mieć świadomość, iż nigdy nie będzie ona pełna – w przypadku masowych badań praktycznie niemożliwa jest ocena „trafności” odpowiedzi, bowiem w badaniach tego typu nie uzyskuje się lub uzyskuje się zbyt mało informacji o reakcji respondenta na dane pytanie. Tym samym ujęcie to jest, jak się wydaje, mało przydatnym narzędziem ilościowej oceny zebranego materiału. Podobne stwierdzenie sformułować można pod adresem pozostałych procedur oceny *ex post* uzyskanego materiału. Dotyczy to przede wszystkim, o czym wspominałem, możliwości oceny wpływu warunków społecznych i sytuacyjnych wywiadu na wartość informacji (zob. Sztabiński F. 1996), ale także w pewnym zakresie również analizy efektu ankierskiego. Wynika to

⁷¹ Związane jest to z przyjęciem wspomnianego, wąskiego rozumienia narzędzia badawczego, zgodnie z którym jest nim pojedyncze pytanie kwestionariuszowe.

⁷² Próbę ilościowej oceny istotności odpowiedzi w badaniach kwestionariuszowych przedstawiłem w książce *Ankieta pocztowa i wywiad kwestionariuszowy* (Sztabiński F. 1997: 104–118).

choćby z tego, iż w standardowych badaniach praktycznie nigdy nie jest spełniony warunek losowego przydziału próby dla ankierów.

Praktyczne możliwości oceny wielkości błędu poszczególnych rodzajów nie są jednakowe. Największe są one w przypadku błędów związanych z próbą, w szczególności błędu losowania oraz braku odpowiedzi (zarówno na poziomie całego wywiadu, jak i poszczególnych pytań). Wielkość tego ostatniego błędu określa się praktycznie w każdym badaniu, obliczając choćby odsetek realizacji⁷³. Relatywnie duże możliwości oceny wielkości błędu istnieją również w przypadku błędów opracowania: kodowania i wprowadzania. Jak wspominałem wcześniej, błąd ten można nie tylko oszacować, ale także usunąć, powtarzając te czynności oraz uzgadniając wyniki. W przypadku błędów obserwacji możliwości te są niestety niewielkie: praktycznie jedynie w przypadku błędu ankiera możemy ilościowo określić jego rozmiar. Na przykład, analizy prowadzone przez Pawła Sztabińskiego (1997) pokazują, że w przypadku niektórych pytań błąd ten jest na tyle duży, iż cechy społeczne i opinie ankierów wyjaśniają większy odsetek zróżnicowania odpowiedzi respondenta niż cechy respondenta. W przypadku pozostałych błędów związanych z obserwacją, a więc błędów techniki, pytań i respondenta, możliwości ich szacowania liczbowego są raczej niewielkie⁷⁴. Sytuację tę dobrze charakteryzują Hans Hippler i Norbert Schwarz (1988: 102) pisząc, iż „Pomimo popularności badań surveyowych (...) nigdy nie została rozwinięta »teoria zadawania pytań«. W rezultacie współczesna me-

⁷³ Przyjmuje się, że jeśli odsetek realizacji osiąga poziom 70%, to praktycznie można mówić o braku „błędu braku odpowiedzi” (*ESS R. 5: Specification for participating countries*, 2011. Por. Stoop, Billiet, Koch, Fitzgerald 2010).

⁷⁴ Pomijam tutaj sytuacje „quasi-eksperymentalne”, w których ocena taka jest możliwa. Mam tutaj na myśli choćby badania, które zrealizowali Clark i Tiffit (1966). W badaniach tych respondenci odpowiadali na pytania dotyczące zachowań dewiacyjnych, a następnie uprzedzono ich, że zostaną poddani badaniu z wykorzystaniem wariografu (wykrywacza kłamstw). Przedtem mogą jednak w warunkach pełnej dyskrecji poprawić udzielone wcześniej odpowiedzi, tak aby były w pełni prawdziwe. Wszyscy respondenci dokonali zmiany przynajmniej jednej odpowiedzi, przy czym trzy czwarte zmian polegało na wskazaniu większej liczby zachowań dewiacyjnych, reszta zaś na jej zmniejszeniu. Po przebadaniu za pomocą wariografu, ujawniono liczne dodatkowe zaniżenia objętych badaniami zachowań. Z drugiej jednak strony rzeczą „oczywistą” jest elektroniczny pomiar audytorium telewizyjnego (Olszewski 2004), a coraz częściej mówi się także o elektronicznym pomiarze czytelnictwa prasy (Mattlin 2002)

Metodologia surveyowa charakteryzuje się rygorystyczną wiedzą w zakresie procedur związanych z próbami z jednej strony i zaskakującym brakiem wiedzy o sztuce (*sic!*) zadawania pytań z drugiej strony. Niestety, badania empiryczne wskazują, że błąd nie związany z próbą powoduje istotne ograniczenia użyteczności wyników badań surveyowych⁷⁵. Dobrą ilustrację powyższego stwierdzenia stanowią rozważania Pawła B. Sztabińskiego (2003) na temat wpływu badacza na wyniki badania. W przypadku badań surveyowych zawsze bowiem stajemy przed decyzjami, które na gruncie obecnej wiedzy metodologicznej są tak samo uprawnione, a które powodować mogą określony „efekt” i rzutować na uzyskany wynik badania. Chodzi tutaj choćby o wybór techniki badawczej (PAPI, CATI, CAPI, ale także ankieta pocztowa i roznoszona), sposób przygotowania kwestionariusza oraz sformułowania pytań (Dillman 1978; Heller i Roth 1994; Krosnick 1991; de Leeuw, Hox i Snijders 1995; Locander, Sudman i Bradburn 1976; Marx 1998, 2000; Marx i de Langhe 1997; Moser i Kalton 1971; Schuman i Presser 1981; Sztabiński F. 1997; Sztabiński P. 1999, 2001, 2003; Sztabiński i Sztabiński 1997 i inni). Jeśli chodzi o technikę badania, to istotne różnice w uzyskiwanych rezultatach zaobserwować można przede wszystkim między PAPI/CAPI i CATI oraz ankietami samodzielnie wypełnianymi przez respondenta: ankietą pocztową oraz roznoszoną przez ankieterów (Dillman 1978; Moser i Kalton 1971; Roberts, Jakle, Lynn 2007; Sztabiński F. 1997; Sztabiński P. 1999). Dotyczą one nie tyle odpowiedzi na pytania faktualne, co odpowiedzi na pytania o opinie (Locander, Sudman i Bradburn 1976) oraz odpowiedzi na pytania otwarte (w ankiecie uzyskuje się obszerniejsze odpowiedzi)⁷⁵. Podobnie jest ze sformułowaniem i językiem pytań: okazuje się, iż nawet „niewielkie różnice w sformułowaniu pytania mogą powodować znaczące skutki dla rozkładów odpowiedzi”, nawet powyżej 20 punktów procentowych (Schuman, Presser 1981: 275; Gremy 1995; Domański 2004). Podobnie jest w przypadku prezentowanych skal (Gremy 1995; Ostrom, 1988; Schwarz, Hippler, Deutsch i Strack 1985; Schwarz 1996, 1997; Sułek 1999; Sztabiński P. 2002) oraz zamieszczania, bądź nie, odpowiedzi „trudno powiedzieć” (Białecki i Sztabiński 1997; Domański 2000; Kros-

⁷⁵ W przypadku wywiadu kwestionariuszowego (PAPI) oraz ankiety pocztowej różnice w pytaniach o opinie sięgać mogą 18 punktów procentowych (Sztabiński F. 1997), natomiast w wywiadzie PAPI i CATI, w pytaniach o wiedzę, aż 30 (Sztabiński P. 1999).

nick i Fabrigar 1997; Schuman i Presser 1981). Co to oznacza w praktyce? Jedyną odpowiedzią jest stwierdzenie, że badacz jest „współautorem” wyników badania (Sztabiński P. 2003)⁷⁶.

Jakie są z kolei możliwości praktycznego wykorzystania podejścia „zewnętrznego”, a więc koncepcji trafności i rzetelności oraz weryfikacji i walidacji do oceny wartości uzyskiwanych informacji? Z pewnością duże, biorąc pod uwagę, że ocena ta może być dokonywana zarówno przed, jak i po badaniu, a więc *ex ante* i *ex post*. W przypadku oceny trafności *ex ante* mam tutaj na myśli przede wszystkim ocenę trafności fasadowej i treściowej, a więc trafności, którą Trochim nazywa translacyjną. *Ex post*, stosować możemy praktycznie wszystkie inne metody oceny trafności, łącznie z metodą MTMM, o ile przewidziano to w procedurze badania. Jeśli chodzi o rzetelność pomiaru, jest oczywiste, iż dokonuje się jej przede wszystkim *ex post*, po badaniu. Niemniej, rzetelność pomiaru ocenić można również *ex ante*, o ile planujemy realizację badania próbnego, poprzedzającego zasadnicze. Możemy wtedy ocenić rzetelność pomiaru w sensie jego homogeniczności lub równoważności. W praktyce jednak, przede wszystkim ze względu na koszty, procedury oceny rzetelności *ex ante* są wykorzystywane w niewielkim stopniu w badaniach socjologicznych i tym samym mają dość ograniczoną przydatność (Domański, Dukaczewska 1996).

Praktyczne możliwości weryfikacji i walidacji uzyskanych danych są jeszcze mniejsze. Jeśli myślimy o walidacji w rozumieniu Parry’ego i Crosly, to ma ona bardzo ograniczone możliwości, choćby ze względu na zakres oraz dostępność danych, a także koszty. Poza tym, jak walidować zmienne psychologiczne? Czy w ogóle jest to możliwe? Ten typ analiz ma sens jedynie w przypadku zmiennych, w przypadku których wartość „prawdziwa” istnieje, a więc z pewnością nie zmiennych psychologicznych. Jeśli chcielibyśmy to robić w przypadku tego typu zmiennych, to nie byłaby to walidacja, a więc sprawdzanie zgodności deklaracji z danymi dokumentalnymi, lecz szeroko rozumiana weryfikacja.

⁷⁶ Paweł Sztabiński mówi o „współautorstwie”, mając na myśli: wybór techniki badania, sposób przygotowania kwestionariusza, sposób sformułowania pytań, oraz porządek bloków pytań i ich kolejność w ramach bloków.

ROZDZIAŁ 2.

Ocena jakości danych: podejście pośrednie i bezpośrednie

W rozdziale 1 omówiłem dwa „tradycyjne” podejścia do oceny jakości danych z badań surveyowych. Pierwsze z nich, nazwane przeze mnie „wewnętrznym”, może ogniskować się na: (i) analizie przebiegu procesu badawczego, (ii) analizie poszczególnych jego elementów oraz (iii) błędach występujących w procesie badawczym. Drugie podejście, nazwane „zewnętrznym”, w ocenie jakości danych abstrahuje od procesu badawczego oraz ewentualnych błędów i skupia się na: (i) analizie trafności i rzetelności danych oraz (ii) określeniu ich „prawdziwości” w drodze weryfikacji lub walidacji danych. Każde z tych podejść, o czym wspomniałem, ma swoje ograniczenia i zalety. Ograniczenia sprowadzają się głównie do możliwości ich praktycznego zastosowania. Wymagają bowiem dodatkowych nakładów, nie tylko finansowych.

Do wyjątków należą projekty, w których prowadzi się analizy trafności i rzetelności oparte na metodzie MTMM, walidację na podstawie danych dokumentalnych (pomijając zmienną, jaką jest rok urodzenia respondentów, w przypadku prób imiennych), analizy przy zastosowaniu współczynnika korelacji wewnątrzklasowej (*intraclass correlation coefficients*, ICC), która jest miarą jednorodności obserwacji wewnątrz klas w próbie zrealizowanej, itp. W większości badań działania służące ocenie uzyskanych danych nie są podejmowane (może z wyjątkiem pewnych elementów z zakresu *quality control*), a wynik badania przyjmuje się z „dobrodziejstwem inwentarza”. Tak jednak być nie musi: nie trzeba „udawać”, że nasz pomiar jest „idealny” i „bezbłędny”.

Jan Lutyński (1972), opisując proces badawczy, wymienia jako jego uczestników badacza (zespół badawczy), ankieterów, respondentów i koderów. Schemat ten pomija jednak niezwykle istotnego uczestnika owego procesu, a mianowicie instytut realizujący badanie w terenie. Między instytutem (wykonawcą) a badaczem (zleceniodawcą) może bowiem występować, zgodnie ze wspomnianą teorią agencji, „konflikt interesów” (Jensen 1994). Jednemu i drugiemu zależy oczywiście na właściwej realizacji badania, na tym, aby ankieterzy pracowali poprawnie itd. Faktycznie jednak wykonawcy (instytutowi) zależy przede wszystkim na tym, aby dobrze wykonać zlecenie (zrealizować badanie), natomiast zleceniodawcy (badaczowi) na tym, aby uzyskane materiały (dane) były zgodne z rzeczywistością. Innymi słowy, instytut jest zainteresowany procesem zbierania danych, a badacz jego efektem, końcowym rezultatem. Można nawet zaryzykować stwierdzenie, iż badacza nie interesuje jak pracowali ankieterzy, czy w ogóle realizowali wywiady, byle ostateczny rezultat był „prawdziwy”, wiarygodny. Ów „konflikt interesów” wyznacza dwa, przyjęte przeze mnie, podejścia w ocenie uzyskanych materiałów: pośrednie i bezpośrednie. Różnica między nimi jest szczególnie widoczna w badaniach rynkowych: o ile ośrodek jest zainteresowany przede wszystkim, czy ankieterzy pracowali uczciwie, czy nie, o tyle klienta interesuje wynik badania, na przykład: „jaki odsetek badanych jest użytkownikami produktu X”. Mamy tutaj do czynienia z dwiema różnymi „rzeczywistościami”.

Poniżej przedstawiam propozycję sposobu oceny jakości uzyskanych danych w badaniach surveyowych, która jest wolna od wspomnianych ograniczeń, a więc nie wymaga dodatkowych nakładów finansowych i czasowych ze strony instytutu badawczego. Jak wspomniałem we wstępie, wszystkie kodeksy badań opinii i rynku mówią o konieczności realizacji badań kontrolnych jako nieodzownej, koniecznej fazy procesu badawczego. Innymi słowy, realizacja każdego badania winna podlegać kontroli. Wydaje się, iż tego typu „zapis” umożliwi ocenę jakości zbieranych danych. Równocześnie nie precyzuje się w nich przedmiotu, zakresu i metod realizacji badania kontrolnego. Zazwyczaj – tradycyjnie – ogranicza się ona do pytania: „czy ankieter zrealizował wywiad”. Moja propozycja znacznie wykracza poza to pytanie i łączy pewne elementy scharakteryzowanych wcześniej perspektyw: wewnętrznej i zewnętrznej, a konkretnie *quality control* oraz rzetelności pomiaru. Sprowadza się ona nie tylko do odpowiedzi na pytanie, „czy ankieter zrealizował wywiad”, „przestrzegaj

procedury” itd., ale także „jaka jest wartość odpowiedzi respondenta” (czy są one rzetelne, to znaczy w jakim stopniu replikowalne).

Rozróżnienie „interesu” wykonawcy (instytutu) i zleceniodawcy (badacza) jest zbieżne z dwiema funkcjami kontroli: pragmatyczno-szkoleniowa i walidacyjna (Sztabiński F. 2005: 350–353; por.: Lavrakas 2008: 712, 888–890). Ta pierwsza ma na celu ocenę pracy ankietera, druga zaś – wyniku badania.

Funkcja pragmatyczno-szkoleniowa ma na celu ocenę pracy ankieterów uczestniczących w badaniu. Jej przedmiotem jest przede wszystkim stwierdzenie, czy i jakie błędy popełniali ankieterzy w trakcie pracy w terenie, czy pracowali rzetelnie, a więc czy nie fałszowali wywiadów, nie opuszczali pytań itp. Można wskazać zasadnicze kategorie aktywności ankietera, będące przedmiotem kontroli. Są to (Frey i Oishi 1995: 142–146; Boeswillwald 1991: 301–303):

1. Efektywność ankietera, to znaczy liczba przeprowadzonych wywiadów w stosunku do czasu poświęconego na ich realizację.

2. Odsetek przypadków niezrealizowania wywiadu przez ankietera, to znaczy odsetek odmów udzielenia wywiadu, niedotarcia do respondenta itp.

3. Jakość wypełnienia kwestionariusza, a więc: przybliżony czas trwania wywiadu; sposób prowadzenia wywiadu – prezentacja materiałów (kart respondenta); kompletność wypełnienia kwestionariusza, czytelność zapisu itp.

4. Poprawność prowadzenia wywiadu (na przykład poprawność sondowania).

W tym ujęciu celem kontroli jest maksymalnie szczegółowe ustalenie okoliczności zbierania informacji. Przy takim rozumieniu funkcji kontroli jednostką „analizy”, a właściwie kontroli, jest ankieter. Wyniki są oczywiście wykorzystywane nie tylko do bieżącej oceny ankieterów, ale także perspektywicznie, stanowiąc podstawę szkolenia ankieterów i współpracy z siecią. W ten sposób działania kontrolne, podejmowane po badaniach (*ex post*), wyznaczają niejako zakres i kierunek działań zmierzających do poprawy jakości realizacji (*ex ante*). Z punktu widzenia instytutu realizującego badanie ocena „jakości” pracy poszczególnych ankieterów jest niezwykle istotna i użyteczna. Z jednej strony pozwala bowiem eliminować z sieci nierzetelnych, nieuczciwych ankieterów, z drugiej zaś wskazuje ewentualne ich „braki” czy też sfery „nieznajomości” zasad prowadzenia wywiadów.

Druga funkcja kontroli to funkcja oceniająca zebrany materiał (walidacyjna), rozumiana jako oszacowanie błędu pomiaru. Błąd można określać dwojako – jako pośredni i bezpośredni. Mówiąc o ocenie pośredniej, mam na myśli ocenę jakości uzyskanych danych na podstawie oceny jakości pracy ankietera. Mówiąc z kolei o ocenie bezpośredniej, mam na myśli ocenę jakości danych uzyskanych poprzez powtórzony pomiar (powtórne zadanie części pytań z kwestionariusza zasadniczego w badaniu kontrolnym). Jest to więc ocena rzetelności pomiaru. Upraszczając, można powiedzieć, że jeśli pytamy o to samo dwa razy (po raz pierwszy w wywiadzie zasadniczym i po raz drugi w wywiadzie kontrolnym) i uzyskujemy rozbieżne opinie, to wartość uzyskanych informacji staje się problematyczna. Dwie rozbieżne opinie nie mogą być bowiem prawdziwe. W dalszej części przedmiotem moich zainteresowań będzie jedynie walidacyjna funkcja kontroli, pozwalająca określić jakość uzyskanych danych poprzez powtórzony pomiar.

Należy zatem się zastanowić, w jaki sposób możemy określić pośrednio lub bezpośrednio jakość uzyskanych danych. W tym celu omówię możliwe do wykorzystania metody realizacji badań kontrolnych. Poniższe rozważania mają oczywiście charakter instrumentalny w stosunku do głównego wątku rozważań, wydają się jednak niezbędne.

2. 1. Ocena danych: metody nieterenowe i terenowe

Niezależnie od funkcji, przedmiotu i zakresu kontroli, a więc tego, co i ile chcemy kontrolować, podstawę naszych wniosków stanowić musi analiza materiałów zebranych w badaniach zasadniczych (wypełnionych kwestionariuszy, formularzy doboru, ankiet do ankietera itp.), kontrolnych bądź też analiza okoliczności i warunków, w jakich powstały. Informacje te uzyskuje się przy zastosowaniu terenowych lub tak zwanych nieterenowych metod kontroli. Te drugie nie wymagają uczestnictwa kontrolera w sytuacji badawczej ani ponownego kontaktu z respondentem (Sztabiński P. 1978: 200–215). Pierwsze zaś wymagają uprzedniego lub ponownego kontaktu kontrolera z respondentem.

2. 1. 1. Nieterenowe metody oceny

Wśród metod nieterenowych wymienić należy przede wszystkim techniki, które bazują na merytorycznej analizie wypełnionych kwestionariuszy oraz analizie opisu sytuacji wywiadu, której dokonuje ankieter (są to zatem materiały „przygotowywane” przez samych ankieterów). Określa się je mianem *data analytic methods* (Lavrakas 2008: 267–270). Wymieniam je poniżej.

(i) Merytoryczna analiza wypełnionych kwestionariuszy. Może ona być prowadzona pod kątem kompletności zadanych pytań, kompletności, jasności i czytelności zapisu, jego adekwatności do zapotrzebowania badawczego, prawidłowości sondażu, logicznej spójności itd. Analizując wewnętrzną spójność odpowiedzi, musimy jednak pamiętać, iż ostre sprzeczności w ramach kwestionariusza nie świadczą o fałszerstwie – wywiad sfalszowany jest bowiem z reguły bardzo logiczny i konsekwentny (Noelle 1963). Innym rodzajem kontroli jest kontrola logiczna zależności między odpowiedziami na różne pytania, które w mniejszym lub większym stopniu związane są z kryteriami losowania lub doboru. Przykładowo, z reguły analizie poddawany jest zawód i płeć respondenta. Przypadki kobiet „górników” czy „kierowców autobusu” lub mężczyzn pracujących jako „pielęgniarki” bądź „salowe” – zawsze stanowią sygnał, że wywiad danego ankietera należy przyjrzeć się bliżej. Wspomniane niespójności są z reguły konsekwencją pomyłek ankieterów przy zaznaczaniu w kwestionariuszu płci respondenta, będących przejawem niestaranności. Niekiedy jednak przeprowadzona weryfikacja dowodzi, że nie zgadza się nie tylko płeć, ale cały szereg informacji wpisanych przez ankietera do kwestionariusza. Wtedy może być to przejawem nie tyle sporadycznej pomyłki, lecz świadomej podmiany osoby respondenta.

(ii) Kontrola „wewnętrzna”, która polega na porównaniu prawdopodobieństw występowania określonych układów odpowiedzi. Metodę tę stosuje się zazwyczaj do pytań tabelarycznych. Prawdopodobieństwa w metodzie liczone są dla poszczególnych stwierdzeń (układ odpowiedzi dla danego stwierdzenia we wszystkich wywiadach zrealizowanych przez ankietera). Następnie porównywani są dla poszczególnych ankieterów (zob. Sawiński, Sztabiński 2005). Inaczej mówiąc, jeśli sekwencja odpowiedzi na 7 pytań w układzie: „1-2-3-2-2-2-3” pojawiła się w całym badaniu na próbie 1000-osobowej tylko w pięciu przypadkach oraz sekwencja „8-8-8-3-2-2-8” także tylko w sześciu przypadkach, to obie

te sekwencje uznać należy za bardzo „rzadkie” ze względu na częstotliwość. Dodatkowo, jeśli sekwencje te pojawiły się w kwestionariuszach tego samego ankietera, to stanowi to podstawę wysunięcia zastrzeżeń co do autentyczności zapisu w kwestionariuszu dokonanego przez ankietera. Najprawdopodobniej pominął on ten blok pytań, a odpowiedzi na nie wpisał sam (mało prawdopodobne jest bowiem, aby dwie sekwencje odpowiedzi pojawiające się na 1000 wywiadów wystąpiły u tego samego ankietera). Tego typu analizy, oparte na metodach statystycznych, pozwalające na identyfikację elementów zbioru odbiegających od „normy” prowadzili także Carl Pierchala i Jyoti Surti (2009).

(iii) Analiza odpowiedzi na pytania otwarte. Jeśli charakteryzują się one wyraźną monotonią lub bardzo wybujałą „fantazją”, może to świadczyć o fałszerstwie (Noelle 1963).

(iv) Procedura „walidacji”, która polega na porównaniu danych metryczkowych i faktualnych zapisanych w kwestionariuszu z analogicznymi, uzyskanymi z innych źródeł – najczęściej dokumentalnych (dane z próby). Ankieter z reguły otrzymuje wyłącznie informacje o imieniu i nazwisku wylosowanej osoby oraz o jej adresie. Tymczasem instytut badawczy może dodatkowo znać rok urodzenia (lub nawet dokładną datę) tej osoby czy pewne charakterystyki składu jej gospodarstwa domowego – na przykład płeć i wiek poszczególnych domowników. Zakres dodatkowych informacji jest w różnych badaniach niejednakowy i zależy od zawartości operatu oraz od metody doboru próby. W każdym konkretnym przypadku ankieterowi trudno jest przewidzieć, jakimi dodatkowymi danymi dysponuje instytut. Dodatkowe informacje o wylosowanych osobach z reguły nie są także udostępniane instruktorom, aby uniemożliwić jakikolwiek niekontrolowany ich przepływ. Jeśli ankieter nie skontaktuje się z wylosowaną osobą, to praktycznie nie ma żadnych szans wpisania właściwego wieku respondenta lub wymyślenia wiarygodnych odpowiedzi w przypadku innych kontrolowanych cech. Fakt ten ujawnia się natychmiast w fazie weryfikacji dokonywanej przez instytut badawczy. Jeśli próba losowa nie ma charakteru imiennego, lecz adresowy, to wszystkie opisane techniki weryfikacji również mają zastosowanie. W części prób adresowych instytut także może dysponować pewnymi informacjami dotyczącymi gospodarstwa domowego (płeć, wiek poszczególnych członków). Trudno sobie wyobrazić, aby nieuczciwy ankieter był w stanie dane te wygenerować samodzielnie lub uzyskać z innych źródeł. Weryfikacja ma status rutynowej procedury kontroli da-

nych. Za pomocą odpowiedniego oprogramowania porównywane jest komputerowy zapis odpowiedzi badanych uzyskanych przez ankietatorów z oryginalnymi informacjami pozyskanymi z operatu doboru próby. Wszelkie niespójności są wychwytywane i poddawane drobiazgowej analizie.

(v) Analiza opisu sytuacji wywiadu, tak zwane historie wywiadu (na przykład na podstawie ankiety do ankietera, kart doboru/raportu z pracy ankietera) [Sawiński, Sztabiński 2005: 373–375]. Pozwala ona na określenie czasu trwania wywiadów, ocenę wartości uzyskanych materiałów poprzez ocenę warunków wywiadu, a więc między innymi wpływu osób trzecich, miejsca i atmosfery wywiadu, nastawienia respondenta do rozmowy, czasu realizacji wywiadu itp. W tej grupie wskazać także należy na analizy zmierzające do określenia wpływu oraz efektu ankieterskiego. Jednakże w odniesieniu do tego typu analiz wysunąć można daleko idące zastrzeżenie związane z wiarygodnością tych danych (Sztabiński F. 1996: 37–58). Nieco inne możliwości stwarza analiza specjalnych kart, z których jedna zawiera deklarację zgody, druga zaś – odmowę udziału w badaniu (tę ostatnią wylosowana osoba winna odesłać pocztą). Najprostszą metodą kontroli rzetelności doboru udziałowego jest zapis przez ankietera danych personalnych oraz numeru telefonu respondenta po przeprowadzeniu wywiadu. Zakłada się, że w przypadku rzetelnego doboru respondenta ankietery dołoży wszelkich starań, aby uzyskać te dane i nie będzie miał obaw związanych z wynikiem ewentualnej weryfikacji (osobistej lub telefonicznej). Z kolei nierzetelni ankieterzy, dążąc do uniknięcia kontroli, nie będą uzyskiwać tych informacji. W przypadku techniki CAPI bardzo duże możliwości stwarza analiza międzyczasów w realizowanych wywiadach. Z kolei w technice CATI – możliwość „podglądu” i „podśluchu” wywiadu.

(vi) Analiza efektywności ankietera, o czym wcześniej wspominałem, określana z liczbą przeprowadzonych wywiadów w stosunku do czasu poświęconego na ich realizację oraz odsetek przypadków niezrealizowania wywiadu przez ankietera, to znaczy odsetek odmów udzielenia wywiadu, niedotarcia do respondenta itp. (Frey i Oishi 1995: 142–146; Boeswillwald 1991: 301–303).

Analizę porównawczą metod ułatwi tabela 1, w której uwzględniono cztery podstawowe techniki kontroli nieterenowej: (1) analizę wypełnionych kwestionariuszy; (2) kontrolę „wewnętrzną”; (3) weryfikację oraz (4) analizę opisu sytuacji wywiadu.

Tabela 1. *Porównanie nieterenowych metod kontroli ze względu na wymiary oceny*

Przedmiot kontroli	Metoda kontroli			
	(1)	(2)	(3)	(4)
a) dotarcie pod wylosowany adres	–	–	–	+
b) liczba kontaktów	–	–	–	+
c) poprawny dobór adresu dodatkowego	–	–	–	+
d) poprawny dobór respondenta	+	–	o	+
e) realizacja wywiadu	+	+	+	+
f) realizacja wywiadu z właściwym respondentem	+	–	+	+
g) określenie swojej roli przez ankietera	–	–	–	–
h) podanie informacji o badaniach	–	–	–	–
i) kompletność zadanych pytań	+	o	–	–
j) kolejność zadanych pytań	–	–	–	–
k) poprawność odczytania pytania	–	–	–	–
l) sondowanie	+	–	–	–
m) sugerowanie odpowiedzi	–	–	–	–
n) wykorzystanie materiałów dodatkowych	–	–	–	–
o) zgodność zapisu z deklaracją respondenta	–	–	–	–
p) sposób zakończenia rozmowy	–	–	–	–
q) atmosfera wywiadu	–	–	–	+
r) ocena ankietera	–	–	–	–
s) czas trwania wywiadu	o	–	–	+
t) miejsce wywiadu	–	–	–	+
u) stosunek respondenta do wywiadu	–	–	–	+
v) właściwy ankieter	–	–	–	+
w) nastawienie ankietera do wywiadu	–	–	–	–

Oznaczenia symboli w tabeli:

- „+” – metoda dostarcza pełnych informacji na dany temat
- „o” – metoda dostarcza niepełnych informacji na dany temat
- „–” – metoda nie dostarcza informacji na dany temat

Powyższa tabela umożliwia także uszeregowanie poszczególnych metod kontroli z punktu widzenia ich potencjalnej informatywności. Przypisując poszczególnym symbolom wartości liczbowe od 1 do 0, otrzymamy następujący szereg:

Tabela 2. *Uszeregowanie nieterenowych metod kontroli ze względu na potencjalne wymiary oceny*

Procedura kontroli	Liczba punktów
analiza opisu sytuacji wywiadu	11,0
analiza wypełnionych kwestionariuszy	5,5
weryfikacja	2,5
kontrola „wewnętrzna”	1,5

Jak widać z powyższej tabeli, najwięcej informacji na temat różnego rodzaju zachowań ankietera przynosi analiza opisu sytuacji wywiadu, następnie analiza wypełnionych kwestionariuszy oraz weryfikacja, najmniej zaś „kontrola wewnętrzna”. Celowe wydaje się jednak określenie, jakiego rodzaju informacji mogą dostarczać poszczególne procedury. Analiza opisu sytuacji wywiadu dostarcza informacji zarówno na temat rzetelności pracy ankietera, jak i prowadzenia wywiadu, aczkolwiek informacje te są niepełne. Analiza wypełnionych kwestionariuszy mówi z kolei znacznie więcej o sposobie prowadzenia wywiadu niż rzetelności ankietera. Trzecia i czwarta procedura – weryfikacja i kontrola „wewnętrzna” – mogą potencjalnie dostarczyć przede wszystkim informacji na temat rzetelności pracy ankietera, nic niemal nie mówiąc o sposobie prowadzenia wywiadu.

Poszczególne nieterenowe metody różnią się więc ze względu na potencjalny zakres oceny pracy ankietera: jedne dostarczają informacji służących przede wszystkim do oceny rzetelności jego pracy, inne – sposobu prowadzenia wywiadu.

2. 1. 2. Terenowe metody oceny

Znacznie więcej możliwości stwarzają pod tym względem terenowe metody kontroli. Wśród nich wyróżnić należy takie, które wymagają wcześniejszego porozumienia badacza z respondentem (*observational*

methods) oraz takie, które nie wymagają uprzedniego porozumienia (*recontact methods*) [Sztabiński P. 1978: 200–215; Sztabiński F. 2005: 356–356; Lavrakas 2008: 267–270]. Wśród tych pierwszych wymienić należy przede wszystkim techniki:

(i) ukrytego nagrywania magnetofonowego wywiadu (oczywiście bez wiedzy ankietera);

(ii) podstawionego respondenta. Zakłada się tutaj wykorzystanie respondenta jako kontrolera. Stosując tę metodę, należy szczegółowo wyjaśnić respondentowi, na czym polega jego rola jako kontrolera oraz na które zachowania ankietera należy zwrócić szczególną uwagę;

(iii) wywiadu podsłuchiwanego, to znaczy takiego, w którym kontroler (bez wiedzy ankietera) przysłuchuje się prowadzonej rozmowie z respondentem.

Druga grupa terenowych metod kontroli obejmuje te, które nie wymagają uprzedniego porozumienia się z respondentem. Najczęściej stosowane techniki to:

(iv) wywiad obserwowany – kontroler wraz z ankieterem udaje się do respondenta i jest obecny podczas rozmowy (Boeswillwald 1991);

(v) jawne nagrywanie magnetofonowe – ankieter zobowiązany jest do rejestracji w całości prowadzonej przez siebie rozmowy na dyktafonie;

(vi) wywiad kontrolny osobisty (WKO) – kontroler udaje się do respondenta i powtórnie przeprowadza z nim wywiad – w całości lub w części, a także uzyskuje niezbędne informacje na temat wizyty ankietera (Boeswillwald 1991; Noelle 1963);

(vii) wywiad kontrolny telefoniczny (WKT) – jest to technika zbliżona do procedury powtórnego wywiadu osobistego, przy czym zakłada się tu nie bezpośredni, lecz telefoniczny kontakt kontrolera z respondentem. Pewną odmianą wywiadu telefonicznego jest wywiad realizowany w systemie CATI;

(viii) kontrolna ankieta pocztowa (KAP) – do respondenta wysyłany jest kwestionariusz ankiety z prośbą o odpowiedź na zawarte w nim pytania i zwrot pocztą. Odmianą ankiety pocztowej jest pocztowa karta kontrolna (KKP).

Rozważania dotyczące potencjalnego zakresu oceny pracy ankietera ograniczymy jedynie do wybranych spośród wskazanych powyżej metod oceny pracy ankietera. Potencjalny zakres oceny pracy ankietera przy zastosowaniu terenowych metod kontroli prezentuje tabela 3. Uwzględniono w niej następujące techniki kontroli terenowej: (1) ukryte nagry-

Tabela 3. Porównanie terenowych metod kontroli ze względu na potencjalne wymiary oceny

Przedmiot kontroli	Metoda kontroli									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a) dotarcie pod wylosowany adres	+	+	+	+	+	o	+	+	+	+
b) liczba kontaktów	-	+	+	+	o	-	o	-	o	-
c) poprawny dobór adresu dodatkowego	-	-	-	+	+	-	-	-	-	-
d) poprawny dobór respondenta	+	+	+	+	+	-	+	+	+	o
e) realizacja wywiadu	+	+	+	+	+	+	+	+	+	+
f) realizacja wywiadu z właściwym respondentem	+	+	+	+	+	-	+	+	+	o
g) określenie swojej roli przez ankietera	+	+	+	+	o	-	o	o	o	-
h) podanie informacji o badaniach	+	+	+	+	o	-	o	o	o	-
i) kompletność zadanych pytań	+	+	+	+	o	+	o	o	o	-
j) kolejność zadanych pytań	+	+	+	+	-	+	-	-	-	-
k) poprawność odczytania pytania	+	o	+	+	-	+	-	-	-	-
l) sondowanie	+	o	+	+	-	+	-	-	-	-
m) sugerowanie odpowiedzi	+	+	+	+	-	+	-	-	-	-
n) wykorzystanie materiałów dodatkowych	+	+	+	+	+	+	o	o	o	-
o) zgodność zapisu z deklaracją respondenta	+	o	+	o	o	+	o	o	o	-
p) sposób zakończenia rozmowy	o	+	+	+	o	-	o	o	o	-
q) atmosfera wywiadu	+	+	+	+	+	+	+	+	+	o
r) ocena ankietra	-	+	+	+	+	-	+	+	+	o
s) czas trwania wywiadu	+	+	+	+	+	+	+	+	+	+
t) miejsce wywiadu	+	+	o	+	+	-	+	+	+	-
u) stosunek respondenta do wywiadu	+	o	+	+	o	+	o	o	o	-
v) właściwy ankieter	+	+	o	+	o	+	o	o	o	o
w) nastawienie ankietera do wywiadu	+	+	+	+	o	+	o	o	o	-

Oznaczenia symboli w tabeli:

- „+” – metoda dostarcza pełnych informacji na dany temat
- „o” – metoda dostarcza niepełnych informacji na dany temat
- „-” – metoda nie dostarcza informacji na dany temat

wanie magnetofonowe; (2) podstawiony respondent; (3) wywiad podsłuchiwany; (4) wywiad obserwowany; (5) wywiad kontrolny osobisty (WKO); (6) jawne nagrywanie magnetofonowe; (7) wywiad kontrolny telefoniczny (WKT); (8) kontrolna ankieta pocztowa (KAP); (9) wywiad kontrolny CATI (CATI); (10) kontrolna karta pocztowa (KKP).

Powyższa tabela pozwala na uszeregowanie poszczególnych metod kontroli pod względem potencjalnego zakresu przedmiotu oceny. Jeśli symbolowi „+” przypiszemy wartość 1, symbolowi „o” – wartość 0,5, natomiast symbolowi „-” – 0, możemy obliczyć wskaźnik „informatywności” poprzez zsumowanie punktów. Porównując wartości obliczone dla poszczególnych metod, będzie można z kolei określić w sposób „syntetyczny” potencjalny zakres przedmiotu oceny. Powyższy sposób analizy posiada jednak dość istotne ograniczenie. Przyjmuje się bowiem, iż waga poszczególnych informacji jest taka sama, podczas gdy w rzeczywistości może być, i zazwyczaj jest, różna. Przykładowo, w przypadku pytań faktualnych (na przykład o historię zawodową respondenta), czasami istotniejsza jest kompletność zadanych pytań niż ich kolejność.

Tabela 4. *Uszeregowanie terenowych metod kontroli ze względu na potencjalne wymiary oceny*

Procedura kontroli	Liczba punktów
wywiad obserwowany	22,5
wywiad podsłuchiwany	21,0
podstawiony respondent	20,0
ukryte nagrywanie magnetofonowe	19,0
wywiad kontrolny osobisty (WKO)	14,5
jawne nagrywanie magnetofonowe	13,5
wywiad kontrolny telefoniczny (WKT)	13,0
wywiad kontrolny Cati (CATI)	13,0
kontrolna ankieta pocztowa (KAP)	12,5
kontrolna karta pocztowa (KKP)	5,5

Jak widać z powyższej tabeli, najwięcej informacji na temat różnego rodzaju zachowań ankietera przynoszą techniki kontroli wymagające nawiązania kontaktu z respondentem przed wywiadem, a więc następu-

jące: wywiad obserwowany, następnie wywiad podsłuchiwany, technika podstawionego respondenta, ukryte nagrywanie magnetofonowe. Relatywnie mniej możliwości stwarzają: wywiad kontrolny osobisty (WKO), jawne nagrywanie magnetofonowe, wywiad kontrolny telefoniczny (WKT), wywiad kontrolny Cati (CATI), kontrolna ankieta pocztowa (KAP), najmniej zaś kontrolna karta pocztowa (KKP).

Celowe wydaje się określenie, jakiego rodzaju informacji mogą dostarczać poszczególne procedury. Otóż procedury ukrytego nagrywania magnetofonowego, podstawionego respondenta, wywiadu podsłuchiwanego, wywiadu obserwowanego oraz jawnego nagrywania magnetofonowego mogą dostarczyć bez mała pełnych informacji zarówno na temat rzetelności pracy ankietera, jak i prowadzenia wywiadu. Procedury wywiadu kontrolnego osobistego (WKO), wywiadu kontrolnego telefonicznego (WKT), kontrolnej ankiety pocztowej (KAP) oraz wywiadu kontrolnego CATI (CATI) mogą także przynieść informacje na temat rzetelności pracy ankietera i poprawności realizacji wywiadu, aczkolwiek ostatnie z nich są niepełne. Z kolei ostatnia z procedur – kontrolna karta pocztowa (KKP) – może potencjalnie dostarczyć nam przede wszystkim informacji na temat rzetelności pracy ankietera, nic praktycznie nie mówiąc, a przynajmniej bardzo mało, na temat sposobu prowadzenia wywiadu.

Poszczególne terenowe metody różnią się więc ze względu na możliwe wymiary oceny pracy ankietera: jedne z nich dostarczają przede wszystkim informacji służących do oceny rzetelności jego pracy, inne sposobu prowadzenia wywiadu. Nie jest zatem tak, iż metody terenowe służą wyłącznie lub przede wszystkim kontroli rzetelności pracy ankietera, a nieterenowe – poprawności realizacji wywiadu. Z punktu widzenia potencjalnego zakresu oceny nie można więc w sposób jednoznaczny przyporządkować procedur kontrolnych celom, jakie mogą one spełniać.

Reasumując, poszczególne techniki kontroli znacznie różnią się między sobą i każda z nich ma zalety i ograniczenia. Niektóre dostarczają informacji dotyczących przede wszystkim prawidłowości prowadzenia wywiadu przez ankietera, inne jego rzetelności, jedne przynoszą mniej, inne więcej informacji danego typu, niektóre dostarczają informacji bardziej pewnych, obiektywnych, inne mniej, niektóre wreszcie stwarzają większe możliwości praktycznego zastosowania, inne mniejsze. Analizując poszczególne techniki kontroli, zauważymy, że rodzaj i zakres

dostarczanych przez nie informacji oraz ich pewność są odwrotnie proporcjonalne do możliwości ich praktycznego stosowania w badaniach masowych. Technika podstawionego respondenta przynosi bardzo dużo informacji na temat rzetelności ankietera i prawidłowości prowadzenia wywiadu, jednak jest praktycznie niemożliwa do zastosowania na szerzą skalę. Z kolei w przypadku kontroli techniką pocztową nie występują praktycznie żadne warunki ograniczające jej stosowalność, ale jej informatywność jest raczej niska: dzięki niej można bowiem stwierdzić, czy ankieter przeprowadził wywiad, czy też nie, niewiele natomiast możemy powiedzieć na temat samej poprawności jego realizacji.

Przy wyborze metody oceny wartości materiałów, poza kwestią przedmiotu oceny, jej zakresu (chodzi o odsetek faktycznie skontrolowanych przypadków), kierować się należy jej wiarygodnością. Innymi słowy, zawsze musimy odpowiedzieć sobie na pytanie, jaka jest wartość danych służących do oceny wartości danych w badaniach zasadniczych. Odpowiedź na to pytanie wydaje się niestety znacznie bardziej skomplikowana, niż się wydaje. Paradoksalnie bowiem zazwyczaj nie dostrzegamy konieczności określania wiarygodności poszczególnych procedur kontrolnych i uzyskiwanych przy ich zastosowaniu informacji. Z metodologicznego punktu widzenia jest to dość poważne uchybienie.

Celowe zatem wydaje się zamieszczenie kilku uwag na temat trzech najczęściej stosowanych terenowych procedur kontrolnych, a mianowicie: karty kontrolnej, ankiety kontrolnej oraz wywiadu kontrolnego. Refleksje te dotyczyć będą dwóch problemów: faktycznego zakresu kontroli przy zastosowaniu danej procedury, to znaczy odsetka faktycznie skontrolowanych przypadków, oraz wiarygodności informacji uzyskiwanych przy ich zastosowaniu. Bazują one na analizie materiałów z kontroli pocztowej i osobistej przeprowadzonej przez Ośrodek Realizacji Badań Socjologicznych IFiS PAN w 1993 i 1994 roku w związku z realizacją trzech badań: „Sytuacja życiowa a opinie i poglądy” (kierowanego przez K. M. Słomczyńskiego), Polski Generalny Sondaż Społeczny – Edycja II (prowadzonego pod kierownictwem B. Cichomskiego i Z. Sawińskiego) oraz „Problemy życiowe osób niepełnosprawnych” (kierownikiem projektu była A. Ostrowska)¹.

¹ Wymienione powyżej metodologiczne badania kontrolne realizowane były w ramach grantu KBN „Ankieter jako źródło zniekształceń w procesie badawczym”, kierowanego przez Pawła B. Sztabińskiego. Ze względu na swój charakter nie zosta-

W badaniach „Sytuacja życiowa a opinie i poglądy” kontrolą objęto ogółem 426 przypadków (co piąty adres przydzielony każdemu ankieterowi). Podczas kontroli zastosowano procedurę wywiadu osobistego i telefonicznego. Z kolei w badaniach Polski Generalny Sondaż Społeczny (PGSS) kontrola prowadzona była dwuetapowo. W pierwszym objęła 1000 adresów (50% próby założonej). Wśród nich znalazło się 833, w których zrealizowano wywiad, oraz 167, w których nie udało się go zrealizować. Pod wszystkie adresy wysłano ankietę kontrolną. W drugim etapie, spośród pozostałych adresów (1000) wylosowano dalszych 600. Pod 500 z nich wysłano kartę kontrolną (wśród nich znalazło się 395, w których zrealizowano wywiad, oraz 105, w których nie udało się go zrealizować); pod pozostałe 100 adresów wysłano natomiast ankietę kontrolną z badań „Problemy życiowe osób niepełnosprawnych”. Wreszcie, w badaniach „Problemy życiowe osób niepełnosprawnych” (NP), spośród wszystkich zrealizowanych wywiadów wylosowano co piąty każdego ankietera (ogółem 347 wywiadów). Niezależnie od tego, dodatkowo wylosowano 200 adresów, w których nie udało się zrealizować wywiadu. Do wszystkich tych osób wysłano ankietę kontrolną.

Pierwszym, zupełnie oczywistym wnioskiem jest to, że faktyczny zakres kontroli, a więc liczba rzeczywiście skontrolowanych przypadków, jest zależny od zastosowanej procedury: największy jest w przypadku wywiadu (80,0%), następnie karty kontrolnej (63,6%), najmniejszy zaś przy wykorzystaniu ankiety kontrolnej (44,8% – PGSS i 58,1% – NP). Po drugie, faktyczny zakres kontroli jest różny w przypadku respondentów, a więc osób, z którymi udało się zrealizować wywiad, oraz nierespondentów, to znaczy tych, z którymi wywiadu nie zrealizowano. W procedurach pocztowych (ankieta, karta), wśród nierespondentów faktyczny zakres kontroli jest 2–3-krotnie mniejszy niż w przypadku respondentów. Podobnie jest w wywiadzie, aczkolwiek dysproporcja między kategorią respondentów i nierespondentów nie jest aż tak duża. Dane te wskazują, że prowadząc kontrolę pocztową, sprawdzamy przede wszystkim te przypadki, w których faktycznie zrealizowano wywiad; osoby, z którymi wywiad nie został przeprowadzony, po prostu nie odpisują, a przynajmniej robią to znacznie rzadziej. Oznacza to, że kontrola pocztowa służy raczej „potwierdzeniu” faktu realizacji wywiadu niż jego „zaprzeczeniu” (Sztabiński F. 1995).

Źródło: Wykazane w Aneksie 1. „Zestawie badań i projektów będących przedmiotem analizy”.

Wnioski dotyczące wiarygodności danych z kontroli ograniczymy jedynie do tych przypadków, w których, zgodnie z relacją ankietera, nie udało się zrealizować wywiadu. Jest to uzasadnione tym, iż tylko w odniesieniu do takich przypadków relacje ankietera można uznać za w pełni wiarygodne. Mało prawdopodobna wydaje się bowiem sytuacja, w której ankieter twierdzi, że nie zrealizował wywiadu, podczas gdy faktycznie został on przeprowadzony. Otóż okazuje się, iż nierespondenci wielokrotnie twierdzą, iż brali udział w badaniach i realizowano z nimi wywiad. Przypadki takie są znacznie częstsze w procedurach pocztowych (sięgają nawet 20%) niż w wywiadzie (potwierdzenie „realizacji” niezrealizowanego wywiadu zasadniczego miało miejsce tylko w 1 przypadku na 25). Jeśli chodzi o wywiad (osobisty lub telefoniczny) wiarygodność danych jest oczywiście uzależniona od rzetelności kontrolera. Opisany powyżej system pozwala jednak w pewnym zakresie „kontrolować kontrolera”. Jest tak, ponieważ kontroler nigdy nie jest informowany, czy pod wskazanym adresem został przeprowadzony wywiad czy też nie; kto powinien realizować wywiad oraz czy była to kobieta czy mężczyzna; ile osób zamieszkuje pod wskazanym adresem; jaki jest wiek respondenta; jakie jest usytuowanie jego mieszkania itp. Kontroler zobowiązany jest uzyskać te informacje, a ponieważ badacz z reguły posiada większość z nich, ma możliwość ocenić jego rzetelność. W ten sposób kontroler kontroluje niejako „sam siebie”. Wskazane sytuacje uświadamiają nam, iż do informacji zawartych zarówno w kartach, jak i ankietach kontrolnych, nie można podchodzić bezkrytycznie. Ich występowanie świadczy o pewnej tendencji do „chronienia” czy też „krycia” ankietarów: potwierdzania wydarzeń, które faktycznie nie miały miejsca. Jeśli występuje tego typu tendencja, należałoby postawić pytanie, czy i jak często mamy w kontroli do czynienia ze zjawiskiem przeciwnym: zaprzeczaniem wydarzeniom, które faktycznie miały miejsce.

Podstawę oceny wiarygodności informacji uzyskiwanych w kontroli pocztowej może także stanowić analiza odpowiedzi osób, które nie brały udziału w badaniach będących przedmiotem kontroli (chodzi tutaj o ankietę kontrolną „Problemy życiowe osób niepełnosprawnych” wysłaną do respondentów, którzy brali udział w badaniach PGSS). Otóż, wśród tych, którzy odesłali ankietę, ponad połowa stwierdziła, iż zgłosił się do nich ankieter z prośbą o udzielenie odpowiedzi w wywiadzie „Problemy życiowe osób niepełnosprawnych”, a spośród nich – ponad trzy czwarte, że „udzielało odpowiedzi na te pytania”.

Generalnie zatem, wśród części respondentów występuje, z jednej strony, tendencja do „krycia” ankieterów, z drugiej zaś, nie identyfikują oni tematyki i problematyki badań. Pamiętają jedynie (lub wydaje im się, że pamiętają), że ktoś był i o coś pytał. Biorąc pod uwagę, iż obecnie realizuje się coraz więcej różnych badań, powyższa konstatacja podważa sens kontroli, szczególnie przy zastosowaniu technik pocztowych. Tego typu kontrolę można określić mianem rytualnej (wysoki ogólny odsetek zwrotów, przede wszystkim od tych, którzy brali udział w badaniu).

Powyższe uwagi dotyczące wartości danych z kontroli nie miały na celu wykazania, iż bardziej wskazane jest stosowanie wywiadu kontrolnego niż którejs z procedur pocztowych. Żadna technika nie jest bowiem ani „dobra”, ani „zła”, każda ma pewne ograniczenia i w związku z tym w większym lub w mniejszym stopniu nadaje się do uzyskiwania pewnego rodzaju informacji. Stąd też zastosowanie którejkolwiek musi się odbywać w sposób świadomy, musi być poprzedzone refleksją dotyczącą przedmiotu i zakresu kontroli.

Powyższe uwagi dotyczące wiarygodności powyższych trzech terenowych metod kontroli: wywiadu kontrolnego osobistego, wywiadu kontrolnego telefonicznego/CATI oraz kontrolnych ankiet pocztowych warto odnieść do trzech podstawowych wymiarów kontroli: realizacji wywiadu, realizacji próby i rzetelności uzyskiwanych informacji.

Tabela 5. *Terenowe metody kontroli a podstawowe wymiary kontroli*

Przedmiot oceny	Metody terenowe		
	Wywiad kontrolny osobisty	Wywiad kontrolny telefoniczny/CATI	Kontrolneankiety pocztowe
ocena bezpośrednia			
realizacja wywiadu	+	+	o
realizacja próby	+	o	–
rzetelność informacji	+	+	o
ocena pośrednia			
realizacja wywiadu	+	+	o
realizacja próby	+	o	–

Oznaczenia symboli w tabeli:

- „+” – metoda dostarcza pełnych i wiarygodnych informacji
- „o” – metoda dostarcza niepełnych lub/i mało wiarygodnych informacji
- „–” – metoda nie dostarcza informacji

Należy zatem jeszcze raz podkreślić, że optymalnymi metodami kontrolnymi jest wywiad osobisty lub telefoniczny – kontrola pocztowa, jak wspomniałem, jest kontrolą „rytualną”². Pierwsze dwie techniki mają poza tym najszerszy zakres zastosowania i pozwalają na łączenie podstawowych funkcji kontroli: pragmatyczno-szkoleniowej oraz oceniającej (walidacyjnej).

2. 2. Ocena jakości danych: podejście pośrednie

W przypadku pierwszego z ujęć (podejście pośrednie) ocena wartości uzyskanych materiałów dokonywana jest na podstawie oceny pracy ankieterów. W tym przypadku przyjmuje się, że jeśli ankieterzy pracowali właściwie (poprawnie dobierali respondentów, realizowali z nimi wywiad, w sposób zgodny z instrukcją, wywiad był realizowany w sprzyjających warunkach itd.), to wartość uzyskanych danych jest wysoka. Jeśli natomiast ankieterzy nie pracowali rzetelnie, to jakość danych jest co najmniej dyskusyjna³.

Aby dokonać oceny pracy ankieterów (i na jej podstawie wnioskować o jakości materiałów) musimy monitorować i kontrolować pracę ankieterów. Przykładowo, jeśli na podstawie kontroli stwierdzimy, że w danym badaniu: (i) w 60 przypadkach na 1800 (3%) wystąpiła niezgodność danych zawartych w kartach adresowych bądź kartach i kwestionariuszach, przy czym, co najwyżej w 0,8% przypadków, błędy te mogły mieć wpływ na dobór respondenta, a w konsekwencji na reprezentatywność próby, (ii) na podstawie wyników kontroli terenowej prowadzonej przy zastosowaniu ankiety pocztowej, można wysunąć zastrzeżenia co do rzetelności 2,4% skontrolowanych przypadków, to znaczy 10 wywiadów na 416 (były to wywiady 10 różnych ankieterów), (iii) w 3,8% skontrolowanych

² Warto wspomnieć, iż Central Coordinating Team w projekcie European Social Survey, począwszy od Rundy 3 nie dopuszcza kontroli realizacji wywiadu techniką pocztową.

³ Nie każdy błąd ankietera powoduje wystąpienie błędu pomiaru. Stąd też błędne zachowanie ankietera jest warunkiem koniecznym, ale niewystarczającym zaistnienia błędu rozumianego jako jego wpływ na wynik pomiaru. Podejście to przyjmuje perspektywę wewnętrzną, dynamiczną (w rozumieniu przyjętym w poprzednim rozdziale), a więc oceny dokonujemy tutaj poprzez analizę procesu badawczego.

wywiadów ankieterzy mogli opuszczać poszczególne pytania, ponieważ czas trwania wywiadu był stosunkowo krótki (poniżej 35 min), przy zakładanym 60-minutowym, to z punktu widzenia ośrodka realizującego badania tego typu stwierdzenia są niezwykle istotne i użyteczne. Z jednej strony pozwalają bowiem eliminować nierzetelnych ankieterów, ale także wskazują, w jakim zakresie winni być szkoleni pozostali, z drugiej zaś – dają „ogólny” obraz realizacji badania⁴. W literaturze poświęconej problemom kontroli znaleźć można na przykład stwierdzenie, że merytoryczna analiza wypełnionych kwestionariuszy pozwala wychwycić około 12% ogółu błędów popełnianych przez ankieterów. Są to przede wszystkim: brak sondażu – 39%; zła klasyfikacja odpowiedzi respondenta – 22%; zmiana sensu pytania przez użycie nieprawidłowych sformułowań lub sugerującego sondażu – 19% oraz niekompletna rejestracja odpowiedzi na pytania otwarte – 11% (Willcock 1956: 125–131).

W przypadku oceny pośredniej, zakładając, że jeśli ankieterzy pracowali uczciwie, rzetelnie, to wyniki badania są wiarygodne, przyjmuje się istnienie pewnych standardów poprawnej pracy ankieterów. W rzeczywistości tak jest, aczkolwiek częściej mówi się o błędach ankieterskich. Błędy ankieterskie można rozumieć jednak szerzej – jako świadome (wskazane powyżej) lub nieświadome odstępstwo od pewnych norm, standardów lub zakładanej procedury badania (Kistelski 1981; Lavrakas 2008: 267). Błędy nieświadome wynikają z braków szkolenia ankietera (jego wiedzy). Błędy świadome natomiast to te, które dotyczą sytuacji, w której ankieter wie, jak powinien się zachować, a zachowuje się inaczej. W dalszej części skupię się na błędach świadomych. Paul Lavrakas (2008: 267) wyróżnia następujące ich kategorie⁵:

(i) całkowite sfałszowanie wywiadu (*fabrication*) i częściowe, w sytuacji, gdy ankieter pomija niektóre pytania i później je uzupełnia (*falsification*);

(ii) całkowite lub częściowe fałszowanie informacji dotyczących przebiegu realizacji badania w terenie, na przykład rezultatu kontaktu z respondentem;

(iii) fałszowanie odpowiedzi na pytania w celu pominięcia kolejnych pytań (tak zwane filtrowanie);

⁴ Chodzi więc tu o wspomnianą wcześniej pragmatyczno-szkoleniową funkcję kontroli (Sztabiński F. 1999). Do tego problemu powrócę w dalszej części.

⁵ Por. także: Murphy, Baxter, Eyerman, Cunningham, Kennet 2004.

(iv) realizacja wywiadu z osobami spoza próby.

Wskazane powyżej świadome błędy ankierskie można generalnie podzielić na dwie duże grupy: błędy związane z procedurą doboru respondenta oraz błędy związane z procedurą badania.

2. 2. 1. Nieprawidłowości związane z realizacją próby

Rodzaje błędów związanych z realizacją próby są odmienne w zależności od rodzaju operatu, z którego losuje się próbę do badania (próba imienna – PESEL, próba adresowa – GUS oraz kwotowa – udziałowa).

W przypadku próby imiennej podstawowym błędem jest niepowracanie do respondenta, który był w momencie wizyty ankietera nieobecny. Dziś w standardowych badaniach konsumenckich i opinii publicznej odsetek wywiadów niezrealizowanych z próby zasadniczej (przy zastosowaniu techniki PAPI lub CAPI) wynosi od 60 do 80%. Tymczasem współczynnik realizacji, przy którym nie występuje błąd braku odpowiedzi na poziomie jednostki losowania (*unit non-response bias*), wynosi 70%. Dobrą ilustracją konieczności powracania do respondenta i możliwości minimalizacji tego błędu są dane z Europejskiego Sondażu Społecznego, Runda 1. W badaniu tym ponad 17% wszystkich wywiadów zrealizowano w trakcie trzeciej i kolejnych wizyt. Odpowiednio podczas pierwszej wizyty zrealizowano 56,2% ogółu wywiadów, drugiej – 26,5%, trzeciej – 10,9%, czwartej – 4,4%, piątej – 1,1%, szóstej – 0,5% i siódmej – 0,4%). Innym rodzajem błędu w realizacji próby jest niedotarcie do respondenta i rezygnacja z wywiadu pod „pozorem” odmowy lub niezastania go, mimo kilkukrotnej wizyty. Stąd też, w projekcie Europejski Sondaż Społeczny przyjęto zasadę, iż przedmiotem kontroli są także wywiady niezrealizowane, a więc wspomniane przypadki niezastania respondenta (*non contacts*) oraz odmów (*refusals*).

W przypadku próby adresowej podstawowym błędem ankieterów jest z kolei losowanie respondenta spośród osób obecnych w momencie jego wizyty. W próbach adresowych odsetek osób w wieku emerytalnym jest zazwyczaj wyższy niż w populacji, aczkolwiek w ostatnich 2–3 latach uległo to zmianie (jest niższy). Związane jest to z faktem poczucia zagrożenia tej kategorii osób (osoby starsze obawiają się wpuszczać do domu ankieterów – osoby nieznanne). Inne podstawowe błędy to niepowracanie do respondenta, który był w momencie wizyty ankietera nieobecny oraz niepowracanie pod adres, w którym nikogo nie zastano.

Powyższe dwa rodzaje prób probabilistycznych (imienne i adresowe) wymagają od ankietera realizacji wywiadu ze ściśle określoną osobą. Często zdarza się jednak, że po dotarciu pod wskazany adres ankieter spotyka się z miłym przyjęciem, życzliwością lokatorów i chęcią kooperacji, lecz okazuje, że wylosowanej osoby akurat w domu nie ma. Wymaga to od niego umówienia się na inny termin, co niekiedy wiąże się z koniecznością ponownego przyjazdu do tej samej miejscowości. Stwarza to silną pokusę dokonania „podmiany” wylosowanej osoby na któregoś z obecnych domowników, zwłaszcza gdy manifestują oni wolę wzięcia udziału w badaniu. Osoby te często próbują wręcz skłonić ankietera do zrealizowania wywiadu z kimś innym, a niekiedy twierdzą, że ich udział w wywiadzie jest bardziej uzasadniony niż osoby wylosowanej. Tego typu „podmiana” jest w przypadku prób losowych absolutnie niedopuszczalna w żadnej sytuacji. Ankieter musi mieć świadomość, że realizacja wywiadu z inną osobą jest zawsze kwalifikowana jako świadomy błąd (oszustwo). Wywiad z kimś spoza próby jest bezwartościowy z punktu widzenia celów badania, a co więcej – może wypaczyć jego wyniki. Ankieter, dokonując „podmiany”, nie tylko chce wyłudzić wynagrodzenie za czynność niewykonaną, ale działa na szkodę instytutu badawczego i jego klientów.

Z kolei w przypadku prób nieprobabilistycznych (kwotowych i celowych) dobór respondenta pozostawia się ankieterowi, co oznacza znaczną swobodę wyboru konkretnej osoby jako respondenta (procedura określa jedynie pewne parametry tej osoby, jak płeć, zakres wieku, wykształcenie czy fakt użytkowania określonych produktów). Ankieterzy nierozumiejący istoty doboru udziałowego mogą odnieść mylne wrażenie, że skoro jest obojętne, kto konkretnie zostanie dobrany jako respondent, to i parametry doboru nie są aż tak ważne. Stąd też podstawowym błędem jest realizacją wywiadów z osobami znajomymi (co pociąga za sobą niebezpieczeństwo podobieństwa opinii badanych osób oraz ryzyko udzielania odpowiedzi społecznie akceptowanych). Drugi podstawowy błąd, to tak zwana preselekcja, a więc wykluczenie z góry pewnych osób spełniających kryteria doboru z badania (na przykład realizacja wywiadów w określonych dzielnicach, budynkach, na określonych kondygnacjach, w określonych miejscach: na dworcu, w zakładzie pracy itp., z osobami o sympatycznym wyglądzie). Dobór kwotowy z założenia zawiera pewne elementy subiektywne, wynikające z dowolności doboru osoby spośród spełniających kryteria (jest to właśnie mechanizm preselekcji).

Polega on na tym, iż ankieterzy nie przestrzegają wymogu mówiącego, iż próba kwotowa winna stanowić „miniaturę” badanej zbiorowości, a więc uwzględniać między innymi zróżnicowanie środowiskowe na danym obszarze, na przykład w mieście. Nie wystarcza, iż dobór próby jest zgodny z kryteriami, musi on także umożliwiać każdemu członkowi populacji jednakowe szanse wyboru do próby, podobnie jak w próbie losowej („równe szanse dla każdego”). I wreszcie, tak zwane naciąganie kryteriów doboru respondentów (na przykład wiek, wykształcenie) lub pomijanie niektórych z nich (przy doborze respondentów wszystkie kryteria kwotowe są jednakowo ważne). Im więcej kryteriów doboru skrzyżujemy ze sobą, tym większy lub mniejszy jest krąg potencjalnych badanych. W praktyce oznacza to, iż łatwiej bądź trudniej jest znaleźć osobę kwalifikującą się do uczestnictwa w badaniu. Wybór i liczba kryteriów doboru jest decyzją merytoryczną i może być mniej lub bardziej trafna (można sobie bowiem wyobrazić sytuację, w której praktycznie nie sposób wskazać choćby jednej osoby spełniającej kryteria – czy to ze względu na ich liczbę, czy też ich charakter). Nieprawidłowości pracy ankietera mogą więc być spowodowane wielością kryteriów (ich liczbą), ich charakterem (wykluczają się – na przykład wyższa kadra kierownicza i wieś), nieprecyzyjnym określeniem kryterium doboru (na przykład osoby używające preparaty witaminowe). Próba udziałowa/kwotowa wymaga jasnego i jednoznacznego określenia przez badacza kryteriów doboru⁶.

Do błędów związanych z realizacją próby, niezależnie od jej typu, zaliczyć należy także niezrealizowanie wywiadu ze wskazaną osobą, a więc tak zwany sufit. Określenie to odnosi się do sytuacji, w której ankieter w ogóle nie kontaktuje się z badanym (wylosowaną/wskazaną osobą), lecz cały kwestionariusz wywiadu wypełnia samodzielnie, wymyślając odpowiedź („ankieter patrzy w sufit i wymyśla”). W języku angielskim ten rodzaj fałszerstwa określa się mianem „krawężnika”, tak zwany *curbstone* („ankieter siedzi na krawężniku i wymyśla odpowiedzi”) [za: Lavrakas 2008: 267].

⁶ Należy jednak pamiętać, iż kryteriów tych nie może być zbyt wiele (w jednym z badań respondent musiał spełniać trzynaście kryteriów, aby zakwalifikować się do wywiadu). Jeśli jest ich zbyt wiele, występuje destandardyzacja sposobu próby: każdy ankieter indywidualnie hierarchizuje wagę poszczególnych kryteriów doboru.

2. 2. 2. Nieprawidłowości związane z procedurą badania

Błędy związane z procedurą badania obejmują bardzo obszerny zakres: począwszy od błędów aranżacji wstępnej, poprzez błędy zadawania pytań, zapisu odpowiedzi, a skończywszy na zapisie odpowiedzi respondenta i aranżacji końcowej. Poniżej spróbuję wskazać podstawowe błędy poszczególnych rodzajów (por. Cannell, Lawson, Hausser 1975; Kistelski 1985; Lavrakas 2008: 371–373)⁷.

A. Błędy w aranżacji wstępnej wywiadu:

- pominięcie informacji dotyczących: ankietera realizującego wywiad (imię, nazwisko), instytucji realizującej badanie, tematyki, celu badania, próby, na której realizowane jest badanie, poufności wypowiedzi;
- pokazanie (przed przystąpieniem do wywiadu) lub pozostawienie respondentowi kwestionariusza do wglądu, zmiana procedury badania;
- zaniechanie działań zmierzających do eliminacji obecności osób trzecich;
- informowanie respondenta o długości trwania wywiadu;
- składanie jakichkolwiek obietnic respondentowi celem zmotywowania go do udzielenia zgody na wywiad.

B. Błędy w zadawaniu pytań:

- zmiana procedury badania;
- niewyraźne odczytywanie pytań, zbyt szybkie odczytywanie pytań;
- „poganianie” respondenta;
- niewłaściwa intonacja w trakcie odczytywania pytań;
- odczytywanie pytań niezgodnie z ich treścią wydrukowaną w kwestionariuszu (tak zwane „opowiadanie pytań”), w tym także: odczytywanie kafeterii pytania w niewłaściwej kolejności, niekompletne odczytywanie kafeterii pytania, odczytywanie fragmentów pytania nie przeznaczonych dla respondenta;
- pomijanie instrukcji przeznaczonych dla respondenta;
- zmiana procedury pytania (niewykorzystywanie kart respondenta, skal graficznych, materiałów pomocniczych);

⁷ Problem błędów, a szerzej zachowań ankieterskich, jest przedmiotem zainteresowań nurtu *behaviour coding*. Kodowanie zachowań ankietera i respondenta wykorzystuje się nie tylko w pilotażu narzędzia, ale także przy ocenie pracy ankietera oraz ocenie danych (Lavrakas 2008: 53–55).

- „komentowanie” treści pytań lub robienie własnych „wprowadzeń” do pytania;
- pokazanie respondentowi tekstu pytania;
- odczytanie kategorii odpowiedzi w pytaniu prekategoryzowanym;
- pomijanie pytań z kwestionariusza, nieuzasadnione regułami przejścia;
- zadanie respondentowi pytania, które powinno zostać pominięte na mocy reguły przejścia;
- zmiana kolejności zadawanych pytań lub ich bloków;
- zadawanie kilku pytań jednocześnie (łączenie pytań);
- brak powtórnego odczytania pytania w przypadku nie uzyskania odpowiedzi;
- tłumaczenie treści pytania;
- sugerowanie odpowiedzi (podpowiadanie, pomaganie);
- wymuszanie na respondencie odpowiedzi (zbytne „dociskanie” respondenta);
- brak pogłębienia odpowiedzi respondenta, pomijanie pytań dodatkowych (sondujących), zbędny sondaż;
- błędny sondaż (niewłaściwie ukierunkowany, sugerujący);
- okazywanie przez ankietera braku zainteresowania wypowiedziami respondenta, lekceważenia, brak motywowania respondenta do udzielenia odpowiedzi itp.;
- „ocenie” wypowiedzi respondenta (komentowanie, okazywanie aprobaty lub dezaprobaty);
- zaniechanie działań zmierzających do eliminacji udziału osób trzecich;
- zbytne „usztywnianie” lub „rozluźnianie” atmosfery wywiadu;
- brak kontroli spójności odpowiedzi respondenta.

C. Błędy w zapisie odpowiedzi respondenta:

- posługiwanie się ołówkiem lub długopisem/piórem innego koloru niż niebieski lub czarny;
- nieczytelne notowanie odpowiedzi;
- stosowanie skrótów;
- zapis odpowiedzi w innym miejscu niż przewidziano;
- notowanie „poufnych” odpowiedzi respondenta;
- niekompletne zanotowanie odpowiedzi respondenta (zaznaczenie tylko niektórych punktów kafeterii spośród wskazanych przez respondenta lub w ogóle innych);

- niewłaściwa liczba zaznaczonych odpowiedzi (więcej niż jedna, gdy instrukcja mówi o jednej odpowiedzi);
- zmiana sensu wypowiedzi respondenta;
- „wzbogacanie” odpowiedzi respondenta poprzez jej rozwijanie;
- zapis odpowiedzi respondenta „z pamięci”;
- „zubażanie” odpowiedzi respondenta (brak zapisu dosłownego, pomijanie istotnych elementów wypowiedzi);
- pomijanie istotnych uwag i komentarzy respondenta;
- pozostawianie „pustych miejsc” (brak zapisu, iż nie zadano respondentowi danego pytania, ze względu na regułę przejść; brak zapisu, iż respondent nie udzielił odpowiedzi na pytanie);
- błędna kwalifikacja odpowiedzi respondenta w pytaniach prekategoryzowanych;
- „naciąganie” odpowiedzi respondenta (jeśli nie mieści się ona w żadnej z przewidzianych kategorii);
- brak zaznaczenia rotacji.

D. Błędy w aranżacji końcowej wywiadu:

- niesprawdzenie kompletności wypełnienia kwestionariusza;
- pominięcie pytań dodatkowych, dotyczących danych personalnych respondenta, w celu ewentualnego powtórnego kontaktu;
- brak podziękowania za udzielenie wywiadu;
- załatwianie jakichkolwiek spraw prywatnych ankietera przy okazji wywiadu.

E. Błędy w ankiecie do ankietera:

- wypełnianie ankiety po upływie dłuższego czasu (a nie bezpośrednio po wywiadzie);
- opuszczanie pytań;
- brak nazwiska, podpisu ankietera;
- brak numeru wywiadu (numeru respondenta);
- brak daty i czasu realizacji wywiadu.

F. Błędy w formularzu doboru/raporcie z pracy ankietera:

- niezgodność adresu w formularzu (karcie adresowej/imiennej) z adresem na karcie;
- brak informacji dotyczących respondenta (imię, nazwisko, adres, numer telefonu);
- brak informacji o przebiegu wizyty ankietera lub sprzeczne informacje dotyczące opisu wizyty (przykładowo nie mogą występować po sobie

następujące symbole: „Wywiad zrealizowany” i jakikolwiek inny; „Odmowa wywiadu” i „Błędny adres (nie istnieje)” i jakikolwiek inny; „Nikogo nie zastano w domu” i „Błędny adres (nie istnieje)”; „Brak możliwości udzielenia wywiadu (choroba)” i „Błędny adres (nie istnieje)” itd.);

– przebieg wizyt niezgodny z instrukcją (zbyt mała liczba wizyt w przypadku niezastania respondenta, zbyt krótki odstęp czasu między wizytami);

– brak informacji na temat charakterystyki budynku, wewnętrzna sprzeczność informacji (na przykład w budynku parterowym respondent mieszka na 1. piętrze, respondent mieszka w domu jednorodzinnym na 3. lub wyższym piętrze);

– brak informacji o karcie zapowiedniej;

– brak daty, numeru legitymacji, podpisu ankietera, nazwiska ankietera.

Dodatkowo, w przypadku próby adresowej, w karcie adresowej:

– brak lub błędne informacje o wieku, płci oraz pokrewieństwie (w stosunku do respondenta) osób zamieszkujących w danym lokalu;

– spisanie osób w niewłaściwej kolejności (błędne uszeregowanie – kolejność spisania osób – według wieku w składzie gospodarstwa domowego);

– błędny wybór respondenta;

– brak informacji dotyczącej liczby osób zamieszkujących w lokalu.

Rzeczą oczywistą jest, iż powyższa enumeracja błędów ankietera nie jest wyczerpująca. Nie jest i być nie może, ponieważ różne rodzaje badań, choćby „test produktu”, badania instytucji (*business to business*, „B2B”) stwarzają odmienne sytuacje badawcze i mogą powodować powstawanie innych różnych błędów. Dotyczy ona wyłącznie, jak wspomniałem, badań typu PAPI i CAPI, realizowanych w mieszkaniu respondenta (*in home interview*).

Jakie są konsekwencje scharakteryzowanego powyżej pośredniego podejścia w ocenie jakości danych? Polega ono, jak mówiłem, na ocenie materiałów na podstawie oceny pracy ankieterów. Przyjmuje się, że jeśli ankieterzy pracowali właściwie (poprawnie dobierali respondentów, realizowali z nimi wywiad, w sposób zgodny z instrukcją, wywiad był realizowany w sprzyjających warunkach itd.), to wartość uzyskanych danych jest wysoka. Jeśli natomiast ankieterzy nie pracowali rzetelnie, to ich jakość jest co najmniej dyskusyjna. Z punktu widzenia instytutu

badawczego jest to bardzo istotne, ale z punktu widzenia klienta (zleceńodawcy badania) – zupełnie bez znaczenia.

Dla instytutu badawczego istotne jest bowiem, czy ankieterzy pracowali poprawnie. Wyniki kontroli realizowanej metodami terenowymi, można przedstawić w formie tabeli wskazującej liczbę i charakter zastrzeżeń w stosunku do każdego z kontrolowanych wywiadów danego ankietera. Wyniki te można wykorzystać jednak nie tylko do bieżącej oceny ankieterów, ale także perspektywicznie, jako podstawę szkolenia ankieterów i współpracy z siecią. W ten sposób działania kontrolne, podejmowane po badaniach (*ex post*), wyznaczają niejako zakres i kierunek działań zmierzających do poprawy jakości realizacji (*ex ante*). Pozwalają bowiem eliminować z sieci nierzetelnych, nieuczciwych ankieterów i wskazują ewentualne ich „braki” czy też sfery „nieznajomości” zasad prowadzenia wywiadów. O ile jednak kontrola ma służyć nie tylko bieżącej ocenie pracy ankieterów, jej wyniki muszą być kumulowane i odpowiednio analizowane.

Można wskazać dwie zasadnicze osie tej analizy: czasową oraz „ankieterską”. W rezultacie mamy do czynienia z czterema podstawowymi typami porównania⁸:

1. Porównanie między sobą wszystkich kontrolowanych wywiadów danego ankietera („A”) w danym badaniu („X”). Ten typ analizy obrazowano w tabeli 6. Na jej podstawie można określić najczęściej pojawiające się uchybienia w pracy ankietera. Analizując bowiem cały ciąg wywiadów tego samego ankietera w danym badaniu, wskazać można pewne tendencje w sposobie ich realizacji. Przykładowo może to być pomijanie określonego rodzaju pytań, nieposługiwanie się kartami respondenta przy określonych pytaniach, błędy aranżacji itd. Formułowane wnioski dotyczą jednak wyłącznie danego ankietera oraz danego badania i świadczyć mogą o ewentualnych błędach szkolenia i indywidualnego przygotowania ankietera do tego właśnie badania. Nie oznacza to jednak, iż wnioski z takiej analizy nie mogą być wykorzystywane w szkoleniu danego ankietera do przyszłych badań. Przykładowo tabela obrazująca ten typ porównania może wyglądać następująco:

⁸ Na wagę i zasadność dla instytutów badawczych prezentowanych poniżej analiz wskazują Murphy, Baxter, Eyerman, Cunningham, Kennet (2004). Realizują one przede wszystkim pragmatyczno-szkoleniowe funkcje kontroli, ale mogą i są wykorzystywane do średniej oceny rezultatów badania.

Tabela 6. Wyniki kontroli 3 wywiadów ankietera A (region Aa) w badaniu „X”

Kontrolowane aspekty pracy ankietera A:	Nr wyw. xx – 001	Nr wyw. xx – 002	Nr wyw. xx – 003	Liczba zastrzeżeń
1. Dotarcie pod wybrany adres	–	+	+	1
2. Poprawny dobór respondenta	0	+	+	0
3. Realizacja wywiadu	–	+	+	1
4. Realizacja wywiadu z właściwym respondentem	0	–	+	1
5. Procedura wywiadu:				
a) respondentowi zadano pytanie (pyt. D1)	0	–	–	2
b) wykorzystano zdjęcia produktu (pyt. D1)	0	–	–	2
c) respondentowi zadano pytanie (pyt. I1)	0	–	–	2
d) respondentowi zadano pytanie (pyt. I2)	0	+	+	0
e) demonstracja produktów (pyt. M12)	0	+	+	0
6. Czas wywiadu	0	–	–	2
7. Ocena badania	0	+	+	0
8. Ocena ankietera	0	+	+	0
9. Realizacja wywiadu przez właściwego ankietera	0	+	+	0

Oznaczenia symboli w tabeli:

- symbol „+” oznacza, iż wskazany aspekt pracy ankietera, zgodnie z analizą deklaracją respondenta/informatora, nie budzi zastrzeżeń (ocena pozytywna),
- symbol „–” oznacza, że można wysunąć zastrzeżenia w stosunku do pracy ankietera (ocena negatywna),
- symbol „o” oznacza, iż praca ankietera ze względu na wskazany aspekt nie jest oceniana (nie dotyczy),
- cyfry oznaczają liczbę zastrzeżeń, które można wysunąć, zgodnie z deklaracją respondenta lub informatora, w stosunku do wywiadów danego ankietera.

Zestawienie to jest sumaryczną charakterystyką wielkości błędów „ogniskujących się” wokół procedury badania. Możemy na tej podstawie stwierdzić, iż ankieter A: (a) w ogóle nie zgłosił się pod wskazany adres i nie przeprowadził wywiadu – jeden przypadek; (b) przeprowadził wywiad z niewłaściwym respondentem („zamiana” respondenta) – jeden przypadek; (c) nie zadał pytania D1 – dwa przypadki; (d) nie zadał pytania I1 – dwa przypadki; (e) zrealizował wywiad w zbyt krótkim czasie – dwa przypadki. Ogólnie zatem można stwierdzić, iż ankieter nie pracował

rzetelnie: jeden z wywiadów został sfałszowany, kolejny zrealizowany z niewłaściwym respondentem, a w obu, które zrealizował, opuszczał część pytań. Na tym poziomie interesująca może się także okazać analiza odsetka przypadków niezrealizowania wywiadu przez ankietera, tzn. odsetka odmów udzielenia wywiadu, niedotarcia do respondenta itp.

2. Porównanie wszystkich kontrolowanych wywiadów poszczególnych ankieterów (A, B i C) w danym badaniu („X”). Ten typ analizy obrazuje tabela 6. Umożliwia ona porównanie pracy, a tym samym efektów szkolenia różnych ankieterów. Analizując ich uchybienia w pracy, możemy stwierdzić, czy były one wynikiem błędów szkolenia do badania (jeśli na przykład u wszystkich ankieterów pojawiają się takie same uchybienia) czy też wynikiem braku bądź błędnego przygotowania indywidualnego ankietera (jeśli są one różne i pojawiają się tylko u niektórych ankieterów). Jednocześnie wskazują one, na jakie elementy szkolenia oraz pracy ankietera należy położyć szczególny nacisk. Przygotowując dla każdego z kontrolowanych ankieterów tabele analogiczne, jak zaprezentowane powyżej, otrzymujemy tabele zbiorcze, dające całościowy obraz pracy ankieterów w danym badaniu. Tego typu tabela może wyglądać następująco:

Tabela 7. Wyniki kontroli pracy ankieterów A; B; C w badaniu „X” (region Aa)

Kontrolowane aspekty pracy ankieterów (A, B, C)	Ankieter A (3 wyw.)	Ankieter B (5 wyw.)	Ankieter C (8 wyw.)
1. Dotarcie pod wybrany adres	1	0	0
2. Poprawny dobór respondenta	0	0	1
3. Realizacja wywiadu	1	0	0
4. Realizacja wywiadu z właściwym respondentem	1	0	0
5. Procedura wywiadu:			
a) respondentowi zadano pytanie (pyt. D1)	2	1	0
b) wykorzystano zdjęcia produktu (pyt. D1)	2	1	0
c) respondentowi zadano pytanie (pyt. I1)	2	0	0
d) respondentowi zadano pytanie (pyt. I2)	0	0	0
e) demonstracja produktów (pyt. M12)	0	0	0
6. Czas wywiadu	2	0	0
7. Ocena badania	0	0	1
8. Ocena ankietera	0	0	0
9. Realizacja wywiadu przez właściwego ankietera	0	0	0

Uwaga. Cyfry oznaczają liczbę zastrzeżeń, które można wysunąć, zgodnie z deklaracją respondenta lub informatora, w stosunku do wywiadów danego ankietera.

Odnosząc z kolei liczbę zastrzeżeń do liczby skontrolowanych wywiadów danego ankietera, otrzymamy średnią liczbę zastrzeżeń na jeden skontrolowany wywiad (w przykładzie wynosi on odpowiednio: $11/3 = 3,66$; $2/5 = 0,40$ oraz $2/8 = 0,25$). Wartość ta pozwala stworzyć ogólny ranking ankieterów ze względu na „jakość” ich pracy, rankingu ankieterów w ramach poszczególnych regionów.

3. Porównanie między sobą wszystkich kontrolowanych wywiadów danego ankietera (A) w kolejnych badaniach („X” i „X1”). Ten rodzaj analizy obrazuje z kolei poniższa tabela.

Tabela 8. Wyniki kontroli pracy ankietera A w badaniu „X” i „X1” (region Aa)

Kontrolowane aspekty pracy ankietera A	Badanie „X”	Badanie „X1”
	Ankieter A (3 wyw.)	Ankieter A (7 wyw.)
1. Dotarcie pod wybrany adres	1	0
2. Poprawny dobór respondenta	0	0
3. Realizacja wywiadu	1	1
4. Realizacja wywiadu z właściwym respondentem	1	1
5. Procedura wywiadu:		
a) respondentowi zadano pytanie (pyt. D1)	2	3
b) wykorzystano zdjęcia produktu (pyt. D1)	2	1
c) respondentowi zadano pytanie (pyt. I1)	2	1
d) respondentowi zadano pytanie (pyt. I2)	0	1
e) demonstracja produktów (pyt. M12)	0	1
6. Czas wywiadu	2	2
7. Ocena badania	0	1
8. Ocena ankietera	0	1
9. Realizacja wywiadu przez właściwego ankietera	0	i

Uwaga. Cyfry oznaczają liczbę zastrzeżeń, które można wysunąć, zgodnie z deklaracją respondenta lub informatora, w stosunku do wywiadów danego ankietera.

Sporządzając podobną tabelę, można określić najczęściej pojawiające się uchybienia w pracy ankietera, które można określić mianem „trwałych tendencji”. Analizując bowiem ciągi wywiadów tego samego ankie-

tera w kolejnych badaniach, można wskazać pewne schematy w sposobie ich realizacji. Przykładowo może to być pomijanie określonego rodzaju pytań, nieposługiwanie się kartami respondenta przy określonych pytaniach czy fałszowanie wywiadów lub ich realizacja z niewłaściwym respondentem. Trzeba jednak pamiętać, iż formułowane wnioski dotyczą wyłącznie danego ankietera i właściwie dopiero na tym etapie mogą stać się podstawą do podjęcia określonych działań o charakterze szkoleniowym lub restrykcyjnym.

4. Porównanie wszystkich kontrolowanych wywiadów poszczególnych ankieterów A, B i C w kolejnych badaniach („X” i „X1”). Tego typu analizy dają nam całościowy obraz pracy wszystkich ankieterów w kilku badaniach i pozwalają wychwycić „trwałe tendencje”. Formułowane wnioski dotyczą poszczególnych ankieterów, jednakże relatywizowane są do pracy całej sieci. Nowym aspektem tej analizy jest także to, iż dopiero na tym etapie możemy zaobserwować ewentualne błędy i uchybienia w pracy z całą siecią ankieterską. Jeżeli bowiem okazuje się, iż określone zastrzeżenia „proceduralne” pojawiają się wobec wielu ankieterów, nie świadczy to o ich świadomych błędach, lecz błędach szkolenia lub niedostatecznym przeszkoleniu. Ten rodzaj analizy obrazuje tabela 10.

Powyżej przedstawiłem cztery możliwości analizy wyników badań kontrolnych według ankieterów oraz badań. Oczywiście, analogiczne analizy można prowadzić nie tylko dla poszczególnych ankieterów, lecz także zbiorcze, dla okręgów ankieterskich (zarówno dla danego badania „X”, jak i kilku badań: „X”, „X1” itd.). Podobne zestawienie może mieć następującą postać:

Tabela 9. Wyniki kontroli pracy ankietarów A, B i C w badaniu „X” i „X1” (region Aa)

Kontrolowane aspekty pracy ankietarów (A, B, C) w badaniach „X” i „X1” /	Badanie „X”			Badanie „X1”		
	Ankieter A (3 wyw.)	Ankieter B (5 wyw.)	Ankieter C (8 wyw.)	Ankieter A (7 wyw.)	Ankieter B (10 wyw.)	Ankieter C (12 wyw.)
1. Dotarcie pod wybrany adres	1	0	0	1	0	0
2. Poprawny dobór respondenta	0	0	1	0	0	1
3. Realizacja wywiadu	1	0	0	1	0	0
4. Realizacja wywiadu z właściwym respondentem	1	0	0	1	0	0
5. Procedura wywiadu:						
a) respondentowi zadano pytanie (pyt. D1)	2	1	0	2	1	0
b) wykorzystano zdjęcia produktu (pyt. D1)	2	1	0	2	1	0
c) respondentowi zadano pytanie (pyt. I1)	2	0	0	2	0	0
d) respondentowi zadano pytanie (pyt. I2)	0	0	0	0	0	0
e) demonstracja produktów (pyt. M12)	0	0	0	0	0	0
6. Czas wywiadu	2	0	0	2	0	0
7. Ocena badania	0	0	1	0	0	1
8. Ocena ankietera	0	0	0	0	0	0
9. Realizacja wywiadu przez właściwego Ankietera	0	0	0	0	0	0

Uwaga. Cyfry oznaczają liczbę zastrzeżeń, które można wysunąć, zgodnie z deklaracją respondenta lub informatora, w stosunku do wywiadów danego ankietera.

Tabela 10. Wyniki kontroli pracy ankieterów badaniu „X” w poszczególnych regionach (średnia liczba zastrzeżeń na 1 kontrolowany wywiad)

Okręg	Liczba skontrolowanych wywiadów	Liczba zastrzeżeń	Średnia liczba zastrzeżeń
1. A	16	15	0,94
2. B	25	45	1,80
3. C	20	40	2,00
4. D	15	45	3,00
Ogólna średnia liczba zastrzeżeń	76	145	1,91

Prowadząc jednakże tego typu „rankingi” (ankieterów czy regionów ankieterskich), należy pamiętać o ich podstawowym ograniczeniu, a mianowicie o tym, iż każdemu z zastrzeżeń przypisuje się tę samą wagę. Oznacza to, iż tak samo traktuje się sytuację, w której przykładowo ankieter, zgodnie z deklaracją respondenta, w ogóle nie realizował wywiadu zasadniczego lub zrealizował go z niewłaściwą osobą oraz taką, w której został negatywnie oceniony przez respondenta. W związku z powyższą uwagą celowe wydaje się wyłączenie z rankingu przypadków niezrealizowania wywiadu, odnotowując ten fakt w tabeli⁹.

Mówiąc o pośredniej ocenie uzyskanych w badaniu danych, wskazane rodzaje analiz winny być prowadzone przez ośrodek badawczy nie na poziomie ankieterów czy regionów ankieterskich, a więc na poziomie „makro”. Stanowiąc winny jednak podstawę przyszłych szkoleń ogólnych i problemowych organizowanych centralnie czy też w poszczególnych regionach. Kontrola podejmowana po badaniach (*ex post*), wyznacza więc zakres i kierunek działań zmierzających do poprawy jakości realizacji (*ex ante*). Przy czym, co warto podkreślić, z takich analiz wypływają też wnioski dla ośrodka. Powtarzające się uchybienia na poziomie całej sieci świadczyć mogą bowiem w równym stopniu o niedostatkach szkolenia, jak i błędach przygotowania narzędzia badawczego czy jego

⁹ Skoro tak, to należy rozstrzygnąć, jak postąpić, gdy wywiad danego ankietera nie jest autentyczny. Bez wątplenia wywiad taki należałoby wyeliminować z opracowania. Jednakże w przypadku większości ośrodków przyjęcie takiego trybu „traktowania” wywiadów sfalszowanych jest niemożliwe. Charakter realizowanych przez nie badań (głównie marketingowych i opinii) wymusza bowiem konieczność szybkiego opracowania uzyskiwanych materiałów, co praktycznie wyklucza możliwość wykorzystania wyników kontroli, a więc i ewentualnej eliminacji z analizy wątpliwych przypadków.

niedostosowaniu w sensie społecznym, psychologicznym lub metodologicznym.

Poza wskazanymi funkcjami pragmatyczno-szkoleniowymi i pośredniej oceny danych, kontrola, jak się wydaje, może mieć także jeszcze inne cele, które można określić jako prewencyjne i motywacyjne. Z jednej strony fakt prowadzenia systematycznej kontroli i świadomość bycia kontrolowanym zapobiega pokusie ułatwiania sobie pracy przez ankieterów, a z drugiej skłania ich i motywuje do rzetelnego wykonywania obowiązków i ciągłego doskonalenia swojego warsztatu. Kontrola daje także wielu ankietom podstawę do swoistej „nobilizacji” i pozytywnego „odróżnienia się”. Jednocześnie stanowi przejaw zainteresowania ośrodka ewentualnymi trudnościami i problemami, które napotykają ankierzy z pracy w terenie i sposobami ich rozwiązywania.

2. 3. Ocena jakości danych: podejście bezpośrednie

Jak już wspomniałem, podejście drugie, zwane bezpośrednim, bazuje na ocenie rezultatu pracy ankiera (uzyskanego wyniku), nie zaś samej ocenie jego pracy. Polega ona na ocenie wartości odpowiedzi respondenta na wybrane pytania poprzez powtórne ich zadanie. Badania, jak wiadomo, są typem pomiaru, opartym na subiektywnej ocenie ankiera bądź raportowaniu stanu rzeczy przez respondenta i stąd występujące w nim błędy. Ponieważ rzadko kiedy kwestionariusz wywiadu pozwala na oszacowanie tych błędów poprzez powtórzony pomiar (na przykład w kwestionariuszu zawarte są te same pytania, zadawane w różnych fazach wywiadu), funkcję tę może spełnić powtórne zadanie części pytań z kwestionariusza zasadniczego w badaniu kontrolnym. Kontrola pozwala wówczas ocenić wartość zebranych w badaniu zasadniczym materiałów nie pod kątem pracy ankierów, ale ze względu na ich rzetelność (zgodność udzielanych przez respondentów odpowiedzi). Jest to zatem nic innego jak „powtórzony pomiar”, a więc ocena jego rzetelności rozumianej jako „stabilność”. Takie ujęcie wydaje się w pełni uzasadnione, jeśli przyjmiemy, że rzetelność pomiaru jest koniecznym warunkiem jego trafności. Idea ta wywodzi się z prostej linii z pozytywistycznych korzeni badań ilościowych, których jedną z cech jest replikowalność wyników: wynik uznaje się za pewny, jeśli pomiary przeprowadzone w tych

samych warunkach dają te same rezultaty (Heisenberg 1975: 55; Bryman 1995: 37–38; Jones 2000: 28–29). Ma ona służyć określeniu błędu badania: pomiar jest rzetelny, jeśli jest powtarzalny¹⁰. Aby zatem ocenić rzetelność pomiaru, pomiar ten musimy powtórzyć. W badaniach ilościowych, w większości przypadków sprowadza się on do powtórnego wywiadu z tą samą osobą (*reinterview*), to znaczy powtórnego zadania respondentom wybranych pytań z wywiadu zasadniczego w trakcie wywiadu kontrolnego. Umożliwia to określenie szacunkowej rzetelności pomiaru, a więc jego ewentualny błąd. Mówiąc inaczej, jeśli pytamy dwa razy o to samo (jeden raz w badaniu zasadniczym, drugi raz w wywiadzie kontrolnym) i uzyskujemy rozbieżne odpowiedzi, to wartość tych informacji staje się problematyczna. Dwie rozbieżne opinie jednocześnie nie mogą być bowiem prawdziwe. Przykładowo weźmy pod uwagę odpowiedzi respondentów na dwukrotnie zadane pytanie o palenie papierosów.

Tabela 11. *Odpowiedzi respondentów na pytanie o palenie papierosów zadane w wywiadzie zasadniczym i kontrolnym*

Czy pali P. papierosy?	Pomiar I (wywiad zasadniczy)	Pomiar II (wywiad kontrolny)	Różnica (pkt. %)
TAK	66,7%	69,2%	+ 2,5
NIE	33,3%	30,8%	- 2,5
	100,0%	100,0%	

W efekcie rozkłady odpowiedzi uzyskane w trakcie I i II pomiaru mogą być różne. Ową różnicę określa się mianem błędu netto. W analizowanym przykładzie różnice rozkładów odpowiedzi nie są duże i wynoszą 2,5 punktu procentowego (jest to tak zwany procent błędu netto – por. Hyman 1954). W związku z tym zgodność rozkładów odpowiedzi na poziomie całej zbiorowości można ocenić na 97,5%. Nie oznacza to jednak, iż tylko w 2,5% przypadków respondenci udzielili innej odpowiedzi

¹⁰ Procedura powtórnego wywiadu ma na celu nie tylko ocenę rzetelności pomiaru, ale ogólnie oceny jakości rezultatów badania. W Stanach Zjednoczonych już w latach 50. XX wieku podjęto zakrojony na szeroką skalę program oceny jakości danych z badań realizowanych przez U. S. Bureau of the Census (Bailar 1968: 41–63). O różnych możliwościach wykorzystania z powtórnego wywiadu piszą także: Biemer, Forsman 1992: 915–923; Forsman, Schreiner 1991: 279 i n.

w wywiadzie zasadniczym i wywiadzie kontrolnym. Mogło się bowiem zdarzyć, iż ewentualne zmiany odpowiedzi mogły przebiegać w obydwu kierunkach (z odpowiedzi „tak” na „nie” i odwrotnie) i w efekcie częściowo „się znosić” w całej populacji. Stąd, mimo faktycznie większej rozbieżności odpowiedzi na poziomie jednostkowym, ich rozkłady są zbliżone. Sytuację tę obrazują dane zawarte w tabeli 12.

Tabela 12. *Odpowiedzi respondentów na pytanie o palenie papierosów zadane w wywiadzie zasadniczym i kontrolnym*

Czy pali P. papierosy?	Pomiar II (wywiad kontrolny)		Ogółem
	TAK	NIE	
Pomiar I (wywiad zasadniczy)			
TAK	65,4%	1,3%	66,7%
NIE	3,8%	29,5%	33,3%
Ogółem	69,2%	30,8%	100,0%

O stopniu zgodności obu pomiarów możemy mówić dopiero wtedy, gdy określamy rozbieżności odpowiedzi na poziomie jednostkowym: chodzi o tak zwany błąd brutto (por. Hyman 1954), a więc rozbieżność odpowiedzi tego samego respondenta na to samo pytanie. Stopień zgodności może stanowić podstawę oszacowania rzetelności pomiaru w badaniu zasadniczym. W powyższym przykładzie szacunkowa rzetelność pomiaru wynosi 94,9% (65,4%+29,5%). Stwierdzono 5,1% odpowiedzi rozbieżnych (3,8%+1,3%). Konieczne zatem wydaje się porównywanie rozkładów odpowiedzi uzyskanych w wyniku pierwszego i drugiego pomiaru na poziomie jednostkowym, a nie całej zbiorowości.

W dalszych analizach można obliczyć średni odsetek odpowiedzi zgodnych dla wszystkich pytań powtórzonych. Nazywam go „szacunkową rzetelnością pomiaru” i może to być najprostszy wskaźnik jakości pomiaru. Ilustruje to tabela 13 (dane dla Polski).

Tabela 13. *Odpowiedzi respondentów na wybrane pytania powtórzone zadane w wywiadzie zasadniczym ESS, Runda 4, i wywiadzie kontrolnym (ESS_k r. 4)*

Pytanie	Odsetek odpowiedzi zgodnych
1) Jak by P. określił swoje zainteresowanie polityką? Czy P. polityką ... 1. Bardzo się interesuje 2. Dość się interesuje 3. Niezbyt się interesuje 4. Czy w ogóle się nie interesuje? 8. Trudno powiedzieć	56,4%
2) Jak często – w porównaniu z innymi osobami w P. wieku – uczestniczy P. w spotkaniach, imprezach lub wydarzeniach towarzyskich? 1. Znacznie rzadziej niż większość osób w moim wieku 2. Rzadziej niż większość z nich 3. Mniej więcej tak samo 4. Częściej niż większość z nich 5. Znacznie częściej niż większość osób w moim wieku 8. Trudno powiedzieć	48,9%
3) Na ile P. zdaniem poważnym problemem w Polsce jest dyskryminacja, gorsze traktowanie ludzi ze względu na wiek – to znaczy ze względu na to, że są starzy lub młodzi? 1. Bardzo poważnym problemem 2. Raczej poważnym problemem 3. Niezbyt poważnym problemem 4. Nie jest to wcale poważnym problemem 5. To zależy 6. W Polsce nie ma dyskryminacji ze względu na wiek 8. Trudno powiedzieć	34,3%
4) Czy ma P. swój telefon komórkowy? 1. Tak 2. Nie 8. Trudno powiedzieć	91,3%
Średni odsetek odpowiedzi zgodnych:	57,7%

2. 4. Implikacje praktyczne

Powyżej przedstawiłem dwa podejścia, które można zastosować w ocenie jakości danych. Celem pierwszego z nich jest szczegółowe ustalenie okoliczności zbierania informacji w odniesieniu do danego wywiadu zrealizowanego przez danego ankietera. Przyjmuje się tutaj założenie, że jeżeli ankieterzy pracowali uczciwie, rzetelnie, to wyniki badania są wiarygodne. Z punktu widzenia badacza (klienta instytutu) tego typu pośrednia ocena wydaje się niewystarczająca, ponieważ w niewielkim tylko stopniu pozwala określić wartość uzyskanych informacji. Przyjęte założenie jest bowiem nieco ryzykowne: nie zawsze jest bowiem tak, iż poprawnie pracujący ankieter zawsze uzyskuje wiarygodne informacje. Wniosek, iż na przykład prawdopodobnie 10 wywiadów zostało zrealizowanych nierzetelnie (zostało sfabrykowanych), nie pozostaje bez wpływu na wartość materiałów, ale wpływ ten jest trudny do oszacowania. Można co prawda sytuację taką traktować jako swoisty rodzaj „efektu ankieterskiego”, który należałoby zoperacjonalizować, a tym samym zmierzyć. Zazwyczaj jednak poprzestaje się na wyeliminowaniu z opracowania wszystkich przypadków wątpliwych, jeśli oczywiście istnieje taka możliwość. Nie oznacza to jednak, że pozostała część kontrolowanych materiałów ma wysoką wartość i jest wiarygodna. Co więcej, nie wiadomo, jaki wpływ na wartość materiałów miały inne uchybienia ankieterów, nie zawsze możliwe do stwierdzenia.

Celem drugiego podejścia jest bezpośrednia ocena jakości uzyskanych danych. Dokonuje się jej na podstawie powtórzonego pomiaru, a więc porównania odpowiedzi respondenta na te same pytania, uzyskane w dwóch różnych momentach czasowych. W praktyce oznacza to konieczność ponownego kontaktu z respondentem. Aby nie ponosić dodatkowych nakładów finansowych, cele te można łączyć z kontrolą terenową, do której zobowiązany jest każdy instytut badawczy (PKJPA). Oznacza to, iż w ramach terenowej kontroli pracy ankieterów można dokonać nie tylko pośredniej, ale również bezpośredniej oceny materiału. W ten sposób badania kontrolne mogą łączyć oba podejścia (por. Sztabiński F. 1999). Decyzja dotycząca wyboru metody kontroli, a więc to, jakie to mają być procedury terenowe (na przykład wywiad osobisty, telefoniczny, karta lub ankieta pocztowa), winna być przede wszystkim uzależniona od zapotrzebowania informacyjnego badacza – odpowiedzi

na pytanie, co chcemy kontrolować i w jakim zakresie (Sztabiński F. 1995).

Nie sposób jednak, mimo wszystkich zalet, nie wspomnieć o podstawowych ograniczeniach określania jakości uzyskanych danych metodą powtórzonego pomiaru. Wspominałem o tym wcześniej, pisząc o warunkach nakładanych na badanie rzetelności metodą *test – retest*. Oto te warunki (Styczeń 1968):

- oba pomiary winny być dokonane w takim odstępie czasu, aby mierzone własności nie mogły ulec zmianie (chodzi przede wszystkim o opinie);
- dystans czasowy między dwoma pomiarami winien być na tyle długi, aby na pomiar nie miały wpływu te same, niekontrolowane czynniki (na przykład zapamiętywanie);
- niezależność pomiaru, to znaczy pierwszy pomiar nie może wpływać na przebieg i wynik drugiego pomiaru.

Jeśli warunki te nie są spełnione, procedura *test – retest* prowadzi do błędnych wniosków, ponieważ trudno oddzielić efekt zastosowania procedury od rzetelności pomiaru (Blair, Sudman 1996: 421–426; Groves, Fowler Jr., Couper, Lepkowski, Singer, Tourangeau 2004: 262). Henryk Domański i Aleksandra Dukaczewska (1996: 73–75) stwierdzają, że poziom rzetelności pomiaru, niezależnie od tego, czy jest ona pojmowana jako stabilność odpowiedzi czy ich homogeniczność, stanowi wypadkową wielu czynników:

(a) samego narzędzia badawczego, a konkretnie pytania i poprawności jego sformułowania (na przykład jasności), co sprawia, iż w sytuacji „testu” i „retestu” może być ono przez respondenta różnie rozumiane i w efekcie mogą być udzielane różne odpowiedzi;

(b) braku „niezależności” pomiaru. Pierwszy pomiar „uczula” badanego na narzędzie pomiarowe, co może wpływać na wynik drugiego pomiaru i w efekcie respondent pytany po raz drugi o to samo udziela identycznej odpowiedzi (efekt zapamiętywania);

(c) efektu „pos-testu” (*retest*), który polega na tym, iż respondent pytany po raz drugi o to samo, może udzielić innej odpowiedzi (na przykład uważa, że ze względu na powtórzenie pytania chodzi o inny aspekt zjawiska, lub weryfikację wcześniejszej odpowiedzi);

(d) sytuacji badawczej, a więc: miejsce, pora wywiadu, czas trwania i charakter interakcji, osoba ankietera itp.

(e) czynników zewnętrznnych oddziałujących na pomiar danej cechy, które mogą zmieniać wyniki drugiego pomiaru;

(f) upływu czasu między pomiarami (opinie mogą ulegać zmianie w czasie).

Można wskazać jeszcze inne czynniki wpływające na wynik powtórzonego pomiaru, jak choćby efekt ankieterski, efekt kontekstu (miejsca wśród innych powtarzanych pytań), nie mówiąc oczywiście o samej procedurze badawczej.

Rodzi się więc pytanie, na ile procedura oceny jakości danych poprzez badania kontrolne, oparta na powtórnym pomiarze, spełnia te warunki? Ogólnie biorąc, wydaje się, że tylko nieliczne. Z pewnością zachowany jest warunek dystansu czasowego między pierwszym i drugim pomiarem. Praktycznie nie występuje zatem niebezpieczeństwo wystąpienia efektu zapamiętywania wcześniejszych odpowiedzi. Jednak niekiedy dystans czasowy jest na tyle długi, iż mierzone cechy (przede wszystkim opinie) mogły ulec zmianie. Nie istnieje także niebezpieczeństwo wystąpienia efektu „pos-testu” (respondent pytany po raz drugi o to samo, udziela innej odpowiedzi, uznając, że chodzi na przykład o inny aspekt zjawiska lub weryfikację wcześniejszej odpowiedzi).

Oczywiście, nigdy nie są spełnione warunki związane z sytuacją badawczą (pora wywiadu, czas trwania, ankietów) oraz oddziaływaniem czynników zewnętrznych. Ale czy jesteśmy w stanie zagwarantować spełnienie tych warunków nawet w sytuacji badania laboratoryjnego? Chyba nie. Najpoważniejszym ograniczeniem proponowanego sposobu oceny danych jest jednak, przynajmniej w niektórych przypadkach, brak tożsamości procedury powtórzonego pomiaru. Niekiedy bowiem badanie kontrolne (a więc powtórny pomiar) realizuje się przy zastosowaniu innej techniki badawczej niż pierwotny pomiar. Nie stanowi to większego problemu, gdy wywiad zasadniczy realizowany jest techniką PAPI lub CAPI, a wywiad kontrolny techniką PAPI. Problem pojawia się, gdy pierwszy pomiar jest prowadzony przy zastosowaniu techniki PAPI lub CAPI, a drugi techniką CATI (dzieje się tak często ze względów czysto ekonomicznych, finansowych: technika CATI jest bowiem znacznie tańsza od technik wywiadu F2F). Polega on na tym, że, jak wskazują wyniki badań empirycznych, są to dwie bardzo różniące się między sobą techniki badawcze (Sztabiński 1999; Sztabiński 1999a: 51–66), a różnice dotyczą chyba nie tylko pytań o wiedzę, ale także opinie (wynika to z psychologicznych mechanizmów generowania wypowiedzi w sytuacji rozmowy twarzą w twarz oraz rozmowy przez telefon).

Wspomniane ograniczenia wpływają oczywiście na zakres i siłę wniosków płynących z podanych analiz, choć wydaje się, że ich nie dyskwalifikują. Po pierwsze, „lepiej jest wiedzieć cokolwiek, niż nic nie wiedzieć”, po drugie trzeba pamiętać, że każde badanie w socjologii jest „obrazem”, przybliżeniem rzeczywistości, nie zaś jej wiernym odwzorowaniem.

ROZDZIAŁ 3.

Bezpośrednia ocena jakości danych: analiza struktury pytania

W niniejszym rozdziale skoncentruję się na ocenie wartości danych na podstawie powtórnego pomiaru. Mówiąc o wartości danych, mam na myśli ich replikowalność, a więc rzetelność odpowiedzi (w sensie ich stabilności). Jednym z istotnych elementów moich rozważań empirycznych będzie analiza rzetelności uzyskiwanych odpowiedzi w zależności od struktury pytania kwestionariuszowego i jego przedmiotu. Innymi słowy, odpowiedź na pytanie, czy replikowalność odpowiedzi zależy od sposobu sformułowania pytania oraz tego, czego ono dotyczy (rodzaju poszukiwanej informacji). Można na przykład założyć, iż poziom zgodności odpowiedzi na pytanie o wykształcenie może być różny, w przypadku gdy raz zadajemy to pytanie jako otwarte, a raz jeśli zadajemy je jako pytanie zamknięte z kartą, na której prezentowane są możliwe odpowiedzi. Można by przyjąć, iż prawdopodobieństwo uzyskania jednakowych odpowiedzi na dwukrotnie zadane pytanie o poziom wykształcenia może być wyższe niż w przypadku dwukrotnie zadanego pytania o poczucie bezpieczeństwa osobistego (ta ostatnia odpowiedź może bowiem być uzależniona od sytuacji, miejsca i czasu jego zadania). Stąd też analizę rzetelności (zgodności odpowiedzi na pytania powtórzone) prowadzić będę w zależności od ich struktury i przedmiotu.

Poniższe analizy będą miały na celu wykazanie, iż wyniki pomiaru praktycznie nigdy nie są w pełni rzetelne (zgodne). Postaram się także wskazać ewentualne przyczyny „mniejszej” bądź „większej” zgodności pomiarów. Dodatkowo zebrany materiał pozwoli mi na stworzenie swego rodzaju benchmarku poszczególnych rodzajów pytań.

Pojęcie benchmarku wywodzi się z teorii zarządzania i określa proces oceny różnych aspektów działań w relacji do działań optymalnych – standardów (*best practice*). W *Merriam-Webster Online Dictionary* (2001) zdefiniowano „benchmark” jako „punkt odniesienia dla dokonanego pomiaru” lub „coś, co służy za standard przy ocenie innych pomiarów”. Benchmark jest zatem swoistym testem, który pozwala na ocenę uzyskanego, w szerokim rozumieniu, wyniku.

Poprzez porównanie uzyskanych rezultatów z rezultatami znanymi (istniejącymi) benchmarki mogą być także stosowane jako narzędzie monitorujące lub diagnostyczne, a tym samym traktowane jako punkt odniesienia oceny. Są zatem standardem lub średnią wartością wybranych wskaźników, obliczaną na podstawie wielu kolejnych badań realizowanych według analogicznej metodologii i powtarzalnego schematu badawczego (Good, Irwin, Jackson 2002)¹.

Pojęcie benchmarku stało się w ostatnich latach jednym z podstawowych w badaniach ilościowych w marketingu ze względu na wagę, jaką przykładają badacze i finalni odbiorcy do wyników badań. Umożliwia ono bowiem rozpatrywanie uzyskanego wyniku w kategoriach ilościowych, „obiektywnych” i wypowiadać się, czy obiekt lub aspekt podlegający ocenie jest gorszy czy też lepszy od innych, podobnych, ocenianych wcześniej. Stąd też jego znaczenie, między innymi w procesach decyzyjnych marketingu. Należy podkreślić, iż celem benchmarkingu nie jest identyfikacja „najlepszych” i „najgorszych” przypadków, jest to po prostu baza do porównywania i określania różnic, które mogą lub nie mogą być uzasadnione i prowadzić lub nie do identyfikacji potencjalnej poprawy działania.

W ujęciu statycznym benchmarking pozwala określić, na ile uzyskany przez nas wynik odbiega *in plus* lub *in minus* od średniej wartości pomiaru, a tym samym jego miejsce w hierarchii. W badaniach ilościowych prowadzonych jednokrotnie (*ad hoc*), mamy możliwość odwołania się do benchmarku tylko poprzez standaryzację narzędzia i procedury. Stanowi to teoretyczną ramę, w którą wpisywany jest za każdym razem inny problem badawczy: opinia na dany temat, określone fakty. W ujęciu dynamicznym, w perspektywie czasowej benchmarking pozwala okre-

¹ W moich analizach warunek ten nie do końca jest spełniony, ponieważ analizuję badania, które nie zawsze były realizowane według analogicznej metodologii i powtarzalnego schematu badawczego.

ślać kierunek zmian (postęp lub regres) w kolejnych pomiarach (Sowden 2002; Dyjas-Pokorska, Poławska 2004).

Zdaniem Gooda, Irwina i Jacksona (2002), o ile benchmarki są dość powszechnie stosowane, o tyle raczej mało mówi się o charakterystykach, cechach, które powinny posiadać dobre benchmarki. Jeśli mają być użyteczne, to znaczy służyć do porównań i oceny, powinny spełniać między innymi następujące warunki:

1. Powinny być proste do obliczania ze względu na liczbę przypadków w bazie. Najprostszym i najbardziej właściwym benchmarkiem, spełniającym wszystkie kryteria, jest średnia z poszczególnych badań oraz wartości maksymalna i minimalna. Benchmarki winny odzwierciedlać wartości dla poszczególnych produktów i ich grup².

2. Sposób ich tworzenia powinien być jednoznaczny.

3. Winny być tworzone w sposób umożliwiający porównywalność wskaźników.

4. Powinny być zewnętrzne w stosunku do aktualnej działalności poszczególnych podmiotów (odnosić się do rynku jako całości).

5. Powinny być stabilne w dłuższym okresie, to znaczy niepodatne na krótkoterminowe zmiany.

Wydaje się, iż w świetle sformułowanych przez Gooda, Irwina i Jacksona (2002) warunków uzasadnione wydaje się określenie prowadzonych przeze mnie analiz mianem benchmarkingu. Będą to benchmarki dla: pytań według ich struktury oraz rodzaju, ale także dla tak zwanych zastrzeżeń podstawowych w badaniach, to znaczy przypadków, w których na podstawie kontroli stwierdzono, iż wywiad zasadniczy nie został zrealizowany³. Ponieważ wskaźniki winny być „proste do obliczania”, „możliwe do stosowania” i „porównywalne”, dla dwóch pierwszych rodzajów benchmarków będą się posługiwał następującymi miarami rzetelności pomiaru:

² Mogą to być wskaźniki zarówno pojedyncze dla danej zmiennej (na przykład pytania o opinie na dany temat), jak i zbiorcze (na przykład pytania o opinie w ogóle).

³ Rozróżniam tutaj benchmarking ze względu na strukturę i przedmiot, ponieważ ta sama zmienna może być wskaźnikowana przez różnie sformułowane pytania (ze względu na strukturę), ale jednocześnie to samo pytanie ze względu na strukturę jest zazwyczaj wskaźnikiem różnych zmiennych. Dillman (1978: 94) podaje przykład pytania dotyczącego preferencji wyborczych sformułowanego na cztery różne sposoby.

– odsetek odpowiedzi zgodnych. Jest on najprostszy, w porównaniu z innymi, aczkolwiek jego wielkość zależy może od liczby wyróżnionych w danym pytaniu kategorii odpowiedzi. Stąd też dla analogicznych pytań obliczano tak zwane średnie ważone, a więc sumowano liczbę wszystkich odpowiedzi zgodnych we wszystkich pytaniach i dzielono ją przez łączną liczbę odpowiedzi⁴. Niemniej, jest on jedyną miarą, którą praktycznie można dowolnie przekształcać;

– współczynnik niepewności (wskaźnik ten będę liczył dla każdego rodzaju zmiennych). Jest to miara związku wskazująca proporcjonalną redukcję błędu, w sytuacji gdy jedna zmienna jest używana do predykcji wartości drugiej zmiennej⁵;

– V Cramera, który jest jednym z najczęściej stosowanych mierników siły związku między zmiennymi. Będę się nim również posługiwał w odniesieniu do każdego rodzaju zmiennych. Wskaźnik ten przyjmuje wartość od -1 do $+1$ ⁶;

– współczynnik korelacji τ Kendalla. Zastosuję go do analizy zmiennych porządkowych, interwałowych oraz ilorazowych. Przyjmuje on wartość od -1 do $+1$, przy czym wartości skrajne występują wyłącznie w przypadku tablic czteropolowych;

– współczynnik korelacji liniowej r Pearsona, którego zastosowanie ogranicza się w zasadzie do zmiennych interwałowych oraz ilorazowych⁷.

⁴ Biorąc jednak pod uwagę „wielopoziomowość” analiz, nie można tutaj wykluczyć wystąpienia paradoksu Simpsona, który polega, ogólnie biorąc, na tym, że zmienna ważona, która różni się od wartości określonej indywidualnie dla poszczególnych grup, jest używana do oceny połączonych grup (Sawiński 2010: 33–35).

⁵ Na przykład wartość współczynnika 0,83 mówi nam, że znajomość jednej zmiennej redukuje błąd związany z predykcją wartości drugiej zmiennej o 83%. Wskaźnik ten przyjmuje wartość od 0 do $+1$.

⁶ Nie obliczam tutaj innej miary, jaką jest Φ , ponieważ dla tablic większych niż 4-polowe przyjmuje ona wartości maksymalne większe od $|1|$ i nie daje się prosto interpretować.

⁷ Ponieważ analizuję wiele różnych badań, w których część uzyskiwanych informacji jest identyczna (na przykład wiek, wykształcenie), istnieje potrzeba prezentacji „wskaźnika zbiorczego” dla danej zmiennej. Najlepszym miernikiem jest oczywiście średnia, którą zawsze będę liczył dla wspomnianego „odsetka odpowiedzi zgodnych”. W przypadku pozostałych wskaźników nie jest to możliwe. Dla współczynnika niepewności i V Cramera, jeśli rozkład dla odpowiedzi w wywiadzie zasadniczym i kontrolnym jest zdominowany przez jeden jej typ odpowiedzi (na przykład

Wyniki moich analiz można traktować również jako „zewnętrzne w stosunku do aktualnej działalności poszczególnych podmiotów” oraz „stabilne w dłuższym okresie”. Są one bowiem – jak wspomniałem – tworzone na podstawie analizy wyników badań realizowanych przez różne organizacje badawcze: instytuty badania rynku, badania opinii, ale także placówki akademickie (odnoszą się zatem do rynku badań jako całości). Co więcej, uwzględniają dane z różnych okresów (od 1995 do 2005 roku, a więc 11 lat) i tym samym krótkookresowe wahania nie powinny znacząco wpływać na średnie współczynniki ogólne.

Wszystkie powyższe uwagi dotyczące sposobu analizy oraz benchmarkingu pytań ze względu na ich strukturę, odnoszą się oczywiście także do dalszej części rozważań, poświęconej ocenie jakości danych wyróżnionych ze względu na przedmiot.

3. 1. Dane

Przystępując do analiz empirycznych, rozporządzałem danymi z 420 badań kontrolnych (32 692 przypadki), zrealizowanych na przestrzeni czternastu lat (1992–2005) na zlecenie różnych firm, zarówno akademickich, jak i rynkowych i badania opinii publicznej (większość z nich zrealizowano na zlecenie instytutów prowadzących badania zasadnicze, ale także na zlecenie klientów tych instytutów, bezpośrednich zlecciodawców)⁸.

„tak”), to wskaźniki te są relatywnie niższe niż w sytuacji, gdy rozkład odpowiedzi jest bardziej zrównoważony. W rezultacie średnia byłaby zaniżona, co tworzyłoby fałszywy obraz. W takim przypadku będą prezentować jedynie wartość maksymalną i minimalną. Podobnie, aczkolwiek z innych względów, nie będą prezentować średniej liczonej ze wskaźników *Tau-b* Kendalla i korelacji liniowej *r* Pearsona. Wskaźniki te nie są bowiem addytywne i tym samym średnia wartość współczynników korelacji z wielu próbek nie będzie równa „średniej korelacji” w tych wszystkich próbkach. Wynika to stąd, że współczynnik korelacji nie jest liniową funkcją siły relacji między zmiennymi (Statsoft.pl, 1984–2010). Dla tych dwóch wskaźników również będą prezentować wyłącznie wartość maksymalną i minimalną.

⁸ Pragnę w tym miejscu jeszcze raz podziękować Kierownictwu firmy EPI B. A., która była bezpośrednim wykonawcą większości badań kontrolnych, Kierownictwu ORBS IFiS PAN oraz wszystkich instytutów badania rynku i opinii publicznej za udostępnienie materiałów z realizowanych na ich zlecenie badań kontrolnych.

Jest oczywiste, iż z tak bogatego materiału badawczego do dalszych analiz należało wyselekcjonować tylko pewną część. Chodziło mi o to, aby wybrane badania kontrolne były maksymalnie zróżnicowane i umożliwiały, w pewnym zakresie, generalizowanie wniosków. W związku z tym przyjąłem następujące kryteria selekcji badań kontrolnych do dalszych analiz:

(i) profil instytutu realizującego badanie zasadnicze (instytuty naukowe i akademickie oraz instytuty badania opinii i rynku);

(ii) typ (charakter) badania zasadniczego (badania jednorazowe, realizowane w mieszkaniu respondenta, tak zwane *ad hoc in home*; badania trackingowe – ciągle, realizowane w mieszkaniu respondenta; projekty stałe, tak zwane omnibusy/omnimasy realizowane w mieszkaniu respondenta; projekty panelowe także realizowane w mieszkaniu respondenta oraz badania typu *central location test*);

(iii) rodzaj i zasięg terytorialny próby (badania realizowane na próbie losowej – imiennej; na próbie losowej – adresowej; realizowane na próbie nielosowej – udziałowej oraz badania o zasięgu ogólnopolskim i zrealizowane na próbach regionalnych – o zasięgu lokalnym);

(iv) technika realizacji badania zasadniczego (PAPI, CAPI, CASAQ, PSAQ oraz obserwacja nieuczestnicząca – zewnętrzna).

Jako dodatkowe kryteria selekcji badań przyjęto:

(v) liczbę kontrolowanych przypadków z badania zasadniczego (min. 100 przypadków);

(vi) technikę realizacji badania kontrolnego (wywiad kwestionariuszowy osobisty, wywiad kwestionariuszowy telefoniczny lub wywiad kwestionariuszowy osobisty i telefoniczny).

W rezultacie do analizy dobrano materiały z kontroli terenowej projektów badawczych realizowanych w latach od 1995–2005. Jest to blisko 22 000 wywiadów z 34 badań. Zgodnie z przyjętymi kryteriami są to kontrole badań zasadniczych:

– realizowanych zarówno przez instytuty badania opinii i rynku, jak i instytuty naukowe i akademickie (w sumie 10 instytutów, odpowiednio 6 i 4),

– jednokrotnych, realizowanych w mieszkaniu respondenta, *ad hoc in home* (16 badań); 12 projektów trackingowych (ciągłych), realizowanych w mieszkaniu respondenta (projekty te obejmowały w sumie kontrolę 134 fal); 3 projekty stałe, omnibusy/omnimasy realizowane w mieszkaniu respondenta (ogółem 155 fal); 2 projekty panelowe (4 fale) realizowane także w mieszkaniu respondenta oraz 1 badanie typu *central location test*,

– realizowanych na próbie losowej – imiennej (22 badania); 5 badań, realizowanych na próbie losowej – adresowej oraz 7 badań realizowanych na próbie nielosowej – udziałowej. Większość dobranych do analizy skontrolowanych badań zasadniczych (29 badań) zrealizowana została na próbach, a 5 – na próbach regionalnych,

– realizowanych na próbie o zasięgu ogólnopolskim (29 badań) i regionalnej – o zasięgu lokalnym (5 badań),

– realizowanych techniką PAPI (32 badania), PSAQ (1 badanie – 12 fal projektu trackingowego), obserwacja nieuczestnicząca – zewnętrzna (1 badanie),

– w przypadku których liczebność próby kontrolnej wynosiła minimum 100 przypadków, co stanowiło od 1,7% do 23% wywiadów zrealizowanych w badaniu zasadniczym (średnio było to 10,9%). Próby do kontroli badań jednorazowych (*ad hoc*) liczyły od 103 do 629 losowo dobranych przypadków. Odsetek kontrolowanych wywiadów wahał się od 7,2 do 21 zrealizowanych wywiadów. W badaniach „ciągłych” próby kontrolne liczyły od 133 do 6436 przypadków. Odsetek kontrolowanych wywiadów wahał się od 1,7 do 23,1 zrealizowanych wywiadów. Dobrano też dodatkowo jedno badanie kontrolne, w którym liczba kontrolowanych przypadków nie przekroczyła 100. Wynikało to między innymi z faktu, iż w badaniu tym dysponowano danymi z 3-krotnego pomiaru (chodzi o badanie zatytułowane „Rok Rodziny”),

– w przypadku których kontrola realizowana była techniką wywiadu osobistego (4 badania), techniką wywiadu telefonicznego (1 badanie) oraz wywiadu osobistego i telefonicznego (29 badań).

W sumie zanalizuję blisko 21 892 przypadków, w tym 18 180 skontrolowanych osobiście i 3712 skontrolowanych telefonicznie⁹. W przypadku niektórych badań dysponowałem danymi z trzech róż-

⁹ Zestawienie badań i projektów dobranych do analizy zawarto w Aneksie 1. W tym miejscu należy wspomnieć, iż pewnym ograniczeniem prowadzonych przeze mnie analiz jest fakt, że w przypadku niektórych badań zasadniczych, realizowanych techniką PAPI, kontrola realizowana była techniką PAPI i CATI. Oznacza to, że porównując dane, z których część uzyskana została różnymi technikami, które niekiedy uznawane są za najbardziej odmienne spośród technik surveyowych (Sztabiński, Sztabiński 1997; Sztabiński P. 1999). Przyjmuje się jednak, iż błąd spowodowany różnicą procedur (*mode effect*) jest mniejszy niż błąd spowodowany niepełną realizacją próby i dopuszcza się łączenie różnych procedur, w tym właśnie PAPI i CATI (Roberts 2007; Eva, Widdop 2007; Roberts, Jäckle Lynn 2007).

nych źródeł: z wywiadu zasadniczego, z wywiadu kontrolnego oraz danymi dokumentalnymi, przy czym z reguły dotyczyło to danych faktualnych (płeć i wiek), zaczerpniętych z próby (był to więc schemat: test–retest–walidacja). W przypadku jednego badania dokonano trzykrotnego pomiaru innych cech niż wspomniane¹⁰. Wiązało się to ze specyficznym, quasi-eksperymentalnym schematem badania. W badaniu tym dobrano dwie porównywalne ze względu na płeć, wiek i wykształcenie próby udziałowe (A i B). Z dobranymi osobami z obu z nich zrealizowano wywiad na temat ich sytuacji zdrowotnej oraz opinii o „zdrowym sposobie życia” (Faza 1A i Faza 1B). Następnie jednej z grup (A) zaprezentowano serię filmów na temat „zdrowego sposobu życia” (bodziec). Po upływie 2–3 miesięcy z respondentami obu grup przeprowadzono analogiczny wywiad na temat ich sytuacji zdrowotnej oraz opinii o „zdrowym sposobie życia” (Faza 2). Dodatkowo wśród respondentów obu grup przeprowadzono kontrolę, w trakcie której zadano niektórym respondentom, po raz trzeci, wybrane pytania z kwestionariusza zasadniczego, które dotyczyły nie tylko cech metryczkowych (oczywiście nie mówię tutaj o cechach, które zgodnie z założeniami schematu eksperymentalnego mogły ulec modyfikacji). Dzięki temu dla grupy A i B dysponowałem danymi z trzech pomiarów: z wywiadu Faza 1, wywiadu Faza 2 oraz z wywiadu kontrolnego. W związku z tym mogłem porównywać dla obu grup (A i B) odpowiedzi respondentów uzyskane w fazie 1 i fazie 2, uzyskane w fazie 1 i w kontroli oraz uzyskane w fazie 2 i w kontroli. Był to więc klasyczny trzykrotny pomiar, a więc sytuacja, w której tym samym wybranym osobom zadawano trzy razy te same pytania.

W większości analizowanych badań standardowy kwestionariusz do badania kontrolnego (niezależnie od tego, czy był on realizowany jako wywiad bezpośredni czy przez telefon) zawierał wprowadzenie przypominające tematykę i termin realizacji badania zasadniczego oraz pytania dotyczące (por. American Statistical Association, 2003).

a. Wizyty ankietera, to znaczy czy pod wskazany adres zgłosił się ankieter. W przypadku prób adresowych z rezerwą dobieraną metodą *random route* dodatkowo oceniana była poprawność doboru adresu, na przykład kierunek „poruszania” się ankietera w terenie i poprawność „skoku” ankietera (liczby odliczanych mieszkań).

¹⁰ Chodzi tutaj o badanie „Rok Rodziny” oznaczone w Aneksie 1 numerem 10.

b. Realizacji wywiadu, to znaczy czy pod wskazanym adresem ankieter zrealizował wywiad.

c. Realizacji wywiadu z właściwą osobą, to znaczy z wylosowaną osobą (w przypadku prób adresowych i udziałowych oceniana była dodatkowo poprawność losowania/doboru respondenta).

d. Procedury wywiadu, a więc czy wywiad zasadniczo zrealizowany był zgodnie z przewidzianą techniką (na przykład zamiast PAPI – wywiad przez telefon, zamiast CAPI – wywiad PAPI), czy sytuacja i miejsce realizacji wywiadu było zgodne z założeniami (na przykład wywiad *in-home* realizowany był w kawiarni) itp.

e. Poprawności aranżacji wstępnej, to znaczy czy ankieter podczas aranżacji wstępnej przedstawił się, poinformował, jakiego instytutu jest współpracownikiem, czy pokazał legitymację ankieterską, wyjaśnił cel swojej wizyty, zasady losowania, czy zapewnił o poufności badania i poinformował o statystycznym opracowaniu wyników.

f. Przestrzegania zasad oraz sposobu prowadzenia wywiadu (określonych w podręcznikach dla ankietera, instrukcji ogólnej i instrukcjach zawartych w kwestionariuszu), w tym także czy ankieter odczytywał pytania niewyraźnie i zbyt cicho, sugerował odpowiedzi, poganiał respondenta itd.

g. Tematyki rozmowy i kompletności zadawanych pytań (według modułów i bloków pytań kwestionariusza).

h. Wykorzystania materiałów pomocniczych (kart respondenta) w przypadku poszczególnych pytań/ modułów kwestionariusza.

i. Czasu trwania wywiadu.

j. Przybliżonego terminu wizyty ankietera.

k. Realizacji wywiadu przez właściwego ankietera (chodzi o zgodność płci ankietera realizującego i sygnującego wywiad, ale także wywiady realizowane w tandemie).

l. Ogólnej oceny wywiadu i ankietera.

m. Aranżacji końcowej, a więc czy ankieter sprawdził kompletność wypełnienia kwestionariusza po zakończeniu wywiadu, czy poprosił respondenta o jego dane personalne oraz numer telefonu.

n. Opisu budynku i usytuowania mieszkania respondenta.

Niezależnie od tego, każdy kwestionariusz kontrolny zawierał blok pytań powtórzonych dotyczących:

o. Cech społeczno-demograficznych respondenta (płeć, wiek, wykształcenie, liczba osób i skład gospodarstwa domowego respondenta itp.).

p. Specyficznych zagadnień zawartych w kwestionariuszu zasadniczym (na przykład korupcji, stanu zdrowia, konsumpcji itd.).

Jest oczywiste, iż powyższy schemat kwestionariusza był każdorazowo modyfikowany, w zależności od celu i przedmiotu kontroli oraz charakteru próby. Najbardziej „rozbudowany” kwestionariusz kontrolny stosowano w badaniu Europejski Sondaż Społeczny, ponieważ jego celem była kontrola wywiadów zrealizowanych, wywiadów niezrealizowanych ze względu na odmowę respondenta oraz niezrealizowanych ze względu na brak kontaktu z respondentem. Kwestionariusz ten zamieszczony jest w Aneksie (zob. Aneks 2).

3. 2. Powtórzony pomiar: struktura pytania

Wiadomo, iż pytanie na dany – ten sam – temat sformułowane może być na różne sposoby, należy więc zanalizować rzetelność uzyskiwanych informacji (odsetek zgodności na pytania powtórzone) w zależności od ich struktury formalnej lub logicznej. Jednakże w literaturze metodologicznej zasadniczo nie funkcjonuje jeden, powszechnie przyjęty podział pytań ze względu na sposób ich sformułowania (strukturę). Należy zatem uściślić, o jakich ich rodzajach będzie mowa.

Najbardziej ogólny i powszechnie przyjęty jest podział pytań na otwarte (które nie wyznaczają dokładnie postaci zdań nadających się na odpowiedź) oraz zamknięte, w których wymienione są możliwe odpowiedzi lub dokładnie określona jest ich struktura i zakres (zob. m.in.: Ajdukiewicz 1965: 88; Babbie 2004: 270–271; Dillman 1978: 86–95; Lutyński 1972: 38–45; Sztabiński P. 1975: 259–276, 2005: 181–195). Na tym jednak zasadniczo kończy się zgodność poglądów poszczególnych autorów.

Charakteryzując te dwa typy pytań, poszczególni autorzy stosują odmienne kryteria ich podziału, a co za tym idzie, wskazują odmienne ich rodzaje. Na przykład Dillman (1978), przyjmując za podstawę charakter zadania, które ma wykonać respondent, udzielając odpowiedzi na pytanie, wyróżnia następujące rodzaje pytań zamkniętych: pytania zamknięte z uporządkowanymi alternatywami (*close-ended with ordered choices*), pytania zamknięte bez uporządkowania alternatyw (*close-ended with unordered response choices*) oraz pytania częściowo zamknięte (*partially close-ended*). Z kolei Lutyński (1972) mówi o pytaniach zamknię-

tych typu „czy” (pytania „rozstrzygnięcia”), w których wymienia się właściwe odpowiedzi, oraz pytaniach zamkniętych typu „który” (pytania dopełnienia), w których nie wymienia się właściwych odpowiedzi. Wśród pytań otwartych (w których nie są wymienione możliwe odpowiedzi) wyróżnia on: niedoskonałe pytania zamknięte, pytania żądające podwójnego wyboru, pytania żądające narracji oraz pytania nieokreślające ani zbioru alternatyw, ani liczby wskazań (alternatyw). Sztabiński (1975; 2005), mówiąc o pytaniach zamkniętych, wymienia: pytania dychotomiczne oraz pytania kafeteryjne („wieloodpowiedziowe”), wśród których wyróżnia: pytania wieloalternatywne (żądające wyboru jednej lub wielu alternatyw), pytania z szeregowaniem alternatyw oraz pytania skale (żądające jednego wyboru). Niezależnie od tego jako osobny typ wymienia on pytania prekategoryzowane oraz, oczywiście, otwarte.

Dla potrzeb prezentowanych analiz przyjmę zatem podział pytań na trzy zasadnicze kategorie, biorąc pod uwagę ich strukturę logiczną. Będą to pytania otwarte, zamknięte oraz prekategoryzowane. Jako dodatkowe kryterium przyjmę charakter zadania, które ma wykonać respondent, to znaczy wskazać jedną lub kilka (wszystkie) możliwych odpowiedzi¹¹. Analizie zostanie poddanych osiem rodzajów pytań. Będą to następujące ich rodzaje (nawiasie podaję liczbę analizowanych pytań): otwarte (7); zamknięte dychotomiczne (92); zamknięte kafeteryjne ze skalą (35); zamknięte, tak zwane pytania dopełnienia właściwe (31); zamknięte wieloalternatywne jednodpowiedziowe (13); zamknięte kafeteryjne wieloodpowiedziowe, tak zwane *multiple choice* (86); proste pytania tabelaryczne jednodpowiedziowe zawierające skalę (4); prekategoryzowane (13). Przykłady tych pytań zamieszczam, owawiając poszczególne ich rodzaje.

Analizując rzetelność odpowiedzi na poszczególne pytania wyróżnione ze względu na ich strukturę logiczną, będę się posługiwać mierznikami, o których wspominałem powyżej, omawiając wskaźniki stosowane w benchmarkingu. Będą to zatem: odsetek odpowiedzi zgodnych;

¹¹ Ponieważ moim celem nie jest prezentacja klasyfikacji rodzajów pytań ze względu na ich strukturę, nie wymieniam wszystkich ich rodzajów (pomijam kategorię pytań zamkniętych kafeteryjnych z szeregowaniem alternatyw, złożonych pytań tabelarycznych zawierających skalę itd.). Wskazuję wyłącznie te typy pytań, które znalazły się w analizowanych materiałach. W sumie można bowiem wskazać ponad 20 różnego rodzaju pytań kwestionariuszowych.

symetryczny współczynnik niepewności; współczynnik V Cramera; współczynnik τ Kendalla; współczynnik korelacji liniowej r Pearsona, o ile będą adekwatne dla danego poziomu pomiaru.

3. 2. 1. Pytania otwarte

Mówiąc o pytaniach otwartych, mam na myśli typ pytań, w których rodzaj możliwych odpowiedzi respondenta nie jest dokładnie określony, a one same (odpowiedzi) nie są zamieszczone w kwestionariuszu lub na karcie w postaci gotowych odpowiedzi, spośród których respondent może wybrać jedną lub kilka. Nie narzucają one dokładnie postaci odpowiedzi, respondentowi pozostawia się zupełną swobodę, jeśli chodzi o jej treść. Respondent musi „stworzyć” własną odpowiedź i sformułować ją własnymi słowami. Pytania te traktuję jako kategorię zbiorczą, bez wyróżniania poszczególnych rodzajów, na przykład „niedoskonałych” pytań zamkniętych, pytań żądających podwójnego wyboru czy narracji. Przykładowo tego typu pytanie może mieć następujące brzmienie: „Proszę powiedzieć, co rozumie Pan/Pani przez klasę średnią?” (badanie: „Podziały społeczne w Polsce”).

Odpowiedzi na pytania otwarte, z wywiadu zasadniczego i kontrolnego, zakodowane zostały według tego samego klucza kodowego. Wyniki analizy zgodności odpowiedzi na siedem pytań otwartych zawarto w tabeli 14.

Dla pytań otwartych ogólny odsetek odpowiedzi zgodnych wynosi 68,2%. Zwraca uwagę bardzo wysoka rozpiętość między minimalnym i maksymalnym odsetkiem zgodności, wynosząca ponad 70 punktów procentowych (od 20,0% do 92,1%). Jednocześnie nie jest to zaskakujące, biorąc pod uwagę „mechanizmy” generowania odpowiedzi na tego typu pytania i ich bardziej „spontaniczny” charakter w porównaniu z odpowiedziami na pytania zamknięte. Dodatkowo wynika to stąd, iż pytania te traktuję, o czym już wspominałem, jako kategorię zbiorczą, bez wyróżniania poszczególnych ich rodzajów. Tym samym znalazły się tam pytania bardzo ogólne (na przykład wspomniane pytanie o „rozumienie klasy średniej”), jak „szczegółowe” (na przykład „zawód” lub „markę posiadanego telewizora”).

Tabela 14. Wartości mierników dla pytań wyróżnionych ze względu na ich strukturę: pytania otwarte

1. Pytania otwarte					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	20,0–92,1	0,712–0,932 ^f	0,814–0,931 ^f	0,894** ^f	c
\bar{X}	68,2	a	a	b	c

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

c nie obliczono wartości współczynnika ze względu na jego nieadekwatność do poziomu pomiaru.

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczone ze względu na jego nieadekwatność do poziomu pomiaru).

** korelacja istotna na poziomie 0,01 (test dwustronny).

3. 2. 2. Pytania zamknięte

Analizując pytania zamknięte, biorę pod uwagę jedynie sześć ich rodzajów, najczęściej występujących, a mianowicie:

– pytania dychotomiczne (tak zwane rozstrzygnięcia lub typu „czy”).

Mówiąc o pytaniach dychotomicznych, mam na myśli pytania typu „tak – nie”, a więc zawierające wzajemnie wykluczające się możliwe odpowiedzi (mogą to być także stwierdzenia prezentowane na karcie respondenta). Przykładowo, tego typu pytanie może mieć następujące brzmienie: „Czy głosował Pan/Pani w referendum w sprawie przystąpienia Polski do Unii Europejskiej?: Tak – Nie” (badanie: „Mieszkańcy Warszawy” – SIRS);

– pytania ze skalą. Pytania zamknięte ze skalą zawierają przynajmniej trzy możliwe odpowiedzi, tworzące *continuum*, z których należy wybrać jedną. Może to być skala semantyczna, graficzna, numeryczna lub „mieszana”. Pytanie takie może mieć następujące brzmienie: „Chciałbym się dowiedzieć jak Pan/Pani ocenia różne okresy rozwoju kraju. Czy Pana/Pani zdaniem, zmiany wprowadzone w Polsce po 2001 roku (tj. po objęciu władzy przez SLD i partie koalicyjne) przyniosły większości ludzi: (1) same korzyści; (2) więcej korzyści niż strat; (3) tyle samo korzyści co strat; (4) więcej strat niż korzyści; (5) same straty (badanie: „Przemiany społeczne”);

– pytania kafeteryjne jednoodpowiedziowe to pytania, w których należy dokonać wyboru jednej możliwości spośród przedstawionych (na przykład „najważniejsza przyczyna”). Tego typu pytanie może mieć następujące brzmienie: „Przypuśćmy, że poproszono Pana/Panią o zaliczenie siebie do jednej spośród pięciu klas lub warstw społecznych. Do której z nich zaliczył(a)by Pan/Pani siebie – do klasy: (1) Klasy niższej; (2) Klasy robotniczej; (3) Klasy średniej; (4) Wyższej klasy średniej; (5) Klasy wyższej” (badanie: „Podziały społeczne w Polsce”);

– tak zwane właściwe pytania dopełnienia (pytania typu „który?”). Są to pytania, w których nie wymienia się możliwych odpowiedzi, jednak ich zbiór jest ściśle określony, a one same wzajemnie się wykluczają. Tego typu pytanie może mieć następujące brzmienie: „Proszę powiedzieć, jaki jest Pana/Pani wzrost” (badanie: „Nasze zdrowie, nasze dolegliwości”).

– pytania kafeteryjne wieloodpowiedziowe. Są to pytania, w których spośród możliwych odpowiedzi (minimum trzech) respondent ma za zadanie wybrać więcej niż jedną (dwie, trzy lub też wszystkie, które odnoszą się do jego sytuacji). W tego rodzaju pytaniach możliwe odpowiedzi wzajemnie się nie wykluczają (wybór jednej nie uniemożliwia wyboru którejkolwiek z pozostałych). Mogą one mieć następujące brzmienie: „Proszę wskazać produkty, które spożywa Pan/Pani przynajmniej od czasu do czasu: (1) Jogurt; (2) Kefir; (3) Serek homogenizowany; (4) Deser mleczny; (5) Serek wiejski/twarożek (w kubeczku); (6) Napój mleczny (badanie: „Mediolan”);

– pytania tabelaryczne jednoodpowiedziowe zawierające skale (Babie 2004: 279). Każde z pytań tabelarycznych może być przedstawione jako szereg atomicznych pytań kafeteryjnych, w których zadaniem respondenta jest ocena danego stwierdzenia na skali. Tego typu pytanie może mieć następujące brzmienie (badanie: „Mediolan”):

„Proszę powiedzieć, jak często zazwyczaj jada Pan/Pani ... (odczytywać nazwy produktów, a następnie skale).

1. Codziennie
2. 5–6 razy w tygodniu
3. 3–4 razy w tygodniu
4. 2 razy w tygodniu
5. 1 raz w tygodniu
6. 1 raz na 2 tygodnie
7. 1 raz na miesiąc
8. Rzadziej niż 1 raz w miesiącu

	1	2	3	4	5	6	7	8
1. Jogurt	1	2	3	4	5	6	7	8
2. Kefir	1	2	3	4	5	6	7	8
3. Serek homogenizowany	1	2	3	4	5	6	7	8
4. Deser mleczny	1	2	3	4	5	6	7	8
5. Serek wiejski/twarożek (w kubku)	1	2	3	4	5	6	7	8
6. Napój mleczny	1	2	3	4	5	6	7	8

Dla poszczególnych wskazanych powyżej typów pytań obliczono wskaźniki umożliwiające stworzenie benchmarkingu. Wyniki analiz przedstawiono w tabeli 15.

Średni ogólny odsetek odpowiedzi zgodnych jest najwyższy w przypadku pytań dychotomicznych i wynosi 90,4. Jest to zrozumiałe, ponieważ tego typu pytania są dla respondentów „najprostsze”, wymagają wyłącznie deklaracji: „tak–nie” lub wyboru między dwiema przeciwnymi odpowiedziami. Podobnie wysoki średni ogólny odsetek odpowiedzi zgodnych występuje w przypadku pytań zamkniętych, tak zwanych właściwych pytań dopełnienia. Wynosi on 87,5%. Należy jednak zwrócić uwagę, iż w porównaniu z pytaniami dychotomicznymi waha się on od 27,5 do 100,0, a więc jego rozpiętość jest bardzo duża.

Porównywalne wskaźniki uzyskano w przypadku pytań kafeteryjnych jednoodpowiedziowych. Średni odsetek odpowiedzi zgodnych wynosi bowiem 82,5 (waha się on od 62,4 do 98,3). Z kolei dla pytań kafeteryjnych wieloodpowiedziowych średni odsetek odpowiedzi zgodnych jest jeszcze niższy, ale zbliżony do wcześniej wskazanych. Wynosi on bowiem 77,3, ale waha się on jednak od 7,0 do 100,0. Jest to największa rozpiętość dla wszystkich pytań zamkniętych.

W przypadku pozostałych pytań wyróżnionych ze względu na ich strukturę wskaźniki były znacznie niższe. Dla pytań zamkniętych ze skalą ogólny odsetek odpowiedzi zgodnych wynosi 52,3 (waha się on od 25,2 do 92,2). Z kolei najniższe współczynniki występują w przypadku prostych pytań tabelarycznych zawierających skale. Dla pytań tych ogólny odsetek odpowiedzi zgodnych wynosi jedynie 27,0 i waha się od 21,0 do 35,3.

Tabela 15. Wartości mierników dla pytań wyróżnionych ze względu na ich strukturę: pytania zamknięte

1. Pytania zamknięte dychotomiczne					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	58,5–100,0	0,018–1,000	0,114–1,000	c	c
\bar{X}	90,4	a	a	c	c
2. Pytania zamknięte kafeteryjne ze skalą					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	25,2–92,2	0,026–0,458	0,109–0,745	0,082–0,769**	0,038–0,813**
\bar{X}	52,3	a	a	b	b
3. Pytania zamknięte kafeteryjne jednoodpowiedziowe					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	62,4–98,3	0,070–0,851	0,233–0,921	0,261**–0,955**	c
\bar{X}	82,5	a	a	b	c
4. Pytania zamknięte tzw. właściwe pytania dopełnienia					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	27,5–100,0	0,102–1,000	0,273–1,000	0,664–1,000**	0,451–1,000**
\bar{X}	87,5	a	a	b	b
5. Pytania zamknięte kafeteryjne wieloodpowiedziowe					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	7,0–100,0	0,001–1,000	0,013–1,000	0,751 ^f	c
\bar{X}	77,3	a	a	b	c
6. Proste pytania tabelaryczne jednoodpowiedziowe zawierające skale					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	21,0–35,3	0,059–0,083	0,182–0,438	0,063–0,237	c
\bar{X}	27,0	a	a	b	c

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

c nie obliczono wartości współczynnika ze względu na jego nieadekwatność do poziomu pomiaru.

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczone ze względu na jego nieadekwatność do poziomu pomiaru).

** korelacja istotna na poziomie 0,01 (test dwustronny).

3. 2. 3. Pytania prekategoryzowane

Są to pytania będące swoistym „połączeniem” pytań zamkniętych i otwartych, to znaczy są one zamknięte dla ankietera, lecz otwarte dla respondenta. Oznacza to, iż pytanie zawiera przynajmniej trzy możliwe odpowiedzi, które są zamieszczone w kwestionariuszu, których jednak nie odczytuje się respondentowi (ani nie prezentuje na karcie respondenta) – ankieter sam kwalifikuje odpowiedź respondenta do jednej z kategorii. Pytania te stosuje się, gdy nie chcemy niczego zasugerować respondentowi (podobnie jak w pytaniach otwartych). Pytania te, ze względu na niewielką ich liczbę, traktuję jako kategorię zbiorczą, bez wyróżniania poszczególnych ich rodzajów (pytania prekategoryzowane jednoodpowiedziowe i wieloodpowiedziowe).

Tabela 16. *Wartości mierników dla pytań wyróżnionych ze względu na ich strukturę: pytania prekategoryzowane*

8. Pytania prekategoryzowe					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearso
Min–Max	63,9–100,0	0,160–1,000	0,327–1,000	0,405–0,923	0,728 ^f
\bar{X}	87,4	a	a	b	b

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczone ze względu na jego nieadekwatność do poziomu pomiaru).

W przypadku tych pytań średni ogólny odsetek odpowiedzi zgodnych jest dość wysoki i wynosi 87,4. Jest on niemal identyczny jak analogiczny wskaźnik dla pytań zamkniętych, tak zwanych właściwych pytań dopełnienia. Jest to zaskakujące, ponieważ tego typu pytania są trudne zarówno dla respondentów, jak i ankieterów. Dla respondentów – ponieważ muszą sformułować odpowiedź „własnymi słowami”, co wymaga od nich pewnego wysiłku. Dla ankieterów – ponieważ wymagają od nich „dopytywania” respondentów, aby zakwalifikować ich

odpowieź do właściwej kategorii. Tak wysoka ilość wyników w powtórzonym pomiarze może świadczyć o wysokim poziomie rzetelności tych pytań bądź o błędach ankierów/kontrolerów: zamiast zadawać je jako pytania „otwarte” (dla respondenta), zadawali je jako kafeteryjne, a więc odczytywali możliwe odpowiedzi.

3. 3. Implikacje praktyczne

Powyższe wyniki mogą się stać podstawą benchmarkingu dla pytań wyróżnionych ze względu na ich strukturę. Z jednej strony chodzi o typy pytań (otwarte, zamknięte, prekategoryzowane), z drugiej zaś – o ich rodzaje (pytania otwarte; pytania zamknięte dychotomiczne, kafeteryjne ze skalą, kafeteryjne jednoodpowiedziowe, właściwe pytania dopełnienia, kafeteryjne wieloodpowiedziowe, tabelaryczne jednoodpowiedziowe zawierające skalę; oraz pytania prekategoryzowane). Przyjrzyjmy się im zbiorczo, co ułatwi tabela 17. W pierwszej kolumnie przytoczone są średnie odsetki zgodności dla wszystkich ośmiu rodzajów pytań; w drugiej przedstawiono rangi poszczególnych rodzajów pytań ze względu na średni odsetek odpowiedzi zgodnych. W kolumnie ostatniej – rangi poszczególnych typów pytań ze względu na średni odsetek odpowiedzi zgodnych.

Ogólny średni odsetek odpowiedzi zgodnych w pytaniach wyróżnionych ze względu na strukturę (otwarte, zamknięte, prekategoryzowane) jest różny i wydaje się to oczywiste. Rezultaty moich analiz dotyczących zgodności wyników powtórzonego pomiaru w zależności od struktury pytania są jednak dość zaskakujące. Po pierwsze zaskakuje relatywnie wysoki średni odsetek zgodności w przypadku pytań prekategoryzowanych. Tego typu pytania zazwyczaj są dość rzadko stosowane ze względu na „element subiektywny”, to znaczy ankierę, który dokonuje kwalifikacji odpowiedzi respondenta. Jednak zgodność odpowiedzi w powtórzonym pomiarze sięga niemal 90%. Równie zaskakujący jest wysoki średni odsetek zgodności w przypadku pytań otwartych. Zbliży się on bowiem do 70.

Odsetek odpowiedzi zgodnych na pytania zamknięte także sięga 70 (69,5). Nie oznacza to jednak, iż wyniki dla poszczególnych rodzajów pytań są homogeniczne: jeden z najniższych odsetków zgodności

Tabela 17. Benchmarki ze względu na strukturę pytania. Pytania uszeregowane według: średniego odsetka odpowiedzi zgodnych uzyskanych w powtórny pomiarze dla rodzajów pytań (kol. b) oraz średniego odsetka odpowiedzi zgodnych uzyskanych w powtórny pomiarze dla danego typu pytań (kol. c)

Struktura pytania	%	Ranga 2	Ranga 1
	zgodności	(wg rodzajów)	(wg typu)
	a	b	c
1. Pytania otwarte	68,2	6	3
2. Pytania zamknięte w tym:		69,5	
a) dychotomiczne	90,4	1	
b) kafeteryjne ze skalą	52,3	7	
c) kafeteryjne jednoodpowiedziowe	82,5	4	
d) właściwe pytania dopełnienia	87,5	2	2
e) kafeteryjne wieloodpowiedziowe	77,3	5	
f) tabelaryczne jednoodpowiedziowe zawierające skale	27,0	8	
3. Pytania prekategoryzowane	87,4	3	1

uzyskano w pytaniach zamkniętych ze skalą. Wydawać by się mogło, że są to pytania najbardziej rezystentne, które w powtórzonym pomiarze winny dawać odpowiedzi maksymalnie rzetelne. Niestety tak nie jest. Odsetek odpowiedzi zgodnych na te pytania wynosi jedynie nieco powyżej 50. Nie jest natomiast zaskoczeniem bardzo niski średni odsetek zgodności na pytania tabelaryczne zawierające skale (27). Tego typu pytania „skłaniają” bowiem respondów do udzielania odpowiedzi w sposób „mechaniczny”. Wniosek ten jest zgodny z powyższym, dotyczącym skal: w przypadku pytań ze skalami uzyskujemy relatywnie niski odsetek zgodności (na poziomie 50%) i jest on dodatkowo „zaniżony” w przypadku pytań tabelarycznych, zadawanych według tego samego schematu. Wynika to prawdopodobnie z tego, iż pytania te są nużące dla respondenta.

Ogólnie można powiedzieć, iż uszeregowanie średniego odsetka odpowiedzi zgodnych uzyskanych w powtórzonym pomiarze według poszczególnych rodzajów pytań nie jest zgodne z ich uszeregowaniem według typu (otwarte, zamknięte i prekategoryzowane). Oznacza to, iż

benchmarkingi nie są wewnętrznie spójne i tym samym odnosić je należy do konkretnego rodzaju pytań. Innymi słowy, jeśli tworzymy benchmarki dla zgodności odpowiedzi na pytania wyróżnione ze względu na ich strukturę, nie możemy operować wyłącznie bardzo ogólnym podziałem na pytania otwarte, zamknięte i prekategoryzowane. W szczególności dotyczy to pytań zamkniętych, wśród których odsetek zgodności odpowiedzi na pytania powtórzone waha się od 90 do 27. Musimy odwoływać się do konkretnego rodzaju pytania – na przykład pytań dychotomicznych, dopełnienia, kafeteryjnych itd. – nie zaś ich typów.

Rozdział 4.

Bezpośrednia ocena jakości danych: analiza przedmiotu pytania

W niniejszym rozdziale także skoncentruję się na ocenie wartości danych na podstawie powtórnego pomiaru. Teraz jednakże skupię się na analizie rzetelności uzyskiwanych odpowiedzi w zależności od przedmiotu pytania kwestionariuszowego. Innymi słowy na tym, czy replikowalność odpowiedzi zależy od tego, czego ono dotyczy (jaki jest rodzaj poszukiwanej informacji). Jak wspomniałem, można założyć, iż prawdopodobieństwo uzyskania jednakowych odpowiedzi na dwukrotnie zadane pytanie o poziom wykształcenia, może być wyższe niż w przypadku dwukrotnie zadanego pytania o poczucie bezpieczeństwa osobistego (ta ostatnia odpowiedź może być bowiem uzależniona od sytuacji, miejsca i czasu jego zadania). Stąd też analizę rzetelności (zgodności odpowiedzi na pytania powtórzone) prowadzić będę w zależności od przedmiotu pytania.

Będę chciał nie tylko wykazać, iż wyniki pomiaru praktycznie nigdy nie są w pełni rzetelne (zgodne), lecz zamierzam także wskazać ewentualne przyczyny „mniejszej” bądź „większej” zgodności pomiarów. Dodatkowo zebrany materiał pozwoli mi na stworzenie swojego benchmarku dla poszczególnych rodzajów pytań. W analizach stosuję te same wskaźniki zgodności pomiaru, co w poprzedniej części, mając świadomość ich ograniczeń czy niedoskonałości.

4. 1. Powtórzony pomiar: przedmiot pytania

Ponieważ w literaturze metodologicznej zasadniczo nie funkcjonuje jedna klasyfikacja pytań ze względu na ich przedmiot (to, czego dotyczy

pytanie), należy uściślić, o jakich ich rodzajach będzie mowa. Biorąc pod uwagę przedmiot (poszukiwaną informację), najczęściej wymienia się – przeciwstawiając je sobie – dwa rodzaje pytań: pytania o fakty i o opinie (zob. na przykład Kopczyńska-Jaworska 1971: 62; Przybyłowska 1983: 46). Pierwsze to po prostu pytania dotyczące pewnych „zdarzeń” z życia badanej osoby lub jej otoczenia. Takimi są przykładowo pytania metryczkowe (o wiek, wykształcenie, zawód itp.). Pytania o opinie, to z kolei pytania o to, jak dana jednostka postrzega określone fakty, ocenia, jak zamierza się wobec nich zachować. Przykładem mogą być pytania o ocenę warunków pracy, zadowolenia z pracy, prestiż wykonywanego zawodu. Wskazany podział jest jednak bardzo uproszczony i wydaje się nie spełniać warunków poprawności logicznej: nie jest wyczerpujący (nie obejmuje wszystkich istotnych kategorii) i nie jest rozłączny (istnieją trudności w zakwalifikowaniu konkretnych pytań do jednego i tylko jednego rodzaju). Jak na przykład zakwalifikować pytania o wiedzę, znajomość różnego typu marek produktów (chodzi o tak zwane pytania o znajomość spontaniczną i wspomaganą). Czy pytania o przyszłe zachowania (na przykład zamiar spędzenia urlopu za granicą) jest pytaniem o fakty czy opinie? Trudno też zgodzić się z koniecznością zaliczenia pewnych pytań o cechy respondenta czy cechy kontekstowe (na przykład o aktywność zawodową, skład gospodarstwa domowego) do pytań faktualnych. Podobnie jest w przypadku pytań dotyczących zachowań: czy wszystkie zachowania są faktami?

Nieco bardziej rozbudowany podział pytań ze względu na ich przedmiot funkcjonuje w literaturze anglojęzycznej. Tam dzieli się je według czterech typów informacji: postawy, przekonania, zachowania oraz cechy (zob. na przykład Dillman 1978: 79–85; Fishbein 1967: 257–266). Jest to zatem podział ze względu na przedmiot – to, czego dotyczy pytanie. Pytania o postawy (*attitudes*) to, najogólniej mówiąc, pytania o to, „co ludzie czują w jakiejś sprawie”, a odpowiedzi na nie są wartościujące i odzwierciedlają pogląd respondenta na temat stopnia pożądania przedmiotu postawy. Pytania o przekonania to pytania o to, „co jest prawdą dla ludzi”, a ponieważ dotyczą one zawsze przeszłości, teraźniejszości lub przyszłości, sprowadzają się do stwierdzenia, czy coś istniało, istnieje lub będzie istniało, czy też nie (w związku z tym, w niektórych przypadkach pytania te służą sprawdzeniu wiedzy na dany temat). Kolejny rodzaj pytań, to pytania o zachowania, o to, „co ludzie robią”. Ponieważ dotyczą zachowań przeszłych, teraźniejszych lub przyszłych, w gruncie rzeczy są to pytania

o „przekonania” na temat zachowań. I wreszcie pytania o cechy, które dotyczą tego, „jacy ludzie są”, a więc ich charakterystyk demograficznych lub społecznych (na przykład płeć, wiek, wykształcenie) i służą do analizy zróżnicowania innych typów informacji (postawy, zachowania itp.). Zaproponowany przez tych autorów podział pytań nie jest także wolny od wad. Na jego niedoskonałość wskazują zresztą sami autorzy. W wielu przypadkach trudno jest bowiem rozróżnić pytania „o postawy” od pytań „o przekonania”, czasami pytanie „o zachowanie ” może być w rzeczywistości pytaniem „o przekonania”. Dodatkowo należy wskazać problem z zakwalifikowaniem do jednej z powyższych kategorii wspomnianych pytań o wiedzę lub fakty. Trudno też zgodzić się, że pytanie dotyczące roku rozpoczęcia pracy zarobkowej jest pytaniem „o cechy”.

Dlatego też, na potrzeby mojej analizy przyjmę podział pytań na pięć kategorii (Sztabiński F. 2005: 55–77). Będę zatem mówił o pytaniach o: opinie, fakty, zachowania, cechy oraz o wiedzę. Pytania o opinie to pytania o postawy, przekonania, poglądy itp. Przykładem może być pytanie: „Jak Pan/Pani sądzi, czy ludzie powinni planować, ile będą mieli dzieci?”. Drugi rodzaj, to pytania o fakty (na przykład wspomniane pytanie: „W którym roku rozpoczął Pan/rozpoczęła Pani. pracę zarobkową?”). Przykładem pytania o zachowanie jest następujące: „Jakiej marki gumę żuje Pan/Pani najczęściej?”. Pytania o cechy to z kolei pytania o wiek, wykształcenie, sytuację zawodową itd. Co oczywiste, te ostatnie, podobnie jak pytania o zachowania, mogą dotyczyć zarówno respondenta, jak i innych osób. I wreszcie pytanie o wiedzę, to przykładowo następujące pytanie: „Jakie marki gumy do żucia Pan/Pani zna? Proszę wymienić również te, które Pan/Pani zna tylko z nazwy lub o których Pan/Pani słyszał/a”. Analizując wyniki powtórzonego pomiaru, oprę się na powyższym podziale¹.

Wśród wspomnianych badań kontrolnych, stanowiących bazę materiałową moich rozważań, znalazły się ogółem 232 zmienne². Były to py-

¹ Rzeczą oczywistą jest, iż podział pytań o opinie i kolejne o fakty, zachowania i cechy jest pochodną rodzajów pytań zadawanych w analizowanych badaniach zasadniczych i kontrolnych. Nie przyjmuję podziału pytań na merytoryczne i metryczkowe, ponieważ dla moich rozważań jest on zbyt ogólny i mało przydatny. Oprócz tego z punktu widzenia problematyki badania to samo pytanie może być traktowane jako pytanie metryczkowe lub jako merytoryczne.

² Chodzi tu o rodzaje poszukiwanych informacji. Tym samym liczba analizowanych pytań jest znacznie większa, co wynika z tego, iż dane pytanie było zadawane

tania dotyczące: opinii (25 zmiennych); faktów (90 zmiennych); zachowań (83 zmienne); cech (28 zmiennych) oraz wiedzy (6 zmiennych).

4. 1. 1. Pytania o opinie

Samo określenie „pytania o opinie” jest bardzo ogólne, ponieważ opinie te mogą dotyczyć bardzo wielu różnych spraw. W analizie uwzględniono 25 zmiennych, które podzielono na następujące bloki:

(i) Opinie na tematy społeczno-polityczne (8 pytań). Znalazły się tu między innymi pytania dotyczące zjawisk korupcyjnych; oceny idei prywatyzacji gospodarki; zadań rządu w zakresie walki z ubóstwem oraz przynależności do klasy społecznej.

(ii) Opinie na temat roli kobiet i mężczyzn (7 pytań). Chodzi tutaj o pytania dotyczące opinii na temat roli kobiety w rodzinie i jej obowiązków; zadowolenia z wykonywania prac domowych i prowadzenia gospodarstwa oraz postrzegania roli mężczyzny.

(iii) Opinie na temat pracy zarobkowej (4 pytania). Blok ten obejmował pytania na temat wagi różnych czynników (pewność zatrudnienia, wysokość dochodów itp.) przy wyborze pracy.

(iv) Opinie na temat poczucia bezpieczeństwa osobistego i własnej sytuacji. W bloku tym znalazło się 6 pytań, dotyczących między innymi poczucia bezpieczeństwa osobistego w swojej dzielnicy i najbliższej okolicy oraz oceny sytuacji finansowej gospodarstwa domowego respondenta.

Poniżej dla poszczególnych bloków pytań o opinie przedstawiono odsetki odpowiedzi zgodnych oraz pozostałe mierniki³.

w więcej niż w jednym badaniu. Dalej będę jednak mówił zamiennie o zmiennych i pytaniach. Szczegółowy wykaz analizowanych zmiennych wraz z wartościami współczynników zawarto w Aneksie 3.

³ Należy pamiętać, że w poniższych analizach wartość pomiaru określana jest na podstawie oceny zgodności odpowiedzi na wybrane pytania, które uzyskano w trakcie powtórzonego pomiaru realizowanego w badaniach kontrolnych (w pewnym sensie jest to zatem ocena rzetelności odpowiedzi). Błąd traktowany jest tutaj zatem jako brak replikowalności pomiaru, a więc niezgodność odpowiedzi respondenta na to samo pytanie zadane w trakcie wywiadu zasadniczego i kontrolnego.

Tabela 18. Pytania o opinie: wartości mierników dla poszczególnych bloków pytań wyróżnionych ze względu na przedmiot

1. Opinie na tematy społeczno-polityczne					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	39,4–78,0	0,029–0,268	0,138–0,474	0,107–0,604**	0,099–0,469**
\bar{X}	53,8	a	a	b	b
2. Opinie na temat roli kobiet i mężczyzn					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	26,1–62,8	0,033–0,108	0,152–0,299	0,079–0,336**	0,078–0,394**
\bar{X}	44,6	a	a	b	b
3. Opinie na temat pracy zarobkowej					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	47,3–73,1	0,026–0,916	0,109–0,886	0,082–0,146	0,038–0,200**
\bar{X}	55,8	a	a	b	b
4. Opinie na temat poczucia bezpieczeństwa osobistego i ocena własnej sytuacji					
	% odp. Zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	56,6–93,5	0,152–0,458	0,357–0,745	0,445**– 0,769**	0,447**–0,813**
\bar{X}	72,3	a	a	b	b

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

** korelacja istotna na poziomie 0,01 (test dwustronny).

Jak można zauważyć, wartości współczynników dla poszczególnych bloków zmiennych są dość zróżnicowane. Relatywnie najwyższe są one w przypadku pytań o poczucie bezpieczeństwa osobistego i własną sytuację. W przypadku tych pytań średni odsetek odpowiedzi zgodnych wynosi 72,3% i jest on o 15,7 punktów procentowych wyższy od średniej dla całego bloku pytań o opinie (56,6%). Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 36,9 punktu procentowego, co świadczy o znacznym ich zróżnicowaniu. Relatywnie wyższe wartości współczynników można zaobserwować

w przypadku pytań dotyczących „pośredniej” oceny sytuacji materialnej (na przykład pytanie o „zainteresowanie zakupem samochodu”: 93,5% odpowiedzi zgodnych). Z kolei niższe wskaźniki występują w przypadku pytań o „bezpośrednią” ocenę sytuacji finansowej gospodarstwa domowego respondenta (66,2%) oraz poczucie bezpieczeństwa osobistego po zmroku w okolicy zamieszkania (57,1%). Blisko 30-punktowa różnica zgodności odpowiedzi między „pośrednią” i „bezpośrednią” oceną sytuacji materialnej zdaje się wynikać z różnic w sposobie operacjonalizacji tej zmiennej. W pierwszym przypadku (ocena „pośrednia”) pytanie było sformułowane bardzo konkretnie i brzmiało: „Proszę powiedzieć, czy w ciągu najbliższych 3 miesięcy planuje się w Pana/Pani gospodarstwie domowym zakup samochodu? Tak/Nie” (badanie: „Panel C – 113, 114). W drugim zaś zastosowano pytanie z 4-punktową skalą słowną, od: „nie mamy żadnych problemów finansowych” do „utrzymujemy się z trudem” (badanie: „Mieszkańcy Warszawy” – SIRS).

Niższe wartości współczynników zaobserwować można w przypadku pytań dotyczących opinii na temat pracy zarobkowej oraz opinii na tematy społeczno-polityczne. Średni odsetek odpowiedzi zgodnych na pytania w tych blokach wynosi odpowiednio 55,8 i 53,8 i jest on niewiele niższy od średniej dla całego bloku pytań o opinie (odpowiednio 0,8 i 2,8 punktu procentowego). Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tych blokach wynosi 25,8 i 38,6 punktu procentowego, co także świadczy o znacznym ich zróżnicowaniu. W bloku pytań o pracę zarobkową relatywnie wyższe wartości współczynników można zaobserwować w przypadku tych dotyczących wykonywanego zawodu w przyszłości (73,1 % odpowiedzi zgodnych), niższe natomiast w pytaniach o wagę różnych elementów przy wyborze pracy (na przykład „możliwość łączenia obowiązków rodzinnych i zawodowych” – 47,3% odpowiedzi zgodnych). Być może związane jest to z „hipotetycznym” charakterem tych pytań. Było ono bowiem sformułowane w następujący sposób: „Gdyby dokonywał/a Pan/Pani wyboru pracy, na ile ważna byłaby dla Pana/Pani osobiście każda z następujących spraw?”; badanie Europejski Sondaż Społeczny, Runda 2).

W przypadku bloku pytań o kwestie społeczno-polityczne relatywnie wyższe wartości współczynników można zaobserwować w przypadku pytań dotyczących wprowadzenia górnej granicy dochodów oraz przynależności do klasy społecznej (odpowiednio: 78,0% oraz 66,8 % odpowiedzi zgodnych). Najniższe wartości współczynników można natomiast

zaobserwować w pytaniu o ocenę słuszności idei prywatyzacji polskiej gospodarki (39,4% odpowiedzi zgodnych) oraz pytaniach dotyczących zjawisk korupcyjnych (chodzi o pytania dotyczące: „częstotliwości załatwiania sprawy »po znajomości«” oraz „oceny działań państwa w zakresie zwalczania korupcji”). W przypadku tych ostatnich pytań odsetek odpowiedzi zgodnych wynosi odpowiednio 42,3 i 43,8. Tak niski odsetek zgodności może być związany z przedmiotem pytania. W pierwszym przypadku dotyczyło ono spraw bardzo ogólnych, abstrakcyjnych i trudnych dla respondentów – idei prywatyzacji. Stąd też mogli oni udzielać zupełnie przypadkowych odpowiedzi, różnych w pierwszym i powtórny pomiarze. W drugim przypadku pytania dotyczyły spraw drażliwych – korupcji. Owa drażliwość problematyki, ale także odmiennosc charakteru kontaktu ankietera z respondentem oraz kontrolera z respondentem mogły powodować tak znaczne różnice uzyskiwanych w powtórny pomiarze odpowiedzi. W badaniu kontrolnym relacje te mają charakter bardziej oficjalny (ze względu na krótkotrwałość kontaktu) i w większym stopniu generują odpowiedzi bardziej zrównoważone i bardziej neutralne, ostrożne. Świadczą o tym niższe odsetki odpowiedzi mówiących o dużym zasięgu tych zjawisk oraz znaczny – nawet trzykrotny wzrost odpowiedzi „trudno powiedzieć”

Najniższe wskaźniki zgodności odpowiedzi zaobserwowano w przypadku pytań o rolę kobiet i mężczyzn. Średni odsetek odpowiedzi zgodnych wyniósł tu tylko 44,6 (waha się on od 26,1 do 62,8) i jest o 12 punktów procentowych niższy od średniej ogólnej dla całego bloku pytań o opinie. Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 36,7 punktu procentowego, co świadczy o dużym ich zróżnicowaniu. Relatywnie wyższe wartości współczynników można zaobserwować w przypadku pytań dotyczących „postrzegania roli respondenta w rodzinie, gdy wchodził w dorosłe życie” (odsetek odpowiedzi zgodnych – 62,8) oraz opinii na temat „postrzegania prowadzenia gospodarstwa domowego jako pracy” (53,3% odpowiedzi zgodnych). Najniższe współczynniki zgodności występują w przypadku pytań o „zadowolenie respondenta z wykonywanych prac domowych” (26,1%) oraz postrzeganie „prac domowych jako formę wykorzystywania kobiet” (36,0%). Tak niska „replikowalność” odpowiedzi na te pytania może się wiązać z wpływem czynników sytuacyjnych, zmiennych, które mogą mieć wpływ na wygłaszane opinie. Jeśli więc na przykład respondent dzień przed wywiadem miał do wykonania bardzo dużo „ciężkich”

prac domowych (sprzątanie, prasowanie, gotowanie itp.), to z pewnością nie zgodził się ze stwierdzeniem, że: „Czynności domowe są źródłem mojego zadowolenia”. I odwrotnie, jeśli ostatnio intensywność ich wykonywania była „niższa” i nie były one zbyt uciążliwe, to można sądzić, że respondent skłonny był zgodzić się z tym stwierdzeniem.

Na zakończenie części poświęconej analizie pytań o opinie warto przedstawić wyniki powtórzonego pomiaru z dwóch fal badania, które dotyczyło opinii na temat marek słodczy. W pytaniu tym respondent miał za zadanie wskazać marki produktów (14 i 13 marek), do których pasuje dane stwierdzenie (powtórzono 8 stwierdzeń)⁴. Poniżej przedstawiono wyniki (odsetki zgodności odpowiedzi) powtórzonego pomiaru. Dla każdego stwierdzenia analizowałem: (a) zgodność liczby wskazanych marek, do których pasuje dane stwierdzenie, oraz (b) zgodność liczby i nazw marek produktów. Taki sposób analizy materiałów wydaje się konieczny, ponieważ porównując dane z dwóch pomiarów, może się okazać, iż przy zgodnej liczbie wskazanych marek nie są zgodne ich nazwy (respondent wskazał tyle samo marek, ale są to inne marki).

Tabela 19. *Odsetek zgodności liczby wskazanych marek produktów oraz zgodności marek produktów wskazanych w dwóch falach badania*

Opinie	% zgodności liczby wskazanych marek produktów		% zgodności liczby i nazw wskazanych marek produktów	
	Fala 1	Fala 2	Fala 1	Fala 2
1. Mają bardzo dobre reklamy	20,7	11,1	13,8	8,3
2. Mają doskonały smak	20,7	13,9	17,2	3,9
3. To moja ulubiona marka	13,8	38,1	3,2	19,9
4. Są produktem dla całej rodziny	10,3	16,7	3,2	11,1
5. Są delikatne i miękkie	13,8	13,9	6,9	5,6
6. Rozpływają się w ustach	13,8	27,8	6,9	5,6
7. Mają mnóstwo świetnych, smacznych dodatków	10,3	27,8	3,4	16,7
8. Są produktem na co dzień	37,9	27,8	6,9	8,3
Średni %	17,7	22,2	7,7	11,2

⁴ Pytanie to nie zostało uwzględnione w powyższych analizach.

Wyniki analizy tego pytania są dość zaskakujące. Okazuje się bowiem, że zgodność liczby wskazanych produktów waha się od 17,7% do 22,2% (średnia: 20%), a odsetek odpowiedzi zgodnych (zgodność liczby i nazw marek produktów) w obu falach badania był bardzo niski i oscylował w granicach 10 (odpowiednio: 7,7 i 11,2, średnia wynosi 9,5). Odsetek ten jest praktycznie równy odsetkowi odpowiedzi rozbieżnych na niektóre pytania omówione wyżej, przykładowo pytanie o zainteresowanie zakupem samochodu, czy opinii na temat prowadzenia samochodu po spożyciu piwa (odsetek odpowiedzi rozbieżnych odpowiednio: 6,5 i 7,8). Oznacza to, iż wartość danych z tych dwóch fal badania jest bardzo niska. Można sądzić, iż tak duży błąd pomiaru spowodowany został błędem narzędzia (pytanie zawierało 13 nazw marek produktu oraz ponad 30 stwierdzeń), który generował błąd respondenta (zmęczenie i znużenie prowadzące do udzielania odpowiedzi przypadkowych). Być może jest jednak tak, że w przypadku pytań dotyczących spraw istotnych dla respondentów, w których mają oni wyrobioną, ugruntowaną opinię, zgodność odpowiedzi, a tym samym stabilność pomiaru jest wysoka, niska zaś – w przypadku spraw mniej ważnych dla respondentów, takich właśnie jak ocena marek produktów. Problem polega jednak na tym, iż rzadko można stwierdzić, czy dana opinia jest ważna, ugruntowana, stała czy też nie. Jest to jednak problem wykraczający poza założone ramy badań. Warto jednak wspomnieć, że postawy, niezależnie od ich przedmiotu, są „konstruktami temporalnymi”, niezależnie od stopnia ich stabilności czy stałości (Wilson, Hodges 1992: 37–65). Co więcej, zależne są od kontekstu, sytuacji, w której się manifestują, a w przypadku badań surveyowych – deklarowane.

4. 1. 2. Pytania o fakty

Podobnie jak określenie „pytania o opinie”, termin „pytania o fakty” jest bardzo ogólny. Dlatego też należy analizowane pytania podzielić na bloki. W analizie uwzględniłem 90 zmiennych, które zgrupowałem w następujące bloki tematyczne:

(i) Posiadanie przez respondenta świadectwa ukończenia szkoły danego poziomu/nabycia kwalifikacji (4 pytania), w tym: (a) otrzymanie przez respondenta świadectwa ukończenia szkoły podstawowej, średniej szkoły zawodowej lub liceum ogólnokształcącego (3 pytania); (b) otrzymanie przez respondenta świadectwa ukończenia szkoły danego poziomu (1 pytanie).

(ii) Sytuacja zawodowa (5 pytań), w tym: (a) zawód, stanowisko respondenta (2 pytania); (b) sytuacja zawodowa innych osób: zawód ojca oraz matki, gdy respondent miał 14 lat (2 pytania) oraz (c) pracę któregoś z domowników w określonej branży (1 pytanie).

(iii) Posiadanie dóbr trwałego użytku i ich charakterystyka (31 pytań), w tym: (a) posiadanie telefonu stacjonarnego/komórkowego, jego rodzaj (telefon prywatny/służbowy); wyposażenia gospodarstwa domowego w telefon stacjonarny i możliwość korzystania z automatycznej sekretarki, operatorzy telefonii stacjonarnej, z usług których korzysta gospodarstwo domowe respondenta (7 pytań); (b) posiadanie i charakterystyka samochodu (marka i model, rok dopuszczenia samochodu do ruchu – pierwszej rejestracji, zakupu samochodu nowego/używanego oraz wspomniane pytania dotyczące obecnego przebiegu samochodu i średniego rocznego przebiegu (7 pytań); (c) posiadanie innych dóbr, w tym między innymi: zamrażarka, zmywarka do naczyń, pralka automatyczna, wieża hi-fi, komputer, dostęp do internetu itd. (17 pytań).

(iv) Charakterystyka sprzętu TV (17 pytań), w tym: (a) liczba odbiorników TV w gospodarstwie domowym (2 pytania); (b) marka, rodzaj, typ, wyposażenie w pilota, podłączenie do magnetowidu telewizora „głównego” (telewizor 1) [5 pytań]; (c) posiadanie anteny satelitarnej, jej rodzaj (indywidualna/zbiorcza) i typ (z pozycjonerem/bez pozycjonera), posiadanie tunera satelitarnego i podłączenie telewizora 1 do tunera satelitarnego (6 pytań); (d) możliwość odbioru ogólnopolskich i regionalnych stacji telewizyjnych (2 pytania); (e) posiadanie TV kablowej i podłączenie telewizora 1 do telewizji kablowej (2 pytania).

(v) Stan zdrowia respondenta: przebyte choroby i dolegliwości (25 pytań), w tym: (a) przebyte choroby (12 pytań); (b) dolegliwości reumatologiczne (13 pytań).

(vi) Inne fakty dotyczące osoby respondenta (8 pytań), w tym: (a) fakty z okresu uczęszczania do szkoły podstawowej (3 pytania); (b) wakacje i podróże lotnicze (2 pytania); (c) liczba wypalanych papierosów i inne (3 pytania).

Średni ogólny odsetek odpowiedzi zgodnych dla wszystkich powtórzonych pytań o fakty (niezależnie od ich tematyki) wynosi 87,5. Porównując wartości współczynników dla pytań o fakty ze współczynnikami dla wszystkich pytań (a więc o opinie, fakty, zachowania itd.), zauważyć można, iż są one, z wyjątkiem wartości minimalnej, ogólnie nieco wyż-

sze. Przykładowo, średni odsetek odpowiedzi zgodnych dla pytań o fakty jest wyższy od średniej ogólnej o 4,2 punkty procentowe.

Poniżej przedstawiono odsetki odpowiedzi zgodnych oraz pozostałe współczynniki dla poszczególnych bloków pytań o fakty.

Tabela 20. *Pytania o fakty: wartości mierników dla poszczególnych bloków pytań wyróżnionych ze względu na ich przedmiot*

1. Posiadanie przez respondenta świadectwa ukończenia szkoły danego poziomu/ nabycia kwalifikacji					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	87,2–100,0	0,415–0,651	0,706–0,785	b	b
\bar{X}	94,6	a	a	b	b
2a. Sytuacja zawodowa respondenta					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	74,6–99,3	0,424–0,932	0,614–0,966	b	b
\bar{X}	87,6	a	a	b	b
2b. Sytuacja zawodowa innych osób					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	88,2–99,6	0,001–0,693	0,013–0,814	b	b
\bar{X}	95,9	a	a	b	b
3a. Posiadanie dóbr trwałego użytku i ich charakterystyka: telefon stacjonarny i komórkowy					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	76,6–100,0	0,028–1,000	0,101–1,000	b	b
\bar{X}	92,9	a	a	b	b
3b. Posiadanie dóbr trwałego użytku i ich charakterystyka: samochód					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	27,5–100,0	0,461–1,000	0,689–1,000	0,706**–0,901***†	0,791**–0,974***†
\bar{X}	82,9	a	a	c	c
3c. Posiadanie dóbr trwałego użytku i ich charakterystyka: inne dobra					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	80,0–100,0	0,171–0,752	0,344–0,910	b	b
\bar{X}	91,9	a	a	b	b
4. Charakterystyka sprzętu TV					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	37,5–99,2	0,002–0,886	0,013–0,956	0,686**–0,836***†	0,728**–0,848***†
\bar{X}	90,5	a	a	c	c
5a. Stan zdrowia: przebyte choroby					
	<i>% odp. zgod.</i>	<i>Wsp. niepew.</i>	<i>V Cramera</i>	τ <i>Kendalla</i>	<i>r Pearsona</i>
<i>Min–Max</i>	86,1–99,7	0,291–0,839	0,542–0,923	b	b
\bar{X}	94,1	a	a	b	b

cd. tabeli 20

5b. Stan zdrowia: dolegliwości reumatologiczne					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	35,3–95,8	0,094–0,471	0,239–0,736	0,341**–0,497**	0,418**–0,579**
\bar{X}	60,6	a	a	b	b
6. Inne fakty					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	27,0–97,8	0,243–0,924	0,355–0,922	0,681**–0,838**	0,571**–0,800**
\bar{X}	78,2	a	a	c	c

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b nie obliczono wartości współczynnika ze względu na jego nieadekwatność do poziomu pomiaru.

c średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczono ze względu na jego nieadekwatność do poziomu pomiaru).

** korelacja istotna na poziomie 0,01 (test dwustronny).

Jak można zauważyć, wartości współczynników dla poszczególnych bloków zmiennych faktualnych są dość zróżnicowane (różnice średniego odsetka odpowiedzi zgodnych sięgają blisko 20 punktów procentowych), aczkolwiek wysokie, w porównaniu z analogicznym, dotyczącym pytań o opinie. Relatywnie najwyższe są one w przypadku pytań o posiadanie świadectwa ukończenia szkoły i nabycie określonych kwalifikacji. W przypadku tych pytań średni odsetek odpowiedzi zgodnych wynosi 94,6 i jest o 7,1 punktu procentowego wyższy od średniej dla całego bloku pytań o fakty (87,5%). Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 12,8 punktu procentowego, co świadczy o nieznacznym ich zróżnicowaniu. Relatywnie wyższe wartości współczynników można jednak zaobserwować w przypadku pytań dotyczących otrzymania świadectwa ukończenia szkoły danego typu (na przykład liceum, średniej szkoły zawodowej) niż świadectwa ukończenia szkoły danego poziomu (na przykład szkoły średniej). Różnice nie są duże, ale sięgają 7–9 punktów procentowych (odpowiednio: 94,8%, 96,5% oraz 87,2%).

Nieco niższe wartości współczynników można zaobserwować w przypadku pytań dotyczących sytuacji zawodowej respondenta i innych osób. Średni odsetek odpowiedzi zgodnych w przypadku tych pytań wynosi 90,6 i jest on nieco wyższy od średniej dla całego bloku pytań o fakty (3,1 punktu procentowego). Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 25,0 punktów procentowych, co świadczy o dość dużym ich zróżnicowaniu. Wynika to przede wszystkim z różnic wartości współczynników zgodności sytuacji zawodowej respondenta oraz innych osób. Wbrew oczekiwaniom okazuje się, że ogólnie wyższy odsetek zgodności odpowiedzi uzyskuje się w przypadku pytań o sytuację zawodową innych osób niż przy pytaniach o sytuację zawodową respondenta (odpowiednio średni odsetek zgodności wynosi: 95,9 i 87,6)⁵.

Warto tutaj jednak zwrócić uwagę na trzy sprawy. W przypadku pytań dotyczących respondenta odsetek zgodności „zaniża” pytanie o jego zawód i stanowisko: wynosi on bowiem tylko 74,6. Być może wynika to z różnic kompetencji osób pytających: wiadomo bowiem, iż jest to jedna ze zmiennych, na której wartość ma wpływ sposób zadawania pytań. W przypadku pytań dotyczących sytuacji zawodowej innych osób odsetek zgodności „zawyża” pytanie o pracę kogokolwiek z gospodarstwa domowego respondenta w określonej branży (średni odsetek odpowiedzi zgodnych wynosi 97,5). Ostatnia sprawa to wyższy (blisko o 5 punktów procentowych) odsetek zgodności odpowiedzi na pytanie o zawód ojca, gdy respondent miał 14 lat, w porównaniu z takim samym pytaniem dotyczącym matki. Nie zamierzam tutaj interpretować tego faktu, tym bardziej że różnica nie jest duża, wydaje się wszakże, iż może się to wiązać z tradycyjnym postrzeganiem roli mężczyzny i kobiety oraz ich miejsca w rodzinie.

Kolejnym blokiem pytań, w którym odsetek odpowiedzi zgodnych przekroczył średnią dla wszystkich pytań faktualnych, był blok dotyczący charakterystyki sprzętu TV. Średni odsetek odpowiedzi zgodnych był o 3,0 punkty procentowe wyższy od średniej dla wszystkich pytań o fakty. Odsetek zgodności na te pytania waha się od 37,5 do 99,2. Tak

⁵ Inne współczynniki, to znaczy współczynnik niepewności symetryczny oraz V Cramera, wskazują na odwrotną tendencję, to znaczy są wyższe w przypadku pytań dotyczących respondenta, ale może być to związane z tym, że odpowiedzi na dane pytanie są zdominowane przez jeden jej typ.

duża rozpiętość (ponad 60 punktów procentowych) nie świadczy jednak o dużym jego zróżnicowaniu: jeśli bowiem pominiemy tę wartość, to rozpiętość odsetka zgodności wynosi od 81,0 do 99,2. Ów niski odsetek zgodności, który aż tak zaniża średnią, uzyskano w pytaniu dotyczącym możliwości odbioru przez respondenta ogólnopolskich stacji telewizyjnych⁶. Być może został on spowodowany sposobem definiowania „zgodności odpowiedzi”. Jako zgodność odpowiedzi traktowano bowiem zgodność liczby stacji i ich nazwy. Przykładowo, gdy respondent w wywiadzie zasadniczym wymienia trzy stacje, a w wywiadzie kontrolnym tylko dwie spośród wskazanych wcześniej, to sytuację taką traktowano jako niezgodność odpowiedzi. Jednakże w pytaniu o możliwość odbioru przez respondenta regionalnych stacji telewizyjnych odsetek odpowiedzi zgodnych, liczonych w analogiczny sposób, wyniósł 81,0. Jedyne wyjaśnienie, które się nasuwa, jest takie, że liczba stacji ogólnopolskich jest znacznie większa niż regionalnych, które respondent może odbierać. Tym samym istnieje większe prawdopodobieństwo, że w wywiadzie zasadniczym i kontrolnym wymieni on wszystkie stacje regionalne (tzn. jedną lub dwie), niż że wymieni wszystkie stacje ogólnopolskie.

Oprócz tych dwóch bloków pytań odsetki zgodności poniżej średniej (tzn. poniżej 90%) uzyskano w przypadku pytań dotyczących charakterystyki anteny, do której podłączony jest telewizor (średnia 85,7%) oraz, co może być zaskakujące, liczby odbiorników telewizyjnych w gospodarstwie (średni odsetek zgodności wynosi 85,9). W pierwszym przypadku wyjaśnienie jest proste: nie każdy dokładnie wie, czy ma antenę wewnętrzną, zewnętrzną, zbiorczą czy indywidualną, satelitarną z pozycjonerem czy bez niego itd. W drugim przypadku poza „odwiecznym” problemem definiowania przez respondenta gospodarstwa domowego, pojawia się drugi: wliczać czy nie wliczać telewizory nieużywane, zepsute (które jednak można naprawić) itd. Tego typu „dylematy” powodują, iż odsetek odpowiedzi zgodnych na pytanie o liczbę telewizorów jest relatywnie niski. Wyższe wartości współczynników (tzn. powyżej średniej, rzędu 95% odpowiedzi zgodnych) uzyskano w przypadku pytań dotyczących charakterystyki sprzętu, to znaczy marki, rodzaju i typu „głównego” telewizora, jego wyposażenia w pilot, podłączenia do magnetowidu itd. W przypadku tych pytań odsetki zgodności sięgają nawet 99.

⁶ Wynik ten nie wydaje się przypadkowy, ponieważ rozporządzano danymi z 1075 wywiadów kontrolnych.

W kolejnym bloku pytań, który dotyczył posiadania dóbr trwałego użytku i ich charakterystyki, odsetek odpowiedzi zgodnych także przekroczył średnią dla wszystkich pytań faktycznych (różnica wynosi 2,9 punktu procentowego). Odsetek zgodności na te pytania waha się od 27,5 do 100,0 (najniższy odsetek zgodności 27,5 oraz 42,3 uzyskano w pytaniach dotyczących posiadanego samochodu: jego obecnego przebiegu oraz średniego przebiegu rocznego). Tak duża rozpiętość (ponad 70 punktów procentowych) nie świadczy jednak o dużym jego zróżnicowaniu: jeśli bowiem pominiemy wartość najniższą, to otrzymamy rozpiętość 76,6–100,0. Wydaje się, iż można to wyjaśnić jedynie tym, że większość posiadaczy samochodów „mniej więcej” wie, ile kilometrów przejeżdża w ciągu roku, natomiast ogólny przebieg samochodu zna „w dużym przybliżeniu”.

Wspomniałem, iż w bloku pytań dotyczących posiadania dóbr trwałego użytku i ich charakterystyki wprowadzono podział na baterie pytań o telefon stacjonarny i komórkowy, samochód oraz o inne dobra.

Pierwszym blokiem pytań o dobra trwałego użytku były pytania o telefon stacjonarny i komórkowy. Dla wszystkich tych pytań łącznie ogólny średni odsetek odpowiedzi zgodnych wynosi 92,9 i jest wyższy o ponad 5 punktów procentowych od średniej dla wszystkich pytań o fakty. Waha się on od 76,6 do 100,0. Najwyższy średni odsetek zgodności występuje w przypadku pytań dotyczących wyposażenia gospodarstwa domowego w telefon stacjonarny i możliwość korzystania z automatycznej sekretarki. Relatywnie wysoki jest również odsetek odpowiedzi zgodnych w pytaniu o operatorów telefonii stacjonarnej, z usług których korzysta gospodarstwo domowe respondenta. Wynosi on 93,5. Niewiele niższe są współczynniki dla pytań o posiadanie telefonu komórkowego i jego rodzaj (telefon prywatny/służbowy): średni odsetek zgodności wynosi 92,2. Najniższy średni odsetek zgodności występuje w przypadku pytań dotyczących posiadania telefonu komórkowego przez inną osobę w gospodarstwie domowym i jej rodzaj (telefon prywatny/służbowy). Jest to zrozumiałe ze względu na to, że respondent występuje tutaj jako informator o faktach dotyczących innych osób i nie zawsze dysponuje odpowiednią wiedzą na dany temat.

Dla pytań dotyczących samochodu ogólny średni odsetek odpowiedzi zgodnych wynosi 82,9 i jest niższy o blisko 5 punktów procentowych od średniej dla wszystkich pytań o fakty (uwzględniając dwa wspomniane pytania o obecny przebieg samochodu oraz średni przebieg

roczny). Wbrew oczekiwaniom, zgodność odpowiedzi na powtórzone pytanie o posiadanie samochodu nie jest pełne: wynosi jedynie 91,8%. I znowu, jak się wydaje, można to tłumaczyć problemem definiowania przez respondenta gospodarstwa domowego (samochód teścia, z którym respondent mieszka, może być uznany za samochód „wspólny”, lub nie), ale także wątpliwościami, jak traktować samochód służbowy do wyłącznej dyspozycji respondenta. Znacznie mniej wątpliwości pojawia się w sytuacji, gdy pytamy o markę i model samochodu oraz czy jest to samochód nowy czy używany (średnia odpowiedzi zgodnych wynosi 98,1% i waha się od 95,7% do 100,0%). W pytaniach o rok dopuszczenia samochodu do ruchu (pierwsza rejestracja) odsetek zgodności jest nieco niższy (79,3). Jak wspomniałem wcześniej, współczynniki dla pytań o obecny przebieg oraz średni przebieg roczny samochodu są najniższe.

Pozostałe pytania faktualne obejmowały informacje o wyposażeniu gospodarstwa domowego w kuchenkę mikrofalową, pralkę automatyczną, suszarkę automatyczną oraz posiadanie pomocy domowej (w pełnym lub w niepełnym wymiarze godzin); sprzęt audio-wideo, to znaczy wieżę hi-fi, walkmana, odtwarzacz CD, magnetowid, kamerę wideo, aparat fotograficzny, oraz sprzęt AGD, to znaczy lodówkę, zamrażarkę, zmywarkę do naczyń, komputer, w tym dostęp do internetu i miejsce korzystania przez respondenta z internetu. Dla wszystkich tych pytań łącznie ogólny średni odsetek odpowiedzi zgodnych wynosi 91,9 i jest on wyższy o ponad 4 punkty procentowe od średniej dla wszystkich pytań o fakty. Waha się on od 80,0 do 100,0.

Najwyższy średni odsetek zgodności występuje w przypadku pytań dotyczących wyposażenia gospodarstwa w sprzęt AGD oraz posiadanie pomocy domowej (w pełnym lub w niepełnym wymiarze godzin). Wynosi on 94,8, natomiast odsetek zgodności waha się od 85,7 (pytanie o wyposażenie gospodarstwa domowego w zamrażarkę) do 100,0 (posiadanie pomocy domowej w pełnym wymiarze godzin). Niewiele niższe są współczynniki dla pytań o posiadanie komputera, dostęp do internetu i miejsce korzystania przez respondenta z internetu. Średni odsetek zgodności wynosi 91,5. Odsetek zgodności waha się od 85,8 (pytanie o dostęp do internetu) do 96,6 (wyposażenie gospodarstwa domowego w komputer). Najniższy średni odsetek zgodności występuje w przypadku pytań dotyczących wyposażenia gospodarstwa domowego w sprzęt audio-wideo. Wynosi on 88,1%. Odsetek zgodności waha się

od 80,0 do 96,6 (odpowiednio, wyposażenie gospodarstwa domowego w odtwarzacz CD i kamerę wideo).

W przypadku pytań o wyposażenie materialne gospodarstwa domowego respondenta wydaje się, że jedynym z błędów może być „wpływ ankierski” i związana z tym waloryzacja: w niektórych środowiskach „wstyd” jest się przyznać, że nie posiada się telefonu lub samochodu (dobra te mogą stanowić według niektórych respondentów wyznacznik statusu społecznego lub materialnego – są to tak zwane dobra prestiżowe). Wpływ ankierski na uzyskiwane odpowiedzi może jednak mieć „odwrotny” efekt. Respondent może nie przyznać się do posiadania pewnych dóbr (samochód) z obawy przed zagrożeniem bezpieczeństwa osobistego (Sztabiński, Sztabiński 1994). Błąd pomiaru tej cechy może być także skutkiem obiektywnych zmian w wyposażeniu gospodarstwa: zakupem samochodu – pierwszego lub drugiego bądź jego/ich sprzedażą (Sztabiński F. 2001).

Kolejny blok pytań dotyczył stanu zdrowia respondenta i przebytych chorób. Średni odsetek odpowiedzi zgodnych jest niższy od średniej dla wszystkich pytań faktualnych (różnica wynosi 10,1 punktu procentowego) i wynosi 77,4. Odsetek zgodności na te pytania waha się od 35,3 do 99,7. Tak duża rozpiętość (blisko 65 punktów procentowych) świadczy o dużym zróżnicowaniu tego bloku: najniższe odsetki zgodności (poniżej 50) uzyskano w pytaniach o częstotliwość bólów mięśni, ścięgien i więzadeł, bóle odcinka lędźwiowego kręgosłupa oraz bóle kolan.

W bloku tych pytań wprowadzono podział na pytania dotyczące przebytych chorób oraz dolegliwości reumatologicznych. Pytania dotyczące przebytych przez respondenta chorób obejmowały informacje na temat: zawału serca; chorób serca lub ataku serca; nadciśnienia; wysokiego poziomu cholesterolu; raka płuc; chorób układu oddechowego; uszkodzenia wątroby; raka odbytu; perforacji przewodu pokarmowego; krwawienia z przewodu pokarmowego; choroby wrzodowej żołądka, dwunastnicy oraz otyłości, nadwagi. Dla pytań tych ogólny średni odsetek odpowiedzi zgodnych wynosi 94,1 i jest wyższy o 6,6 punktu procentowego od średniej dla wszystkich pytań i fakty. Wyższy odsetek zgodności (powyżej 95) zaobserwować można w przypadku chorób, które respondenci uznają za „poważne”, a więc takich jak rak płuc, rak odbytu, zawał serca, perforacja lub krwawienie z przewodu pokarmowego czy choroba wrzodowa. W przypadku zaś takich chorób jak nadciśnienie, wysoki poziom cholesterolu, czy otyłość, nadwaga,

odsetki zgodności są niższe i wahają się od 86,1 (otyłość), do 90,7 (nadciśnienie).

W przypadku pytań dotyczących dolegliwości reumatologicznych współczynniki zgodności są generalnie niższe. Pytania te obejmowały informacje na temat wizyty u lekarza z powodu dolegliwości reumatologicznych; rozpoznania osteoporozy; częstotliwości bólów odcinka szyjnego kręgosłupa; częstotliwości bólów odcinka piersiowego kręgosłupa; częstotliwości bólów odcinka lędźwiowego kręgosłupa; częstotliwości bólów stawu biodrowego; częstotliwości bólów kolan; częstotliwości bólów rąk; częstotliwości bólów innych stawów; częstotliwości bólów mięśni, ścięgien, więzadeł; częstotliwości bolesnych obrzęków stawów rąk; częstotliwości bolesnych obrzęków stawów stóp oraz częstotliwości bolesnych obrzęków stawów kolan. Dla pytań tych ogólny średni odsetek odpowiedzi zgodnych wynosi 60,6 i jest niższy o prawie 27 punktów procentowych od średniej dla wszystkich pytań i fakty. Odsetek zgodności waha się od 35,3 do 95,8 (odpowiednio było to pytanie o częstotliwość bólów mięśni, ścięgien i więzadeł oraz o rozpoznania osteoporozy). Wyższy odsetek zgodności (powyżej 80) wystąpił jedynie w przypadku dwóch pytań: o rozpoznanie osteoporozy oraz wizyty u lekarza z powodu dolegliwości reumatologicznych. Niższy odsetek, ale powyżej średniego (63,9–68,2) wystąpił w przypadku pytań dotyczących częstotliwości bolesnych obrzęków stawów.

Ostatnim z wyróżnionych bloków pytań są pytania o inne fakty związane z osobą respondenta. Jak wspomniałem, znalazły się tutaj pytania dotyczące posiadania przez respondenta własnego pokoju, gdy miał 14 lat; liczby uczniów w ostatniej klasie szkoły podstawowej; miejsca zamieszkania respondenta, gdy uczęszczał do szkoły podstawowej; okresu zamieszkiwania przez respondenta w danej w okolicy; liczby palonych przez respondenta papierosów; miejsca dokonywania większości zakupów żywnościowych (chodzi o rodzaj sklepu); miejsca spędzenia ostatnich wakacji przez respondenta (nazwa kraju); odbytych przez respondenta podróży lotniczych w ciągu ostatnich 12 miesięcy oraz wysokości średniego miesięcznego rachunku za telefon stacjonarny.

Średni odsetek odpowiedzi zgodnych na te pytania jest niższy od średniej dla wszystkich pytań faktualnych i wynosi 78,2%. Odsetek zgodności waha się od 27,0 do 97,8. Tak duża rozpiętość (ponad 70 punktów procentowych) nie świadczy jednak o dużym zróżnicowaniu tego bloku: bardzo niski odsetek zgodności wystąpił tylko w przypadku

pytania o wysokość średniego miesięcznego rachunku za telefon stacjonarny. Jeśli pominiemy to pytanie, najniższy odsetek zgodności wynosił 63,9 (w pytaniu o miejsce dokonywania większości zakupów żywnościowych). Najwyższe odsetki zgodności wystąpiły w przypadku pytań dotyczących miejsca spędzenia ostatnich wakacji przez respondenta w kraju) i odbytych przez respondenta niego lotniczych w ciągu ostatnich 12 miesięcy (odpowiednio: 90,7 i 97,8) oraz pytań, co jest zaskakujące, o liczbę uczniów w ostatniej klasie szkoły podstawowej (z dokładnością do 5 osób) i miejsce zamieszkania respondenta, gdy uczęszczał do szkoły podstawowej (odpowiednio: 94,0 i 97,5).

Jak można było sądzić, współczynniki dla pytań o fakty są znacznie wyższe niż dla pytań o opinie. Najwyższe są w pytaniach dotyczących posiadania przez respondenta świadectwa ukończenia szkoły danego poziomu, sytuacji zawodowej respondenta i innych osób, charakterystyki sprzętu TV oraz posiadania dóbr trwałego użytku. We wszystkich tych blokach pytań średni odsetek zgodności odpowiedzi respondenta w wywiadzie zasadniczym i kontrolnym przekracza 90. Nieco niższe średnie odsetki zgodności (77–78) wystąpiły w przypadku pytań o stan zdrowia oraz inne fakty dotyczące respondenta. W przypadku tych pierwszych należy jednak pamiętać o znacznym zróżnicowaniu współczynników dla chorób uznawanych za „ciężkie” (na przykład rak) oraz „łżejsze” (na przykład nadciśnienie).

Jednocześnie rodzi się pytanie, czy rzeczywiście odsetki zgodności rzędu 90–95 należy uznać za wysokie? Dlaczego w pytaniach o fakty współczynniki nie osiągają wartości maksymalnych? Respondent powinien wiedzieć, czy ma zmywarkę do naczyń albo kuchenkę mikrofalową. Nie są to przecież fakty dotyczące odległej przeszłości, w przypadku których w grę wchodzić mogą błędy respondenta spowodowane zawodnością pamięci. Niestety okazuje się, że w przypadku żadnej zmiennej nie możemy oczekiwać pełnej zgodności. Tezę tę potwierdzają wyniki „powtórzonego pomiaru” w badaniu, w którym zastosowano obserwację nieuczestniczącą, zewnętrzną (poniższe dane nie dotyczą więc deklaracji respondentów, lecz danych obserwowalnych). Chodzi o badanie, w którym ankieter, a następnie kontroler miał dokonać opisu nośników reklamy zewnętrznej (tak zwanych billboardów). Zadanie każdego z nich polegało między innymi na identyfikacji nośnika reklamy zewnętrznej oraz zapisaniu informacji dotyczących: jego adresu (ulica oraz numeru budynku); liczby nośników w grupie; liczby nośników według rodzajów; jego lokalizacji (parking, na budynku itp.); infrastruktury komunikacyjnej

wokół nośnika (parking, przystanek tramwajowy); otoczenia w odległości do 100 m (banki, stacje benzynowe, obiekty użyteczności publicznej, hiper- i supermarkety itp.) oraz otoczenia w odległości od 100 do 500 m, w tym: obiekty użyteczności publicznej (biblioteki, kościoły, szkoły itp.). Dla powyższych informacji średni odsetek zgodności wyniósł 94,4 i był wyższy jedynie o niespełna 7 punktów procentowych od średniego odsetka dla wszystkich pytań o fakty. Wahał się on od 86,4 do 99,9. Dla takich informacji jak adres nośnika (ulica oraz numeru budynku), liczba nośników w grupie oraz infrastruktura komunikacyjna wokół nośnika, odsetek zgodności wyniósł poniżej 90. Najwyższy jest w przypadku informacji na temat obiektów znajdujących się w otoczenia nośnika w odległości do 100 m, takich jak banki, towarzystwa ubezpieczeniowe, stacje benzynowe, obiekty użyteczności publicznej itd. (średni odsetek 99,2, dla porównania, w niektórych pytaniach o „poważne” choroby odsetek zgodności wynosi 99,7). Przytoczone dane wskazują, że wyniki obserwacji prowadzonej przez różne osoby mogą się między sobą różnić⁷. Oznacza to, że „dwoje ludzi, oglądając to samo, może widzieć różne rzeczy”. Nie powinniśmy zatem oczekiwać, że „ten sam człowiek pytany dwa razy o to samo, zawsze udzieli takiej samej odpowiedzi”.

4. 1. 3. Pytania o zachowania

Analizowane pytania o zachowania (83 zmienne) także podzielono na bloki. W analizie uwzględniono następujące bloki tematyczne:

(i) Zachowania „społeczne”, a więc udział w wyborach, referendum itp (3 pytania).

(ii) Zachowania konsumpcyjne (50 pytań), w tym konsumpcji: (a) kawy/herbaty (8 pytań); (b) piwa (4 pytania); (c) alkoholu (13 pytań); (d) słodyczy i przekąsek (8 pytań) (e) palenia wyrobów tytoniowych (4 pytania) oraz (f) dokonywania zakupów żywnościowych i konsumpcji produktów mlecznych (13 pytań).

⁷ Podobne różnice wyników obserwacji stwierdziłem podczas tak zwanych wywiadów obserwowanych. Polegały one na tym, iż ankieterowi towarzyszył obserwator. Po zakończonym wywiadzie obydwaj wypełniali ankietę do ankietera, która dotyczyła terminu wywiadu, jego długości itd. Jedynie w trzech przypadkach na 68 wszystkie informacje opisane w ankiecie przez ankietera i obserwatora były zgodne (Sztabiński F. 1996). Można zatem postawić pytanie, czy obserwacja dostarcza bardziej pewnych, wiarygodnych informacji niż deklaracje respondentów?

(iii) Zachowania „edukacyjne”, a więc uszesczczanie do placówek edukacyjnych różnego poziomu (9 pytań).

(iv) Zachowania „kulturowe”, a więc oglądanie telewizji itp. (13 pytań).

(v) Inne zachowania, wyjazdy wakacyjne, korzystanie z Internetu itp. (8 pytań).

W tabeli 21 przedstawiono odsetki odpowiedzi zgodnych oraz pozostałe współczynniki dla poszczególnych bloków pytań.

Tabela 21. Pytania o zachowania: wartości mierników dla poszczególnych bloków pytań, wyróżnionych ze względu na ich przedmiot

1. Zachowania „społeczne”					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	85,0–96,6	0,278–0,692	0,476–0,848	b	b
\bar{X}	93,7	a	a	b	b
2a. Konsumpcja kawy i herbaty					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendall a	r Pearsona
Min–Max	25,2–96,8	0,069–0,371	0,303–0,646	0,323**–0,559**	b
\bar{X}	68,5	a	a	c	b
2b. Konsumpcja piwa					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	67,3–91,5	0,052–0,511	0,202–0,772	0,265**	b
\bar{X}	84,3	a	a	c	b
2c. Konsumpcja alkoholu					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	65,6–94,2	0,018–0,92	0,114–0,719	b	b
\bar{X}	73,9	a	a	b	b
2d. Konsumpcja słodyczy i przekąsek					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	58,6–79,4	0,018–0,083	0,132–331	b	b
\bar{X}	65,7	a	a	b	b
2e. Palenie papierosów					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	90,6–95,3	0,495–0,722	0,762–0,904	b	b
\bar{X}	93,8	a	a	b	b

cd. tabeli 21

2f. Dokonywanie zakupów żywnościowych i konsumpcja produktów mlecznych					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	21,0-85,6	0,014-0,231	0,128-0,515	0,063-0,237**	b
\bar{X}	56,9	a	a	c	b
3. Zachowania „edukacyjne”					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	83,6-100,0	0,133-0,828	0,351-0,945	d	d
\bar{X}	93,3	a	a	d	d
4. Zachowania „kulturowe”					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	61,3-86,1	0,031-0,291	0,179-0,608	d	d
\bar{X}	76,3	a	a	d	d
5. Inne zachowania					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	58,5-98,3	0,080-0,466	0,305-0,589	0,515** ^f	d
\bar{X}	80,1	a	a	0,515** ^f	d

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zanizania spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b nie obliczono wartości współczynnika ze względu na jego nieadekwatność do poziomu pomiaru.

c średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczono ze względu na jego nieadekwatność do poziomu pomiaru).

** korelacja istotna na poziomie 0,01 (test dwustronny).

Średni ogólny odsetek odpowiedzi zgodnych dla wszystkich powtórzonych pytań o zachowania (niezależnie od ich tematyki) wynosi 75,6% i jest o blisko 8 punktów procentowych niższy od średniej ogólnej dla wszystkich analizowanych pytań, niezależnie od ich przedmiotu.

Jak można zauważyć, wartości współczynników dla poszczególnych bloków zmiennych są dość zróżnicowane (różnice średniego odsetka odpowiedzi zgodnych wynoszą 23 punkty procentowe). Relatywnie najwyższe są one w przypadku pytań o zachowania „społeczne”. Blok ten

obejmował pytania o udział respondenta w referendum w sprawie przystąpienia Polski do Unii Europejskiej; udział respondenta w wyborach parlamentarnych; udział respondenta lub kogoś z domowników w badaniach ankietowych (kiedykolwiek lub w ciągu ostatnich 6 miesięcy).

W przypadku tych pytań średni odsetek odpowiedzi zgodnych wynosi 93,7% i jest o 18,1 punktu procentowego wyższy od średniej dla całego bloku pytań o zachowania. Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 11,6 punktu procentowego, co świadczy o nieznacznym ich zróżnicowaniu. Relatywnie wyższe wartości współczynników można zaobserwować w przypadku pytań dotyczących udziału respondenta w referendum i wyborach parlamentarnych (odpowiednio 96,4% i 94,9%) niż udziału w badaniach ankietowych, aczkolwiek różnice są minimalne, rzędu 3–4 punktów procentowych.

Praktycznie identyczne wartości współczynników wystąpiły w pytaniach dotyczących zachowań „edukacyjnych”. Blok ten obejmował pytania o uczęszczanie przez respondenta do przedszkola; klasy zerowej; szkoły podstawowej; zasadniczej szkoły zawodowej; średniej szkoły zawodowej; liceum ogólnokształcącego; szkoły policealnej/pomaturalnej; na studia licencjackie oraz studia wyższe.

W bloku tym średni odsetek odpowiedzi zgodnych wynosi 93,3 i jest wyższy od średniej dla całego bloku pytań o zachowania o blisko 18 punktów procentowych (17,7). Maksymalna różnica między odsetkiem zgodności odpowiedzi na pytania w tym bloku wynosi 16,4 punktów procentowych, co świadczy o pewnym ich zróżnicowaniu. Wynika to przede wszystkim z różnic wartości współczynników dla pytań o różne poziomy edukacji. Są one wyższe, gdy pytamy o uczęszczanie do szkół od poziomu podstawowego do policealnego/pomaturalnego, niższe natomiast w pytaniach o edukację przedszkolną (przedszkole i klasa zerowa) oraz studia licencjackie i wyższe (średni odsetek zgodności wynosi odpowiednio: 96,9 i 88,8).

Znacznie niższy odsetek odpowiedzi zgodnych (różnica 13 punktów procentowych) wystąpił w bloku pytań dotyczących zachowań „kulturowych”. Blok ten obejmował pytania o regularne oglądanie następujących pozycji telewizyjnych: filmy pełnometrażowe; wiadomości; programy sportowe; programy historyczno-naukowe; programy kulturalne; seriale filmowe; komedie; programy dla dzieci; programy muzyczne/teledyski; filmy rysunkowe; programy religijne; programy dla majsterkowiczów oraz programy dramatyczne.

Średni odsetek odpowiedzi zgodnych w tym bloku jest praktycznie równy średniej dla wszystkich pytań o zachowania (mediana i dominanta: 77,0 i 85,0). Odsetek zgodności na te pytania waha się od 61,3 do 86,1. Znaczna jego rozpiętość (blisko 25 punktów procentowych) wskazuje na jego zróżnicowanie. Odsetek zgodności poniżej średniej dla tego bloku wystąpił w przypadku pytań o regularne oglądanie filmów pełnometrażowych, programów historyczno-naukowych oraz komedii (średni odsetek 62,5). Odsetek zgodności równy średniej dla tego bloku (76,3) wystąpił w pytaniu o oglądanie: programów muzycznych i teledysków, wiadomości, seriali firmowych oraz programów dramatycznych i kulturalnych. Odsetki zgodności powyżej średniej wystąpiły w przypadku programów: sportowych, religijnych, programów dla dzieci i majsterkowiczów oraz filmów rysunkowych (średni odsetek zgodności dla tych pytań wynosi 84,4).

W kolejnym bloku pytań, który dotyczył zachowań konsumpcyjnych, odsetek odpowiedzi zgodnych wynosi 70,6 i jest niższy od średniej dla wszystkich pytań o zachowania. Odsetek zgodności na te pytania waha się od 21,0 do 96,8,0. Tak duża rozpiętość (ponad 75 punktów procentowych) świadczy o dużym jego zróżnicowaniu. Aby ułatwić analizę, wszystkie zmienne dotyczące konsumpcji (ogółem 50) zgrupowano, jak wspomniałem, w baterii pytań dotyczących konsumpcji: kawy i herbaty, piwa, alkoholu, konsumpcji słodyczy i przekąsek, palenia wyrobów tytoniowych oraz dokonywania zakupów żywnościowych i konsumpcji produktów mlecznych.

Pytania dotyczące konsumpcji kawy i herbaty obejmowały informacje na temat picia przez respondenta: kawy (w ogóle); kawy naturalnej (mielonej) od czasu do czasu; częstotliwości picia kawy mielonej; kawy rozpuszczalnej od czasu do czasu; częstotliwości picia kawy rozpuszczalnej; kawy cappuccino od czasu do czasu; częstotliwości picia kawy cappuccino oraz picia herbaty.

Dla pytań dotyczących konsumpcji kawy i herbaty ogólny średni odsetek odpowiedzi zgodnych wynosi 68,5 i jest niższy o 7,1 punktu procentowego od średniej dla wszystkich pytań o zachowania i waha się on od 25,2 do 96,8. Wysokość odsetka zgodności dla pytań o picie poszczególnych rodzajów kawy jest różny dla różnych rodzajów kawy. Najniższy jest dla picia kawy cappuccino (69,5), następnie – kawy rozpuszczalnej (79,7), kawy naturalnej, mielonej (83,0) i najwyższy w pytaniach o picie kawy w ogóle (85,9). W przypadku pytań o częstotliwość picia poszcze-

gólnych rodzajów kawy uporządkowanie odsetka zgodności odpowiedzi jest identyczne. Najniższe odsetki zgodności występują dla kawy cappuccino, wyższe – dla kawy rozpuszczalnej i najwyższe – dla naturalnej, mielonej (odpowiednio: 25,2, 35,2 i 47,1). Wszystkie współczynniki dotyczące konsumpcji kawy są jednak niższe od współczynnika dotyczącego picia herbaty (odsetek zgodności wynosi 96,8).

Pytania dotyczące konsumpcji piwa obejmowały informacje na temat picia piwa przez respondenta: kiedykolwiek; w okresie ostatnich 3 miesięcy (w ogóle); w ciągu ostatniego miesiąca oraz częstotliwość picia piwa. Dla pytań tych ogólny średni odsetek odpowiedzi zgodnych jest dość wysoki i wynosi 84,3 (jest on wyższy o 8,7 punktu procentowego od średniej dla wszystkich pytań o zachowania). Waha się on od 67,3 do 91,5. Tak jak w przypadku konsumpcji kawy jest on najniższy w przypadku pytania o częstotliwość picia piwa (67,3). O ponad 20 punktów procentowych wyższe są odsetki zgodności w pytaniach o picie piwa kiedykolwiek (88,1), w ciągu ostatnich 3 miesięcy/w ogóle (85,5%) oraz w ciągu ostatniego miesiąca (91,5).

Kolejna bateria w bloku pytań o zachowania konsumpcyjne dotyczyła picia alkoholu. Były to pytania na temat picia kiedykolwiek przez respondenta: drinków alkoholowych; whisky; ginu; koniaku/winiaku; tequili; likieru; w ogóle wódki bez dodatków; miodu pitnego; szampana, wina musującego; wina; wina zagranicznego; wina krajowego owocowego oraz innych napojów alkoholowych.

Dla pytań o konsumpcję alkoholu ogólny średni odsetek odpowiedzi zgodnych nie jest tak wysoki, jak w przypadku konsumpcji piwa, i wynosi 73,9 (jest on nieco niższy, o 1,7 punktu procentowego od średniej dla wszystkich pytań o zachowania). Waha się on od 65,6 do 94,2. Relatywnie najniższe odsetki zgodności, poniżej 70, występują w przypadku pytań o picie: likieru, whisky, wina zagranicznego i krajowego oraz drinków alkoholowych. Nieco wyższe odsetki, to znaczy powyżej 70,0 zgodności, ale poniżej średniej dla tego bloku pytań (tzn. 73,9), występują w przypadku pytań o konsumpcję kiedykolwiek: ginu, koniaku, winiaku, miodu pitnego, szampana lub wina musującego oraz wódki bez dodatków. Najwyższe odsetki zgodności odpowiedzi występują natomiast w przypadku picia tequili i innych napojów alkoholowych (odpowiednio: 91,3 i 94,2).

Kolejna bateria w bloku pytań o zachowania konsumpcyjne dotyczyła słodyczy i przekąsek. Były to pytania o kupowanie i konsumpcję przez

respondenta w ciągu ostatnich 3 miesięcy: chipsów ziemniaczanych lub kukurydzianych; chrupek i innych snacków; orzeszków; czekolady i batonów czekoladowych; słodkich ciastek; słonych ciastek; cukierków oraz lodów pojedynczych i w dużych opakowaniach. Dla pytań o kupowanie i konsumpcję słodczy i przekąsek ogólny średni odsetek odpowiedzi zgodnych jest relatywnie niski i wynosi 65,7 (jest on niższy o 9,9 punktu procentowego od średniej dla wszystkich pytań o zachowania). Waha się on od 58,6 do 79,4. Średnie odsetki zgodności nie są zbyt zróżnicowane: najniższy jest w przypadku pytania o kupowanie i konsumpcję chrupek oraz innych snacków (chipsy, orzeszki, ciastka, lody, cukierki) i waha się od 58,6 do 70,1 odpowiedzi zgodnych. Wyraźnie wyższy jest on jedynie w przypadku pytania o kupno i konsumpcję czekolady i batonów czekoladowych (79,4 odpowiedzi zgodnych).

Znacznie wyższe i bardzo zbliżone do siebie odsetki zgodności występują w bloku pytań dotyczącym palenia papierosów i wyrobów tytoniowych. Blok ten obejmował pytania o palenie przez respondenta: wyrobów tytoniowych (papierosy, fajka, cygara, cygaretki itd.); papierosów obecnie; papierosów przez ponad 6 miesięcy oraz papierosów kiedykolwiek. Dla pytań tych średni odsetek zgodności wynosi 93,8 i jest on wyższy o ponad 18 punktów procentowych od średniej dla wszystkich pytań o zachowania. Waha się on od 90,6 do 95,4. Jak już wspomniałem, odsetki zgodności odpowiedzi na poszczególne pytania są bardzo wysokie i zbliżone. Dla pytań o palenie (obecnie) średni odsetek zgodności wynosi 90,6, natomiast dla pytań o palenie papierosów – 95,3. Porównując powyższe dane z poprzednimi, dotyczącymi konsumpcji alkoholu, zaskakująca jest wielkość rozbieżności średniego odsetka odpowiedzi zgodnych: różnica wynosi 20 punktów procentowych. Wydaje się, iż jedynym wyjaśnieniem jest fakt, iż tytoń jest używką, która nie budzi tak silnych negatywnych skojarzeń, jak alkohol, i w związku z tym ludzie nie kryją się z tym, że palą.

Następna bateria w bloku pytań o zachowania konsumpcyjne dotyczy dokonywania przez respondenta zakupów żywnościowych i konsumpcji produktów mlecznych. Były to pytania na temat zakupów żywnościowych oraz konsumpcji (w ogóle, od czasu do czasu i w ostatnim czasie): jogurtu, kefiru, serka wiejskiego, serka homogenizowanego, deseru mlecznego i napojów mlecznych.

Dla pytań tych ogólny średni odsetek odpowiedzi zgodnych jest dość niski i wynosi 56,9 (jest on o 18,7 punktów procentowych niższy od

średniej dla wszystkich pytań o zachowania) i waha się od 21,0 do 85,6. Najniższe odsetki zgodności można zaobserwować w przypadku pytań o spożywanie w ostatnim czasie: jogurtu, kefiru, serka homogenizowanego oraz serka wiejskiego (odsetek zgodności waha się od 21,0 do 35,3). Znacznie wyższe odsetki zgodności występują w przypadku pytań o spożywanie tych produktów od czasu do czasu (odsetek zgodności waha się od 62,8 do 85,6, najwyższe współczynniki występują w przypadku pytań dotyczących konsumpcji jogurtu). Niski odsetek zgodności odpowiedzi na te pytania wydaje się efektem sformułowania pytania, a konkretnie niejasności określenia „od czasu do czasu”.

Relatywnie wysoki jest natomiast odsetek zgodności w pytaniu o częstotliwość dokonywania przez respondenta zakupów żywnościowych: wynosi on 74,1.

Omawiając pytania dotyczące zachowań konsumpcyjnych, warto przedstawić wyniki powtórzonego pomiaru z badania dotyczącego konsumpcji słodczy (batonów czekoladowych i cukierków). W pytaniach tych respondenci mieli za zadanie wskazać marki produktów (20 marek batonów i 20 marek cukierków), które (a) konsumowali kiedykolwiek i (b) konsumowali w ciągu ostatnich 12 miesięcy⁸. Pytania te brzmiały następująco:

„Które z tych słodczy kiedykolwiek Pan/Pani jadł/a? (poprosić respondenta, aby wybrał zdjęcia produktów. Zaznaczyć wszystkie wskazania respondenta).

„A które z tych słodczy jadł/a Pan?Pani w ciągu ostatnich 12 miesięcy? (poprosić respondenta, aby wybrał zdjęcia produktów. Zaznaczyć wszystkie wskazania respondenta).

Ze względu na tematykę (słodyczne) oraz i liczbę „itemów”, pytania te przypominają omawiane wcześniej, dotyczące opinii, a nie zachowań (por. s. 166). Zastosuję zatem tutaj ten sam schemat analizy: najpierw wskażę (a) odsetek zgodności liczby wskazanych marek, które respondent konsumował kiedykolwiek/w ciągu ostatnich 12 miesięcy, a następnie (b) zgodność liczby i nazw marek produktów, które respondent konsumował kiedykolwiek/w ciągu ostatnich 12 miesięcy.

W tabeli 22 przedstawiono wyniki (odsetki zgodności odpowiedzi) powtórzonego pomiaru.

⁸ Pytania te nie zostały uwzględnione w powyższych analizach.

Tabela 22. *Odsetek zgodności liczby wskazanych marek produktów oraz zgodności nazw wskazanych marek produktów*

Zachowania	% zgodności liczby wskazanych marek produktów	% zgodności liczby i nazw wskazanych marek produktów
Marki batonów konsumowane kiedykolwiek	11,8	6,9
Marki batonów konsumowane w ciągu ostatnich 12 miesięcy	7,0	3,5
Marki cukierków konsumowane kiedykolwiek	16,5	5,3
Marki cukierków konsumowane w ciągu ostatnich 12 miesięcy	10,6	2,3
Średni %	11,5	4,5

Wyniki analizy tych pytań są bardzo podobne do tych, które dotyczyły opinii na temat słodyczy, aczkolwiek odsetki są jeszcze niższe. Okazuje się bowiem, że zgodność liczby wskazanych produktów waha się od 7,0% do 11,8% (średnia: 11,5%), a odsetek odpowiedzi zgodnych (zgodność liczby i zgodność nazw marek produktów) jest jeszcze niższy i waha się od 2,3 do 6,9 (średnia wynosi 4,5). Odsetek ten jest praktycznie równy odsetkowi odpowiedzi rozbieżnych na niektóre pytania omówione powyżej, na przykład pytania o picie herbaty czy palenie papierosów (odsetek odpowiedzi rozbieżnych odpowiednio: 3,2 i 4,7). Oznacza to, iż wartość danych z tych dwóch badań jest bardzo niska: jedynie niespełna 5% respondentów (co dwudziesty) udzieliło w powtórny pomiarze takiej samej odpowiedzi. Przypuszczać można, iż tak duży błąd pomiaru spowodowany został błędem narzędzia (pytania dotyczyły, jak wspominałem, 20 marek batonów i 20 marek cukierków, spożywanych kiedykolwiek/w ciągu ostatnich 12 miesięcy), który mógł generować błąd respondenta (zmęczenie i znużenie powodujące udzielanie odpowiedzi przypadkowych).

Ostatnim z wyróżnionych bloków są pytania o inne zachowania respondenta. Znalazły się tutaj pytania dotyczące: przemieszczania się poza najbliższą okolicę danego dnia i godzina (pora) wyjścia z domu; wyjazdów co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajowe i zagraniczne); wyjazdów wakacyjnych w ciągu ostatnich 12 miesięcy;

korzystania z internetu i częstotliwość korzystania z internetu; wykonywania połączeń z telefonu stacjonarnego na telefony komórkowe oraz korzystania z automatów telefonicznych w miejscach publicznych. Dla pytań tych ogólny średni odsetek odpowiedzi zgodnych jest dość wysoki i wynosi 80,1 (jest on o 4,5 punktu procentowego wyższy od średniej dla wszystkich pytań o zachowania). Waha się on od 58,5 do 98,3. Najniższe odsetki zgodności w tym bloku występują w przypadku pytań o korzystanie z automatów telefonicznych w miejscach publicznych oraz wykonywanie połączeń z telefonu stacjonarnego na telefony komórkowe (odsetek zgodności wynosi odpowiednio: 58,5 i 72,0, średnia: 65,3). W przypadku pytań o korzystanie i częstotliwość korzystania z internetu średni odsetek zgodności jest wyższy (75,6). Co oczywiste, odsetek ten jest wyższy w przypadku pytania o korzystanie z internetu niż pytania o częstotliwość korzystania (odpowiednio: 88,7 i 62,4. Najwyższe współczynniki w tym bloku można zaobserwować w przypadku pytań dotyczących wyjazdów (średni odsetek zgodności wynosi 85,9) oraz w pytaniach przemieszczanie się poza najbliższą okolicę danego dnia i godzinę (porę) wyjścia z domu (odpowiednio: 98,3 i 88,7, średnia: 93,5).

Analizując zgodność odpowiedzi na powtórzone pytania dotyczące zachowań respondenta, wskazywałem na ich zróżnicowanie. Dość zaskakujące okazało się, że najniższe współczynniki zgodności zaobserwować można w przypadku pytań o zachowania konsumpcyjne. Zazwyczaj przyjmuje się, że pytaniami drażliwymi, w przypadku których uzyskiwane odpowiedzi są obciążone błędem, są między innymi pytania o wykształcenie i konsumpcję kulturową. Nie są to co prawda pytania wzbudzające niepokój, ale skłaniające do udzielania społecznie akceptowanych odpowiedzi. W ich przypadku można wskazać normy społeczne określające, co jest pożądane, a co nie, i stąd ryzyko udzielania odpowiedzi zgodnej z normami (Sztabiński P. 1997: 80–89). Średni odsetek zgodności odpowiedzi na te pytania jest wyższy od średniego odsetka zgodności odpowiedzi na pytania o zachowania w ogóle (zachowania „edukacyjne” – 93,3, zachowania „kulturowe” – 76,3, zachowania w ogóle 75,6). Jednocześnie średni odsetek zgodności w pytaniach o zachowania „konsumpcyjne” jest o 5 punktów procentowych niższy od średniej ogólnej (70,6). Z pewnością nie są to pytania drażliwe (poza, być może, pytaniami o konsumpcję alkoholu). Dlaczego więc wartość odpowiedzi na te pytania jest relatywnie niska? Trudno jednoznacznie odpowiedzieć

na to pytanie. Być może jest to związane z postrzeganym przez respondentów „użytkowym” charakterem badania i jego „statusem”, być może z wagą przywiązywaną do odpowiedzi na tego typu pytania czy znużeniem spowodowanym „schematycznością” podobnych pytań. Najprawdopodobniej jednak czynniki te oddziałują równocześnie i wzajemnie się wzmacniają.

Innym szerszym problemem związanym z pytaniami o zachowania przeszłe, i to nie tylko zachowania, także wydarzenia, a więc tak zwane *retrospective questions*, jest pamięć respondenta. Pytania te rodzą bardzo duże problemy (Fink 1960: 143–148; Gaskell, Wright, O’Muirheartaigh 2000: 77–89; Brennan Mike, Joanna Chan, Dean Hini, Don Esslemont 1996: 20–29). Do spraw związanych z raportowaniem w wywiadzie spraw przeszłych, tak zwanych *recall data*, powrócę w dalszej części rozważań.

4. 1. 4. Pytania o wiedzę

W przypadku pytań o wiedzę baza materiałowa jest niestety znacznie uboższa niż w przypadku omówionych wcześniej pytań. Dysponuję bowiem danymi jedynie z pięciu badań, w których pytania takie były zadawane. Były to pytania o: „rozumienie pojęcia »klasa średnia«”; „znajomość marek batonów”; „znajomość marek cukierków”; „znajomość marek chrupek/chipsów widzianych przez respondenta w promocji/ofercie specjalnej”; „znajomość marek piwa” oraz „znajomość nazw firm telekomunikacyjnych”.

Ze względu na fakt, iż nie dysponuję danymi źródłowymi, a jedynie danymi wynikowymi z porównania odpowiedzi respondentów, zastosuję tu ten sam schemat analizy co poprzednio (por. s. 166 i 185–186). A zatem, najpierw przedstawię wyniki analizy: (a) odsetka zgodności liczby wskazanych wymiarów rozumienia pojęcia „klasa średnia” oraz liczby znanych respondentowi marek produktów/firm, a następnie (b) zgodności liczby oraz wskazanych wymiarów pojęcia/nazw marek produktów/firm. W tabeli 23 przedstawiono wyniki (odsetki zgodności odpowiedzi) powtórnego pomiaru.

Tabela 23. *Pytania o wiedzę: odsetek zgodności liczby wskazanych wymiarów pojęcia/marek produktów/firm oraz zgodności wymiarów/nazw marek wskazanych produktów*

Wiedza	% zgodności liczby wymiarów/marek/firm	% zgodności liczby i nazw wymiarów/marek/firm
Rozumienie przez respondenta pojęcia „klasa średnia”	40,7	8,1
Znajomość przez respondenta marek batonów	20,4	17,9
Znajomość przez respondenta marek cukierków	10,1	5,1
Marki chrupek/chipsów widziane przez respondenta w promocji/ofercie specjalnej	20,3	6,5
Znajomość przez respondenta marek piwa	7,1	0,8
Znajomość przez respondenta firm telekomunikacyjnych	42,0	38,9
Średni %	23,4	12,9

Ogólny odsetek odpowiedzi zgodnych dla wszystkich powtórzonych pytań o wiedzę wyniósł 12,9⁹. Waha się on od 0,8 do 38,9, a więc jest bardzo zróżnicowany. Maksymalna różnica między odsetkiem zgodności odpowiedzi na te pytania wynosi 38,1 punktów procentowych. Wyniki te nie powinny być zaskakujące, biorąc pod uwagę przedstawione wyżej wyniki analizy pytań o opinie na temat słodczy oraz konsumpcji słodczy, w przypadku których odsetek ten był bardzo niski (w przypadku tych pytań, wskaźniki zgodności wynosiły bowiem odpowiednio 8,3 i 4,5). Zaskakujące jest natomiast to, iż odsetek odpowiedzi zgodnych, zarówno jeśli chodzi o liczbę wymiarów/marek, jak i liczbę oraz nazwy wymiarów/marek jest znacząco wyższy w przypadku pytań o wiedzę niż pytań o zachowania (odpowiednio 12,9 i 4,7). Równie zaskakujące jest to, że pytania o opinie przy tego typu analizie plasują się „pośrodku”.

⁹ Mówiąc o odpowiedziach zgodnych, mam tutaj na myśli odpowiedzi „w pełni zgodne”, to znaczy zgodność liczby i nazw wymiarów/marek produktów.

4. 1. 5. Pytania o cechy

Analizowane pytania o cechy (28 pytań) także podzielono na bloki. W analizie uwzględniono dwa bloki tematyczne:

(i) Pytania o cechy respondenta (12 pytań), a więc: płeć, wiek, wykształcenie, stan cywilny, ale także waga i wzrost.

(ii) Pytania o charakterystyki dotyczące gospodarstwa domowego respondenta – tak zwane cechy kontekstowe (16 pytań). W bloku tym znalazły się między innymi pytania o liczbę osób w gospodarstwie, posiadanie i liczba dzieci oraz pewne ich charakterystyki.

W tabeli 24 przedstawiono odsetki odpowiedzi zgodnych oraz pozostałe współczynniki dla poszczególnych bloków pytań o cechy.

Tabela 24. *Pytania o cechy: wartości mierników dla poszczególnych bloków pytań, wyróżnionych ze względu na ich przedmiot*

1a. Cechy demograficzne respondenta: płeć					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	94,4-100,0	0,667-1,000	0,879-1,000	b	b
\bar{X}	99,8	a	a	b	b
1a. Cechy demograficzne respondenta: wiek					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendal'a	r Pearsona
Min–Max	82,4-100,0	0,917-1,000	0,889-1,000	0,826**-1,000**	0,738**-1,000**
\bar{X}	96,3	a	a	c	c
1b. Cechy społeczne respondenta: wykształcenie					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	65,9-91,5	0,519-0,846	0,624-0,901	0,749**-0,955**	b
\bar{X}	80,8	a	a	c	b
1b. Cechy społeczne respondenta: stan cywilny					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	96,5-98,2	0,844-0,856	0,894-0,957	b	b
\bar{X}	97,3	a	a	b	b
1c. Cechy fizyczne respondenta: waga/ wzrost					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Min–Max	33,3-83,5	0,618-0,900	0,633-0,909	0,801**-0,955**	0,886**-0,987**
\bar{X}	58,5	a	a	c	c

1d. Inne cechy respondenta					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
<i>Min-Max</i>	39,9-99,3	0,102-0,917	0,273-0,972	0,439**,-0,902** ^f	0,593**,-0,961** ^f
\bar{X}	73,5	a	a	c	c
2a. Charakterystyka gospodarstwa domowego respondenta: liczba osób					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
<i>Min-Max</i>	72,5-93,2	0,500-0,878	0,702-0,938	0,709**,-0,921**	0,451**,-0,942**
\bar{X}	84,0	a	a	c	c
2b. Charakterystyka gospodarstwa domowego respondenta: posiadanie/ liczba dzieci					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
<i>Min-Max</i>	87,1-97,2	0,624-0,822	0,745-0,923	0,832**,-0,948**	0,835**,-0,958**
\bar{X}	92,9	a	a	c	c
2c. Charakterystyka gospodarstwa domowego respondenta: inne cechy					
	% odp. zgod.	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
<i>Min-Max</i>	15,9-93,1	0,307-0,929	0,627-0,910	0,581**,-0,884** ^f	-0,089- 0,913** ^f
\bar{X}	76,0	a	a	c	c

Oznaczenia w tabeli:

a średnia wartość dla współczynnika nie jest liczona ze względu na prawdopodobieństwo jego zaniżenia spowodowane nierównomiernym rozkładem odpowiedzi (zdominowanie rozkładu przez jeden typ odpowiedzi).

b nie obliczono wartości współczynnika ze względu na jego nieadekwatność do poziomu pomiaru.

c średnia wartość dla współczynnika nie jest liczona ze względu na jego nieaddytywność (średnia wartość współczynnika z wielu prób nie jest równa „średniej korelacji” we wszystkich próbach).

f współczynniki liczone wyłącznie dla zmiennej z niektórych badań (dla pozostałych nie liczone ze względu na jego nieadekwatność do poziomu pomiaru).

** korelacja istotna na poziomie 0.01 (test dwustronny).

Ogólny średni odsetek odpowiedzi zgodnych dla wszystkich powtórzonych pytań o cechy (niezależnie od ich przedmiotu) wyniósł 88,4. Porównując wartości współczynników dla pytań o cechy respondenta i gospodarstwa domowego ze współczynnikami dla wszystkich pytań, można zauważyć, iż wszystkie są ogólnie nieco wyższe (dla wartości minimalnej są równe). Przykładowo średni odsetek odpowiedzi zgodnych na pytania o cechy jest wyższy od średniej ogólnej (dla wszystkich pytań powtórzonych) o 5,1 punktu procentowego.

Porównując współczynniki dla pytań o cechy respondenta i pytań o charakterystykę gospodarstwa domowego, można zauważyć, iż są one

zawsze wyższe w przypadku tych pierwszych (różnica między średnimi odsetkami zgodności wynosi 3,8 punktu procentowego). Jest to zrozumiałe, ponieważ respondent wydaje się bardziej kompetentnym informatorem, jeśli chodzi o własną osobę niż cechy gospodarstwa domowego. Nie oznacza to jednak, że jest tak w każdym przypadku. Średni odsetek zgodności dotyczący cech respondenta wynosi 89,7, podczas gdy na przykład średni odsetek zgodności w pytaniach dotyczących telefonu stacjonarnego i komórkowego to 92,9, w pytaniach o zachowania społeczne, to znaczy udział w wyborach i referendum – 93,7%. Przeanalizujemy zatem dokładniej, w których pytaniach o cechy respondenta współczynniki są relatywnie niższe.

Blok dotyczący respondenta obejmował pytania o jego cechy: demograficzne (płeć i wiek z dokładnością do jednego roku); społeczne (wykształcenie, stan cywilny); fizyczne (waga – z dokładnością do 1 kg i wzrost – z dokładnością do 1 cm) oraz inne cechy respondenta: dochód (z dokładnością do 1 zł), poczucie związku z religią, związek z religią, posiadanie rodzeństwa oraz posiadanie znajomych wśród imigrantów.

Jak wskazałem, średni odsetek zgodności w tych pytaniach wynosi 89,7 (jest o 1,3 punktu procentowego wyższy od średniej dla wszystkich pytań o cechy), waha się od 33,3 do 100,0. Najwyższy odsetek zgodności w tym bloku występuje w przypadku pytań o cechy demograficzne: płeć i wiek respondenta i wynosi odpowiednio 99,8 oraz 96,3. Jest wyższy od średniej dla wszystkich pytań o cechy respondenta o 10,1 i 6,6 punktu procentowego. Współczynniki te są nieco wyższe w przypadku pytań o płeć respondenta niż w pytaniach o wiek. Należy jednak podkreślić, iż w przypadku pytań o płeć zgodność odpowiedzi nie jest pełna: średni odsetek zgodności waha się od 94,4 do 100,0¹⁰. Cecha ta jest tak zwaną zmienną indywidualną, występującą w każdym badaniu, niezależnie od jego celu i przedmiotu. Wydawać by się mogło, że ze względu na jej obserwowalny (jawny) charakter nie jest możliwy błąd pomiaru. Nie-

¹⁰ Odsetek zgodności płci respondenta rzędu 94 wystąpił w badaniu, w którym ankiet w ramach wylosowanego gospodarstwa domowego sam miał dobrać „kompetentnego informatora w danej sprawie”. Zdarzały się zatem przypadki, w których ankiet zapisywał, iż osobą tą był na przykład mężczyzna w wieku 40 lat, podczas gdy faktycznie ankiet rozmawiał z jego matką – kobietą w wieku 70 lat, która z pewnością nie spełniała kryterium „kompetentnego informatora”.

stety tak nie jest. W jednym z analizowanych badań błąd taki wystąpił i stwierdzono go w 1,0% wywiadów (2 przypadki). W pozostałych badaniach nie zaobserwowano przypadków niezgodności tej cechy¹¹.

W przypadku wieku respondenta maksymalny odsetek rozbieżności w powtórzonym pomiarze wynosi 17,6, natomiast średni odsetek zgodności, jak wspomniano, 96,3. Z całą pewnością można stwierdzić, iż poza przypadkami „klinicznymi” błąd pomiaru tej cechy nie jest spowodowany błędami pamięci ani obiektywnym brakiem wiedzy. Wbrew jednak powszechnemu przekonaniu, iż zmienna ta jest bardzo podatna na tak zwany wpływ ankieterski (obserwowalne i nieobserwowalne cechy ankietera), jej pomiar uznać należy za rzetelny. Średni odsetek odpowiedzi rozbieżnych rzędu niespełna 4 punktów procentowych jest bowiem niski w porównaniu z analogicznym wskaźnikiem dla innych zmiennych. Należy jednak pamiętać, iż potencjalny błąd, którego źródłem jest ankieter, może się pojawić tylko w sytuacji, gdy współwystępuje z nim świadoma bądź nie tendencja respondenta do waloryzowania sytuacji i udzielania społecznie akceptowanych odpowiedzi. Powyższe zjawiska nie przesądzą o kierunku niezgodności pomiaru: w praktyce zawyżenia wieku respondenta równoważą liczbę przypadków zaniżeń (Sztabiński F. 2001).

Znacząco niższy odsetek zgodności występuje w przypadku pytań o cechy społeczne respondenta, a więc wykształcenie oraz stan cywilny. Wynosi on 82,1% i jest niższy od średniej dla wszystkich pytań o cechy respondenta o 7,6 punktu procentowego. Przy czym współczynniki te są wyższe w przypadku pytań o stan cywilny respondenta niż w pytaniach o wykształcenie (odpowiednio: średni odsetek odpowiedzi zgodnych – 97,3 i 80,8)¹².

W przypadku pytań o wykształcenie źródła błędów mogą być, jak się wydaje, dwojakiego rodzaju. W pewnym stopniu mogą stanowić efekt obiektywnego braku wiedzy (przykładowo, respondenci nie zawsze wiedzą, do której kategorii zaliczyć tzw. „czteroklasówkę”, przedwojenne gimnazjum czy „małą maturę”). Z pewnością cecha ta jest także bardzo

¹¹ Podstawowym źródłem tego błędu jest ankieter, a konkretnie nieświadome błędy (pomyłki) oraz świadome (intencjonalne) zafałszowania. Ta ostatnia sytuacja ma miejsce przede wszystkim w przypadku badań realizowanych na próbach udziałowych, w którym jednym z kryteriów doboru jest płeć.

¹² Warto podkreślić, iż średni odsetek odpowiedzi zgodnych na pytanie o stan cywilny respondenta jest także wyższy od odsetka zgodności na pytanie o wiek respondenta: różnica nie jest jednak duża i wynosi 1 punkt procentowy.

podatna na „wpływ ankierski”. Wykształcenie jest bowiem w naszym społeczeństwie wartością bardzo wysoko waloryzowaną. Stąd też zaobserwować można znaczną przewagę przypadków zawyżeń poziomu wykształcenia nad przypadkami zaniżeń. Tendencję do zawyżania poziomu wykształcenia przy powtórzonym pomiarze można dodatkowo wytłumaczyć dwoma czynnikami: charakterem kontaktu ankietera *vs* kontrolera z respondentem oraz błędami ankietera. W pierwszym przypadku można przyjąć, iż w sytuacji powtózonego pomiaru relacje te mają charakter krótkotrwały, bardziej oficjalny (w porównaniu z kontaktem z ankieterem) i w związku z tym w większym stopniu generuje odpowiedzi społecznie pożądane. Z kolei błędy ankierskie sprowadzają się przede wszystkim do zmiany procedury pytania: nieodczytania kafeterii lub niewręczenia karty respondenta¹³.

Najniższy odsetek odpowiedzi zgodnych występuje w przypadku pytań o cechy fizyczne respondenta, a więc jego wzrost i wagę. Wynosi on 58,5 i jest niższy od średniej dla wszystkich pytań o cechy respondenta o ponad 30 punktów procentowych (31,2 punktów). Waha się on od 33,3 do 83,5. Współczynniki te są wyższe w przypadku pytań o wzrost respondenta niż w pytaniach o jego wagę (odpowiednio: średni odsetek odpowiedzi zgodnych – 85,0 i 50,0). Interesujące, że w przypadku pomiaru tych dwóch cech wystąpił wyraźnie efekt sponsora i tematyki badania: znacząco wyższe współczynniki można bowiem zaobserwować w przypadku badań akademickich o charakterze epidemiologicznym niż badań marketingowych. Wynika to, jak się wydaje, ze sposobu postrzegania przez respondentów związku zadawanych pytań z problematyką badania: jeśli wywiad dotyczy na przykład konsumpcji produktów mlecznych, to respondenci nie widzą bezpośredniego ich związku z pytaniami o wagę, wzrost itp., i udzielają odpowiedzi „przybliżonych”. Jeżeli jednak badanie dotyczy na przykład problemów zdrowia, to związek ten jest dla nich oczywisty i udzielają odpowiedzi bardziej precyzyjnych – replikowanych.

Najbardziej zróżnicowane są jednak odsetki odpowiedzi zgodnych na takie pytania jak dochód respondenta, posiadanie rodzeństwa i najzo-

¹³ W obu przypadkach efekt jest ten sam: respondent nie wie, w jakich kategoriach powinien udzielać odpowiedzi, jak szczegółowa winna być jego odpowiedź (chodzi przede wszystkim o poziom wykształcenia: ukończone – nieukończone oraz kierunek: średnie ogólnokształcące – średnie zawodowe).

mych wśród imigrantów oraz poczucie związku z religią. Na pytania te średni odsetek odpowiedzi zgodnych jest co prawda dość wysoki (73,5), ale jest on bardzo zróżnicowany (waha się od 39,9 do 99,3), co nie jest zaskoczeniem, biorąc pod uwagę zróżnicowane przedmioty tych pytań. Najniższy jest on w przypadku pytania o dochód (33,9), najwyższy zaś w przypadku pytania o posiadanie rodzeństwa (99,3). Zaskakujące są natomiast różnice odsetka zgodności odpowiedzi na pytania dotyczące „poczucia związku z daną religią” i „związku z daną religią” (odpowiednio: 93,5 odpowiedzi zgodnych i 59,4).

Blok dotyczący gospodarstwa domowego respondenta obejmował pytania o liczbę osób w gospodarstwie domowym; posiadanie/liczba dzieci oraz inne cechy gospodarstwa domowego. Średni odsetek odpowiedzi zgodnych na pytania dotyczące gospodarstwa domowego wynosi 85,9 i jest o 2,5 punktu procentowego niższy od średniej dla wszystkich pytań o cechy. Waha się on od 15,9 do 97,2.

Najwyższy odsetek zgodności w tym bloku występuje w przypadku pytań o posiadanie i liczbę dzieci w gospodarstwie domowym respondenta. Wynosi on 92,9 i jest wyższy o 7 punktów procentowych od średniej dla wszystkich pytań o charakterystykę gospodarstwa. Waha się on od 87,1 do 97,2. Odsetek zgodności jest nieco wyższy w przypadku pytań o posiadanie dzieci niż pytań o ich liczbę (różnica wynosi 2,4 punktu procentowego). Interesujące jest także to, że w pytaniach tych odsetek odpowiedzi jest wyższy, jeśli dotyczą one dzieci młodszych. Na przykład w pytaniu o liczbę dzieci w wieku do lat trzech, wynosi on 96,5, o dzieci w wieku 4–7 lat – 92,%, a w wieku 8–15 lat 90,8. Być może wiąże się to z tym, iż rodzice „lepiej pamiętają”, ile lat mają młodsze dzieci, wymagające większej troski (a więc, w którym przedziale wiekowym się mieszczą), natomiast w przypadku starszych granice wiekowe ulegają „rozmyciu”: wydają się one „młodsze” lub „starsze”¹⁴.

Nieco niższe współczynniki zgodności uzyskano w przypadku pytań dotyczących wielkości gospodarstwa domowego respondenta (liczby osób w gospodarstwie) oraz pytań o liczbę osób wspólnie zamieszkujących z respondentem i liczbę osób pełnoletnich w gospodarstwie. W pierwszym z pytań (liczba osób w gospodarstwie domowym) średni

¹⁴ Chodzi mi tutaj o zjawiska *backward* oraz *forward telescoping* opisane między innymi w: Auriat (1996), Brennan, Chan, Hini, Esslemont (1996: 20–29), Gaskell, Wright, O’Muirheartaigh (2000: 77–89), Lavrakas (2008: 886–887).

odsetek zgodności wynosi 84,0 i jest niższy o 1,9 punktu procentowego od średniej dla wszystkich pytań o charakterystykę gospodarstwa. Waha się on od 72,5 do 93,2. Średni odsetek zgodności jest niższy o 5,5 punktu procentowego od analogicznego, dla pytania o liczbę osób pełnoletnich w gospodarstwie. Jednocześnie jest nieco wyższy (1,5 punktu) niż w pytaniu o liczbę osób wspólnie zamieszkujących z respondentem.

Zastanawiające wydaje się, dlaczego w pytaniach tych odsetek zgodności jest relatywnie niski (zbliżony do tych z pytań o wykształcenie respondenta, które określa się mianem pytań skłaniających do udzielania społecznie akceptowanych odpowiedzi, w odniesieniu do których normy społeczne określają, co jest pożądane, a co nie jest). W pytaniach o wielkość gospodarstwa domowego tego typu „obciążenie” nie powinno występować. Co więcej, w treści pytania zazwyczaj zamieszcza się definicję gospodarstwa domowego, wyjaśniając, iż „chodzi o wszystkie osoby wspólnie z Panem/Panią zamieszkujące i utrzymujące się ze wspólnych funduszy”. Dlaczego zatem odsetek zgodności sięga jedynie 84? Można, jak się wydaje, wskazać dwie przyczyny. Pierwsza to „kumulacja błędów”: błędu związanego z „posiadaniem dzieci” (odsetek zgodności 94,9) oraz błędu związanego z „liczbą posiadanych dzieci” (odsetek zgodności 90,4). Druga z przyczyn to „subiektywne” definiowanie przez respondenta „gospodarstwa domowego” (mimo zamieszczonej definicji), szczególnie w przypadku rodzin wielopokoleniowych. Analizy na poziomie jednostkowym, poszczególnych wywiadów, wyraźnie wskazują, że ten sam respondent w jednym wypadku wlicza do gospodarstwa domowego na przykład rodziców/teściów, w drugim zaś ich wyklucza. Sytuację tę niełatwo wyjaśnić – być może owe sprzeczne deklaracje były spowodowane chwilowymi emocjami związanymi z relacjami interpersonalnymi osób wspólnie zamieszkujących¹⁵. Nie można także zapominać o innym jeszcze „błędogennym” czynniku, mianowicie faktycznej zmianie składu gospodarstwa domowego respondenta w okresie między pomiarem zasadniczym a powtórzonym. Przykładowo dorosły syn mógł

¹⁵ Na występowanie tego typu przypadków wskazuje analiza składu gospodarstwa domowego, w którym respondent proszony jest o wymienienie wszystkich osób, podając ich płeć, rok urodzenia oraz stopień pokrewieństwa. Osobną sprawą, w którą nie chcę tutaj wnikać, są błędy ankietera na poziomie losowania respondenta w przypadku prób adresowych (w składzie gospodarstwa domowego celowe nieuwzględnianie pewnych osób, aby wylosowanym respondentem była któraś z osób obecnych w trakcie wizyty ankietera lub skłonna do udziału w badaniu).

się wyprowadzić od rodziców, mógł „wprowadzić” swoją żonę, mogło się urodzić dziecko, ktoś mógł umrzeć. Wszystko to są „naturalne ruchy” w ramach gospodarstwa domowego i nikt za błąd nimi spowodowany nie jest odpowiedzialny.

Inną sprawą jest zgodność płci członków gospodarstwa domowego respondenta. Błąd pomiaru tej cechy nie jest tak powszechny, jak poprzednie – wystąpił tylko w niektórych badaniach. Przy czym, co należy podkreślić, pominięto tutaj wszystkie przypadki niezgodności liczby osób w gospodarstwie domowym, a dodatkowo brano pod uwagę niezgodność płci z członków gospodarstwa. Jeśli pominiemy przypadki niezgodności liczby i płci osób w gospodarstwie domowym, błąd pomiaru wieku jego członków jest znacznie częstszy: maksymalny odsetek odpowiedzi rozbieżnych w powtórzonym pomiarze wynosi 23,1, minimalny – 9,8, natomiast średni odsetek zgodności – 82,9. Źródła błędów pomiaru tej cechy należy upatrywać przede wszystkim w błędach pamięci respondenta (lub obiektywnym braku wiedzy) oraz w niewielkim stopniu we „wpływie ankierskim”. Błędy pamięci lub braku wiedzy dotyczą przede wszystkim osób z pokolenia starszego niż respondent (rodzice, dziadkowie), aczkolwiek mogą również wystąpić w informacjach na temat wieku młodszego/starszego rodzeństwa/dzieci. Błędy te często mają związek także z wielkością gospodarstwa domowego respondenta: im jest ono mniejsze, tym „łatwiej wszystko zapamiętać”. Tego typu błąd w odniesieniu do wieku współmałżonka/dzieci może być efektem społecznej waloryzacji. Na przykład respondent zawyża/zaniża wiek współmałżonka (ponieważ „ma zbyt młodego/starego współmałżonka”) lub też wiek dzieci („w moim wieku mam zbyt małe/duże dzieci”). Ogólnie jednak, niezależnie od źródeł błędów, nie można uznać, że respondent jest w pełni kompetentnym informatorem o innych członkach swojego gospodarstwa domowego.

Analizując zmienne dotyczące gospodarstwa domowego, należy podkreślić, iż „łączny” błąd ich pomiaru przekracza 30%. Oznacza to, iż jedynie w niespełna 70 przypadkach na 100 uzyskujemy rzetelne informacje dotyczące liczby osób w gospodarstwie domowym oraz płci i wieku poszczególnych jego członków. Oczywiście skrajne są sytuacje, w których „nic się nie zgadza”. Powodują je czynniki sytuacyjne, przejściowe (na przykład emocje), błędy pamięci, ale także błędy ankierskie (błąd zapisu). W tabeli 25 przedstawiono odpowiedzi jednego z respondentów na pytanie o wiek i płeć wszystkich członków jego gospodarstwa domo-

wego uzyskane w wywiadzie zasadniczym (pierwotny pomiar) i wywiadzie kontrolnym (pomiar powtórzony).

Tabela 25. *Płeć i wiek członków gospodarstwa domowego (poza respondentem) – przykład*

Pomiar zasadniczy	Pomiar powtórzony
1. Kobieta, ur. 1973 r. (żona respondenta)	1. Kobieta, ur. 1976 r. (żona respondenta)
2. Kobieta, ur. 1998 r. (córka respondenta)	2. Mężczyzna, ur. 1997 r. (syn respondenta)
	3. Kobieta, ur. 1940 r. (teściowa respondenta)
	4. Mężczyzna, ur. 1940 r. (teść respondenta)

Najniższy odsetek zgodności w tym bloku uzyskano w przypadku pytań o inne charakterystyki gospodarstwa domowego respondenta, takie jak identyfikacja osoby, będącej głową gospodarstwa domowego, dochód gospodarstwa oraz cechy głównego użytkownika samochodu w gospodarstwie (średni odsetek zgodności wynosi 76,0 i jest niższy o 9,9 punktów procentowych od średniej dla wszystkich pytań o charakterystykę gospodarstwa). Odsetek ten waha się od 15,9 do 93,1. Jeśli bliżej przyjrzeć się odsetkom zgodności w tym bloku, okazuje się, iż identyfikacja i charakterystyki głównego użytkownika samochodu są bardziej stabilne niż identyfikacja głowy gospodarstwa domowego (średni odsetek odpowiedzi zgodnych wynosi odpowiednio 92,3 i 83,6). Być może jest to związane z pewną „subiektywnością” pojęcia „głowa gospodarstwa domowego”: nawet dla tego samego respondenta może to być osoba wnosząca największy dochód lub – innym razem – osoba podejmująca kluczowe dla gospodarstwa decyzje. Co więcej, odsetek odpowiedzi zgodnych na pytanie o wykształcenie głównego użytkownika samochodu jest wyższy o prawie 13 punktów procentowych od odsetka odpowiedzi zgodnych na pytanie o wykształcenie respondenta (odsetki wynoszą odpowiednio: 93,5 i 80,8).

Pytaniem, w którym odsetek odpowiedzi zgodnych był najniższy, jest pytanie o dochód gospodarstwa (odsetek odpowiedzi zgodnych 15,9, wartość współczynnika r Pearsona wynosi $-0,089$). Warto pod-

kreślić, iż jest to najniższy odsetek spośród wszystkich analizowanych pytań w ogóle i jedyny przypadek, w którym wartość współczynnika r Pearsona przyjmuje wartość ujemną. Im więc na przykład wyższe kwoty podawali respondenci w przypadku pierwszego badania, tym niższe deklarowali przy powtórzonym pomiarze. To nie powinno zaskakiwać: pytanie o dochody jest jednym z pytań określanych mianem wzbudających niepokój. Zgodnie z normami społecznymi pytania dotyczące zarobków, dochodów, oszczędności – ogólnie spraw finansowych, nie są tematami, które mogą stanowić przedmiot rozmowy. Dodatkowo wpływ na deklaracje mógł mieć wspomniany wielokrotnie charakter kontaktu z ankierem/kontrolerem (chodzi o czas jego trwania), a także skojarzenia wizyty tego drugiego z instytucją tak zwanego donosu obywatelskiego w związku z nieujawnianymi dochodami.

Analizując źródła błędów w pomiarze powyższych cech, wskazywałem wielokrotnie na respondenta. Cechą pytań metryczkowych jest to, iż ich przedmiotem są zazwyczaj wydarzenia przeszłe (na przykład data urodzenia lub wiek poszczególnych członków gospodarstwa domowego, historia zawodowa respondenta, przebieg edukacji, dochody w określonym przedziale czasu) oraz fakty będące ich konsekwencją lub pochodną (na przykład stan cywilny respondenta, wykształcenie jego lub współmałżonka). Innymi słowy, w badaniach surveyowych respondenci bardzo często są pytani nie tylko o to, co się w ich życiu wydarzyło, ale także kiedy miało to miejsce. Są to zatem pytania retrospektywne, wymagające odwołania się do pamięci. Jednocześnie od dawna wiadomo, że odpowiadając na pytania retrospektywne, respondenci popełniają błędy: zapominają, że pewne wydarzenia miały miejsce, lub przypominają sobie wydarzenia, które nie miały miejsca (ta ostatnia sytuacja jest znacznie rzadsza). Są to tak zwane błędy „pominięcia” i „popełnienia”. Relatywnie najczęstsze są jednak sytuacje, w których respondenci pamiętają, iż dane wydarzenie miało miejsce, ale błędnie umiejscawiają je w czasie (tak zwany efekt teleskopowy). Błędne umiejscowienie w czasie danego wydarzenia może polegać na wydłużaniu dystansu czasowego, a więc raportowaniu, iż miało ono miejsce wcześniej niż faktycznie (wsteczny efekt teleskopowy, *backward telescoping*) lub skracaniu dystansu czasowego (kompresja) – raportowaniu, iż miało ono miejsce później niż faktycznie (uprzedni efekt teleskopowy, *forward telescoping*) [zob. m.in. Auriat 1996: 24–33; Gaskel, Wright, O’Muircheartaigh 2000: 77].

Trzeba jednak pamiętać również o błędach związanych bezpośrednio z ankierem i jakością jego pracy. Innymi słowy, osobnym źródłem zniekształceń pomiaru tych cech jest niska jakość pracy ankiera. Podstawowym czynnikiem jest błędne przekonanie o „łatwości” pytań metryczkowych („takie same pytania lub podobne zamieszczane są przecież w kwestionariuszu do każdego badania”) oraz mniejszym ich znaczeniu („są na końcu, a więc mniej ważne”) [Sztabiński F. 2001]. Taki sposób postrzegania pytań metryczkowych prowadzi do pomijania niektórych z nich lub ich „opowiadania” zamiast odczytywania. Podobne podejście ankierów do „metryczki” może bezpośrednio rzutować na stosunek respondentów do pytań. Respondenci, nie uświadamiając sobie znaczenia i nie widząc związku między pytaniami merytorycznymi i metryczkowymi, nie zadają sobie trudu udzielania na nie „prawdziwych” odpowiedzi (Sztabiński, Sztabiński 1994). Niewątpliwie nie bez znaczenia jest tu również zmęczenie i znużenie ankiera wywiadem. W efekcie, jak wykazują wyniki badań metodologicznych, liczba błędów ankierskich w tej części kwestionariusza jest zazwyczaj najwyższa (Sztabiński F. 1997: 104–118).

Podsumowując wyniki powyższych analiz, należy stwierdzić, iż niezgodność odpowiedzi w pomiarze powtórzonym dotyczy wszystkich rodzajów pytań „metryczkowych”. Przy czym „prawdopodobieństwo” wystąpienia niezgodności jest różne i zależy od przedmiotu pytania. Wbrew pozorom nie jest tak, iż znacznie częściej występuje on przy pomiarze cech kontekstowych (dotyczących gospodarstwa domowego), w przypadku których respondent informuje o innych. Powstaje zatem pytanie, dlaczego właśnie w przypadku takich cech jak płeć, wiek członków gospodarstwa domowego błąd pomiaru jest aż tak duży. Być może wynika to z faktu, iż charakterystyczną cechą tych pytań jest, jak już wspomniano, to, iż odwołują się one do wydarzeń przeszłych (na przykład daty narodzin, posiadanie dzieci), których zapamiętywanie różni się od zapamiętywania innych wydarzeń – codziennych (na przykład zakupów, wizyty u lekarza). Są to jednak wydarzenia, o których trudno zapomnieć (pomijam tu przypadki amnezji, choroby Alzheimera), ponieważ sprowadzają się one do raportowania dat (Poulain, Riandey, Firdion 1991: 65–88). Czy wspomniane rozbieżności są jednak specyfiką analizowanych przeze mnie badań? Aby odpowiedzieć na to pytanie, należy odwołać się do innych badań o charakterze walidacyjnym. Jak bowiem wspominałem, innym sposobem określenia wiarygodności informacji

jest, poza powtórny z zadaniem tych samych pytań, odwołanie się do danych dokumentalnych (Parry, Crossley 1950: 61–80). W Polsce tego typu badania realizowane były wielokrotnie (Daniłowicz, Sztabiński 1986; Koniarek 1975: 525–548; Lutyńska 1975: 117–188; Sztudynger-Kaliszewicz 1966: 257–281), w ramach projektu analiz weryfikacyjnych danych uzyskiwanych przy zastosowaniu wywiadu kwestionariuszowego, którego autorem był Lutyński (1975: 319–392). Porównajmy zatem wyniki uzyskane w powtórny pomiarze z wynikami jednego ze wspomnianych badań walidacyjnych (na podstawie dokumentów). Chodzi o badanie Daniłowicza i Sztabińskiego (1986). Porównanie wyników tego badania z analizowanymi przeze mnie zawiera poniższa tabela.

Tabela 26. *Odsetki zgodności w badaniach kontrolnych i badaniu weryfikacyjnym*

Zmienna	Odsetek zgodności	
	Badania zasadnicze vs kontrolne 1995–2005	Badanie zasadnicze vs walidacyjne 1976–1977
Płeć respondenta	99,9	100,0
Wiek respondenta	93,9	96,4
Wykształcenie respondenta	79,8	72,4
Posiadanie samochodu	90,4	95,2
Zarobki/dochód respondenta	33,9	0,0
Dochód gospodarstwa domowego R.	15,9	0,0
Średnia	68,9	60,7

Odsetki dla poszczególnych zmiennych są zbliżone: różnice wahają się od 0,1 do 7,4 punktu procentowego (pomijając zmienne „zarobki” i „dochody”). W przypadku zmiennej „wiek respondenta” różnica między wynikami powtórny pomiaru i badania walidacyjnego jest niewielka, sięga 2,5 punktu procentowego. Interpretując wielkość rozbieżności w pomiarze tej cechy, Daniłowicz i Sztabiński stwierdzają, że „rozkłady wieku (...) trafnie odzwierciedlają rozkłady rzeczywiste” (1986: 98). Wydaje się, iż to samo można powiedzieć w odniesieniu do powtórny pomiaru tej cechy.

W przypadku zmiennej wykształcenie, różnica między powtórny pomiar i walidacją jest znacznie większa i wynosi 7,4 punktu procen-

towego. Wspomniani autorzy zaobserwowali niższy poziom zgodności, choć uznają, iż „istnieje bardzo silna zależność pomiędzy informacjami z wywiadu a danymi dokumentalnymi. (...). Można zatem orzekać o podobieństwie struktur badanych prób pod względem wykształcenia” (tamże: 104).

Z kolei wyższy odsetek zgodności stwierdzili oni w przypadku posiadania samochodu, aczkolwiek różnica nie jest wysoka – wynosi 4,8 punktu procentowego (tamże: 164).

Największe różnice w wynikach pojawiają się w przypadku dwóch zmiennych: zarobki/dochód respondenta oraz dochód gospodarstwa domowego respondenta. Wynikają one z tego, że respondenci zazwyczaj wskazują kwoty przybliżone, zaokrąglając je do „pięciu” lub „dziesięciu”¹⁶. W takim przypadku prawdopodobieństwo, że respondent wymieni tę samą kwotę, jest większe niż w sytuacji, gdy deklaracje porównujemy z dokumentami, w których widnieją kwoty „niezaokrąglone”.

Warto wspomnieć, iż autorzy ci analizowali, poza wspomnianymi, również inne cechy metryczkowe: wykształcenie współmałżonka respondenta, zawód respondenta i współmałżonka, zarobki współmałżonka oraz, warunki mieszkaniowe i zasoby (tamże: 98–166). W przypadku wszystkich tych zmiennych stwierdzili oni rozbieżności deklaracji respondentów i danych dokumentalnych, co potwierdza wyniki moich analiz i pozwala sformułować tezę o „powszechności” błędów.

Tabela 27. *Odsetek odpowiedzi zgodnych z danymi dokumentalnymi na pytania o cechy społeczno-demograficznych (w rozbiciu na deklaracje mężczyzn i kobiet oraz „uzgodnione”)*

Cecha	Odsetek odpowiedzi zgodnych z danymi dokumentalnymi		
	Mąż	Żona	„Uzgodnione”
Data zawarcia związku małżeńskiego	94,1	98,9	99,6
Daty urodzenia dzieci	93,8	97,9	98,3
Rok usamodzielnienia się dzieci	44,2	52,2	53,6

¹⁶ W jednym z analizowanych badań na 1087 przypadków jedynie 135 respondentów w badaniu zasadniczym i 202 w badaniu kontrolnym jako swój dochód wskazali kwotę „niezaokrągloną” (tzn. na przykład 497 zł), z czego jedynie w 46 przypadkach (4,2%) obie deklaracje były zgodne.

Podobny wniosek można wysnuć z badań Michel Poulain, Benoit Riandey i Jean Marie Firdion (1991: 65–88), realizowanych na próbie 500 gospodarstw domowych. Autorzy ci stwierdzają, iż mężczyźni nieco częściej niż kobiety popełniają błędy w raportowaniu pewnych faktów, a co więcej, błędy te są popełniane nawet wtedy, gdy istnieje możliwość uzgodnienia odpowiedzi z drugą osobą¹⁷. W tabeli 27 przedstawiono wyniki tego badania w odniesieniu do trzech cech społeczno-demograficznych.

4. 2. Przedmiot a struktura pytania

Można by się pokusić o szczegółową analizę związku przedmiotu pytania (pytanie o cechy, zachowania, fakty itd.), a jego strukturą (zamknięte, otwarte, prekategoryzowane). Innymi słowy, mam na myśli problem, czy, na przykład, pytanie zamknięte ze skalą o cechy daje wyższe wskaźniki „reprodukowalności” niż pytanie prekategoryzowane o cechy? Niestety nie jest to możliwe. Po pierwsze dlatego, że kategorie szczegółowe przedmiotu pytania mają zbyt duży wpływ na wartość współczynników zgodności. Po drugie, gdybym chciał przyjąć owe szczegółowe kategorie (przedmiot pytania oraz rodzaj pytania), analizowane zbiory okazałyby się wbrew pozorom zbyt mało liczebne.

Dlatego też skupię się jedynie na jednym rodzaju pytań, mianowicie pytaniach ze skalami semantycznymi. Analizując rzetelność odpowiedzi w zależności od struktury pytania, stwierdziłem, iż odpowiedzi na pytania zawierające skale są – wbrew oczekiwaniom – w relatywnie niskim stopniu „reprodukowane” (ok. 50%). Być może jednak jest to efekt „skali”: jeśli skala jest zbyt „czuła”, zawiera zbyt wiele punktów, to właśnie dlatego jej „reproduktywność” jest tak niska. Interesujące zatem wydaje się sprawdzenie hipotezy, na ile spłaszczenie skali 7-, 5- lub 4-punktowej, do 3- lub 2-punktowej zmniejsza wielkość błędu. Mówiąc

¹⁷ W badaniu tym wywiady prowadzono w gospodarstwach domowych równoległe z mężem i żoną (w osobnych pomieszczeniach, bez możliwości wzajemnej konsultacji), a następnie wywiad w „diadzie” (mąż i żona), z możliwością wzajemnej konsultacji. W kolejnym kroku odpowiedzi uzyskane od poszczególnych osób oraz „wspólne” były porównywane z danymi dokumentalnymi.

Tabela 28. *Pytania o opinie: odsetek odpowiedzi zgodnych według skali zawartej w pytaniu oraz według skali „skróconej”*

Cecha	Skala wg pytania (% zgodności)	Skala „skrócona” (% zgodności)”
Prowadzenie gospodarstwa jako obowiązek kobiety (skala 4 pkt. – skala 2 pkt.)	42,0	71,5
Prowadzenie gospodarstwa dom. to praca (skala 4 pkt. – skala 2 pkt.)	53,2	82,1
Prace domowe jako forma wykorzystywania kobiet (skala 4 pkt. – skala 2 pkt.)	36,0	61,3
Zadolenie z wykonywania czynności domowych (skala 4 pkt. – skala 2 pkt.)	44,0	60,7
Poczucie bezpieczeństwa respondenta w swojej dzielnicy (skala 4 pkt. – skala 2 pkt.)	70,2	96,2
Poczucie bezpieczeństwa respondenta po zmroku w okolicach domu (skala 4 pkt. – skala 2 pkt.)	57,1	73,8
Czynności domowe jako obowiązek (skala 4 pkt. – skala 2 pkt.)	48,2	85,9
Częstotliwość załatwiania sprawy „po znajomości” (skala 4 pkt. – skala 2 pkt.)	42,3	75,4
Ocena słuszności idei prywatyzacji gospodarki polskiej (skala 4 pkt. – skala 2 pkt.)	39,4	62,0
Ocena sytuacji finansowej GD. (skala 4 pkt. – skala 3 pkt.)	66,2	80,1
Częstotliwość załatwiania sprawy „jeśli się da” (skala 5 pkt. – skala 2 pkt.)	48,0	91,3
Ocena stanu zdrowia respondenta (skala 5 pkt. – skala 3 pkt.)	63,4	78,2
Ocena działań państwa w zakresie zwalczania korupcji (skala 5 pkt. – skala 3 pkt.)	43,8	70,6
Na ile ważne byłyby dla respondenta wysokie dochody przy wyborze pracy (skala 5 pkt. – skala 3 pkt.)	54,6	88,6
Na ile ważna byłaby dla respondenta pewność zatrudnienia przy wyborze pracy (skala 5 pkt. – skala 3 pkt.)	48,0	90,7
Na ile ważna byłaby dla respondenta możliwość łączenia obowiązków rodzinnych i zawodowych przy wyborze pracy (skala 5 pkt. – skala 3 pkt.)	47,3	77,7
Częstotliwość występowania w Polsce korupcji (skala 5 pkt. – skala 3 pkt.)	52,6	81,9
Walka z ubóstwem jako cel władzy i rządu (skala 4 pkt. – skala 2 pkt.)	55,9	87,0
Zadolenie respondenta z wykonywania prac domowych (skala 7 pkt. – skala 3 pkt.)	26,1	45,1

o „skróceniu” lub „spłaszczeniu” skali, mam na myśli zmniejszenie liczby jej „punktów”, z czterech do dwóch, pięciu do trzech itp. Przykładowo skalę 4-punktową typu: „1. bardzo często; 2. raczej często; 3. raczej rzadko; 4. bardzo rzadko” można skrócić do 2-punktowej: „1. często; 2. rzadko”. Podobnie, skalę 5-punktową typu: „1. bardzo często; 2. raczej często; 3. ani często, ani rzadko; 4. raczej rzadko; 5. bardzo rzadko” można skrócić do 3-punktowej: „1. często; 2. ani często, ani rzadko; 3. rzadko”. Podobnie, skalę: „1. bardzo dobry; 2. dobry; 3. średni; 4. zły; 5. bardzo zły” można zrekodować na skalę: „1. bardzo dobry/ dobry; 2. średni; 3. zły/ bardzo zły”.

Problem ten dotyczy nie tylko skal, ale także poziomu „dokładności” innych zmiennych, na przykład wieku, liczby lat skolaryzacji (na przykład zamiast dokładności +/- 1 rok, tolerancja +/- 3 lata). Z pewnością, „spłaszczenie” skali daje efekt zwiększenia odsetka odpowiedzi zgodnych, ale powstaje pytanie, na ile jest to wzrost istotny.

Zacznijmy od pytań o opinie. Poniższa tabela zawiera odsetki odpowiedzi zgodnych w powtórzonym pomiarze dla pytań ze skalą zastosowaną w pytaniu „źródłowym” oraz dla pytań ze skalą „skróconą” („spłaszczoną”).

W przypadku powyższych pytań o opinie odsetek odpowiedzi zgodnych, mierzonych według kategorii odpowiedzi zastosowanych w pytaniach kwestionariuszowych, jest bardzo zróżnicowany i waha się od 26% do 70%. Jeśli jednak „spłaszczymy” te skale do 2-, 3-punktowych, odsetek odpowiedzi zgodnych wzrasta o ponad 20 punktów i sięga 96 (w przypadku niektórych pytań wzrasta on niemal dwukrotnie). Należy podkreślić, iż zabieg ten w mniejszym stopniu podnosi zgodność odpowiedzi, gdy respondenci mają ugruntowaną opinię w danej sprawie, na przykład na temat stosunków seksualnych osób tej samej płci, stosunków seksualnych przed ślubem (nie zostało to uwzględnione w powyższej tabeli).

Przejdźmy do pytań o fakty. Czy i na ile „spłaszczenie” skali lub „zwiększenie przedziału tolerancji” powoduje wzrost odsetka zgodności odpowiedzi w powtórzonym pomiarze na pytania o fakty? Tabela 29 zawiera procentowe rozkłady odpowiedzi zgodnych dla pytań ze skalą zastosowaną w pytaniu oraz dla pytań ze skalą „skróconą” (rekodowaną). Zawarto w niej zasady „zwiększania tolerancji” i „skracania” skal.

Tabela 29. *Pytania o fakty: odsetek odpowiedzi zgodnych według skali zawartej w pytaniu oraz według skali „skróconej”*

Cecha	Skala wg pytania (% zgodności)	Skala „skrócona” (% zgodności)”
Wyposażenie gosp. dom.: Rok dopuszczenia do ruchu samochodu-pierwszej rejestracji (dokładnie – z dokładnością +/-1 rok)	79,3	92,1
Wyposażenie gosp. dom.: Średni przebieg roczny samochodu (dokładnie – z dokładnością +/-2 tys. km)	42,3	57,7
Wyposażenie gosp. dom.: Obecny przebieg samochodu (dokładnie – z dokładnością +/-5 tys. km)	27,5	73,3
Sytuacja zawodowa respondenta według kategorii pytania – według: pracuje – nie pracuje	87,1	92,7
Stan zdrowia respondenta: Bóle stawu biodrowego (skala 6 pkt. – skala 4 pkt.)	57,0	69,4
Stan zdrowia respondenta: Bóle rąk (skala 6 pkt. – skala 4 pkt.)	52,7	71,1
Stan zdrowia respondenta: Bóle odcinka szyjnego kręgosłupa (skala 6 pkt – skala 4 pkt.)	50,8	61,4
Stan zdrowia respondenta: Bóle odcinka piersiowego kręgosłupa (skala 6 pkt. – skala 4 pkt.)	51,9	63,0
Stan zdrowia respondenta: Bóle odcinka lędźwiowego kręgosłupa (skala 6 pkt. – skala 4 pkt.)	46,3	54,7
Stan zdrowia respondenta: Bóle mięśni, ścięgien, więzadeł (skala 6 pkt. – skala 4 pkt.)	35,3	47,9
Stan zdrowia respondenta: Bóle kolan (skala 6 pkt. – skala 4 pkt.)	48,9	58,9
Stan zdrowia respondenta: Bóle innych stawów (skala 6 pkt. – skala 4 pkt.)	64,4	73,4
Stan zdrowia respondenta: Bolesne obrzęki stawów stóp (skala 6 pkt. – skala 4 pkt.)	63,9	69,6
Stan zdrowia respondenta: Bolesne obrzęki stawów rąk (skala 6 pkt. – skala 4 pkt.)	67,2	71,4
Stan zdrowia respondenta: Bolesne obrzęki stawów kolan (skala 6 pkt. – skala 4 pkt.)	68,2	71,4

W przypadku pytań o fakty odsetek odpowiedzi zgodnych jest zróżnicowany i waha się od 27 do 87, o ile analizujemy je według kategorii odpowiedzi zastosowanych w pytaniu kwestionariuszowym. Jeśli jednak

„spłaszczymy” te skale do dwóch trzecich punktów lub założymy wyższy poziom „tolerancji”, odsetek odpowiedzi zgodnych wzrasta o ponad 20 punktów i sięga 93 (w przypadku niektórych informacji: obecny przebieg samochodu – dokładnie/z dokładnością +/-5 tys km – wzrasta on niemal trzykrotnie, z 27 do 73). Należy podkreślić, iż zabieg ten w mniejszym stopniu podnosi zgodność odpowiedzi, gdy pytania dotyczą na przykład stanu zdrowia respondenta, w szczególności tak zwanych poważnych dolegliwości (zawał serca, perforacja, krwawienie z przewodu pokarmowego, choroba wrzodowa żołądka/dwunastnicy, rozpoznana osteoporoza).

Przejdźmy do pytań o zachowania. Na ile są one podatne na „spłaszczenie” skali lub „zwiększenie przedziału tolerancji”? Tabela 30 zawiera procentowe rozkłady odpowiedzi zgodnych dla pytań ze skalą zastosowaną w pytaniu oraz dla pytań ze skalą „skróconą” (rekodowaną).

Tabela 30. *Pytania o zachowania: odsetek odpowiedzi zgodnych według skali zawartej w pytaniu oraz według skali „skróconej”*

Cecha	Skala wg pytania (% zgodności)	Skala „skrócona” (% zgodności)”
Konsumpcja respondenta w ostatnim czasie: serek wiejski (skala 7 pkt. – skala 3 pkt.)	27,1	67,5
Konsumpcja respondenta w ostatnim czasie: serek homogenizowany (skala 7 pkt. – skala 3 pkt.)	24,6	72,9
Konsumpcja respondenta w ostatnim czasie: kefir (skala 7 pkt. – skala 3 pkt.)	21,0	58,0
Konsumpcja respondenta w ostatnim czasie: jogurt (skala 7 pkt. – skala 3 pkt.)	61,5	79,7
Częstotliwość picia kawy rozpuszczalnej przez respondenta (skala 10 pkt. – skala 4 pkt.)	35,2	62,8
Częstotliwość picia kawy mielonej przez respondenta (skala 10 pkt. – skala 4 pkt.)	47,7	85,9
Częstotliwość picia cappucino przez respondenta (skala 10 pkt. – skala 4 pkt.)	25,2	44,2
Częstotliwość korzystania przez respondenta z Internetu (skala 7 pkt. – skala 4 pkt.)	62,4	69,4

W przypadku pytań o zachowania odsetek odpowiedzi zgodnych jest różnicowany i waha się od 21 do 62, o ile analizujemy je według kategorii odpowiedzi zastosowanych w pytaniu kwestionariuszowym. Jeśli

jednak „spłaszczymy” te skale do 4 lub 3 punktów, odsetek odpowiedzi zgodnych wzrasta o ponad 20 punktów procentowych i sięga 86% zgodności (w przypadku niektórych informacji: konsumpcja w ostatnim czasie serka homogenizowanego – wzrasta on niemal trzykrotnie, z 24 do 73).

I wreszcie pytania o cechy. Związek przedmiotu pytania i jego struktury (charakteru zastosowanej w pytaniu skali) przedstawia tabela 31.

Tabela 31. *Pytania o cechy: odsetek odpowiedzi zgodnych według skali zawartej w pytaniu oraz według skali „skróconej”*

Cecha	Skala wg pytania (% zgodności)	Skala „skrócona” (% zgodności)”
Wykształcenie respondenta (według 9 kategorii pytania – według 4 poziomów wykształcenia)	80,8	87,1
Wiek respondenta (dokładnie – z dokładnością +/-1 rok)	96,2	97,8
Wiek głównego użytkownika samochodu (dokładnie – z dokładnością +/-1 rok)	90,3	92,7
Waga respondenta (dokładnie – z dokładnością +/- 5 kg)	60,0	91,0
Liczba lat nauki respondenta (dokładnie – z dokładnością dokładnością +/-1 rok)	70,8	92,5

W przypadku pytań o cechy efekt spłaszczania skal powoduje – podobnie jak w przypadku innych pytań – wzrost odsetka zgodności odpowiedzi. W przypadku tych pytań wzrost ten jest relatywnie niski: od 1,6 punktu procentowego (wiek respondenta) do 21,7 punktu (liczba lat nauki). Warto tutaj wspomnieć o efekcie „poszerzania” kohort wiekowych w przypadku respondenta: przyjmując 5-letnie kohorty wiekowe, odsetek zgodności odpowiedzi wynosi jedynie 85,1%, a więc jest niższy niż wtedy, gdy analizujemy wiek respondenta z dokładnością +/- 1 rok (97,8) [Sztabiński 2001: 91–114].

Przedstawiony powyżej efekt „spłaszczania skal” lub efekt przyjęcia wyższego poziomu „tolerancji” pomiaru pewnych zmiennych z pewnością nie jest zaskoczeniem. Im szersze kategorie analizy stosujemy, tym bardziej kategorie są spójne (Lozano, Garcia-Cueto, Muniz 2008: 73–79). Nikt tej ogólnej tendencji nie podważa. Czy wiadomo jednak, jak jest to w przypadku różnych pytań, w zależności od ich przedmiotu? Na

ile „spłaszczanie skal” lub „podwyższenie poziomu tolerancji” wpływa na zgodność wyniku? Wydaje się, że nie.

4. 3. Trzykrotny pomiar

Dotychczas rozważaliśmy sytuacje, w których ocena wartości danych dokonywana była na podstawie powtórnego pomiaru. Umożliwia to wskazanie rozbieżności – zbieżności między tymi pomiarami, ale nie przesądza, który z nich jest bardziej wiarygodny: zasadniczy czy kontrolny. Spróbuję zatem odpowiedzieć na pytanie, czy powtórzenie pomiaru po raz trzeci lub odwołanie się do danych dokumentalnych rozstrzyga problem rzetelności.

W przypadku oceny bezpośredniej dysponujemy danymi z dwóch źródeł: z wywiadu zasadniczego i z wywiadu kontrolnego. Pozwoli to na ocenę szacunkowej rzetelności uzyskanych danych. Niekiedy jednak – ze względu na schemat badania – dysponujemy dwukrotnymi deklaracjami badanych z wywiadu oraz uzyskanymi po raz trzeci w wywiadzie kontrolnym. Dość często dysponujemy również danymi z trzech różnych źródeł: danymi z wywiadu zasadniczego, kontrolnego oraz danymi dokumentalnymi. W pierwszym przypadku możemy mówić o „klasycznym” trzykrotnym pomiarze, w drugim zaś – o testowaniu rzetelności poprzez walidację (wynik powtórnego pomiaru oceniamy bowiem ze względu na dane dokumentalne).

4. 3. 1. Deklaracje respondentów

Na pytanie o trzykrotny pomiar deklaracji respondentów spróbuję udzielić odpowiedzi, analizując wyniki wspomnianego wcześniej badania, zatytułowanego „Rok Rodziny”. Przypomnijmy, że badanie to było realizowane na podstawie specyficznego, quasi-eksperymentalnego schematu: do badania dobrano dwie porównywalne, ze względu na płeć, wiek i wykształcenie próby udziałowe (A i B). Z osobami z obu prób zrealizowano wywiad na temat ich sytuacji zdrowotnej oraz opinii o „zdrowym sposobie życia” (Faza 1A i Faza 1B). Następnie jednej z grup (A) zaprezentowano serię filmów na temat „zdrowego sposobu życia” (bodziec). Po upływie 2–3 miesięcy z respondentami obu grup przeprowa-

dzono analogiczny wywiad na temat ich sytuacji zdrowotnej oraz opinii o „zdrowym sposobie życia” (Faza 2A i Faza 2B). Niezależnie od tego wśród respondentów obu grup przeprowadzono kontrolę, w trakcie której zadawano niektórym respondentom, po raz trzeci wybrane pytania z kwestionariusza zasadniczego, które dotyczyły nie tylko cech metryczkowych (przy czym nie były to pytania, na które odpowiedzi, zgodnie z założeniami schematu eksperymentalnego, mogły ulec modyfikacji). Dzięki temu dysponowałem danymi z trzech pomiarów: z wywiadów zrealizowanych w fazie 1 (Fz1), wywiadów w fazie 2 (Fz2) oraz z wywiadu kontrolnego (K). Uzyskane wyniki przedstawiono w tabeli 32 (ze względu na niskie liczebności, nie przedstawiam odsetka, lecz jedynie stosunek liczebności wyniku pomiaru 1, 2 i 3).

Tabela 32. *Proporcje między liczbą odpowiedzi zgodnych a liczbą wszystkich odpowiedzi na dane pytanie (w nawiasie proporcje te wyrażono w postaci ułamka dziesiętnego)*

Przedmiot	Pomiar 1: Fz1 – Fz2	Pomiar 2: Fz1 – K	Pomiar 3: Fz2 – K
Płeć respondenta	50/50 (1,00)	45/45 (1,00)	45/45 (1,00)
Wiek respondenta. Z dokładnością do 1 roku	50/50 (1,00)	43/45 (0,96)	43/45 (0,96)
Wykształcenie respondenta według kategorii pytania	41/50 (0,82)	25/45 (0,56)	29/45 (0,64)
Wzrost respondenta z dokładnością do 1 cm	36/50 (0,72)	30/42 (0,71)	27/42 (0,64)
Liczba osób w GD z dokł. do 1 osoby	46/50 (0,92)	35/44 (0,80)	32/44 (0,73)
Stan zdrowia respondenta: choroba/atak serca	45/50 (0,90)	41/43 (0,95)	40/43 (0,93)
Stan zdrowia respondenta: nadciśnienie	47/50 (0,94)	41/43 (0,95)	42/43 (0,98)
Stan zdrowia respondenta: wysoki poziom cholesterolu	46/48 (0,96)	42/43 (0,98)	42/43 (0,98)
Stan zdrowia respondenta: choroba układu oddechowego	43/50 (0,86)	39/43 (0,91)	41/43 (0,95)
Stan zdrowia respondenta: otyłość, nadwaga	45/50 (0,90)	39/43 (0,91)	36/43 (0,84)
Stan zdrowia respondenta: uszkodzenie wątroby	47/50 (0,94)	41/43 (0,95)	41/43 (0,95)
Stan zdrowia respondenta: rak odbytu	50/50 (1,00)	43/43 (1,00)	43/43 (1,00)
Stan zdrowia respondenta: rak płuc	49/49 (1,00)	43/43 (1,00)	42/42 (1,00)
Palenie papierosów przez respondenta	47/49 (0,96)	39/42 (0,93)	38/43 (0,88)

Analizując powyższe dane, można zauważyć, że w przypadku trzech zmiennych: płeć oraz stwierdzone choroby nowotworowe, wszystkie trzy pomiary charakteryzują się pełną zgodnością. W przypadku pozostałych zmiennych (z wyjątkiem trzech: wykształcenia, wzrostu i liczby osób w gospodarstwie domowym) uzyskano podobne rezultaty: liczba odpowiedzi zgodnych jest zbliżona. Przy czym w przypadku schorzeń powszechnie uważanych za „poważniejsze” (choroby krążenia, uszkodzenie wątroby) zgodność pomiarów jest wyższa niż w przypadku „łżejszych” (choroba układu oddechowego, otyłość, nadwaga). Potwierdza to wspomniana pełna zgodność odpowiedzi na pytania o choroby nowotworowe.

W przypadku trzech zmiennych: wykształcenia, wzrostu oraz liczby osób w gospodarstwie domowym, w poszczególnych pomiarach liczba odpowiedzi zgodnych jest znacznie niższa, a w niektórych przypadkach uzyskano odmienne wartości w trzech pomiarach (oznacza to, iż zadając to samo pytanie temu samemu respondentowi, uzyskano trzy różne odpowiedzi). Sytuacja taka wystąpiła w przypadku trzech wspomnianych pytań o:

- wykształcenie respondenta: w jednym przypadku uzyskano trzy różne odpowiedzi (wykształcenie podstawowe; zasadnicze zawodowe oraz średnie zawodowe),

- wzrost respondenta w centymetrach: w trzech przypadkach uzyskano trzy różne odpowiedzi (wzrost 161/165/164; 168/175/164 i 173/172/174) oraz

- liczbę osób w gospodarstwie domowym respondenta: w jednym przypadku uzyskano trzy różne odpowiedzi (5 osób, 4 osoby oraz 1 osoba).

Co to oznacza? Chyba nic innego niż sytuację przedstwioną przez Akiro Kurosawę w filmie *Rashōmon* z 1950 r., w którym różne wersje popełnionego mordu i gwałtu prezentowane są tak, że nie wiadomo, kto jest sprawcą, kto ofiarą i co prawdą.

4. 3. 2. Deklaracje respondentów a dane dokumentalne

Analizując niektóre badania, wykorzystałem dane z trzech różnych źródeł: z wywiadu zasadniczego, z wywiadu kontrolnego oraz dane dokumentalne, przy czym dotyczyło to danych faktualnych – demograficznych: płeć, wiek – zaczerpnięte z próby. Było to więc połączenie: powtórnego pomiaru oraz walidacji danych. W przypadku zmiennej

„płeć” zgodność jest oczywiście stuprocentowa. Inaczej jednak wygląda to w przypadku wieku.

Tabela 33. Średni odsetek zgodności odpowiedzi w badaniu kontrolnym i zasadniczym (K/Z), kontrolnym i danymi dokumentalnymi (K/D) oraz w badaniu zasadniczym i danymi dokumentalnymi (Z/D)

Cecha	Pomiar 1–2 (K/Z)	Pomiar 1–3 (K/D)	Pomiar 2–3 (Z/D)
Wiek respondenta z dokładnością do 1 roku	96,2	89,6	87,8
Wiek respondenta z dokładnością +/-1 rok	97,8	91,4	90,5

Co bardzo dziwne, okazuje się, że największe rozbieżności występują między deklaracjami w wywiadzie zasadniczym i danymi dokumentalnymi. Najmniejsze zaś między deklaracjami w wywiadzie kontrolnym i zasadniczym. Świadczyć to może o niskiej rzetelności danych dokumentalnych, ale równie dobrze o wspomnianej wcześniej „ułomności” ludzkiej pamięci. Jest to jednocześnie świadectwo rzetelności pomiaru, oczywiście w ograniczonym zakresie. To samo dotyczy danych z wywiadu zasadniczego oraz kontrolnego: dane te są wyraźnie bliższe niż wszystkie pozostałe. Nie potrafię tego wyjaśnić.

4. 4. Wpływ rozbieżności informacji na wynik badania

Wcześniej wskazałem, iż praktycznie w przypadku każdej zmiennej występuje różnica między pomiarem zasadniczym i kontrolnym. Oznacza to, iż rzetelność uzyskanych informacji jest dyskusyjna. Jakie są jednak konsekwencje posługiwania się danymi nie w pełni rzetelnymi dla wyników badań? Najistotniejszą kwestią jest wpływ błędów na siłę zależności między rozpatrywanymi zmiennymi, które dla badacza są podstawą wnioskowania o mechanizmach i zjawiskach społecznych. Gdyby okazało się, iż na przykład korelacje między odpowiedziami respondentów na pytanie X a ich odpowiedziami na pytanie Y oraz analogicznymi uzyskanymi w powtórny pomiarze (tzn. X' i Y') różnią się, można by wówczas zakwestionować trafność wniosków odwołujących się do zależności między tymi zmiennymi.

W celu rozpoznania tej kwestii, to znaczy wpływu rzetelności pomiaru na siłę zależności między zmiennymi, wykorzystam dane z badań, w których liczba przypadków była znacząca. Będą to: Omnibus (6436 przypadków), Parasol (2263), TGI (1564), Omnimas (1507 przypadków), Tenpies (629), Szklanka (571) oraz Tankard (464). W swoich analizach uwzględniłam cztery podstawowe zmienne demograficzne i społeczne („metryczkowe”), a mianowicie: płeć, wiek, wykształcenie oraz sytuacja zawodowa respondenta. Zmienne te stanowią stałą i integralną część każdego kwestionariusza badań ilościowych i w analizach pełnią rolę tak zwanych zmiennych niezależnych, ze względu na które analizuje się zmienne „merytoryczne”.

Zacznijmy od analizy związku płci z innymi zmiennymi demograficznymi i dotyczącymi konsumpcji. W tabeli 34 przedstawiłem wartości współczynnika Cramera dla zależności między płcią a wybranymi zmiennymi dla pierwszego i powtórnego pomiaru¹⁸. W przypadku zależności ze zmienną „wiek” odwołuje się do danych z siedmiu badań, w przypadku „wykształcenia” – sześciu, „sytuacji zawodowej” – trzech, a w przypadku zależności ze zmienną „dochód” do jednego badania. Wynika to stąd, że w badaniach kontrolnych nie zawsze zamieszczany był ten sam zestaw pytań powtórzonych.

¹⁸ Posłużyłem się współczynnikiem Cramera, ponieważ jest on powszechnie stosowanym miernikiem siły związku między zmiennymi reprezentującymi „najśłabszy” poziom pomiaru. Wydaje się on najbardziej adekwatnym wskaźnikiem korelacji w moich analizach, w sytuacji gdy wiele z uwzględnionych zmiennych „kryterialnych” to zmienne nominalne, a względy porównywalności przemawiają za zastosowaniem jednolitego miernika.

Tabela 34. *Wartość współczynników korelacji między odpowiedziami respondenta na wybrane pytania w pierwszym i powtórzonym pomiarze, a płcią respondenta*

Zmienna 1	Pozostałe zmienne	Wywiad zasadniczy	Wywiad kontrolny
Płeć	Wiek respondenta z dokładnością do 1 roku	$V = 0,125$	$V = 0,124$
		$V = 0,222$	$V = 0,226$
		$V = 0,178$	$V = 0,181$
		$V = 0,279$	$V = 0,290$
		$V = 0,326$	$V = 0,325$
		$V = 0,354$	$V = 0,360$
	Wykształcenie respondenta	$V = 0,246$	$V = 0,244$
		$V = 0,211$	$V = 0,214$
		$V = 0,160$	$V = 0,158$
		$V = 0,084$	$V = 0,113$
		$V = 0,214$	$V = 0,226$
		$V = 0,248$	$V = 0,276$
	Sytuacja zawodowa respondenta	$V = 0,222$	$V = 0,202$
		$V = 0,287$	$V = 0,295$
		$V = 0,252$	$V = 0,247$
	Dochód respondenta	$V = 0,182$	$V = 0,169$
		$V = 0,414$	$V = 0,371$
	Konsumpcja przez respondenta kawy	$V = 0,014$	$V = 0,116$
		Konsumpcja przez respondenta (od czasu do czasu) kawy naturalnej, mielonej	$V = 0,110$
	Częstotliwość picia przez respondenta kawy naturalnej, mielonej		$V = 0,050$
Konsumpcja przez respondenta (od czasu do czasu) kawy rozpuszczalnej		$V = 0,152$	$V = 0,289$
	Częstotliwość picia przez respondenta kawy rozpuszczalnej	$V = 0,038$	$V = 0,091$
Konsumpcja przez respondenta (od czasu do czasu) kawy cappuccino		$V = 0,022$	$V = 0,038$
	Częstotliwość picia przez respondenta kawy Cappuccino	$V = 0,241$	$V = 0,262$
Konsumpcja przez respondenta (od czasu do czasu) kawy cappuccino		$V = 0,065$	$V = 0,108$
	Częstotliwość picia przez respondenta kawy Cappuccino	$V = 0,057$	$V = 0,119$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle piwa		$V = 0,215$	$V = 0,201$
		$V = 0,117$	$V = 0,098$

Konsumpcja przez respondenta piwa w ciągu ostatniego miesiąca	$V = 0,170$	$V = 0,003$
Konsumpcja przez respondenta piwa w okresie ostatnich trzech miesięcy	$V = 0,332$	$V = 0,395$
Konsumpcja przez respondenta wina	$V = 0,113$	$V = 0,038$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle szampana, wina musującego	$V = 0,135$	$V = 0,024$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle wódki bez dodatków	$V = 0,337$	$V = 0,324$
Konsumpcja przez respondenta: jogurt	$V = 0,034$	$V = 0,040$
Konsumpcja przez respondenta: kefir	$V = 0,032$	$V = 0,017$
Dostęp respondenta do internetu	$V = 0,126$	$V = 0,132$
Korzystanie z internetu przez respondenta	$V = 0,040$	$V = 0,014$
Częstotliwość korzystania przez respondenta z internetu	$V = 0,180$	$V = 0,192$
Posiadanie telefonu komórkowego przez respondenta	$V = 0,157$	$V = 0,134$
Rodzaj telefonu komórkowego respondenta (prywatna)	$V = 0,088$	$V = 0,078$
Rodzaj telefonu komórkowego respondenta (służbowa)	$V = 0,128$	$V = 0,129$
Korzystanie przez respondenta z automatów telefonicznych w miejscach publicznych	$V = 0,073$	$V = 0,067$
Wyjazdy respondenta co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajowe i zagraniczne)	$V = 0,093$	$V = 0,078$
Wyjazdy wakacyjne/urlopowe respondenta w ciągu ostatnich 12 miesięcy	$V = 0,062$	$V = 0,076$

Tabela 35 zawiera współczynniki korelacji V Cramera dla pierwszego i powtórnego pomiaru dla związków z wiekiem, tabele 36 i 37 zaś, odpowiednio, dla pierwszego i powtórnego pomiaru dla związków z wykształceniem i sytuacją zawodową.

Tabela 35. Wartość współczynników korelacji między odpowiedziami respondentów na wybrane pytania w pierwszym i powtórzonym pomiarze, a wiekiem respondenta

Zmienna 2	Zmienna 3 i następane	Wywiad zasadniczy	Wywiad kontrolny
Wiek	Wykształcenie respondenta	$\tau = 0,116^{**}$	$\tau = 0,137^{**}$
		$\tau = 0,069^{**}$	$\tau = 0,074^{**}$
		$\tau = 0,149^{**}$	$\tau = 0,163^{**}$
		$\tau = 0,069^*$	$\tau = 0,075^*$
		$\tau = 0,095^*$	$\tau = 0,150^{**}$
		$\tau = 0,094^{**}$	$\tau = 0,146^{**}$
	Sytuacja zawodowa respondenta	$V = 0,427$	$V = 0,416$
		$V = 0,490$	$V = 0,489$
		$V = 0,456$	$V = 0,435$
	Dochód respondenta	$r = -0,205^{**}$	$r = -0,185^{**}$
	Liczba osób w gospodarstwie domowym respondenta	$r = 0,347^{**}$	$r = 0,329^{**}$
		$r = 0,354^{**}$	$r = 0,339^{**}$
		$r = 0,250^{**}$	$r = 0,230^{**}$
		$r = 0,166^{**}$	$r = 0,101^*$
		$r = 0,205^{**}$	$r = 0,233^{**}$
		$r = 0,309^{**}$	$r = 0,379^{**}$
	Liczba osób wspólnie zamieszkujących z respondentem (także nie tworzących gospodarstwa domowego)	$r = 0,321^{**}$	$r = 0,287^{**}$
		$r = 0,366^{**}$	$r = 0,352^{**}$
	Liczba dzieci w gospodarstwie domowym do lat 18	$r = 0,070$	$r = 0,059$
	Konsumpcja przez respondenta kawy	$V = 0,365$	$V = 0,294$
Konsumpcja przez respondenta (od czasu do czasu) kawy naturalnej, mielonej	$V = 0,344$	$V = 0,307$	
	$V = 0,333$	$V = 0,372$	
Częstotliwość picia przez respondenta kawy naturalnej, mielonej	$r = 0,144^{**}$	$r = 0,178^{**}$	
Konsumpcja przez respondenta (od czasu do czasu) kawy rozpuszczalnej	$V = 0,314$	$V = 0,278$	
	$V = 0,199$	$V = 0,278$	

Częstotliwość picia przez respondenta kawy rozpuszczalnej	$r = 0,147^*$	$r = 0,172^*$
Konsumpcja przez respondenta (od czasu do czasu) kawy cappuccino	$V = 0,338$	$V = 0,328$
	$V = 0,364$	$V = 0,412$
Częstotliwość picia przez respondenta kawy cappuccino	$r = -0,012$	$r = -0,033$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle piwa	$V = 0,415$	$V = 0,384$
Konsumpcja przez respondenta piwa w ciągu ostatniego miesiąca	$V = 0,341$	$V = 0,499$
Konsumpcja przez respondenta piwa w okresie ostatnich trzech miesięcy	$V = 0,376$	$V = 0,375$
Konsumpcja przez respondenta wina	$V = 0,323$	$V = 0,340$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle szampana, wina musującego	$V = 0,345$	$V = 0,355$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle wódki bez dodatków	$V = 0,397$	$V = 0,420$
Konsumpcja przez respondenta: jogurt	$V = 0,495$	$V = 0,392$
Konsumpcja przez respondenta: kefir	$V = 0,345$	$V = 0,357$
Dostęp respondenta do internetu	$V = 0,487$	$V = 0,471$
Korzystanie z internetu przez respondenta	$V = 0,423$	$V = 0,473$
Częstotliwość korzystania przez respondenta z internetu	$r = -0,234^{**}$	$r = -0,076$
Posiadanie telefonu komórkowego przez respondenta	$V = 0,452$	$V = 0,503$
Rodzaj telefonu komórkowego respondenta (prywatna)	$V = 0,584$	$V = 0,543$
Rodzaj telefonu komórkowego respondenta (służbowa)	$V = 0,376$	$V = 0,355$
Korzystanie przez respondenta z automatów telefonicznych w miejscach publicznych	$V = 0,430$	$V = 0,492$
Wyjazdy respondenta co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajowe i zagraniczne)	$V = 0,324$	$V = 0,359$
Wyjazdy wakacyjne/urlopowe respondenta w ciągu ostatnich 12 miesięcy	$V = 0,322$	$V = 0,329$

* Korelacja istotna na poziomie 0,05 (test dwustronny).

** Korelacja istotna na poziomie 0,01 (test dwustronny).

Tabela 36. *Wartość współczynników korelacji między odpowiedziami respondenta na wybrane pytania w pierwszym i powtórzonym pomiarze, a wykształceniem respondenta*

Zmienna 3	Zmienna 4 i następane	Wywiad zasadniczy	Wywiad kontrolny
Wykształcenie	Sytuacja zawodowa respondenta	$V = 0,260$	$V = 0,259$
		$V = 0,244$	$V = 0,264$
		$V = 0,243$	$V = 0,265$
	Dochód respondenta	$\tau = 0,149^{**}$	$\tau = 0,115^{**}$
	Liczba osób w gospodarstwie domowym respondenta	$\tau = -0,042^{**}$	$\tau = -0,034^{**}$
		$\tau = -0,077^{**}$	$\tau = -0,089^{**}$
		$\tau = -0,035$	$\tau = -0,043^*$
		$\tau = -0,104^{**}$	$\tau = -0,120^{**}$
		$\tau = -0,031$	$\tau = -0,030$
		$\tau = -0,007$	$\tau = -,010$
	Liczba osób wspólnie zamieszkujących z respondentem (także nie tworzących gospodarstwa domowego)	$\tau = -0,032$	$\tau = -0,017$
	Liczba dzieci w gospodarstwie domowym do lat 18	$\tau = -0,116^{**}$	$\tau = -0,134^{**}$
	Konsumpcja przez respondenta kawy	$V = 0,133$	$V = 0,138$
	Konsumpcja przez respondenta (od czasu do czasu) kawy naturalnej, mielonej	$V = 0,195$	$V = 0,186$
		$V = 0,165$	$V = 0,176$
	Częstotliwość picia przez respondenta kawy naturalnej, mielonej	$\tau = 0,120^{**}$	$\tau = 0,069$
	Konsumpcja przez respondenta (od czasu do czasu) kawy rozpuszczalnej	$V = 0,202$	$V = 0,183$
		$V = 0,227$	$V = 0,213$
	Częstotliwość picia przez respondenta kawy rozpuszczalnej	$\tau = -0,038$	$\tau = -0,070$
	Konsumpcja przez respondenta (od czasu do czasu) kawy cappuccino	$V = 0,121$	$V = 0,171$
$V = 0,144$		$V = 0,105$	
Częstotliwość picia przez respondenta kawy cappuccino	$\tau = 0,025$	$\tau = 0,090$	

Konsumpcja przez respondenta kiedykolwiek/ w ogóle piwa	$V = 0,134$	$V = 0,195$
Konsumpcja przez respondenta piwa w ciągu ostatniego miesiąca	$V = 0,211$	$V = 0,146$
Konsumpcja przez respondenta piwa w okresie ostatnich trzech miesięcy	$V = 0,239$	$V = 0,209$
Konsumpcja przez respondenta wina	$V = 0,249$	$V = 0,263$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle szampana, wina musującego	$V = 0,283$	$V = 0,266$
Konsumpcja przez respondenta kiedykolwiek/ w ogóle wódki bez dodatków	$V = 0,227$	$V = 0,235$
Konsumpcja przez respondenta: jogurt	$V = 0,253$	$V = 0,296$
Konsumpcja przez respondenta: kefir	$V = 0,203$	$V = 0,148$
Dostęp respondenta do internetu	$V = 0,385$	$V = 0,344$
Korzystanie z internetu przez respondenta	$V = 0,283$	$V = 0,261$
Częstotliwość korzystania przez respondenta z internetu	$\tau = -0,087$	$\tau = 0,138^*$
Rodzaj telefonu komórkowego respondenta (prywatna)	$V = 0,310$	$V = 0,329$
Rodzaj telefonu komórkowego respondenta (służbowa)	$V = 0,197$	$V = 0,236$
Wyjazdy respondenta co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajo- we i zagraniczne)	$V = 0,149$	$V = 0,133$
Wyjazdy wakacyjne/urlopowe respondenta w ciągu ostatnich 12 miesięcy	$V = 0,276$	$V = 0,320$

* Korelacja istotna na poziomie 0,05 (test dwustronny).

** Korelacja istotna na poziomie 0,01 (test dwustronny).

Tabela 37. Wartość współczynników korelacji między odpowiedziami respondenta na wybrane pytania w pierwszym i powtórzonym pomiarze, a sytuacją zawodową respondenta

Zmienna 4	Zmienna 5 i następne	Wywiad zasadniczy	Wywiad kontrolny
Sytuacja zawodowa respondenta	Konsumpcja przez respondenta kawy	$V = 0,142$	$V = 0,107$
	Konsumpcja przez respondenta (od czasu do czasu) kawy naturalnej, mielonej	$V = 0,136$	$V = 0,099$
	Częstotliwość picia przez respondenta kawy naturalnej, mielonej	$V = 0,147$	$V = 0,179$
	Konsumpcja przez respondenta (od czasu do czasu) kawy rozpuszczalnej	$V = 0,184$	$V = 0,125$
	Częstotliwość picia przez respondenta kawy rozpuszczalnej	$V = 0,185$	$V = 0,231$
	Konsumpcja przez respondenta (od czasu do czasu) kawy cappuccino	$V = 0,147$	$V = 0,138$
	Częstotliwość picia przez respondenta kawy cappuccino	$V = 0,205$	$V = 0,253$
	Dostęp respondenta do internetu	$V = 0,467$	$V = 0,479$
	Częstotliwość korzystania przez respondenta z internetu	$V = 0,243$	$V = 0,231$
	Wyjazdy respondenta co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajowe i zagraniczne)	$V = 0,168$	$V = 0,186$

Jakie wnioski można wysnuć z przedstawionej analizy? Okazuje się, że między zależnościami ustalonymi dla pierwszego i powtórzonego pomiaru nie występują wyraźne różnice, a ich kierunek zawsze jest zgodny. Mam świadomość, iż stwierdzenie to jest bardzo ogólne i powinienem zastosować testy statystyczne na istotność różnic między pierwszym i powtórzonym pomiarem. Posługując się jednak współczynnikiem Cramera (z wcześniej wspomnianych względów), nie mogłem zastosować testu statystycznej istotności różnic w odniesieniu do zależności dla pierwszego i powtórzonego pomiaru).

Z przedstawionych ogólnych analiz wynika, że rzetelność pomiaru wydaje się nie rzutować w istotnym stopniu na merytoryczną wartość ostatecznych ustaleń. Wnosek taki można sformułować przynajmniej w odniesieniu do przedstawionego tu materiału, który jest jednak (jak do tej pory) najobszerniejszy z możliwych.

4. 5. Implikacje praktyczne

Średni ogólny odsetek odpowiedzi zgodnych dla wszystkich powtórzonych pytań (niezależnie od ich przedmiotu) wynosi 83,3. Waha się on od jednak od 15,9 do 100,0. Najwyższy jest w przypadku pytań o cechy (88,4% odpowiedzi zgodnych) i pytań o fakty (87,5% zgodności), następnie w pytaniach o zachowania (75,6%), najniższy zaś w pytaniach o opinie (56,6%). Stwierdzenia te mogą stać się podstawą benchmarkingu dla pytań wyróżnionych ze względu na przedmiot.

Tabela 38. *Benchmarki ze względu na przedmiot pytania. Pytania uszeregowane według: średniego odsetka odpowiedzi zgodnych uzyskanych w powtórny pomiarze dla danego rodzaju pytań (kol. b) oraz średniego odsetka odpowiedzi zgodnych uzyskanych w powtórny pomiarze dla wszystkich pytań (kol. c)*

Przedmiot pytania	% zgodności	Ranga 2 (wg rodzajów)	Ranga 1 (wg typów)
	a	b	c
I. Cechy	88,4		
1. Cechy respondenta, w tym:	89,7	16	7
a) demograficzne (płeć, wiek)	99,8/96,3	1/3	
b) społeczne (wykształcenie, stan cywilny)	80,8/97,3	23/2	
c) waga/wzrost	50,0/85,0	39/19	
d) inne cechy	73,5	30	
2. Cechy innych osób/gospodarstwa domowego, w tym:	85,9	18	8
a) liczba osób	84,0	21	
b) posiadanie/liczba dzieci	92,9	10	
c) inne cechy	76,0	28	
II. Fakty	87,5		
1. Posiadanie świadectwa ukończenia szkoły danego poziomu/ nabycia kwalifikacji	94,6	5	1
2. Sytuacja zawodowa, w tym:	90,6	13	4
a) sytuacja zawodowa respondenta	87,6	17	
b) sytuacja zawodowa innych osób	95,9	4	

cd. tabeli 38

3. Posiadanie dóbr trwałego użytku i ich charakterystyka, w tym:	90,4	15	6
a) samochód	82,9	22	
b) telefon stacjonarny i komórkowy	92,9	11	
c) inne dobra	91,9	12	
4. Charakterystyka sprzętu TV	90,5	14	5
5. Stan zdrowia: przebyte choroby i dolegliwości, w tym:	77,4	26	11
a) przebyte choroby	94,1	6	
b) dolegliwości reumatologiczne	60,6	35	
6. Inne fakty dotyczące respondenta	78,2	25	10
III. Zachowania	75,6		
1. Zachowania „społeczne”	93,7	8	2
2. Zachowania konsumpcyjne, w tym:	70,6	32	14
a) palenie papierosów	93,8	7	
b) konsumpcja piwa	84,3	20	
c) konsumpcja alkoholu	73,9	29	
d) konsumpcja kawy/herbaty	68,5	33	
e) konsumpcja słodczy/przekąsek	65,7	34	
f) dokonywanie zakupów żywnościowych i konsumpcja prod. mlecznych	56,9	36	
3. Zachowania „edukacyjne”	93,3	9	3
4. Zachowania „kulturowe”	76,3	27	12
5. Inne zachowania respondenta	80,1	24	9
IV. Opinie	56,6		
1. Opinie na temat poczucia bezpieczeństwa osobistego i ocena własnej sytuacji	72,3	31	13
2. Opinie na temat pracy zarobkowej	55,8	37	15
3. Opinie na tematy społeczno-polityczne	53,8	38	16
4. Opinie na temat roli kobiet i mężczyzn	44,6	40	17

Ogólny średni odsetek odpowiedzi zgodnych w pytaniach wyróżnionych ze względu na przedmiot (cechy, fakty, zachowania i opinie) jest różny, co wyjdaje się oczywiste. Okazuje się, iż relatywnie najwyższy

odsetek odpowiedzi zgodnych w powtórzonym pomiarze otrzymujemy w przypadku pytań faktualnych i o cechy (blisko 90% zgodności), nieco niższy w przypadku pytań o zachowania (rzędu 75%), najniższy natomiast (około 50%) w pytaniach o opinie i wiedzę. Nie oznacza to jednak, iż są one homogeniczne: jeden z najniższych odsetków zgodności uzyskano w pytaniu o wagę respondenta (50% zgodności), a więc w pytaniu o cechy. Nieco bardziej zaskakujące jest to, że dla poszczególnych bloków pytań wyróżnionych ze względu na przedmiot średni odsetek odpowiedzi zgodnych uzyskanych w powtórnym pomiarze jest bardziej zróżnicowany niż w innych. Na przykład najwyższy odsetek zgodności występuje w przypadku pytań ogólnych o cechy (88,4), ale jest on znacznie niższy niż średni odsetek odpowiedzi zgodnych dla pytania o posiadanie świadectwa ukończenia szkoły danego poziomu/nabycia kwalifikacji danego rodzaju (pytanie o fakty) – 94,5% odpowiedzi zgodnych. Również, uszeregowanie średniego odsetka odpowiedzi zgodnych uzyskanych w powtórnym pomiarze dla wszystkich pytań nie jest zgodne z uszeregowaniem według rodzaju pytań. Oznacza to, iż benchmarkingi nie są wewnętrznie spójne i dlatego należy je odnosić do przedmiotu pytania, rodzaju pytania oraz pytań szczegółowych.

Ogólnie różnice siły związku nie są duże. Tam, gdzie są one znaczące, ich struktura pozostaje nienaruszona. Oznacza to, iż relacje są takie same. Na przykład związek między konsumpcją kawy naturalnej, rozpuszczalnej i cappuccino, a cechami społeczno-demograficznymi, jest taki sam w przypadku pytań z wywiadu zasadniczego i kontrolnego. Wniosek ten jest niezwykle istotny dla oceny jakości badań. Oznacza bowiem, iż wskazane różnice w pomiarze (chodzi o jego rzetelność) nie mają większego wpływu na siłę związku między cechami.

Zakończenie

Czy potrafimy odpowiedzieć na pytanie, czy wyniki zrealizowanego przez nas badania są „dobre”, czy „złe”, a więc prawdziwe, zgodne z rzeczywistością, czy nie?

Z pewnością nie, jeśli nie odwołamy się do analizy elementów lub przebiegu procesu badawczego lub też wyników innych badań czy danych dokumentalnych. Pierwszy sposób oceny nazwałem wewnętrznym, drugi zaś zewnętrznym, ponieważ odwołuje się do danych „zewnętrznych” w stosunku do procesu badawczego, będącego przedmiotem oceny. Oba sposoby oceny jakości danych pociągają za sobą pewne nakłady finansowe oraz czasowe i z pewnością – z punktu widzenia instytutu badawczego – są „dodatkowymi” elementami procesu badawczego. Stąd też podobnej oceny dokonuje się bardzo rzadko. Ponieważ moim celem było wskazanie „prostych narzędzi” realizacji metodologicznego postulat oceny jakości uzyskiwanych danych, zaproponowałem podejście „pośrednie” i „bezpośrednie”. Pierwsze z nich sprowadza się do oceny pracy ankietera (błędów w realizacji próby oraz błędów w realizacji wywiadu), drugie – oceny uzyskanego wyniku. Tym, co je łączy, jest wymóg realizacji badania kontrolnego po zakończeniu każdego badania zasadniczego. Wymóg ten formułowany jest nie tylko w pracach metodologicznych, ale także poszczególnych kodeksach branży badawczej. Zaplanowane w określony sposób badanie kontrolne spełnia ten postulat i może odgrywać rolę zarówno pragmatyczno-szkoleniową, jak i oceniającą.

Przedstawione w pracy analizy potwierdzają tezę o powszechności błędu pomiaru, niezależnie od przedmiotu pytania i jego struktury (por. Neter, Wakesberg 1964: 18–55; Silberstein 1989: 125–142; Sudman, Bradburn 1973: 805–815). Jeśli bowiem rozbieżności odpowiedzi w powtórny pomiarze występują w kilku, a w niektórych przypadkach

kilkunastu różnych badaniach, to nie można raczej mówić o błędach jednostkowych, którymi obarczone są tylko niektóre pomiary. Oznacza to, iż błąd pomiaru, rozumiany jako brak pełnej replikowalności, jest powszechny i występuje zawsze. Co więcej, błąd ten występuje niezależnie od sposobu sformułowania pytania (jego struktury) i jego przedmiotu, aczkolwiek jego rozmiary mogą być różne. Jest on relatywnie największy w przypadku pytań otwartych, w których respondent sam formułuje odpowiedź na postawione pytanie. Wbrew pozorom, błąd ten jest większy w przypadku pytań zamkniętych z „długimi” skalami semantycznymi i pytań „tabelarycznych”, w których kilka, kilkanaście, a niekiedy nawet kilkadziesiąt stwierdzeń respondent ocenia na tej samej skali. Oznacza to, że im prostsze sformułowanie pytania, tym większa rzetelność uzyskanych odpowiedzi. Jest to wniosek trywialny, ale zbyt często zapomina się o tej prostej zależności.

Podobnie brzmi drugi wniosek: rzetelność odpowiedzi respondenta zależy od przedmiotu pytania, tego, czego ono dotyczy. Jest oczywiste, iż w najniższym stopniu replikowalne są odpowiedzi na pytania o opinie. Czy jednak bardziej wiarygodne są odpowiedzi na pytania o fakty, cechy czy zachowania? Nie powinny one zależeć od czynników zawnętrznych, winny być „stabilne”, ale tak nie jest; jedną z najmniej rzetelnych informacji jest odpowiedź na pytanie o wagę respondenta.

Wskazanie zróżnicowania oraz poziomu rzetelności pomiaru w zależności od struktury i przedmiotu pytania było pierwszym i podstawowym celem mojej pracy. Drugim było stworzenie benchmarkingu poziomu rzetelności pomiaru ze względu na strukturę pytania oraz jego przedmiot, a tym samym podstawy ogólnej oceny badań surveyowych, niezależnie od charakteru instytutu realizującego badanie.

W niniejszej pracy skupiłem się na benchmarkach dotyczących uzyskanego wyniku. Dlatego teraz chciałbym krótko omówić benchmarki dotyczące pracy ankietera, a konkretnie „fabrykowania” wywiadów. Czy możemy bowiem mówić o ocenie danych, jeśli nie wiemy, czy i w jakim zakresie zostały one sfabrykowane przez ankieterów (chodzi o liczbę „sufitów” – sfałszowanych wywiadów czy „podmian” – wywiadów zrealizowanych z inną osobą niż wylosowana lub też błędnie dobraną). Wszystkie są to tak zwane zastrzeżenia zasadnicze, które winny być podstawą eliminacji przypadku ze zbioru. Poniżej wskażę benchmarking dla kontrolowanych badań także nieomawianych wcześniej: będzie to 68 badań, nie zaś tylko 34, które analizowałem powyżej.

Tabela 39. *Benchmarki dla oceny pracy ankietera (liczba zastrzeżeń podstawowych) w analizowanych badaniach*

	Badanie zasadnicze	Badanie kontrolne
Suma wywiadów zrealizowanych w analizowanych badaniach	291 222	13 335
Maksymalna (badania trackingowe) i minimalna liczba wywiadów zrealizowanych w badaniu	16 000/200	479/58
Maksymalny i minimalny odsetek wywiadów zrealizowanych w badaniu kontrolnym	–	31/0,8
Średni odsetek wywiadów zrealizowanych w badaniu kontrolnym	–	4,6
Maksymalny i minimalny odsetek zastrzeżeń podstawowych	–	24,5/0,0
Średni odsetek zastrzeżeń podstawowych	–	8,6

Biorąc pod uwagę liczbę analizowanych badań kontrolnych, w których powtórnie dotarto do 13 335 respondentów spośród 291 222 (średnio 4,6% w każdym badaniu), można pokusić się o sformułowanie pewnych ogólnych wniosków dotyczących pracy ankieterów. Średni odsetek zastrzeżeń podstawowych w badaniu wynosi 8,6%. Oznacza to, iż średnio co najmniej 8 wywiadów na 100 zostało sfabrykowanych lub przeprowadzonych z niewłaściwym respondentem. Jest to znacznie powyżej tego, o czym pisała Elisabeth Noelle (1963). Równocześnie aż w 31 badaniach (na 68 analizowanych) odsetek tego typu przypadków przekracza średnią. Należy podkreślić, że niektóre instytuty plasują się wyłącznie powyżej tego progu. Pozostawiam to bez komentarza, aczkolwiek w tych przypadkach, bez wątpienia, różne aspekty faktycznych działań instytutów w relacji do działań optymalnych – standardów (*best practice* lub *golden standards*) wymagają usprawnienia.

Powróćmy jednak do problemu benchmarkingu poziomu rzetelności pomiaru ze względu na strukturę pytania oraz jego przedmiot. Jak wspomniałem, odsetek zgodności odpowiedzi na pytania powtórzone zależy przede wszystkim od struktury pytania. Jak można było oczekiwać jest on najwyższy w przypadku pytań dychotomicznych (powyżej 90% odpowiedzi zgodnych), następnie właściwych pytań dopełnienia

(88%) i – co stanowi zaskoczenie – pytań prekategoryzowanych (87%). Dalej, ze względu na odsetek odpowiedzi zgodnych plasują się pytania kafeteryjne jednoodpowiedziowe (83% odpowiedzi zgodnych), kafeteryjne wieloodpowiedziowe (77%), pytania otwarte (68%), co zaskakuje, oraz pytania zamknięte kafeteryjne ze skalą (52%). Najniższy odsetek zgodności odpowiedzi na pytania powtórzone uzyskano w przypadku pytań tabelarycznych jednoodpowiedziowych zawierających skale (jedynie 27%).

Jak wspominałem, powyższe wyniki analiz winny stanowić dla badaczy tworzących kwestionariusze wskazówkę, jakie pytania, biorąc pod uwagę ich strukturę, należy preferować. Bowiem, jak pisał Dillman (1978), pytając o to samo, można to robić na różne sposoby: pytania możemy formułować w różny sposób.

Poziom rzetelności pomiaru zależy również od przedmiotu pytania. Według tego kryterium wyróżniłem cztery kategorie pytań: o cechy, fakty, zachowania oraz opinie (pomiąłem tutaj podział na pytania merytoryczne i metryczkowe, ponieważ w jednym badaniu pytanie o „palenie papierosów” jest pytaniem metryczkowym, w innym zaś może być „merytoryczne”). Także w tym przypadku podjąłem próbę stworzenia benchmarków.

Jak można było oczekiwać, najwyższy odsetek zgodności w powtórzonym pomiarze uzyskano w pytaniach o cechy (powyżej 88), pytaniach o fakty (88), zachowania (76), najniższy zaś w pytaniach o opinie (57). Owe ogólne odsetki są zróżnicowane w zależności od „szczegółowego” przedmiotu. Nie będę tutaj powtarzał wniosków przedstawionych wcześniej, musimy wszakże pamiętać, iż odsetek zgodności odpowiedzi na powtórzone pytania w ramach poszczególnych rodzajów pytań jest zróżnicowany.

W przypadku pytań o cechy zależy on od tego, czy dotyczą one respondenta (jego cech społecznych i demograficznych: płci, wieku, wykształcenia, stanu cywilnego itp., ale też fizycznych: wagi/wzrostu), czy gospodarstwa domowego (liczby osób; posiadania/liczby dzieci). W przypadku pytań o fakty zależy on od tego, czy pytania dotyczą sytuacji respondenta (posiadania świadectwa ukończenia szkoły, sytuacji zawodowej), stanu jego zdrowia (przebytych chorób i dolegliwości) czy też posiadania dóbr trwałego użytku (samochód, telefon stacjonarny, komórkowy, charakterystyka sprzętu RTV itd.). Z kolei w przypadku pytań o zachowania odsetek zgodności zależy od rodzaju zachowań, o które

pytamy, a więc tego, czy są to pytania dotyczące zachowań „społecznych”, konsumpcji (papierosy, piwo, alkohol, kawa, herbata, przekąski itp.), zachowań „edukacyjnych” czy „kulturowych”. I wreszcie w przypadku pytań o opinie zależy on od ich przedmiotu, a więc przykładowo, czy dotyczą one opinii na temat poczucia bezpieczeństwa osobistego i oceny własnej sytuacji, pracy zarobkowej czy oceny sytuacji społeczno-politycznej.

Mówiąc o benchmarkach dotyczących realizacji wywiadu oraz rzetelności pomiaru, należy wspomnieć o ich zmienności w czasie. Dotyczy to przede wszystkim benchmarków odnoszących się do pytań o opinie. Pytania o fakty, cechy oraz zachowania wydają się relatywnie mniej podatne na zmiany w czasie (Good, Irwin, Jackson 1998). Mówiąc o „zmienności w czasie” benchmarków, należy pamiętać jeszcze o jednym problemie. Otóż w przypadku niektórych pytań wyróżnionych ze względu na przedmiot, tracą one sens. Jest tak na przykład w przypadku informacji na temat korzystania przez respondenta z automatów telefonicznych w miejscach publicznych. Za kilka lat żadnego zlecniodawcy problem ten nie będzie interesował, ponieważ zjawisko będzie miało charakter wyłącznie historyczny.

Mając świadomość wszystkich ograniczeń związanych ze sposobem tworzenia i podobnych benchmarków, warto próbować je konstruować. Dzięki nim bez dodatkowych nakładów można ocenić zebrane dane. Gdy zaś tworzenie benchmarków uznamy za celowe, trzeba pamiętać o ich „aktualizowaniu”.

Bibliografia*

- Ajdukiewicz, Kazimierz. 1975. *Logika pragmatyczna*, Warszawa: PWN.
- American Statistical Association. 2003. *Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects*, <http://www.amstat.org/sections/SRMS/falsification.pdf>, dostęp z dnia 24. 11. 2010.
- Antoine, Jacques. 1990. *Le sondage outil du marketing*, Paris: Bordas.
- Arleck, Pamela L., Robert B. Settle. 1985. *The Survey Research Handbook*, Homewood, Ill: Irwin.
- Auriat, Nadia. 1993. „My Wife Knows Best”. *A Comparison of Event Dating Accuracy Between the Wife, the Husband, the Couple, and the Belgium Population Register*, „Public Opinion Quarterly”, t. 57, nr 2, s. 165–190.
- Auriat, Nadia. 1996. *Les Defaillances de la Memoire Humaine. Aspects cognitifs des enquetes retrospectives*, Paris: INED, PUF.
- Babbie, Earl. 2004. *Badania społeczne w praktyce*, Warszawa: Wydawnictwo Naukowe PWN.
- Baddeley, Alan. 2004. *The Limitations of Human Memory: Implications for the Design of Retrospective Surveys*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Bailar, Barbara A. 1968. *Recent Research in Reinterview Procedure*, „Journal of American Statistical Association” t. 63, nr 321, s. 41–63.
- Białecki, Ireneusz, Paweł B. Sztabiński. 1997. *Eurobarometr 38.1.*, „Zagadnienia Naukoznawstwa” 4.

* Bibliografia zawiera pozycje przywołane w tekście, ale także inne, wykorzystane w trakcie pisania pracy, które odnoszą się bezpośrednio do poruszanej problematyki.

- Biemer, Paul P., Gösta Forsman. 1992. *On the Quality of Reinterview Data with Application to the Current Population Survey*, „Journal of the American Statistical Association”, t. 87, nr 420, s. 915–923.
- Biemer, Paul P., Lars E. Lyberg. 2003. *Introduction to Survey Quality*, New Jersey: John Wiley & Sons Inc.
- Biemer, Paul P., Rachel Caspar. 1994. *Continuous Quality Improvement for Survey Operations: Some General Principles and Applications*, „Journal of Official Statistics”, t. 10, nr 3, s. 307–326.
- Blair, Ed, Seymour Sudman, Norman M. Bradburn, Carol Stocking. 2004. *How to ASK Questions About Drinking and Sex: Response Effects in Measuring Consumer Behavior*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Blair, Johnny, Seymour Sudman. 1996. *The Survey Reinterview: Respondent Perceptions and Response Strategies*, „Journal of Official Statistics”, t. 12, nr 4, s. 421–426.
- Boeswillwald, Elisabeth. 1991. *L’expérience du CESP en matière de qualité des mesures d’audience. W: La qualité de l’information dans les enquêtes. Journées d’Etudes*, 13–15 czerwca, Paris: Télécom.
- Bohrnstedt, George W. 1970. *Reliability and Validity Assessment in Attitude Measurement*, w: G. Summers (red.), *Attitude Measurement*. Chicago: Rand McNally.
- Bon, Frederic, Michel-Antoine Burnier, Nonna Mayer. 1994. *Les Sondages. Peuvent-ils se tromper?* Paris: Calmann Levy.
- Borucki, Andrzej. 1978. *Kontrola pracy ankietowanych i weryfikacja danych kwestionariuszowych jako elementy analizy procesu badawczego*, „Przegląd Socjologiczny”, t. XXX, s. 175–196.
- Bowling, Ann. 2005. *Mode of questionnaire administration can have serious effects on data quality*, „Journal of Public Health”, t. 27, s. 281–291.
- Bradburn, Norman M., Seymour Sudman, Ed Blair, Carol Stocking. 2004. *Question Threat and Response Bias*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Bradburn, Norman M., Seymour Sudman. 1979. *Improving interview method and questionnaire design*, San Francisco: Jossey-Bass Publishers.
- Bradburn, Norman M., Seymour Sudman. 1988. *Polls and Surveys. Understanding What They Tell Us*, San Francisco: Jossey-Bass Publishers.

- Brennan, Mike, Joanna Chan, Dean Hini, Don Esslemont. 1996. *Improving the Accuracy of Recall Data: A Test of Two Procedures*, „Marketing Bulletin”, nr 7, s. 20–29.
- Brinberg David, Joseph E. McGrath. 1985. *Validity and Research Process*, Newbury Park: Sage Publications.
- Bryman, Alan. 1989. *Research Methods and Organization Studies*, London: Unwin Hyman.
- Bryman, Alan. 1995. *Quantity and Quality in Social Research*, London: Routledge.
- Brzeziński, Jerzy. 1978. *Metodologiczne i psychologiczne wyznaczniki procesu badawczego w psychologii*, Poznań: UAM.
- Brzeziński, Jerzy. 1984. *Elementy metodologii badań psychologicznych*, Warszawa: PWN.
- Bulmer, Martin. 2004. *Editor's Introduction: Questionnaires – An Overview*, w: M. Bulmer (red.), *Questionnaires*, London: Sage.
- Cahalan, Don. 1968-1969. *Correlates of Respondent Accuracy in the Denver Validity Survey*, „Public Opinion Quarterly”, s. 607–621.
- Campbel, Donald T., Donald W. Fiske. 1959. *Convergent and Discriminant Validation by the Multitrait-multimethod Matrix*, „Psychological Bulletin”, nr 56.
- Cannel, Charles F., Floyd J. Fowler, Graham Kalton, Lois Oksenberg, Katherine Bischooping. 2004. *New Quantitative Technique for Pre-testing Survey Questions*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Cannell, Charles F. 1977. *A Summary of Research Studies of Interviewing Methodology*, „National Center for Health Statistic”, Series 2.
- Cannell, Charles F. 1985. *Experiments in the Improvement of Response Accuracy*, w: T. W. Beed, R. J. Stimson (red.). *Survey Interviewing. Theory and Techniques*, Sydney: George Allen & Unwin.
- Cannell, Charles F., Sally A. Lawson, Doris L. Hausser. 1975. *A Technique for Evaluating Interviewer Performance*, Ann Arbor, Michigan: The University of Michigan.
- Carmines, Edward G., Richard A. Zeller. 1988. *Reliability and Validity Assessment*, Beverly Hills: Sage Publications.
- Carton, Ann. 1995. *Towards an Effective Evaluation of Interviewers: Working Within the Context of a Permanent Network of Interviewers*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1-4 kwietnia, Bristol: American Statistical Association.

- Centrum im. Adama Smitha. 2004. *Ocena zgodności wyników sondaży z rzeczywistymi wynikami wyborów*, www.pgb.org.pl/DOKUMENTY/SMITH_dokladnosc_sondazy, dostęp z dnia 20.12.2005.
- Centrum im. Adama Smitha. 2005. *Ocena zgodności wyników sondaży z rzeczywistymi wynikami wyborów*, www.pgb.org.pl/DOKUMENTY/SMITH_dokladnosc_sondazy, dostęp z dnia 20.12.2005.
- Chavenon, Hugues, Françoise Dupont, Helene Haering, Muriel Libs (red.). 1999. *Le Mediagraphe 98/99. Audit de la mesure d'audience des Medias*, Paris: CESP.
- Chavenon, Hugues, Joseph Khoury. 2002. *Quality Control of TV peoplemeter panels*, Konferencja ESOMAR, WAM: Week of Audience Measurement – Television, Cannes, 9–14 czerwca.
- Cichomski, Bogdan, Tomasz Jerzyński, Marcin Zieliński. 2008. *Polskie Generalne Sondáže Społeczne: struktura skumulowanych wyników badań 1992–2008*, Warszawa: Instytut Studiów Społecznych, Uniwersytet Warszawski.
- Clark, Jonh P. i Larry. L. Tifft. 1966. *Polygraph and Interview Validation of Self-Reported Deviant Behaviour*, „American Sociological Review”, t. 31.
- Collins, Martin, Wendy Sykes. 1995. *Survey Quality – the User's View*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1–4 kwietnia, Bristol: American Statistical Association.
- Converse, Jean M. 2004. *Strong Arguments and Weak Evidence: The Open/Closed Questioning Controversy of the 1940s*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Converse, Jean M., Stanley Presser. 1981. *Survey Questions. Hand-crafting the Standardized Questionnaire*, New York: Academic Press.
- Cook, Thomas D. 2000. *Toward a Practical Theory of External Validity*, w: L. Bickman (red.), *Validity and Social Experimentation: Donald Campbell's Legacy*, Thousand Oaks: Sage Publications.
- Crespi, Leo P. 2003. *The Cheater Problem in Polling*, w: N. Fielding (red.), *Interviewing*, t. III, London: Sage Publication.
- Cronbach, Lee J., Paul E. Meehl. 1955. *Construct validity in psychological tests*, „Psychological Bulletin”, t. 52, s. 281–302.
- Czaja Ronald F., Blair Johnny. 2005. *Designing Surveys: A Guide to Decisions and Procedures*, Thousand Oaks: Sage Publications.

- Daniłowicz, Paweł, Franciszek Sztabiński. 1992. *Nowe spojrzenie na ankietę pocztową. Jak uzyskano 70% zwrotów*, w: Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. IX. Wrocław: Zakład Narodowy im. Ossolińskich.
- Daniłowicz, Paweł, Paweł B. Sztabiński. 1986. *Zmienne osobowe w badaniach socjologicznych. Analizy weryfikacyjne*, Wrocław: Ossolineum.
- Daniłowicz, Paweł. 1981. *Źródła – materiał – dane*, w: K. Lutyńska (red.), *Szkice metodologiczne*, Łódź.
- Dąbała, Wiesława. 2005. *Próby CBOS oraz szacowanie na ich podstawie parametrów populacji w badaniach sondażowych po roku 2000*, Konferencja: Badania sondażowe po roku 2000. Polska demokracja sondażowa, Łódź, grudzień.
- de Leeuw Edith, D. 2004. *New technologies in data collection, questionnaire design and quality*, <http://www.eustat.es/prodserv/datos/sem44.pdf>.
- de Leeuw Edith, D. 2009. *Mixed Mode Surveys – 6th ESS Training Course*, 30 kwietnia–1 maja, Ljubljana.
- de Leeuw Edith, D., Martin Collins. 1997. *Data Collection Methods and Survey Quality: An Overview*, w: L. Lyberg, P. Biemer, M. Collins, E. D. de Leeuw, C. Dippo, N. Schwarz i D. Trewin (red.), *Survey Measurement and Process Quality*, New York: John Wiley & Sons Inc.
- de Leeuw Edith, D., Wim de Heer. 2002. *Trends in Household Survey Nonresponse: A Longitudinal and International Comparison*, w: R. M. Groves, D. A. Dillman, J. L. Eltinge, R. J. A. Little (red.), *Survey Nonresponse*, New York: Wiley.
- de Leeuw Edith, D., Don A. Dillman, Joop J. Hox. 2008. *The Cornerstones of Survey Research*, w: E. D. de Leeuw, J. J. Hox, D. A. Dillman (red.), *International Handbook of Survey Methodology*, New York, London: Lawrence Erlbaum Associates.
- de Leeuw, Edith D., Joop J. Hox, Ger Snijkers. 1995. *The effect of computer-assisted interviewing on data quality. A review*, „Journal of the Market Research Society”, t. 37, nr 4.
- de Leeuw, Edith D., William Nicholls II. 1996. *Technological Innovations in Data Collection: Acceptance, Data Quality and Costs*, „Sociological Research Online”, t. 1, nr 4, <http://www.socresonline.org.uk/socresonline1/4/leeuw.html>.

- de Leeuw, Edith, Johannes van der Zouwen. 1991. *Data Quality and Mode of Data Collection*, w: *La qualité de l'informations dans les enquêtes, Journées d'Etudes*, 13–15 czerwca, Paris: Télécom.
- de Leeuw, Edith, Joop J. Hox, Don A. Dillman. 2008. *Mixed Mode: When and Why?* w: E. D. de Leeuw, J. J. Hox, D. A. Dillman (red.), *International Handbook of Survey Methodology*, New York: Lawrence Erlbaum Associates.
- de Maio, Teresa. 2009. *Attempting to Evaluate Cognitive Testing Methodology*, Conference of ESRA, lipiec, Warsaw.
- Dijkstra, Wil, Johannes van der Zouwen. 1988. *Styles of Interviewing and the Social Context of the Survey-Interview*, w: H.-J. Hippler, N. Schwarz, S. Sudman (red.), *Social Information Processing and Survey Methodology*, New York: Springer-Verlag.
- Dillman Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*, New York: John Wiley & Sons Inc.
- Dillman Don A., Jolene D. Smith, Leah M. Christian. 2009. *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, New York: John Wiley & Sons Inc.
- Dillman, Don A. 1978. *Mail and Telephone Surveys. The Total Design Method*, New York: J. Wiley & Sons.
- Domański, Henryk, Aleksandra Dukaczewska. 1996. *Stabilność odpowiedzi w badaniach socjologicznych*, „ASK. Research&Methods”, nr 1, s. 71–90.
- Domański, Henryk, Paweł B. Sztabiński, Franciszek Sztabiński. 1995. *Badania społeczne w latach dziewięćdziesiątych. Kontynuacje i zmiany*, „ASK”, nr 1, s. 7–24.
- Domański, Henryk, Zbigniew Sawiński, Kazimierz M. Słomczyński (red). 2007. *Nowa klasyfikacja i skale zawodów. Socjologiczne wskaźniki pozycji społecznej w Polsce*, Warszawa: Wydawnictwo IFiS PAN.
- Domański, Henryk. 2000. *Wpływ kategorii „trudno powiedzieć” na wyniki analiz*, „ASK”, nr 9.
- Domański, Henryk. 2004. *Wybór między alternatywnymi wersjami pytań w kwestionariuszu wywiadu*, „Przegląd Socjologiczny”, nr LIII/1, s. 97–116.
- Duncan, Otis D. 1975. *Introduction to Structural Equation Models*, New York: Academic Press.
- Dyjas-Pokorska, Anna, Małgorzata Poławska. 2004. *Trakings jako specyficzny rodzaj pomiaru w badaniach marketingowych*, w: P. B. Sztabiński (red.), *Marketingowe badania społeczne*, Warszawa: Wydawnictwo IFiS PAN.

- biński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.
- European Social Survey, Round 5. *Specification for participating countries*. 2011. www.europeansocialsurvey.org/fieldwork_documentation/project_specification, dostęp z dnia 26.05.2011.
- European Social Survey. *Fortnightly progress reports from survey organisations*. 2002. www.europeansocialsurvey.org/fieldwork_documentation/progress_checks, dostęp z dnia 26.05.2011.
- European Social Survey: Field Procedures in the ESS R3. *Enhancing Response Rate*. 2006. www.europeansocialsurvey.org/fieldwork_documentation/response_rate, dostęp z dnia 26.05.2011.
- Europejski Sondaż Społeczny Runda 4. 2008. <http://ess.nsd.uib.no/ess/round4/>, dostęp 05. 08. 2011.
- Eva, Gillian, Widdop Sally. 2007. *Mixed Mode Data Collection in Europe, 2nd European Survey Research Association Conference*, Prague.
- Fink, Arlene. 2003. *The Survey Handbook*, Thousand Oaks: Sage.
- Fink, Raymond. 2004. *The Retrospective Question*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Fishbein, Martin. 1967. *A Consideration of Beliefs, and Their Role in Attitude Measurement*, w: Fishbein Martin (red.), *Readings in Attitude Theory and Measurement*, New York: J. Wiley & Sons.
- Foddy, William. 2004. *The In-Depth Testing of Survey Questions: A Critical Appraisal of Methods*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Forsman, Gösta, Irwin Schreiner. 1991. *The Design and Analysis of Re-interview: An Overview*, w: P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, S. Sudman (red.), *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Fowler, Floyd J., Jr Thomas W. Mangione. 1990. *Standardized Survey Interviewing. Minimizing Interviewer-Related Error*, Newbury Park: Sage Publications.
- Frey, James. H., Sabine Mertens Oishi. 1995. *How to Conduct Interview by Telephone and in Person*. Thousand Oaks, London, New Delhi: Sage Publications.
- Gaskell, George D., Daniel B. Wright, Colm A. O’Muircheartaigh. 2000. *Telescoping of Landmark Events. Implications for Survey Research*, „Public Opinion Quarterly”, t. 64, s. 77–89.

- Giedymin, Jerzy. 1964. *Problemy, założenia, rozstrzygnięcia: studia nad logicznymi podstawami nauk społecznych*, Poznań: PWN.
- Good, Darrel, L., Scott H. Irwin, Thomas E. Jackson. 1998. *Development of a Market Benchmark Price for AgMAS Performance Evaluations. Research Reports*. <http://www.farmdoc.uiuc.edu/agmas/reports/9802>, dostęp z dnia 24. 11. 2010.
- Goode William J. i Paul K. Hatt. 1952. *Methods in Social Research*, New York: Mc Graw-Hill.
- Goode, William J., Paul K. Hatt. 1952a. *Scaling Techniques: The Basic Problem*, w: W. J. Goode, P. K. Hatt (red.), *Methods in Social Research*, New York: Mc Graw-Hill.
- Goodyear Mary. 1998. *Qualitative Research*, w: C. McDonald, Ph. Vangelder (red.), *ESOMAR Handbook of Market and Opinion Research*, Amsterdam: ESOMAR.
- Gostkowski, Zygmunt. 1961. *Z zagadnień socjologii wywiadu*, „Przeegląd Socjologiczny”, t. XV/2.
- Gostkowski, Zygmunt. 1965. *Badania adekwatności technik badawczych w krajach na drodze rozwoju*, „Przeegląd Socjologiczny”, t. IXX/2.
- Gostkowski, Zygmunt. 1966. *O założeniach i potrzebie empirycznych badań nad technikami i procedurami badawczymi w socjologii*, w: Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. I, Wrocław: Zakład Narodowy im. Ossolińskich.
- Gottschalk, Louis. 1945. *The Historian and the Historical Document*, w: Gottschalk Louis, Kluckohn Clyde, Angell Robert, *The Use of Personal Documents in History, Anthropology and Sociology*, Social Science Research Council, „Bulletin”, nr 53, s. 43–44.
- Grawitz, Madeleine. 1972. *Méthodes des sciences sociales*, Paris: Dalloz.
- Gremy, Jean-Paul. 1995. *Wpływ sposobu formułowania pytań kwestionariuszowych na uzyskiwane odpowiedzi*, „ASK”, nr 2.
- Groves R.M., Couper M.P. 1998. *Nonresponse in Household Interview Surveys*, New York: John Wiley.
- Groves R.M., McGonagle K. 2001. *A theory-guided interviewer training protocol regarding survey participation*, „Journal of Official Statistics”, t. 17, s. 249–265.
- Groves Robert N., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, Roger Tourangeau. 2004. *Survey Methodology*, New York: John Wiley & Sons.

- Groves, Robert M. 1989. *Survey Errors and Survey Costs*, New York: John Wiley & Sons.
- Groves, Robert M. 1991. Measurement Error Across the Disciplines, w: P. P. Biemer, R. M. Goves, L. E. Lyberg, N. A. Mathiowetz, S. Sudman (red.), *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Grzeszkiewicz-Radulska, Katarzyna. 2009. *Respondenci niedostępni w badaniach sondażowych*, w: *Analizy i próby technik badawczych w socjologii*, t. XII, Łódź: Łódzkie Towarzystwo Naukowe, Instytut Socjologii UŁ. Wydawnictwo Uniwersytetu Łódzkiego.
- Harkness, Janet A. 2003. *Questionnaire Translation*, w: J. A. Harkness, F. van de Vijver, P. Ph. Mohler (red.), *Cross-Cultural Survey Methods*, New York: John Wiley and Sons.
- Harkness, Janet A. 2007. *Rounds 3 ESS Translation Strategies and Procedures*. www.europeansocialsurvey.org/methodological_research/translation_strategies/, dostęp z dnia 25. 03. 2007.
- Harvey, Bill, Tony Jarvis, Russ Booth. 2002. *Better Television Audience Measurement through Cable and Satellite Set Top Boxes*, [Digital set top boxes (DSTBx)], Konferencja ESOMAR, WAM: Week of Audience Measurement - Print, France: Cannes, 9–14 czerwca.
- Heeringa Steven G., Kristen H. Alcer. 1995. *Survey Measurement Error in Population Case/Control Studies: Retrospective Reports of Risk Factor Exposures*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1–4 kwietnia, Bristol: American Statistical Association.
- Heise, David R. 1969. *Separating Stability and reliability in test-retest correlations*, „American Sociological Review”, nr 34, s. 93–110.
- Heise, David R. 2000. *Social measurement, classification and scaling*, <http://www.indiana.edu/~socy/papers/Measurement.htm> (dostęp 27. 03. 2008).
- Heisenberg, Werner .1975. *The great tradition: end of an epoch?*, „Encounter”, t. 44, nr 3, s. 52–58.
- Heller, Jean-Luc, Nicole Roth. 1994. *CAPI et l'Enquete Emploi en France: le choix d'une methode et ses enseignements*, INSEE (tekst powielony).
- Hess, Jennifer, Eleanor Singer, John Bushery. 2004. *Predicting Test-Retest Reliability From Behavior Coding*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.

- Hippler, Hans-J, Gabriele Hippler. 2004. *Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the Face" Technique*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Hippler, Hans-J., Norbert Schwarz. 1988. *Response Effects in Surveys*, w: H.-J. Hippler, N. Schwarz, S. Sudman (red.), *Social Information Processing and Survey Methodology*, New York: Springer-Verlag.
- Hornowska Elżbieta, Władysław J. Paluchowski. 2004. *Kulturowa adaptacja testów psychologicznych*, w: J. Brzeziński (red.), *Metodologia badań psychologicznych*, Warszawa: Wydawnictwo Naukowe PWN.
- Hyman, Herbert H. 1954. *Interviewing in Social Research*, Chicago: University of Chicago Press.
- Improving Representativeness and Response*. 2006. [www.europeansocialsurvey.org/Methodological Research/Improving Response](http://www.europeansocialsurvey.org/Methodological%20Research/Improving%20Response), dostęp z dnia 24. 11. 2010
- Jahoda Marie, Morton Deutsch, Stuart W. Cook. 1951. *Research Methods in Social Relations*, New York: The Dryden Press.
- Jensen Michael.C., Meckling William H. 1976. *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, „Journal of Financial Economics”, t. 3, 4.
- Jensen, C. 1994. *Self-interest, Altruism, Incentives, & Agency Theory*, „Journal of Applied Corporate Finance”, t. VII, nr 2.
- Jones Russel A. 2000. *Méthodes de Recherche en Sciences Sociales*, Paris: PUF.
- Kahn, Robrt L., Cannell Charles F. 1957. *The Dynamics of Interviewing*, New York: John Wiley & Sons.
- Kasprzyk, Daniel. 2001. *Measuring and Reporting Data Quality in Federal Data Collection Programs*, w: *Measuring and Reporting Sources of Error in Surveys*, http://www.fcs.gov/01papers/SPWP31_final.pdf, dostęp z dnia 20. 05. 2002.
- Kish, Leslie. 1965. *Survey Sampling*, New York: John Wiley & Sons.
- Kistelski, Krzysztof. 1981. *Problem błędów ankieterskich w wywiadzie kwestionariuszowym w socjologii. Próba klasyfikacji*, w: Z. Gostkowski (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 7, Warszawa: IFiS PAN.
- Kistelski, Krzysztof. 1985. *Rola ankietera w wywiadzie kwestionariuszowym w socjologii*, w: J. Lutyński (red.), „Acta Universitatis Lodziensis – Folia Sociologica”, z. 11.

- Koniarek Jerzy. 1975. *Weryfikacja zewnętrzna informacji uzyskanych w wywiadzie kwestionariuszowym*, w: Zygmunt Gostkowski i Jan Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Ossolineum.
- Krak, Joanna. 1978. *Style przeprowadzania wywiadu i konsekwencje ich zastosowania*, „Przegląd Socjologiczny”, t. XXX.
- Krosnick, Jon A. 1991. *Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys*, „Applied Cognitive Psychology”, nr 5.
- Krosnick, Jon A., Leandre R. Fabrigar. 1997. *Designing Rating Scales for Effective Measurement in Surveys*, w: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, D. Trewin (red.), *Survey Measurement and Process Quality*, New York: John Wiley & Sons Inc.
- Kubiak, Anna, Ilona Przybyłowska. 1999. *Metodologia pragmatyczna Jana Lutyńskiego – propozycje klasyfikacji pytań kwestionariuszowych*, „Przegląd Socjologiczny”, t. XLVIII, s. 102–112.
- Kulej, Małgorzata. 2004. „Praca ankietera w badaniach kwestionariuszowych w ujęciu teorii agencji”, praca licencjacka napisana pod kierunkiem dr. F. Sztabińskiego, Warszawa: Collegium Civitas.
- Lamers, Leida M. 1995. *Validating Survey Data on Medical Consumption by Comparison with Administrative Data*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1-4 kwietnia, Bristol: American Statistical Association.
- Laniel, Normand, Pierre Lavallée. 1991. *L'assurance de la qualité dans les enquêtes auprès des entreprises. Un survol*, w: *La qualité de l'information dans les enquêtes. Journées d'Etudes*, 13–15 czerwca, Paris: Télécom.
- LaPiere, Richard T. 2004. *Attitudes vs Actions*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Laurent, Andre, Charles F. Cannell, Kent. M. Marquis. 1972. *Reporting Health Events in Household Interviews: Effects of an Extensive Questionnaire and a Diary Procedure*, *Vital and Health Statistics. Data Evaluation and methods Research*, Seria 2, nr 49, Rockville: National Center for Health Statistics.
- Lavrakas, Paul J. 2008. *Encyclopedia of Survey Research Methods*, Thousand Oaks: Sage Publications, Inc.
- Lee, Raymond M. 2004. *Asking Sensitive Questions on Survey*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.

- Lejeune, Michel, Denis Bied-Charreton. 1991. *Comparaison des modes postal et en face à face dans une enquête lourde*, w: *La qualité de l'informations dans les enquêtes. Journées d'Etudes*, 13–15 czerwca, Paris: Télécom.
- Lininger, P.H. 1975. *The Sample Survey: Theory and Practice*, New York.
- Litwak, Eugene. 2004. *A Classification of Biased Questions*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Litwin, Mark S. 1995. *How to Measure Survey Reliability and Validity*, Thousand Oaks: Sage Publications.
- Locander, William, Seymour Sudman, Norman M. Bradburn. 1976. *An Investigation of Interview Method, Threat and Response Distortion*, „Journal of the American Statistical Association”, t. 71.
- Locander, William, Seymour Sudman, Norman M. Bradburn. 1976. *An Investigation of Interview Method, Threat and Response Distortion*, „Journal of the American Statistical Association”, t. 71.
- Loftus, Elizabeth F. 2004. *Leading Questions and the Eyewitness Report*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Loftus, Elizabeth F., Wesley Marburger. 2004. *Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving The Accuracy of Retrospective Reports With Landmark Events*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Loosveldt, Geert. 1995. *Interviewer – Respondent Interaction Analysis as a Diagnostic and Validating Instrument*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1-4 kwiecień, Bristol: American Statistical Association.
- Lozano, Luis M., Eduardo Garcia-Cueto, Jose Muniz. 2008. *Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales*, *Methodology*, „European Journal of Research Methods for the Behavioral and Social Sciences”, t. 4 (2), s. 73–79.
- Lutyńska, Krystyna, Neonila Szeszenia. 1972. *Wstępne wyniki zewnętrznej weryfikacji odpowiedzi respondentów w wywiadzie na temat zachowań zdrowotnych*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. IV, Wrocław: Ossolineum.
- Lutyńska, Krystyna. 1968. *Ankieta do ankietera jako źródło wiedzy o wywiadzie kwestionariuszowym*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. II, Wrocław: Ossolineum.

- Lutyńska, Krystyna. 1970. *Osoby trzecie w wywiadzie socjologicznym i ich wpływ na odpowiedzi respondentów*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. III, Wrocław: Ossolineum.
- Lutyńska, Krystyna. 1975. *Pilotaż pogłębiony. Koncepcja, realizacja i analiza materiałów pilotażowych*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Ossolineum.
- Lutyńska, Krystyna. 1975b. *Weryfikacja danych uzyskiwanych z odpowiedzi respondentów na pytania o wysokość zarobków*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Ossolineum.
- Lutyńska, Krystyna. 1978. *Ankieterzy i badacze. Z badań nad wpływem ankieterskim*, „Przegląd Socjologiczny”, t. XXX.
- Lutyńska, Krystyna. 1984. *Wywiad kwestionariuszowy. Przygotowanie i sprawdzenie narzędzia badawczego*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyńska, Krystyna. 1985. *Z problematyki współczesnych badań kwestionariuszowych*, w: J. Kulpińska (red.), *Społeczeństwo a socjologia*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyńska, Krystyna. 1993. *Surveye w Polsce. Spojrzenie socjologiczno-antropologiczne*, Warszawa: Wydawnictwo IFiS PAN.
- Lutyński, Jan 1975b. *Cele, Przebieg i wstępne wyniki badań weryfikacyjnych we Wrocławku w 1973 r.*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyński, Jan. 1968. *Ankieta i jej rodzaje na tle podziału technik otrzymywania materiałów*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. II, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyński, Jan. 1972. *Analiza procesu otrzymywania informacji w badaniach z zastosowaniem wywiadu kwestionariuszowego*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. IV, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyński, Jan. 1975. *Analizy weryfikacyjne w badaniach z zastosowaniem wywiadu kwestionariuszowego, ich rodzaje i możliwości*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Ossolineum.

- Lutyński, Jan. 1978. *Koncepcje pytania kwestionariuszowego, ich zastosowanie w badaniach, konsekwencje i możliwości*, „Przegląd Socjologiczny”, t. XXX.
- Lutyński, Jan. 1979. *Pytanie jako narzędzie w surveyowych badaniach socjologicznych*, „Studia Socjologiczne”, nr 2.
- Lutyński, Jan. 1994. *Metody badań społecznych. Wybrane zagadnienia*, Łódź: ŁTN.
- Lyberg, Lars E., Paul P. Biemer. 2008. *Quality Assurance and Quality Control in Surveys*, w: E. D. de Leeuw, J. J. Hox, D. A. Dillman (red.), *International Handbook of Survey Methodology*, New York: Lawrence Erlbaum Associates.
- Madge, John. 1965. *The Tools of Social Science*, Garden City, New York: Anchor Books Doubleday & Company. Inc.
- Maison, Dominika. 2001. *Zogniskowane wywiady grupowe. Jakościowa metoda badań marketingowych*, Warszawa: Wydawnictwo Naukowe PWN.
- Mangione Thomas W. 1995. *Mail Surveys. Improving the Quality*, Thousand Oaks: Sage.
- Mann, Peter H. 1968. *Methods of Sociological Enquiry*, Oxford: Blackwell.
- Marquis, Kent H., James Marshall, Stuard Oskamp. 2004. *Testimony Validity as a Function of Question Form, Atmosphere, and Item Difficulty*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Martin, Claudia J. 2004. *Monitoring Maternity Services by Postal Questionnaire: Congruity between Mothers' Reports and their Obstetric Records*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Marx, Jean-Louis, Elisabeth de Langhe. 1997. *The CAPI double screen system: a test for magazine press readership survey in France*, Vancouver: Worldwide Readership Research Symposium 8.
- Marx, Jean-Louis. 1998. *Testing the double CAPI for the magazine readership survey*, Audience Etude sur la Presse Magazine (AEPM), EMRO, tekst powielony.
- Marx, Jean-Louis. 2000. *The double screen CAPI readership survey. How it improves data*. Audience Etude sur la Presse Magazine (AEPM), tekst powielony.
- Mattlin, Jay. 2002. *Electronic Measurement of Magazine Reading, ESO-MAR/ARF Conference: Week of Audience Measurement, Cannes, 9–14 czerwca*.

- Mayntz, Renate, Kurt Holm, Peter Hubner. 1985. *Wprowadzenie do metod socjologii empirycznej*, Warszawa: PWN.
- Menner, Peter. 2004. *Retrospective Data in Survey Research*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Menon, Geeta, Barbara Bickart, Seymour Sudmn, Johnny Blair. 1995. *How Well Do You Know Your Partner? Strategies for Formulating Proxy-Reports and Their Effets on Convergence to Self-Reports*, „Journal of Marketing Research”, t. XXXII, s. 75–84.
- Menon, Geeta. 1993. *The Effects of Accessibility of Information in Memory on Judgements of Brhavioral Frequencies*, „Journal of Consumer Research”, t. 20, s. 431–441.
- Merriam-Webster Dictionary*. 2010. Merriam_Webster Online Dictionary, www.Merriam-Webster.com, dostęp z dnia 24. 11. 2010.
- Miller, Norman, William C. Pedersen, Vicki E. Pollock. 2000. *Discriminative Validity*, w: L. Bickman (red.), *Validity and Social Experimentation: Donald Campbell's Legacy*, Thousand Oaks: Sage Publications.
- Miller, Peter V., Robert M. Groves. 1985. *Matching survey response to official records: an exploration of validity in victimization reporting*, „Public Opinion Quarterly”, t. 49, nr 3, s. 366–380.
- Montgomery, Andrew C., Kathleen S. Crittenden. 2004. *Improving Coding Reliability for Open-Ended Questions*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Moser, Claus. A., Graham Kalton. 1971. *Survey Methods in Social Investigation*, London: Heinemann Educational Books.
- Murphy, Joe, Rodney Baxter, Joe Eyerman, David Cunningham, Joel Kennet. 2004. *A System for Detecting Interviewer Falsification*, American Association for Public Opinion Research, 59th Annual Conference May 13–16, 2004 Phoenix, Arizona, <http://www.rti.org/pubs/paper-detectingintrvfalse.pdf>, dostęp z dnia 24. 11. 2010.
- Neter John, John Wakesberg. 1964. *A study of response errors in expenditures data from household interviews*, „Journal of American Statistical Association”, t. 59, s. 18–55.
- Noelle, Elizabeth. 1963. *Umfragen in der Massengesellschaft – Einführung in die Methoden der Demoskopie*, Hamburg: Rowohlt.
- Nowak, Stefan. 1965. *Metody badań socjologicznych*, Warszawa: PWN.
- Nowak, Stefan. 1970. *Metodologia badań socjologicznych*, Warszawa: PWN.

- Olszewski, Andrzej. 2004. *Badania telemetryczne*, w: P. B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.
- Ostrom, Thomas M. 1988. *Bipolar survey items: an information processing perspective*, w: H.-J. Hippler, N. Schwarz, S. Sudman (red.), *Social Information Processing and Survey Methodology*, New York: Springer-Verlag.
- Parry Hugh J., Helen M. Crossley. 1950. *Validity of Responses to Survey Questions*, „Public Opinion Quarterly”, Spring, nr 1, s. 61–80.
- Patchen, Robert H., Beth M. Webb, 2002. *The future is now. The latest findings from the market launch of the portable people meter in Philadelphia (PPM) [Portable People Meter]*, Konferencja ESO-MAR, WAM: Week of Audience Measurement - Radio, Cannes, 9–14 czerwca.
- Payne, Stanley L. 2004. *Who Left it Open? A Description of the Free-Answer Question and its Demerits*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Pełka-Sługocka Maria D., Leszek Sługocki. 1975. *Analiza zgodności między wypowiedziami o faktach, uzyskanymi w wywiadzie kwestionariuszowym a dokumentami*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. V, Wrocław: Osolineum.
- Pett, Richard E., Jon A. Krosnick. 1995. *Attitude Strength. Antecedents and Consequences*, Mahwah, N.J: Lawrence Erlbaum Associates.
- Pierchala Carl E., Jyoti Surti, 2009. *Control Charts as a Tool for Data Quality Control*, „Journal of Official Statistics”, t. 25, s. 167–191.
- Polskie Towarzystwo Badaczy Rynku i Opinii, 2011 Katalog PTBRiO 2010/2011.
- Poulain Michel, Benoit Riandey i Jean-Marie Firdion. 1991. *Enquête biographique et registre belge de population: une confrontation des données*, „Population”, nr 1, s. 65–88.
- Program Kontroli Jakości Pracy Ankieterów. 2010. www.ofbor.pl/PKJPA, dostęp z dnia 25. 03. 2011.
- Program PISA: przygotowanie i przebieg badań. www.ifispan.waw.pl/ifis/badania, dostęp z dnia 25. 03. 2011.
- Przybyłowska, Ilona. 1975. *Wywiad o wywiadzie jako metoda otrzymywania informacji o reakcji wewnętrznej respondenta na pytanie kwestionariuszowe*, w: Zygmunt Gostkowski i Jan Lutyński (red.), *Ana-*

- lizy i próby technik badawczych w socjologii*, t. V, Wrocław: Zakład Narodowy im. Ossolińskich.
- Pyy-Martikainen Marjo, Ulrich Rendtel. 2009. *Measurement errors in retrospective reports of event histories. A validation study with Finnish register data*. „Survey Research Methods”, t. 3, nr 3, s. 139–155, <http://www.surveymethods.org>.
- Reynolds, Nina, Adamantios Diamantopoulos, Bodo Schlegelmilch. 2004. *Pretesting the Questionnaire Design: A Review of the Literature and Suggestions for Further Research*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Riesman, David. 1958. *Some observations on the interviewing in the Teacher Apprehension Study*, w: P. F. Lazarsfeld, W. Thielens, *The Academic Mind. Social Scientists in a Time of Crisis*, Glencoe, Ill: The Free Press.
- Roberts, Caroline, Annette E. Jäckle, Peter Lynn. 2007. *Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-face and Telephone Surveys*. Prague: 2nd European Survey Research Association Conference, czerwiec.
- Roberts, Caroline. 2007. *Mixing Modes of Data Collection in Surveys*. Prague: 2nd European Survey Research Association Conference, czerwiec.
- Rosenthal Robert, Ralph L. Rosnow. 1969. *The Volunteer Subjects*, w: Robert Rosenthal i Ralph. L. Rosnow (red.), *Artifacts in Behavioral Research*, New York: Academic Press.
- Rostocki, Włodzimierz. A. 1992. *Kłamstwo obronne w ujęciu M. Ossowskiej a etyczne problemy wywiadu kwestionariuszowego w socjologii*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. IX. Warszawa: Wydawnictwo IFiS PAN.
- Sampling for the European Social Survey – Round III: Principles and requirements. 2006. [www.europeansocialsurvey.org/Methodological Research/Sampling Designs](http://www.europeansocialsurvey.org/MethodologicalResearch/SamplingDesigns), dostęp z dnia 25. 03. 2007.
- Sanchez, Maria E. 2004. *Effects of Questionnaire Design on the Quality of Survey Data*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Saris Willem, Irmtraud N. Gallhofer. 2007. *Design, Evaluation, and Analysis of Questionnaire for Survey Research*, New York: John Wiley & Sons Inc.
- Saris, Willem E. 2002. *Report on the MTMM experiments In the pilot studies and Proposal for Round 1 of the ESS*. www.europeansocial-

- survey.org/methodological_research/question_improvement, dostęp z dnia 25. 03. 2007.
- Sawiński Zbigniew. 1988. *Błędy pomiaru w badaniach osiągnięć społeczno-zawodowych*, „Sudia Socjologiczne”, nr 2 (109), s. 45–62.
- Sawiński Zbigniew. 2010. *Zastosowanie tablic w badaniach zjawisk społecznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sawiński, Zbigniew, Sztabiński, Franciszek. 2005. *Czy ankieterzy oszukują? Jak można to sprawdzić?* w: P. B. Sztabiński, Z. Sawiński, F. Sztabiński (red.), *Fieldwork jest sztuką*, Warszawa: Wydawnictwo IFiS PAN.
- Sayer Andrew. 1992. *Method in Social Science. A realist approach*, London: Routledge.
- Scherpenzeel, Annette C. 1995. *A Question of Quality. Evaluating Survey Questions by Multitrait-multimethod Studies*, Royal PTT.
- Schuman, Howard, Stanley Presser. 1981. *Questions and Answers in Attitude Surveys. Experiment on Question Form, Wording and Context*, New York: Academic Press.
- Schuman, Howard, Stanley Presser. 2004. *The Open and Closed Question*, w: M. Bulmer (red.), *Questionnaires*, t. I, London: Sage.
- Schuman, Howard, Stanley Presser. 2004a. *The Acquiescence Quagmire*, w: M. Bulmer (red.), *Questionnaires*, t. II, London: Sage.
- Schuman, Howard. 2004. *The Random Probe: A Technique for Evaluating the Validity of Closed Questions*, w: M. Bulmer (red.), *Questionnaires*, t. III, London: Sage.
- Schwarz, Norbert, Fritz Strack i Hans-Peter Mai. 1991. *Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis*, „Public Opinion Quarterly”, t. 55.
- Schwarz, Norbert, Hans-J. Hippler, Brigitte Deutsch, Fritz Strack. 1985. *Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments*, „Public Opinion Quarterly”, t. 49.
- Schwarz, Norbert, Hans-J. Hippler. 1987. *What response scales may tell your respondents: Informative functions of response alternatives*, w: H.-J. Hippler, N. Schwarz, S. Sudman (red.), *Social information processing and survey methodology*, New York: Springer-Verlag.
- Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

- Selltiz, Claire, Lawrence S. Wrightsman, Stuart W. Cook. 1977. *Les Méthodes de Recherche en Sciences Sociales*, Montréal: Les éditions H.R.W.
- Selltiz, Claire, Marie Jahoda, Morton Deutsch, Stuart W. Cook. 1959. *Research Methods in Social Relations*, New York: Holt, Rinehart and Winston.
- Silberstein, Adriana. 1989. *Recall effects in the U.S. Consumer Expenditure Survey*, „Journal of Official Statistics”, t. 5, nr 2, s. 125–142.
- Sitek, Wojciech. 1976. *Ankieter humanista i ankieter technik*, w: Z. Gostkowski (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 4, Warszawa: Wydawnictwo IFiS PAN.
- Słomczyński, Kazimierz M. 1968. *Granice stosowalności ankiety audytoryjnej w środowisku robotniczym*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. II, Wrocław: Zakład Narodowy im. Ossolińskich.
- Słomczyński, Kazimierz M. 1983. *Wpływ oficjalnej i prywatnej sytuacji wywiadu na odpowiedzi respondentów w środowisku inteligencji*, w: K. Lutyńska, A. P. Wejland (red.), *Wywiad kwestionariuszowy. Analizy teoretyczne i badania empiryczne*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Słomczyński, Kazimierz M. 2004. *Polska metodologia socjologicznych badań empirycznych na tle międzynarodowym*, w: P. B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sowden, Philip. 2002. *Benchmarking in the context of Centres of Research and Technology Development*, w: *A Methodology for Benchmarking RTD Organisations In Central and Eastern Europe*. Budapest University of Technology and Economics: The Brighton Proceedings of the RECORD Thematic Network, Brighton 2002.
- Staszyńska, Katarzyna. 1985. *Społeczny kontekst badań ankietowych*, w: K. Lutyńska (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 8, Warszawa: Wydawnictwo IFiS PAN.
- Statsoft.pl 1984–2005. *Statystyki podstawowe. Czy współczynniki korelacji są „addytywne”?*
- Stefanowska, Małgorzata. 1988. *Odbiorcy kultury. Deklaracje i rzeczywistość*, Warszawa: PWN.
- Stein, Jess, Laurence Urdang. (red.). 1983. *The Random House Dictionary of the English Language*, New York: Random House.

- Steinmann Matthias, Manuel Dahler, Knut Hackbart. 2002. *Radiocontrol in Switzerland. One year of experience with audiometers measuring radio listening. [Radiocontrol (RC)]*, Konferencja ESOMAR, WAM: Week of Audience Measurement - Radio, Cannes, 9–14 czerwca.
- Stoop, Ineke, 2007. *No Time, too Busy: Time Strain and Survey Cooperation*, w: G. Loosveldt, M. Swyngedouw, B. Cambré (red.), *Measuring Meaningful Data in Social Research*, Leuven, Voorburg: Acco.
- Stoop, Ineke, Jaak Billiet, Achim Koch, Rory Fitzgerald. 2010. *Improving Survey Response. Lesson learned from the European Social Survey*, Chichester: John Wiley&Sons.
- Stoop, Ineke. 1995. *An Overview of Quantitative Key Informant Studies*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1–4 kwietnia, Bristol: American Statistical Association.
- Stoop, Ineke. 2005. *The Hunt for the Last Respondent. Nonresponse in Sample Surveys*, The Hague: Social and Cultural Planning Office of the Netherlands.
- Stoop, Ineke. 2008. *Event data collection Round 4. Guidelines for ESS National Coordinators*, http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=88&Itemid=359, dostęp z dnia 26.05.2011.
- Styczeń, Marek. 1968. *O trafności i rzetelności testów*, w: Klemens Szaniawski (red.), *Metody statystyczne w socjologii. Wybrane zagadnienia*, Warszawa: PWN.
- Sudman Seymour , Norman M. Bradburn. 1982. *Asking Questions. A Practical Guide to Questionnaire Design*, San Francisco: Jossey-Bass Inc. Publishers.
- Sudman Seymour, Norman M. Bradburn. 1973. *Effects of Time and Memory Factors on Response in Surveys*, „Journal of American Statistical Association”, t. 64, nr 344, s. 805–815.
- Sułek, Antoni. 1990. *O rzetelności i nierzetelności badań sondażowych w Polsce*, w: A. Sułek, *W terenie, w archiwum, w laboratorium*, Warszawa: IS UW.
- Sułek, Antoni. 1999. „Ilu jest Żydów w Polsce?” *Eksperymentalne studium wpływu skali na odpowiedzi ankietowe*, w: H. Domański, K. Lutyńska, A. W. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa: Wydawnictwo IFiS PAN.

- Sulek, Antoni. 2001. *Demokratyczna zmiana ustroju a rzetelność badań sondażowych. Doświadczenia polskie*, w: *Sondaż polski. Przygarść rozpraw o badaniach ankietowych*, Warszawa: Wydawnictwo IFiS PAN.
- Sulek, Antoni. 2002. *O rzetelności i nierzetelności badań sondażowych Polsce socjalizmu realnego*, w: *Ogród metodologii socjologicznej*, Warszawa: Scholar.
- Sulek, Antoni. 2002a. *Wiarygodność źródeł i rzetelność danych urzędowych*, w: *Ogród metodologii socjologicznej*, Warszawa: Scholar.
- Sztabiński Franciszek, Paweł B. Sztabiński, 1994. „Ocena metodologiczna wyników pilotażowego badania kohortowego »Nasze życie codzienne i zdrowie«”, Warszawa, maszynopis.
- Sztabiński Paweł B., Franciszek Sztabiński, Dariusz Przybysz, 2009, *How does length of fieldwork period influence non-response? Findings from ESS 2 in Poland*, „ASK, Research and Methods”, nr 18, s. 67–95.
- Sztabiński, Franciszek, Paweł B. Sztabiński. 1994. *Polscy respondenci początku lat '90. Stosunek do badań i jego uwarunkowanie*, „Studia Socjologiczne”, nr 2, s. 141–164.
- Sztabiński, Franciszek, Paweł B. Sztabiński. 1997. *Wartość odpowiedzi w wywiadach telefonicznych ze wspomaganiami komputerowym (CATI)*, „ASK”, nr 1–2.
- Sztabiński, Franciszek. 1990. *Wywiad kwestionariuszowy i ankieta pocztowa. Analiza przebiegu procesu uzyskiwania informacji i ich rezultaty*. „Przegląd Socjologiczny”, t. 37: 235–263.
- Sztabiński, Franciszek. 1995. *Kontrola w badaniach syurveyowych: pytania i odpowiedzi*, „ASK”, nr 1, s. 49–60.
- Sztabiński, Franciszek. 1996. *Ankieta do ankiera: źródło informacji czy dezinformacji o wywiadzie kwestionariuszowym*, „ASK”, nr 2 (4), s. 37–58.
- Sztabiński, Franciszek. 1997. *Ankieta pocztowa i wywiad kwestionariuszowy*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Franciszek. 1999. *Kontrola w badaniach ilościowych: ocena pracy ankierów czy ocena wartości uzyskanych materiałów?* w: H. Domański, K. Lutyńska, W. A. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Franciszek. 2001. *Co wiemy o respondencie? Rozmiar i źródła błędów w pomiarze cech metryczkowych*, „ASK”, nr 10, s. 91–114.

- Sztabiński, Franciszek. 2003. *Logika badacza i logika respondenta. Problem adekwatności narzędzia badawczego*, „ASK”, nr 12, s. 147–175.
- Sztabiński, Franciszek. 2004. *Jak dotrzeć do respondenta i skłonić go do wzięcia udziału w badaniu?*, „Przegląd Socjologiczny”, t. LIII, nr 1.
- Sztabiński, Franciszek. 2005a. *Narzędzie badawcze i jego elementy. Czy tylko kwestionariusz?*. w: P. B. Sztabiński, Z. Sawiński, F. Sztabiński (red.), *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Franciszek. 2005b. *Szkolenie do badania (tzw. briefing) i przygotowanie ankietera do wywiadu*, w: P. B. Sztabiński, Z. Sawiński, F. Sztabiński (red.), *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 1975. *Próba weryfikacji wiarygodności odpowiedzi na pytania wieloalternatywne*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław: Osolineum.
- Sztabiński, Paweł B. 1997. *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 1999. *Wywiad telefoniczny ze wspomaganie komputerowym (CATI) a wywiad kwestionariuszowy. Problem porównalności wyników*, w: H. Domański, K. Lutyńska, A. W. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 1999a. *Wywiad telefoniczny ze wspomaganie komputerowym (CATI): Co zyskujemy, co tracimy?*, „ASK”, nr 8, s. 51–66.
- Sztabiński, Paweł B. 1999b. *Wywiad telefoniczny ze wspomaganie komputerowym: nowa technika czy wywiad kwestionariuszowy przez telefon?*, „Przegląd Socjologiczny”, t. 48, nr 1.
- Sztabiński, Paweł B. 2001. *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywiście idealna technika?*, „ASK”, nr 10.
- Sztabiński, Paweł B. 2002. *Punkt „środkowy” w skalach bipolarnych: zamieszczać czy nie zamieszczać?*, „ASK”, nr 11.
- Sztabiński, Paweł B. 2003. *Badacz jako współautor wyników badania*, „ASK”, nr 12.
- Sztabiński, Paweł B. 2004. *Metodologia badania Europejski Sondaż Społeczny*, „ASK”, nr 13, s. 27–37.

- Sztabiński, Paweł B. 2005. *Dlaczego należy rygorystycznie przestrzegać zasad pracy ankieterskiej? Wywiad kwestionariuszowy jako technika standaryzowana*, w: P. B. Sztabiński, Z. Sawiński, F. Sztabiński (red.), *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 2005a. *Zadawanie pytań i notowanie odpowiedzi. Rodzaje pytań kwestionariuszowych*, w: P. B. Sztabiński, Z. Sawiński, F. Sztabiński (red.), *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 2006. *Dlaczego respondenci uczestniczą lub nie uczestniczą w badaniach? Porównanie 1994–2004*, „ASK”, nr 15.
- Sztabiński, Paweł B., Franciszek Sztabiński, Dariusz Przybysz, 2007a. *Are Non-Respondents Similar to Respondents? Findings from the ESS-2004 Project in Poland*, „ASK”, nr 16, s. 25–55.
- Sztabiński, Paweł B., Franciszek Sztabiński, Dariusz Przybysz. 2007. *How does length of fieldwork period influence non-response phenomenon and differences between Respondents and Non-Respondents? Findings from the ESS Project in Poland*. 2nd European Survey Research Association Conference, Prague, 25–29 czerwca.
- Sztabiński, Paweł B., Franciszek Sztabiński, Zbigniew Sawiński (red.). 2004. *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B., Franciszek Sztabiński. 1997. *Wartość odpowiedzi w wywiadach telefonicznych ze wspomaganie komputerowym (CATI)*, „ASK”, nr 1–2, s. 73–92.
- Sztabiński, Paweł B., Franciszek Sztabiński. 2005. *Jak połączyć pilotaż z badaniem próbnym? Przykład Europejskiego Sondażu Społecznego 2004*. „ASK”, nr 14, s. 55–75.
- Sztabiński, Paweł B. 1978. *Metody kontroli pracy ankietów w badaniach z zastosowaniem wywiadu kwestionariuszowego*, „Przegląd Socjologiczny”, t. XXX, s. 197–225.
- Sztabiński, Paweł B. 1995. *Dlaczego respondenci zgadzają się na wywiad? Problemy aranżacji wywiadu w badaniach surveyowych*, „ASK”, nr 2 (4).
- Sztabiński, Paweł B., Anna Dyjas-Pokorska, Teresa. Żmijewska-Jędrzejczyk. 2008. *Understanding Refusals*. „ASK. Research&Methods”, nr 17, s. 39–84.

- Sztaudynger-Kaliszewicz Anna. 1966. *Badania wiarygodności zeznań w spisie powszechnym*, w: Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. I, Wrocław: Ossolineum.
- Śmiłowski Eugeniusz, 1990. *Efekt sponsora w świetle metodologicznych analiz i eksperymentów*, „Kultura i Społeczeństwo”, nr 4.
- Topolski, Jerzy. 1984. *Metodologia historii*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Tourangeau Roger, Kenneth Rasinski, Jared B. Jobe, Tom W. Smith, William F. Pratt. 1997. *Sources of Error in a Survey on Sexual Behavior*, „Journal of Official Statistics”, t. 13, nr 4, s. 341-365.
- Tourangeau Roger, Lance J. Rips, Kenneth Rasinski. 2000. *The Psychology of Survey Response*, Cambridge: University Press.
- Tourangeau Roger. 1988. *Attitude Measurement: a Cognitive Perspective*, w: H-J. Hippler, N. Schwarz, S. Sudman (red.), *Social Information Processing and Survey Methodology*, New York: Springer-Verlag.
- Trochim, William M. K. 2006b. *Measurement Validity Types*, w: *Research Methods Knowledge Base*, <http://www.socialresearchmethods.net/kb/measval.php>, dostęp z dnia 26. 10. 2010.
- Tuchańska, Barbara. 1983. *Model wewnętrznej reakcji respondenta na pytanie*, w: K. Lutyńska, A. P. Welland (red.), *Wywiad kwestionariuszowy. Analizy teoretyczne i badania empiryczne*, Wrocław: Zakład Narodowy im. Ossolińskich.
- van der Vaart, Wander, Wil Dijkstra, Johannes van der Zouwen. 1995. *The Effects of Change on Recalling Attitudes and Usual Behavior*, w: *Proceedings of the International Conference on Survey Measurement and Process Quality*, 1-4 kwietnia, Bristol: American Statistical Association.
- Wagenaar, Willem A. 2004. *My memory: A Study of Autobiographical Memory over Six Years*, w: M. Bulmer (red.), *Questionnaires*, t. IV, London: Sage.
- Weaver, David A. 2000. *The Accuracy of Survey-Reported Marital Status: Evidence from Survey Records Matched to Social Security Records*, „Demography”, t. 37, nr 3, s. 395-399. .
- Weisberg, Herbert. F., 2005. *The Total Survey Error Approach. A Guide to the New Science of Survey Research*, Chicago-London: The University of Chicago Press.
- Weiss, Carol H. 1968-69. *Validity of welfare mothers' interview responses*, „Public Opinion Quarterly”, Winter, s. 622-633.

- West, Stephen G. 2004. *Poza eksperyment laboratoryjny – plany eksperymentalne oraz quasi-eksperymentalne w otoczeniu naturalnym*, w: J. Brzeziński (red.), *Metodologia badań psychologicznych*, Warszawa: Wydawnictwo Naukowe PWN.
- Willcock, H.D. 1956. *Field Observation: A Progress Report*, w: M. Harris (red.), *Selected Papers on Interviewers and Interviewing*, London.
- Wilson, Timothy. D., Sara D. Hodges. 1992. *Attitudes as Temporary Constructions*, w: Leonard, Martin, Abraham Tesser (red.), *The construction of social judgment*, Hillsdale, NJ: Erlbaum.
- Wojciszke, Bogdan. 2004. *Systematycznie Modyfikowane Autoreplikacje: logika programu badań empirycznych w psychologii*, w: J. Brzeziński (red.), *Metodologia badań psychologicznych*, Warszawa: Wydawnictwo Naukowe PWN.
- Zetterberg, Hans L. 1965. *On Theory and Verification in Sociology*, Totowa: Bedminster Press.
- Zurn, Marek. 1976. *Próba pomiaru efektu ankieterskiego oraz błędów popełnianych przez ankieterów w wywiadzie kwestionariuszowym*, w: Z. Gostkowski (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 4. Warszawa: Wydawnictwo IFiS PAN.

Aneks 1.

Zestawienie badań i projektów będących przedmiotem analizy

Nr bad.	Rok realizacji	Nazwa badania	Firma badawcza	Typ badania	Rodzaj próby/ (N)	Zasięg terytorialny	Metoda badania ^(a)	Badanie kontrolne (N)	Metoda kontroli Osob. Tel.	
01	1995–2001	Omnibusy (97 fal)	SMG/KRC	Stale (in home)	Imienna/ adresowa (97 000)	Ogólnopolska (m/w)	PAPI	6436	5836	600
02	1996–2000	R2'96, R3-B1/B2'97, M1-B1/B2'98-'99, Szklana Pogoda M6'1/2'00 (9 fal)	SMG/KRC	Tracking (in home)	Adresowa (24 649)	Ogólnopolska (m/w)	PAPI	1488	898	590
03	1996–2000	Parasol (44 fale)	SMG/KRC	Tracking (in home)	Imienna (43 000)	Ogólnopolska (m/w)	PAPI	2263	2145	118
04	1997	Słodycze D-35 (2 fale)	SMG/KRC	Tracking (in home)	Imienna (1000)	Ogólnopolska (m/w)	PAPI	133	130	3
05	1997	Evita P-29	SMG/KRC	Ad hoc (in home)	Celowa (1500)	Ogólnopolska (m)	PAPI	150	103	47
06	1997	Global PR. 97076	MEMRB	Ad hoc (in home)	Celowa (1200)	Ogólnopolska (m/w)	PAPI	200	190	10
07	1997	Prywatni przedsiębiorcy	IFiS PAN	Ad hoc (in home)	Imienna (800)	Ogólnopolska (m/w)	PAPI	117	35	82
08	1997	Samochody P-65	SMG/KRC	Ad hoc (in home)	Adresowa/ celowa (3000)	Ogólnopolska (m)	PAPI	158	39	119
09	1998	Podziały społeczne w Polsce	IFiS PAN	Ad hoc (in home)	Imienna (1596)	Ogólnopolska (m/w)	PAPI	281	188	93

Nr bad.	Rok realizacji	Nazwa badania	Firma badawcza	Typ badania	Rodzaj próby/ (N)	Zasięg terytorialny	Metoda badania ^(a)	Badanie kontrolne (N)	Metoda kontroli	
									Osob.	Tel.
10	1998	Rok rodziny (2 fale)	IFiS PAN, Instytut Kardiologii	Panel (in home)	Celowa/imienna (373/305)	Regionalna (m)	PAPI	50	45	5
11	1998–1999	Omnibusy (11 fal)	CBOS	Stałe (in home)	Adresowa (11 000)	Ogólnopolska (m/w)	PAPI	568	568	0
12	1998–1999	Panel C-113/114 (2 fale)	SMG/KRC	Panel (in home)	Imienna (5000)	Ogólnopolska (m/w)	PAPI	139	131	8
13	1998–1999	Batony i cukierki: S-10/S11(11 fal)	SMG/KRC	Tracking (in home)	Imienna/celowa (11 000)	Ogólnopolska (m/w)	PAPI	372	356	16
14	1998–1999	Tankard (7 fal)	SMG/KRC	Tracking (in home)	Imienna/celowa (14 000)	Ogólnopolska (m/w)	PAPI	464	438	26
15	1998–2000	Słone przekąski D-35 (18 fal)	SMG/KRC	Tracking (in home)	Imienna/celowa (5400)	Ogólnopolska (m/w)	PAPI	630	560	70
16	1998–2000	Mim C-50 (3 fale)	SMG/KRC	Tracking (in home)	Celowa (2500)	Ogólnopolska (m/w)	PAPI	281	281	0
17	1998–2000	Szklanka (8 fal)	SMG/KRC	Tracking (in home)	Celowa (3800)	Ogólnopolska (m/w)	PAPI	571	494	77
18	1998–2000	TGI (17 fal)	SMG/KRC	Tracking (in home)	Imienna/adresowa (91 600)	Ogólnopolska (m/w)	PSAQ	1564	1358	206
19	1999	Mediolan D-188	SMG/KRC	Ad hoc (in home)	Imienna (2000)	Ogólnopolska (m/w)	PAPI	416	371	45

Nr bad.	Rok realizacji	Nazwa badania	Firma badawcza	Typ badania	Rodzaj próby/ (N)	Zasięg terytorialny	Metoda badania ^(a)	Badanie kontrolne (N)	Metoda kontroli Osob.	Tel.
20	2000	Galeria D-02/05	SMG/KRC	Ad hoc (CLT)	Celowa (1000)	Regionalna (m)	PAPI	147	47	100
21	2000	Opinie o rozmaitych zwyczajach oraz o przestrzeganiu prawa	IFiS PAN	Ad hoc (in home)	Imienna (1209)	Ogólnopolska (m/w)	PAPI	219	192	27
22	2000	Sprawy kraju	BSM	Ad hoc (in home)	Imienna (1000)	Ogólnopolska (m/w)	PAPI	124	44	80
23	2000	Tenpies	SMG/KRC	Ad hoc (in home)	Celowa (4000)	Ogólnopolska (m/w)	PAPI	629	119	510
24	2001	Losy szkolne i zawodowe rocznika 1977–1978	IFiS PAN, UW	Ad hoc (in home)	Imienna/ celowa (2500)	Ogólnopolska (m/w)	PAPI	309	163	146
25	2001	CITA	SMG/KRC	Ad hoc (in home)	Imienna (1677)	Regionalna (m)	PAPI	120	119	1
26	2001–2002	Odbiór telewizji w Polsce	GfK Polonia	Ad hoc (in home)	Adresowa (3000)	Ogólnopolska (m/w)	PAPI	122	50	72
27	2001–2002	OVIM	SMG/KRC	Tracking	Nośniki reklamy zewnętrznej (20 221)	Regionalna (m)	Obserwacja	1150	1150	0
28	2002	Europejski Sondaż Społeczny_ R. 1	IFiS PAN	Tracking (in home)	Imienna (2978)	Ogólnopolska (m/w)	PAPI	189* (106/54/29)	106	83
29	2002	Nasze zdrowie, nasze dolegliwości	IFiS PAN, Instytut Reumatologiczny	Ad hoc (in home)	Imienna/ celowa (2010)	Ogólnopolska (m/w)	PAPI	203	110	93

Nr bad.	Rok realizacji	Nazwa badania	Firma badawcza	Typ badania	Rodzaj próby/ (N)	Zasięg terytorialny	Metoda badania ^(a)	Badanie kontrolne (N)	Metoda kontroli Osob.	Tel.
30	2002	Nasze życie codzienne	IFiS PAN	Ad hoc (in home)	Imienna (1000)	Ogólnopolska (m/w)	PAPI	200	87	113
31	2003	Bawaria T-55	SMG/KRC, IFiS PAN	Ad hoc (in home)	Imienna (1010)	Ogólnopolska (m/w)	PAPI	103	0	103
32	2004	Europejski Sondaż Społeczny R. 2	IFiS PAN	Tracking (in home)	Imienna –(2399)	Ogólnopolska (m/w)	PAPI	278* (189/55/34)	170	108
33	2004	Mieszkańcy Warszawy (SIRS)	IFiS PAN	Ad hoc (in home)	Imienna (2248)	Regionalna (m)	PAPI	311* (160/103/ 48)	150	161
34	2004–2005	Omnimasy (31 fal)	TNS OBOP	Stale (in home)	Adresowa/ Random route (31 000)	Ogólnopolska (m/w)	PAPI	1507	1507	0

^(a) Metoda badania: PAPI (Paper and Pencil Interviewing); (Paper Self Administered Questionnaire), Obserwacja nieuczestnicząca (zewnętrzna).

* w podziale na: wywiad zrealizowany, odmowa udzielenia wywiadu, brak kontaktu z wylosowaną osobą.

** w podziale na: wywiad zrealizowany, dobrana osoba nie kwalifikuje się do wywiadu, odmowa udzielenia wywiadu

Aneks 2. Kwestionariusz kontrolny do badania EES R. 2.

Schemat kontroli do badania ESS 2

Każdy przypadek należy traktować indywidualnie, dlatego przedstawiony schemat jest pomocny w początkowej fazie przygotowania do badania w terenie. We wszystkich Częściach opisane są wszystkie możliwe sytuacje oraz reguły przejść i następstw stosowne do nich.

Ogólny schemat kontroli (bez uwzględnienia sytuacji nietypowych):

Część 1 – sterująca		
↓	↓	↓
Część 2 – podany w karcie kontrolnej adres nie istnieje	Wywiad został zrealizowany/ nie został zrealizowany	Część 3 – pod podanym adresem nie mieszka wskazany respondent
	↓	
	Część 4 – kwestionariusz kontrolny	
	Część 5 – opis i charakterystyka sąsiedztwa	

Szczegółowy schemat kontroli (z uwzględnieniem sytuacji nietypowych):

Kontrolowane kategorie wg deklaracji ANKIETERA	Kontrolowane kategorie wg deklaracji RESPONDENTA	Kolejność wypełniania poszczególnych Części
I. wywiad zrealizowany • wywiady zrealizowane	1. wywiad zrealizowany z respondentem	1, 4, 5
	2. zamiana respondenta	
	3. sufit - ankietera w ogóle nie było	1, 4 (pyt. 15-20), 5
	4. sufit - ankieter był, ale z nikim nie przeprowadził wywiadu	1, 4 (pyt. 14-20), 5
II. odmowa • odmowa udzielona ankieterowi • odmowa telefoniczna, faxem lub osobiście w ORBS	1. ankietera w ogóle nie było (wpisał odmowę)	1, 4 (pyt. 15-20), 5
	2. ankieter był, respondent wskazał inny termin (wpisał odmowę)	
	3. odmowa udzielona ankieterowi przez respondenta	1, 4 (pyt. 14-20), 5
	4. odmowa udzielona ankieterowi przez inną osobę	

III. brak kontaktu <ul style="list-style-type: none"> • nie było nikogo w domu • R. nie było w umówionym terminie • R. był w domu, ale nie otworzył drzwi • brak bezpośredniego dostępu do mieszkania/ budynku • ustalono, że R. przeniósł się w inne miejsce, wyprowadził się • R. jest chwilowo nieobecny (na przykład wychodzi/ na wakacjach/ chory). • R. nie jest w stanie udzielić wywiadu przez cały okres trwania badania ze względu na swój stan psychiczny lub fizyczny • R. jest niedostępny przez cały okres trwania badania z innych przyczyn (na przykład wyjechał, jest zagranicą) • Inna przyczyna 	1. ankietera w ogóle nie było (wpisany brak kontaktu)	1, 4 (pyt. 15-20), 5
	2. ankieter nie stawił się w umówionym terminie	1, 4 (pyt. 14-20), 5
	3. respondenta nie było w umówionym terminie	
	4. podany adres nie istnieje	1, 2
	5. pod podanym adresem nie mieszka wskazany respondent	1, 3, 5

03. A czy w badaniu tym brała udział inna osoba z P. gospodarstwa domowego. Chodzi o sytuację, w której ankieter zadawał pytania innej osobie, NIE P.? (Jeżeli respondent nie jest pewien dodać: Proszę spytać o to P. domowników).

(1) tak → kto? (wpisać stopień pokrewieństwa do respondenta)
 Rozmawiać z osobą wskazaną przez respondenta - faktycznym respondentem →
 (przejsć do CZĘŚCI 4)
 Jeżeli nie ma faktycznego respondenta pytania Części 4 należy zadać dotychczasowemu rozmówcy i opisać zaistniałą sytuację.

(2) nie, nie wiem

jeżeli Respondent:

W PYT. 01 odpowiedział „TAK” – „zgłosił się ankieter”

W PYT. 02 odpowiedział „NIE” – „odmówiłem udziału w badaniu”

W PYT. 03 odpowiedział „NIE” – „z nikim z gospodarstwa nie został zrealizowany wywiad → przejsć do CZĘŚCI 4 PYT. 14-20

jeżeli Respondent:

W PYT. 01 odpowiedział „NIE” – „nie zgłosił się ankieter”

W PYT. 03 odpowiedział „NIE” – „z nikim z gospodarstwa nie został zrealizowany wywiad” → przejsć do CZĘŚCI 4 PYT. 15-20

Dodatkowe uwagi dotyczące przebiegu wywiadu:

nr kontr. wywiadu

I _ I _ I I _ I _ I _ I _ I

data wywiadu

I _ I _ I - I _ I _ I - 2004

Część 2 – Podany w Karcie kontrolnej adres nie istnieje
<i>Poniższe punkty wypełniamy po upewnieniu się, iż wskazany adres nie istnieje lub pod tym adresem nie ma mieszkań prywatnych. Wszystkie sytuacje należy traktować indywidualnie, poniższe zagadnienia są schematem wg którego należy zbierać informacje. Jeżeli istnieje taka możliwość należy zasięgnąć informacji na przykład u dozorczy, lokatorów ect.</i>
1. Adres wskazany w Karcie Kontrolnej (wpisać)
2A. Podany adres nie istnieje
2. Nie istnieje ulica o wskazanej nazwie (wpisujemy ulice o podobnych nazwach):
3. Nie istnieje podany numer budynku (opisujemy dokładnie co znajduje się w tym miejscu):
• <i>Od jakiego numeru zaczyna się, a na jakim kończy numeracja w pobliżu wskazanego adresu</i>
• <i>Od jakiego numeru zaczyna się i na jakim kończy ulica (opisujemy stronami - parzysta, nieparzysta)</i>
4. Nie ma podanego numeru mieszkania:
• <i>Opisujemy od jakiego numeru zaczyna się numeracja mieszkań a na jakim kończy (opisujemy klatkami)</i>
• <i>Na domofonie lub w spisie lokatorów sprawdzamy, czy jest nazwisko wskazanego na Karcie Kontrolnej respondenta</i>
• <i>Tak → spisujemy numer mieszkania, kontaktujemy się → realizujemy od CZĘŚCI I PKT. C</i>
• <i>Nie → KONIEC</i>
2B. Budynek niemieszkalny/ nie zamieszkały
<i>Zaznaczyć odpowiedni punkt /opisać</i>
1 = <i>Budynek/ mieszkanie jest opuszczony /zdeastowany lub zburzony</i>
2 = <i>Budynek nie jest jeszcze wybudowany do końca, nie jest gotowy do zamieszkania</i>
3 = <i>Adres nie jest zamieszkały (drugie mieszkanie, obiekt wakacyjny itp.)</i>
4 = <i>Adres nie jest adresem mieszkalnym: znajdują się pod nim tylko firmy (biura) lub obiekty przemysłowe</i>
5 = <i>Adres nie jest adresem mieszkalnym: znajdują się pod nim tylko instytucje (domy opieki, szpitale, obiekty wojskowe, kościelne itp.)</i>
88 = <i>Inna sytuacja (opisać)</i>
→ KONIEC

.....	
Dodatkowe uwagi dotyczące przebiegu wywiadu:	
<i>Stwierdzam, że wywiad kontrolny przeprowadziłem osobiście</i>	
<i>imię i nazwisko kontrolera:</i>	<i>podpis kontrolera:</i>
<i>nr kontr. wywiadu</i> I _ I _ I I _ I _ I _ I _ I	<i>data wywiadu</i> I _ I _ I - I _ I _ I - 2004

Część 3 – Pod podanym adresem nie mieszka wskazany respondent	
<p><i>Poniższe punkty wypełniamy w trakcie rozmowy z lokatorem mieszkania, wszystkie sytuacje należy traktować indywidualnie; poniższe pytania są schematem wg którego należy zbierać informacje.</i></p> <p>Należy wypełnić wszystkie punkty – zadajemy wszystkie pytania od 1 do 5 bez względu na odpowiedzi</p>	
1. Imię i nazwisko respondenta wskazanego w Karcie Kontrolnej (wpisać)	
2. Imię i nazwisko lokatora, z którym rozmawiamy (odmowa: po rozmowie z lokatorem spisujemy z wizytówki, listy lokatorów, domofonu – <u>zapisać skąd jest informacja</u>).	
3. Czy lokator dostał List Zapowiedni z ORBS informujący o badaniu i wizycie ankietera?	
• Tak → odbieramy (jeśli wyrzucił – zapisać)	
• Nie	
4. Czy z lokatorem lub inną osobą z gospodarstwa domowego był przeprowadzony wywiad na temat postaw społecznych i ich zmian zatytułowany „Europejski Sondaż Społeczny”?	
• Tak → z kim? (zapisać imię i nazwisko) I z tą osobą zrealizować → CZĘŚĆ 4	
• Nie → (opisujemy osobę, z którą rozmawialiśmy: płeć, wiek)	
5. Czy lokator zna wskazaną osobę?	
• Tak → wpisać adres, telefon → skontaktować się ze wskazaną osobą → zrealizować od CZĘŚCI 1	
• Nie → przeczytać poniższe uwagi i zastosować się do nich	
JEŚLI POD PODANYM ADRESEM	
a) z nikim nie został zrealizowany wywiad	
b) lokatorzy nie znają wskazanej osoby → ZAKOŃCZYĆ ROZMOWĘ → WYPEŁNIĆ CZĘŚĆ 5	
JEŚLI POD PODANYM ADRESEM	
a) z nikim nie został zrealizowany wywiad	
b) lokatorzy znają wskazaną osobę lecz nie znają/ nie chcą podać jej adresu → DOPYTAĆ OD KIEDY NIE MIESZKA POD PODANYM ADRESEM → ZAKOŃCZYĆ ROZMOWĘ → WYPEŁNIĆ CZĘŚĆ 5	
Wskazana osoba nie mieszka od:	
dzień <u> </u> <u> </u> <u> </u> miesiąc <u> </u> <u> </u> <u> </u> rok <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	
Dodatkowe uwagi dotyczące przebiegu wywiadu:	
nr kontr. wywiadu <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	data wywiadu <u> </u> <u> </u> <u> </u> <u> </u> - <u> </u> <u> </u> <u> </u> <u> </u> - 2004



Instytut Filozofii i Socjologii Polska Akademia Nauk

Europejski Sondaż Społeczny 2
Październik–Listopad 2004

Warszawa, ul. Nowy Świat 72

Część 4 – Kwestionariusz kontrolny

04. Chciałbym P. bardzo serdecznie podziękować za współpracę i pomoc w realizacji naszych badań. Między innymi właśnie dzięki P. odpowiedziom uzyskaliśmy wiele cennych i ciekawych materiałów.

Rzetelność tych badań w dużym stopniu zależy jednak od jakości pracy naszych współpracowników - ankierów. Z tego względu musimy sprawdzać, czy docierali oni do właściwych osób i czy postępowali prawidłowo. Dlatego też bardzo proszę o odpowiedź na kilka pytań dotyczących przebiegu rozmowy z naszym ankierem. Jeśli czegoś nie będzie mógł P. sobie przypomnieć, nie będzie P. pamiętał, jakich odpowiedzi udzielił lub też, czy w ogóle takie pytanie zostało P. zadane podczas rozmowy, proszę mi o tym powiedzieć.

Podobnie jak w przypadku P. rozmowy z naszym współpracownikiem tak i obecnie, wszystkie udzielone przez P. odpowiedzi objęte są całkowitą poufnością.

05. Czy na wstępie swojej wizyty ankier: . . . (odczytać kolejne podpunkty; zaznaczyć odpowiedź do każdego podpunktu)

	1. tak	2. nie	3. nie pamiętam
(a) przedstawił się imieniem i nazwiskiem?	1	2	3
(b) pokazał legitymację ankiera ORBS IFiS PAN?	1	2	3
(c) udzielił informacji na temat badania ESS?	1	2	3
(d) wyjaśnił zasady losowania osób do badania?	1	2	3
(e) zapewnił o poufności badania?	1	2	3
(f) poinformował o statystycznym opracowaniu wyników badania?	1	2	3
(g) poinformował o tym, że prowadzenie badań jest zgodne z Ustawą o ochronie danych osobowych?	1	2	3

06. Czy nasz współpracownik zadawał P. pytania dotyczące: . . . (odczytać kolejne podpunkty; zaznaczyć odpowiedź do każdego podpunktu)

	1. tak	2. nie	3. nie pamiętam
(A1-7) Konsumpcji mediów , to znaczy ile czasu P. poświęca w dzień powszedni na oglądanie telewizji, słuchanie radia, czytanie gazet, korzystania z Internetu?	1	2	3
(B1-3) P. zainteresowania polityką i P. ogólnej opinii na temat polityków?	1	2	3
(B4-10) P. zaufania do różnych instytucji takich jak polski parlament, policja, ONZ?	1	2	3
(B11-12) P. udziału w ostatnich wyborach do Sejmu we wrześniu 2001 r. i wyboru partii, na którą P. głosował?	1	2	3
(B20a-23) P. orientacji politycznych , przynależności lub sympatyzowania z jakąś partią?	1	2	3
(B24-33) P. zadowolenia ze stanu gospodarki, sposobu działania rządu, kształcenia i edukacji oraz służby zdrowia w Polsce?	1	2	3
(B35-40) P. opinii na temat imigracji i osób przyjeżdżających z innych krajów do Polski, aby tutaj żyć i pracować oraz ich wpływu na polską gospodarkę i kulturę?	1	2	3
(C1-8) P. życia towarzyskiego, poczucia bezpieczeństwa osobistego i stanu zdrowia?	1	2	3
(C9-28) P. przynależności religijnej, narodowej, etnicznej oraz kraju urodzenia P. i P. rodziców?	1	2	3
(D1-30) P. stanu zdrowia , wizyt u lekarzy specjalistów, stosowania zabiegów medycyny niekonwencjonalnej, P. opinii na temat relacji lekarza z pacjentem?	1	2	3
(E1-30) Wzajemnych relacji i przestrzegania prawa przez producentów i konsumentów?	1	2	3
(F1-48) Składu P. gospodarstwa domowego , wykształcenia, pracy zawodowej, dochodów P. i P. współmałżonka?	1	2	3
(F49-60) P. rodziców ich wykształcenia oraz sytuacji zawodowej w okresie, kiedy miał P. 14 lat?	1	2	3
(G1-58) P. rodziny , relacji między pracą zawodową a życiem prywatnym, podziałem obowiązków domowych, opieki nad dziećmi, pomocy rodzinnej?	1	2	3
(J1-6) P. zwyczajów żywieniowych wspólnych posiłków rodzinnych, jedzenia poza domem?	1	2	3
(P1-5) P. związku z Polską, Unią Europejską, grupą religijną, regionem?	1	2	3

07. [C6] Na ile bezpiecznie czuje/ czułby się P. chodząc samotnie po zmroku w okolicy swojego domu? Czy czuje/ czułby się P. . . . (odczytać)

(1) całkowicie bezpiecznie	(3) zagrożony
(2) bezpiecznie	(4) czy bardzo zagrożony
	(8) trudno powiedzieć

08. [C7] Jak ocenilby P. stan swojego zdrowia? Proszę wziąć pod uwagę zarówno samopoczucie fizyczne, jak i psychiczne. Czy stan swojego zdrowia określiłby P. jako . . . (odczytać)

(1) bardzo dobry	(4) zły
(2) dobry	(5) czy też bardzo zły
(3) taki sobie	(8) trudno powiedzieć

09. [G] Gdyby dokonywał P. wyboru pracy, na ile ważna byłaby dla P. osobiście każda z następujących spraw? Odpowiadając proszę posłużyć się tą kartą. (wręczyć KARTEĘ 84)

	(1) w ogóle nie- ważna	(2) nieważna	(3) ani ważna, ani nieważna	(4) ważna	(5) bardzo ważna	(8) trudno powie- dzieć
(G11) pewność zatrudnienia	1	2	3	4	5	8
(G112) wysokie dochody	1	2	3	4	5	8
(G115) praca, która dawałaby możli- wość łączenia obowiązków zawodowych i zobowiązań rodzinnych	1	2	3	4	5	8

10. Czy na zakończenie rozmowy nasz współpracownik:

(a) sprawdził kompletność wypełnienia ankiety, tzn. przejrzał ją w P. obecności czy nie pominął jakiegos pytania?	(1) tak	(2) nie	(3) nie pamiętam
(b) poprosił P. o podanie, do celów kontrolnych, danych personalnych, tzn. imienia, nazwiska i numeru telefonu	(1) tak	(2) nie	(3) nie pamiętam
(c) podziękował za przeprowadzoną rozmowę	(1) tak	(2) nie	(3) nie pamiętam

11. Jak ocenia P. sposób prowadzenia rozmowy przez naszego współpracownika? Czy uważa P., że:

(a) zbyt szybko odczytywał on pytania?	(1) tak	(2) nie	(3) trudno powiedzieć
(b) odczytywał pytania niewyraźnie?	(1) tak	(2) nie	(3) trudno powiedzieć
(c) odczytywał pytania zbyt cicho?	(1) tak	(2) nie	(3) trudno powiedzieć
(d) dawał P. wystarczająco dużo czasu do namysłu?	(1) tak	(2) nie	(3) trudno powiedzieć
(e) sugerował P. odpowiedzi?	(1) tak	(2) nie	(3) trudno powiedzieć

12. A jak w ogóle ocenia P. tamto badanie? Czy było ono dla P. ciekawe czy nudne, łatwe czy męczące, potrzebne czy też nie? (dopytać – odpowiedź zapisać dosłownie)

.....

13. Czy przypomina P. sobie jak długo trwała wizyta naszego współpracownika? Proszę przypomnieć sobie co robił P. przed przyjściem naszego współpracownika i co robił P. po jego wyjściu i proszę podać przybliżony czas rozmowy.

minut

--	--	--

14. Czy nasz współpracownik (który z P. rozmawiał) był kobietą czy mężczyzną?

(1) mężczyzną	(2) kobietą	(3) były dwie osoby:	a) 2 mężczyzn	b) 2 kobiety	c) kobieta i mężczyzna
---------------	-------------	----------------------	---------------	--------------	------------------------

Nasz współpracownik zadawał P. różne pytania dotyczące P. gospodarstwa domowego. Służą one do zbadania, czy w opiniach i poglądach na badane tematy występują różnice między kobietami i mężczyznami, osobami w różnym wieku, o różnym wykształceniu itp. Pytania te są więc bardzo ważne i dlatego chciałbym teraz sprawdzić, czy nasz współpracownik dokładnie zapisał wszystkie te informacje.

15. [F1] Ile osób, wliczając dzieci, mieszka tu na stałe, tworząc wspólne gospodarstwo domowe? Proszę wliczyć również siebie.

Wpisać liczbę osób

--	--

(8) trudno powiedzieć

W poniższej tabeli zebrać dane o respondentach (tylko pyt. F2 i F3), a następnie informacje o pozostałych członkach gospodarstwa domowego (pyt. F2-F4), od najstarszych do najmłodszych.

16. [F2] *Zaznaczyć płć.*

17. [F3] Proszę podać jej/ jego rok urodzenia (8888=nie wiem)

18. [F4] Jaki jest stopień pokrewieństwa między P. a tą osobą?

osoba	pyt. F2 pleć		pyt. F3 rok urodzenia	Pyt. F4 symbol pokrewieństwa	pokrewieństwo do respondenta
	M	K			
RESPONDENT	M	K	I _ I _ I _ I _ I		
- Osoba nr 1	M	K	I _ I _ I _ I _ I	I _ I	1. współmałżonek/ partner
- Osoba nr 2	M	K	I _ I _ I _ I _ I	I _ I	2. syn/ córka (w tym dzieci przysposobione, adoptowane, wychowan- kowie, dzieci partnera)
- Osoba nr 3	M	K	I _ I _ I _ I _ I	I _ I	3. rodzic/ teść, teściowa, ojczym/ macocha
- Osoba nr 4	M	K	I _ I _ I _ I _ I	I _ I	
- Osoba nr 5	M	K	I _ I _ I _ I _ I	I _ I	4. brat/ siostra (w tym przyrodni, adoptowany lub wychowywany przez rodziców/ opiekunów)
- Osoba nr 6	M	K	I _ I _ I _ I _ I	I _ I	
- Osoba nr 7	M	K	I _ I _ I _ I _ I	I _ I	5. inny krewny (także dziadkowie)
- Osoba nr 8	M	K	I _ I _ I _ I _ I	I _ I	
- Osoba nr 9	M	K	I _ I _ I _ I _ I	I _ I	6. osoba niespokrew- niona
- Osoba nr 10	M	K	I _ I _ I _ I _ I	I _ I	
- Osoba nr 11	M	K	I _ I _ I _ I _ I	I _ I	88. nie wiem
- Osoba nr 12	M	K	I _ I _ I _ I _ I	I _ I	

19. [F6] Jakie ma P. wykształcenie? Chodzi o ukończoną przez P. szkołę najwyższego szczebla. Odpowiadając proszę posłużyć się kartą. (wręczyć KARTE 42)

(1) nieukończone podstawowe	(6) średnie zawodowe (technikum, liceum zawodowe, techniczne lub profilowane)
(2) ukończone podstawowe	(7) pomaturalne, policealne
(3) gimnazjalne	(8) licencjackie, inżynierskie
(4) zasadnicze zawodowe (także 2-letnia SPR)	(9) wyższe magisterskie lub lekarskie
(5) średnie ogólnokształcące	(10) inne – wpisać jakie?

20. [F8b] Które z określeń najlepiej opisuje P. sytuację. Proszę wybrać tylko jedną odpowiedź (wręczyć KARTE 43)

(1) wykonuję pracę odpłatną (praca najemna, na własny rachunek, w firmie rodzinnej, we własnym gospodarstwie rolnym) – także w przypadku tymczasowej przerwy w pracy	(6) jestem na emeryturze, rencie
(2) uczę się w szkole lub na uczelni (nauka nie jest opłacana przez pracodawcę) – także w przypadku wakacyjnej przerwy w nauce	(7) odbywam służbę wojskową lub służbę zastępczą

(3) jestem <u>bezrobotny</u> i aktywnie poszukuję pracy	(8) zajmuję się <u>domem</u> , opiekuję się <u>dziećmi</u> lub <u>innymi osobami</u>
(4) jestem <u>bezrobotny</u> , <u>nie</u> poszukuję aktywnie pracy, ale chciałbym pracować	(9) inna sytuacja
(5) <u>długotrwała choroba</u> lub <u>niepełnosprawność</u>	(10) trudno powiedzieć
<p>Dziękuję bardzo za udzielenie odpowiedzi na pytania. Jeśli ma P. jakieś uwagi lub spostrzeżenia związane z tym badaniem, którymi chciałby się P. podzielić, bardzo proszę mi je przekazać:</p>	
DZIĘKUJĘ BARDZO	
<i>Przejdź do Części 5</i>	
<p>Dodatkowe uwagi dotyczące przebiegu wywiadu:</p>	
Stwierdzam, że wywiad kontrolny przeprowadziłem osobiście	
imię i nazwisko kontrolera:	podpis kontrolera:
nr kontr. wywiadu I _ I _ II _ I _ I _ I _ I	data wywiadu I _ I _ I - I _ I _ I - 2004

Część 5 – Opis i charakterystyka sąsiedztwa

Należy samodzielnie wypełnić dla każdej wylosowanej osoby, niezależnie od tego, czy wywiad z nią został zrealizowany, czy też nie.

N1. W jakiego typu budynku respondent zamieszkuje?

	1. Wolnostojący budynek mieszkalny/ dom na wsi
Budownictwo jednorodzinne:	2. Dom wolnostojący
	3. Segment dwurodzinny (bliźniak)
	4. Dom w zabudowie szeregowej
	5. Pojedyncze mieszkanie w budownictwie o innym przeznaczeniu (na przykład w biurowcu, kompleksie szpitalnym, akademickim etc.)
Budownictwo wielorodzinne:	6. Mieszkanie w bloku, kamienicy
	7. Mieszkanie, pokoje wynajmowane (także w akademikach)
	8. Mieszkanie w dom opieki społecznej
Inne:	9. Przyczepa, łódka, jacht
	10. Inne:
	88. Trudno powiedzieć

N2. W jakim ogólnym stanie technicznym są budynki/ mieszkania w tej okolicy?

1. W bardzo dobrym stanie	4. W złym stanie
2. W dobrym stanie	5. W bardzo złym stanie
3. W średnim stanie	

N3. Czy wokół budynku, w którym mieszka wylosowana osoba widać leżące wokół odpadki lub śmieci?

1. Tak, bardzo dużo	3. Raczej mało
2. Dość dużo	4. Bardzo mało, nie ma ich

N4. Na ile widoczne są efekty celowego niszczenia budynków: napisy na ścianach (grafiti) potłuczone lub zniszczone oświetlenie budynku, latarnie, domofony, windy itp.?

1. Bardzo widoczne	3. Raczej mało widoczne
2. Dość widoczne	4. Bardzo mało widoczne, nie ma ich

Aneks 3.

1. Pytania o opinie

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
1. Opinie na tematy społeczno-polityczne					
Częstotliwość załatwiania sprawy „po znajomości” (skala 4 pkt.)	42,3	0,029	0,138	0,107	0,122
Częstotliwość załatwiania sprawy „jeśli się da” (skala 5 pkt.)	48,0	0,071	0,300	0,230**	0,250**
Częstotliwość występowania w Polsce korupcji (skala 5 pkt.)	52,6	0,086	0,314	0,295**	0,348**
Ocena działań państwa w zakresie zwalczania korupcji (5 pkt.)	43,8	0,060	0,176	0,128	0,099
Ocena słuszności idei prywatyzacji gospodarki polskiej (skala 4 pkt.)	39,4	0,035	0,177	0,167*	0,179*
Walka z ubóstwem jako cel władzy i rządu (skala 4 pkt.)	59,5	0,149	0,300	0,431**	0,420**
Opinie R. na temat wprowadzenia granicy dochodów	78,0	0,175	0,469	0,469**	0,469**
Przynależność R. do klasy społecznej	66,8	0,268	0,474	0,604**	d
średnia	53,8	0,109	0,294	e	e
2. Opinie na temat roli kobiet i mężczyzn					
Prowadzenie gospodarstwa jako obowiązek kobiety (skala 4 pkt.)	42,0	0,091	0,269	0,336**	0,394**
Zadowolenie z wykonywania czynności domowych (skala 4 pkt.)	44,0	0,064	0,234	0,209**	0,228**
Czynności domowe jako obowiązek (skala 4 pkt.)	48,2	0,033	0,152	0,105	0,106
Prace domowe jako forma wykorzystywania kobiet (skala 4 pkt.)	36,0	0,046	0,208	0,236*	0,300*
Zadowolenie R. z wykonywania prac domowych (skala 7 pkt.)	26,1	0,106	0,236	0,079	0,078
Prowadzenie gospodarstwa domowego to praca (skala 4 pkt.)	53,2	0,108	0,299	0,328**	0,276**

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
Postrzeganie roli R. w rodzinie i pracy, gdy R. wchodził w dorosłe życie	62,8	0,070	0,233	0,261**	d
średnia	44,6	0,074	0,233	e	e
3. Opinie na temat pracy zarobkowej					
Wykonywany zawód, gdy R. będzie miał 30 lat	73,1	0,916	0,886	d	d
Na ile ważna byłaby dla R. pewność zatrudnienia przy wyborze pracy (skala 5 pkt.)	48,0	0,029	0,116	0,082	0,038
Na ile ważne byłyby dla R. wysokie dochody przy wyborze pracy (skala 5 pkt.)	54,6	0,055	0,229	0,146	0,200*
Na ile ważna byłaby dla R. możliwość łączenia obowiązków rodzinnych i zawodowych przy wyborze pracy (skala 5 pkt.)	47,3	0,026	0,109	0,127	0,086
średnia	55,8	0,257	0,335	e	e
4. Opinie na temat poczucia bezpieczeństwa osobistego i ocena własnej sytuacji					
Poczucie bezpieczeństwa R. w swojej dzielnicy (skala 4 pkt.)	70,2	0,292	0,505	0,589**	0,659**
Ocena stanu zdrowia R. (skala 5 pkt.)	63,4	0,330	0,502	0,769** 0,541**	0,813** 0,586**
Ocena sytuacji finansowej GD. (skala 4 pkt.)	66,2	0,253	0,510	0,573**	0,622**
Poczucie bezpieczeństwa R. po zmroku w okolicach domu (skala 4 pkt.)	57,1	0,152	0,357	0,445**	0,477**
Zainteresowanie R. zakupem samochodu	93,5	0,179	0,399	d	d
Opinia na temat prowadzenia samochodu po spożyciu piwa	92,2	0,458	0,745	0,727**	d
średnia	73,8	0,277	0,503	e	e

2. Pytania o fakty

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
1. Posiadanie świadectwa ukończenia szkoły danego poziomu/ nabycia kwalifikacji					
Otrzymanie przez R. świadectwa ukończenia szkoły podstawowej	100	a	a	a	a
Otrzymanie przez R. świadectwa ukończenia średniej szkoły zawodowej	94,8	0,651	0,785	d	d
Otrzymanie przez R. świadectwa ukończenia liceum ogólnokształcącego	96,5	0,605	0,729	d	d
Otrzymanie przez R. świadectwa ukończenia szkoły danego poziomu	87,2	0,415	0,706	d	d
średnia	94,6	0,557	0,740	d	d
2. Sytuacja zawodowa respondenta i rodziców					
Praca R lub któregoś z domowników w branży	97,5	0,347	0,414	d	d
Sytuacja zawodowa R.	90,2	0,677	0,765	d	d
Stanowisko/zawód R.	74,6	0,932	0,931	d	d
Zawód matki, gdy respondent miał 14 lat	88,2	0,416	0,569	d	d
Zawód ojca, gdy respondent miał 14 lat	92,8	0,398	0,672	d	d
średnia	88,7	0,554	0,670	d	d
3. Posiadanie dóbr trwałego użytku i ich charakterystyka					
Posiadanie tel. komórkowego przez R.	90,6	0,460	0,730	d	d

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
Rodzaj tel. komórkowego R. (prywatna/służbowa)	93,8	0,440	0,693	d	d
Wyposażenie GD: telefon komórkowy	91,8	0,573	0,761	d	d
Rodzaj tel. komórkowego innej osoby w GD (prywatna/służbowa)	87,0	0,087	0,259	d	d
Wyposażenie GD.: telefon stacjonarny	96,0	0,726	0,805	d	d
Operatorzy telefonii stacjonarnej, z których korzysta GD R.	93,5	0,738	0,903	d	d
Wyposażenie GD.: automatyczna sekretarka	96,2	0,600	0,795	d	d
Wyposażenie GD: samochód	91,8	0,592	0,785	d	d
Wyposażenie GD.: Marka posiadanego samochodu	100	1,000	1,000	d	d
Wyposażenie GD.: model samochodu	98,6	b	b	b	b
Wyposażenie GD.: Samochód kupiony nowy/używany	95,7	0,743	0,912	d	d
Wyposażenie GD.: Rok dopuszczenia do ruchu samochodu/ 1 rejestracji	79,3	0,703	0,795	0,872**	0,875**
Wyposażenie GD.: Obecny przebieg samochodu	27,5	0,862	0,834	0,901**	0,974**
Wyposażenie GD.: Średni przebieg roczny samochodu	42,3	0,635	0,691	0,706**	0,791**
Wyposażenie GD.: aparat fotograficzny	86,2	0,412	0,715	d	d
Wyposażenie GD.: lodówka	97,8	0,171	0,344	d	d
Wyposażenie GD.: zamrażarka	85,7	0,404	0,708	d	d
Wyposażenie GD.: zmywarka do naczyń	97,9	0,473	0,681	d	d
Wyposażenie GD.: kuchenka mikrofalowa	93,9	0,568	0,795	d	d

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
Wyposażenie GD.: pralka automatyczna	91,9	0,415	0,681	d	d
Wyposażenie GD.: suszarka automatyczna do prania	97,3	0,583	0,753	d	d
Wyposażenie GD.: walkman	83,7	0,354	0,668	d	d
Wyposażenie GD.: odtwarzacz CD	86,0	0,339	0,633	d	d
Wyposażenie GD.: kamera wideo	95,0	0,493	0,729	d	d
Wyposażenie GD.: magnetowid	88,6	0,387	0,680	d	d
Wyposażenie GD.: wieża Hi-Fi	82,5	0,322	0,640	d	d
Wyposażenie GD.: komputer	92,8	0,559	0,798	d	d
Dostęp R. do internetu	85,8	0,397	0,697	d	d
Miejsce korzystania przez R. z internetu	91,9	0,705	0,816	d	d
Posiadanie pomocy domowej w niepełnym wymiarze godzin	96,7	0,384	0,613	d	d
Posiadanie pomocy domowej w pełnym wymiarze godzin	100	a	a	a	a
średnia	88,3	0,522	0,721	e	e
4. Charakterystyka sprzętu TV					
Liczba odbiorników TV w GD	86,0	0,523	0,754	0,686** 0,836**	0,728** 0,848**
Wyposażenie GD: TV kolorowy	97,0	0,002	0,013	d	d
Telewizor 1: Marka	92,1	b	b	b	b
Telewizor 1: Rodzaj (kolorowy-czarno-biały)	98,6	0,669	0,836	d	d

Miara	% odp. zgod.	Wsp. niepraw	V Cramera	τ Kendalla	r Pearsona
Telewizor 1: Typ (przenośny/stacjonarny)	92,4	0,388	0,659	d	d
Telewizor 1: Wyposażenie w pilota	98,3	0,696	0,856	d	d
Telewizor 1: Podłączenie magnetowidu	93,0	0,641	0,862	d	d
Rodzaj używanej anteny/anten	85,7	0,746	0,787	d	d
Możliwość odbioru ogólnopolskich stacji telewizyjnych ^f	37,5	b	b	b	b
Możliwość odbioru regionalnych stacji telewizyjnych ^f	81,0	b	b	b	b
Posiadanie anteny satelitarnej	97,2	0,716	0,892	d	d
Rodzaj anteny satelitarnej (ind/zbiorcza)	96,5	0,517	0,768	d	d
Typ indywidualnej anteny (z pozycjonerem/bez)	87,0	0,327	0,624	d	d
Posiadanie tunera satelitarnego	98,4	0,392	0,573	d	d
Podłączenie telewizora 1 do tunera satelitarnego	99,2	0,886	0,956	d	d
Posiadanie TV kablowej	96,6	0,779	0,928	d	d
Podłączenie telewizora 1 do telewizji kablowej	97,5	0,789	0,925	d	d
średnia	90,2	0,577	0,745	e	e
5. Stan zdrowia (dolegliwości) i przebyte choroby					
Stan zdrowia R.: zawał serca	99,5	0,839	0,923	d	d
Stan zdrowia R.: Choroba serca lub atak serca	91,6	0,401	0,669	d	d
Stan zdrowia R.: Nadciśnienie	90,7	0,458	0,732	d	d
Stan zdrowia R.: Perforacja przewodu pokarmowego	99,5	0,560	0,705	d	d

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
Stan zdrowia R.: Krwawienie z przewodu pokarmowego	95,8	0,384	0,611	d	d
Stan zdrowia R.: Choroba wrzodowa żołądka/ dwunastnicy	95,2	0,565	0,796	d	d
Stan zdrowia R.: Rozpoznana osteoporoza	95,8	0,471	0,736	d	d
Stan zdrowia R.: Wysoki poziom cholesterolu	90,5	0,390	0,673	d	d
Stan zdrowia R.: Choroba układu oddechowego	90,5	0,304	0,580	d	d
Stan zdrowia R.: Otyłość, nadwaga	86,1	0,298	0,598	d	d
Stan zdrowia R.: Uszkodzenie wątroby	93,9	0,291	0,542	d	d
Stan zdrowia R.: Rak odbytu	99,7	a	a	a	a
Stan zdrowia R.: Rak płuc	99,7	a	a	a	a
Stan zdrowia R.: Bóle odcinka szyjnego kręgosłupa (skala 6 pkt)	50,8	0,225	0,382	0,497**	0,579**
Stan zdrowia R.: Bóle odcinka piersiowego kręgosłupa (skala 6 pkt)	51,9	0,129	0,286	0,348**	0,429**
Stan zdrowia R.: Bóle odcinka lędźwiowego kręgosłupa (skala 6 pkt)	46,3	0,158	0,355	0,460**	0,544**
Stan zdrowia R.: Bóle stawu biodrowego (skala 6 pkt)	57,0	0,188	0,322	0,465**	0,549**
Stan zdrowia R.: Bóle kolan (skala 6 pkt)	48,9	0,173	0,322	0,454**	0,526**
Stan zdrowia R.: Bóle rąk (skala 6 pkt)	57,2	0,181	0,336	0,479**	0,563**
Stan zdrowia R.: Bóle innych stawów (skala 6 pkt)	64,4	0,139	0,268	0,393**	0,426**
Stan zdrowia R.: Bóle mięśni, ścięgien, więzadeł (skala 6 pkt)	35,3	0,094	0,239	0,341**	0,428**
Stan zdrowia R.: Bolesne obrzęki stawów rąk (skala 6 pkt)	67,2	0,180	0,304	0,393**	0,418**

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
Stan zdrowia R.: Bolesne obrzęki stawów stóp (skala 6 pkt)	63,9	0,204	0,330	0,493**	0,546**
Stan zdrowia R.: Bolesne obrzęki stawów kolan (skala 6 pkt)	68,2	0,229	0,359	0,484**	0,564**
Wizyta u lekarza z powodu dolegliwości reumatologicznych	81,3	0,262	0,576	d	d
średnia	76,8	0,310	0,506	e	e
6. Inne fakty					
Posiadanie własnego pokoju przez R., gdy miał 14 lat	80,2	0,257	0,578	d	d
Liczba uczniów w klasie R. w ostatnim roku szkoły podstawowej (z dokł. +/- 5)	94,0	0,924	0,911	0,838	0,800
Miejsce zamieszkania R., gdy uczęszczał do szkoły podstawowej	97,5	0,789	0,922	d	d
Okres zamieszkiwania przez R. danej w okolicy (+/- 1 rok)	78,0	0,839	0,816	0,787	0,757
Liczba papierosów palonych przez R. (z dokł. +/- 5)	74,7	0,752	0,803	0,664	0,746
Miejsce dokonywania większości zakupów żywnościowych w GD.	63,9	0,243	0,355	0,405	d
Miejsce spędzenia ostatnich wakacji przez R.	90,7	b	b	b	b
Odbyte podróże lotnicze R. w ciągu ostatnich 12 miesięcy	97,8	0,602	0,793	d	d
średnia	84,6	0,629	0,740	e	e
7. Opis nośnika reklamy zewnętrznej					
Identyfikacja nośnika reklamy zewnętrznej	87,3	c	c	c	c
Adres nośnika reklamy zewnętrznej (ulica)	89,7	c	c	c	c

Miara	% odp. zgod.	Wsp. niepew	V Cramera	τ Kendalla	r Pearsona
Adres nośnika reklamy zewnętrznej (numer budynku)	86,4	c	c	c	c
Liczba nośników reklamy zewnętrznej w grupie	87,0	c	c	c	c
Liczba nośników reklamy zewnętrznej wg rodzajów	97,4	c	c	c	c
Lokalizacja nośnika reklamy zewnętrznej (itp. parking, na budynku)	93,1	c	c	c	c
Infrastruktura komunikacyjna wokół nośnika reklamy zewnętrznej (parking, przystanek tramwajowy)	86,5	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: Banki (do 100 m)	99,1	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: Towarzystwa Ubezpieczeniowe (do 100 m)	99,4	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: stacje benzynowe (do 100 m)	99,4	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: obiekty użyteczności publicznej – biblioteki, kościoły, szkoła itp. (do 100 m)	96,7	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: obiekty użyteczności publicznej – biblioteki, kościoły, szkoła itp. (od 100 do 500 m)	95,1	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: hiper/supermarkety (do 100 m)	99,8	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: hipermarkety „Zrób To Sam” (do 100 m)	99,9	c	c	c	c
Otoczenie nośnika reklamy zewnętrznej: „Cash&Carry” (do 100 m)	99,8	c	c	c	c
średnia	94,4	c	c	c	c

3. Pytania o zachowania

Miara	% odp. zgod.	Wsp. niepraw.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
1. Zachowania „społeczne”					
Udział R. w referendum UE	96,4	0,692	0,848	d	d
Udział R. w wyborach parlamentarnych	94,9	0,611	0,828	d	d
Udział w badaniach R. lub któregoś z domowników kiedykolwiek/w ciągu ostatnich 6 miesięcy	92,3	0,278	0,476	d	d
średnia	94,5	0,527	0,717	d	d
2. Zachowania konsumpcyjne					
Konsumpcja przez R. kawy	85,9	0,339	0,576	d	d
Konsumpcja przez R. (od czasu do czasu) kawy naturalnej/mielonej	83,0	0,176	0,448	d	d
Konsumpcja przez R. (od czasu do czasu) kawy rozpuszczalnej	70,7	0,141	0,424	d	d
Konsumpcja przez R. (od czasu do czasu) kawy Cappuccino	69,5	0,105	0,368	d	d
Częstotliwość picia kawy mielonej przez R. (skala 10 pkt.)	47,1	0,262	0,370	0,559	d
Częstotliwość picia kawy rozpuszczalnej przez R. (skala 10 pkt.)	35,2	0,247	0,376	0,493	d
Częstotliwość picia Cappuccino przez R. (skala 10 pkt.)	25,2	0,213	0,328	0,323	d
Konsumpcja przez R. herbaty	96,8	0,336	0,561	d	d
Konsumpcja przez R. kiedykolwiek piwa	88,1	0,104	0,294	d	d
Konsumpcja przez R. piwa w okresie ostatnich 3 miesięcy/ w ogóle	85,5	0,389	0,678	d	d

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
Konsumpcja przez R. piwa w ciągu ostatniego miesiąca	91,5	0,281	0,549	d	d
Częstotliwość picia przez R. piwa (skala 3 pkt.)	67,3	0,091	0,282	0,265	d
Konsumpcja kiedykolwiek przez R. drinków alkoholowych	69,9	0,920	0,326	d	d
Konsumpcja kiedykolwiek przez R. ginu	70,9	0,163	0,429	d	d
Konsumpcja kiedykolwiek przez R. koniaku/winiaku	70,9	0,145	0,432	d	d
Konsumpcja kiedykolwiek przez R. likieru	66,0	0,132	0,400	d	d
Konsumpcja kiedykolwiek przez R. miodu pitnego	71,8	0,114	0,375	d	d
Konsumpcja przez R. kiedykolwiek whisky	68,0	0,142	0,408	d	d
Konsumpcja przez R. kiedykolwiek Tequili	91,3	0,447	0,719	d	d
Konsumpcja przez R. kiedykolwiek/w ogóle wódki bez dodatków	73,8	0,152	0,420	d	d
Konsumpcja przez R. kiedykolwiek innych napojów alkoholowych	94,2	0,195	0,412	d	d
Konsumpcja przez R. kiedykolwiek/w ogóle szampana, wina musującego	73,6	0,041	0,202	d	d
Konsumpcja przez R. wina	73,2	0,148	0,442	d	d
Konsumpcja przez R. kiedykolwiek wina zagranicznego	68,0	0,107	0,369	d	d
Konsumpcja przez R. kiedykolwiek wina krajowego owocowego	69,9	0,131	0,405	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: chipsy ziemniaczane/kukutydziane	70,1	0,033	0,192	d	d

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: chrupki i inne snacki	58,6	0,024	0,177	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: orzeszki	63,2	0,051	0,263	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: czekolady i batony czekoladowe	79,4	0,019	0,132	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: słodkie ciastka	62,7	0,018	0,151	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: słone ciastka	61,7	0,026	0,187	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: cukierki	61,4	0,022	0,170	d	d
Kupowanie i konsumpcja przez R. w ciągu ostatnich 3 miesięcy: lody pojedyncze i duże	68,5	0,083	0,331	d	d
Palenie przez R. wyrobów tytoniowych	95,2	0,722	0,904	d	d
Palenie papierosów przez R.	95,3	0,712	0,897	d	d
Palenie przez R. papierosów (obecnie)	90,6	0,495	0,762	d	d
Palenie papierosów przez R. ponad 6 miesięcy	94,2	0,677	0,881	d	d
Konsumpcja R.: jogurt	76,9	0,047	0,224	d	d
Konsumpcja R. od czasu do czasu: jogurt	85,6	0,231	0,515	d	d
Konsumpcja R. w ostatnim czasie: jogurt (skala 7 pkt.)	35,3	0,083	0,284	0,213	d
Konsumpcja R.: kefir	61,9	0,065	0,286	d	d

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
Konsumpcja R. od czasu do czasu: kefir	70,7	0,127	0,410	d	d
Konsumpcja R. w ostatnim czasie: kefir (skala 7 pkt.)	21,0	0,059	0,182	0,079	d
Konsumpcja R. od czasu do czasu: serek homogenizowany	73,6	0,096	0,344	d	d
Konsumpcja R. w ostatnim czasie: serek homogenizowany (skala 7 pkt.)	24,6	0,078	0,209	0,237	d
Konsumpcja R. od czasu do czasu: serek wiejski	71,6	0,094	0,348	d	d
Konsumpcja R. w ostatnim czasie: serek wiejski (skala 7 pkt.)	27,1	0,074	0,438	0,063	d
Konsumpcja R. od czasu do czasu: deser mleczny	71,6	0,042	0,222	d	d
Konsumpcja R. od czasu do czasu: napój mleczny	62,8	0,014	0,128	d	d
Dokonywanie przez R. zakupów żywnościowych (skala 3 pkt.)	74,1	0,343	0,547	0,700	d
średnia	69,3	0,195	7,349	e	d
3. Zachowania „edukacyjne”					
Uczęszczanie przez R. do przedszkola	83,6	0,398	0,688	d	d
Uczęszczanie przez R. do klasy zerowej	90,4	0,133	0,351	d	d
Uczęszczanie przez R. do szkoły podstawowej	100	a	a	a	a
Uczęszczanie przez R. do zasadniczej szkoły zawodowej	97,8	0,828	0,945	d	d
Uczęszczanie przez R. do średniej szkoły zawodowej	95,7	0,747	0,914	d	d
Uczęszczanie przez R. do liceum ogólnokształcącego	96,8	0,795	0,935	d	d
Uczęszczanie przez R. do szkoły policealnej/pomaturalnej	94,3	0,587	0,819	d	d

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
Uczęszczanie przez R. na studia licencjackie	90,3	0,356	0,634	d	d
Uczęszczanie przez R. na studia wyższe	91,0	0,551	0,810	d	d
średnia	93,3	0,549	0,762	d	d
4. Zachowania „kulturowe”					
Oglądanie regularnie TV – filmy pełnometrażowe	61,3	0,031	0,205	d	d
Oglądanie regularnie TV – wiadomości	75,8	0,094	0,337	d	d
Oglądanie regularnie TV f programy sportowe	81,7	0,291	0,603	d	d
Oglądanie regularnie TV – programy historyczno-naukowe	61,5	0,040	0,234	d	d
Oglądanie regularnie TV – programy kulturalne	79,7	0,042	0,202	d	d
Oglądanie regularnie TV – seriale filmowe	75,9	0,113	0,374	d	d
Oglądanie regularnie TV – komedie	64,5	0,061	0,287	d	d
Oglądanie regularnie TV – programy dla dzieci	86,1	0,079	0,273	d	d
Oglądanie regularnie TV – muzyczne/teledyski	73,8	0,108	0,368	d	d
Oglądanie regularnie TV – filmy rysunkowe	85,0	0,093	0,306	d	d
Oglądanie regularnie TV – programy religijne	85,0	0,039	0,179	d	d
Oglądanie regularnie TV – programy dla majsterkowiczów	84,0	0,042	0,196	d	d
Oglądanie regularnie TV – programy dramatyczne	77,0	0,094	0,335	d	d
średnia	76,3	0,087	0,300	d	d

Miara	% odp. zgod.	Wsp. niepew.	V-Cramer 'a	τ Kendall 'a	r-Pearson 'a
5. Inne zachowania					
Wyjazdy wakacyjne/urlop R. w ciągu ostatnich 12 miesięcy	81,7	0,254	0,565	d	d
Wyjazdy R. co najmniej na 2 dni w ciągu ostatnich 3 miesięcy (wyjazdy krajowe i zagraniczne)	90,1	0,466	0,589	0,751	d
Przemieszczanie się R. poza najbliższą okolicę danego dnia	98,3	e	e	d	d
Godzina (pora) wyjścia R. z domu danego dnia	88,7	e	e	e	E
Korzystanie z internetu przez R	88,7	0,298	0,574	d	d
Częstotliwość korzystania przez R. z internetu (skala 7 pkt)	62,4	0,267	0,454	0,515	d
Wykonywanie w GD. Połączeń z tel. stacjonarnego na tel. komórkowe	72,0	0,111	0,358	d	d
Korzystanie przez R. z automatów tel. w miejscach publicznych	58,5	0,080	0,305	d	d
średnia	80,1	0,246	0,474	e	d

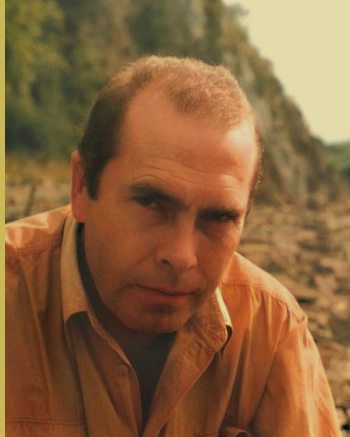
4. Pytania o cechy

Miara	% odp. zgod	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
1. Cechy respondenta					
Liczba lat nauki R.	70,8	0,681	0,746	0,902	0,961
Płeć R	99,9	0,990	0,996	d	d
Dochód R.	65,3	0,489	0,567	0,801	0,852

Miara	% odp. zgod	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Stan cywilny R	97,3	0,849	0,924	d	d
Waga R.	50,0	0,752	0,699	0,899	0,965
Wiek R. Z dokładnością do 1 roku	96,3	0,973	0,970	0,983	0,985
Wykształcenie R.	80,8	0,659	0,731	0,857	d
Wzrost R. (z dokł. +/- 5 cm)	85,0	0,955	0,965	0,934	0,947
Posiadanie przez R. rodzeństwa	99,3	0,917	0,972	d	d
Posiadanie znajomych wśród imigrantów	78,1	0,160	0,327	0,439	d
Poczucie R. związku z religią	93,5	0,274	0,525	d	d
Związek R. z religią	59,4	0,102	0,273	d	d
średnia	81,3	0,650	0,725	e	e
2. Cechy innych osób/gospodarstwa domowego					
Głowa gospodarstwa	83,6	0,358	0,668	d	d
Dochód GD	15,9	0,487	0,623	0,684	-0,089
Liczba dzieci w GD do lat 18	90,4	0,716	0,870	0,887	0,909
Liczba dzieci w GD do lat 3	96,5	0,788	0,862	0,935	0,928
Liczba dzieci w GD w wieku 4–7 lat	92,6	0,671	0,850	0,881	0,882
Liczba dzieci w GD w wieku 8–15 lat	90,8	0,717	0,855	0,909	0,933
Liczba dzieci w GD w wieku 16–18 lat	94,4	0,701	0,768	0,900	0,883
Liczba osób wspólnie zamieszkujących z R.	82,5	0,630	0,728	0,819	0,840

Miara	% odp. zgod	Wsp. niepew.	V Cramera	τ Kendalla	r Pearsona
Liczba osób w GD (z dokładnością do 1 osoby)	84,0	0,669	0,804	0,821	0,829
Liczba osób pełnoletnich w GD z dokładnością do 1 osoby	89,5	0,709	0,816	0,879	0,873
Płeć głównego użytkownika samochodu	93,1	0,558	0,804	d	d
Dzieci do lat 14 w GD (posiadanie)	94,3	0,682	0,884	d	d
Dzieci do lat 18 w GD (posiadanie)	94,9	0,681	0,879	d	d
Wiek głównego użytkownika samochodu (z dokładnością do 1 roku)	90,3	0,929	0,910	0,884	0,913
Wykształcenie głównego użytkownika samochodu	93,5	0,733	0,862	0,919	d
Średni miesięczny rachunek za telefon stacjonarny w GD. (z dokładnością +/-20 zł)	74,1	0,752	0,805	0,778	0,579
średnia	86,9	0,674	0,812	e	e

- a nie zostały wyliczone żadne miary siły związku (jedna ze zmiennych jest stała)
- b dane wprowadzone jako wynik porównania odpowiedzi z wywiadu zasadniczego i kontrolnego (zgodność/ niezgodność)
- c brak danych źródłowych (zbioru) - dane z raportu
- d nie obliczono wartości współczynnika, ze względu na jego nieadekwatność do poziom pomiaru
średnia wartość dla współczynnika nie jest liczona, ze względu jego nie addytywność (średnia wartość współczynnika z wielu prób
- e nie jest równa „średniej korelacji” we wszystkich próbach)
- f jako niezgodność odpowiedzi traktowano niezgodność wskazania choćby 1 stacji telewizyjnej
- * korelacja istotna na poziomie 0,05 (test dwustronny)
- ** korelacja istotna na poziomie 0,01 (test dwustronny)



Franciszek Sztabiński

Doktor socjologii, adiunkt w Instytucie Filozofii i Socjologii PAN. Założyciel i członek redakcji czasopisma „ASK. Research and Methods”. Zajmuje się problemami oceny wartości danych ilościowych, w szczególności ich związku z fazą realizacji terenowej badania. Wykładowca w Szkole Nauk Społecznych w IFiS PAN oraz Collegium Civitas, wcześniej w Instytucie Socjologii UW i Podplomowym Studium Zachowań Rynkowych w Instytucie Psychologii UW. Konsultant metodologiczny licznych projektów badawczych krajowych i międzynarodowych, w tym dla: IFiS PAN, INP PAN, Uniwersytetu Warszawskiego, Uniwersytetu Jagiellońskiego, Szkoły Głównej Handlowej, Centrum Onkologii, Centrum Kardiologii, Instytutu Reumatologicznego, Państwowego Funduszu Rehabilitacji Osób Niepełnosprawnych, Ministerstwa Edukacji Narodowej, Urzędu Miasta Stołecznego Warszawy oraz Ohio State University, University of Michigan, University of Kalifornia, Catholic University of Leuven, City University London i Centre D'Etude Des Supports De Publicity (Paris). We wcześniejszych latach pełnił funkcję kierownika samodzielnych działów w CBOS oraz w Instytucie SMG/KRC Poland Media S.A. Od roku 2001 członek polskiego zespołu w projekcie European Social Survey (ESS). Autor licznych publikacji metodologicznych poświęconych problemom badań surveyowych oraz oceny uzyskiwanych rezultatów, w tym książki „Ankieta pocztowa i wywiad kwestionariuszowy”, współredaktor i współautor trzech podręczników badań ankietowych oraz książki „Nowe metody, nowe podejścia badawcze w naukach społecznych”. Członek European Survey Research Association (od 2005 – daty założenia).

Jest to praca z zakresu metodologii badań społecznych o nachyleniu ilościowym, której głównym celem jest identyfikacja błędów rzetelności pomiaru w najbardziej rozpowszechnionym rozumieniu rzetelności, czyli dostarczania przez narzędzie tych samych wartości mierzonych zjawisk, niezależne od zmieniających się uwarunkowań zewnętrznych. Zaslugą autora było uzyskanie unikalnych danych z kilkudziesięciu badań zawierających informacje o powtórzonem pomiarze w odniesieniu do tych samych zmiennych (pytań kwestionariuszowych - postaw i cech), umożliwiających przeprowadzenie kompleksowej analizy rzetelności, jakiej nikt jeszcze w polskich naukach społecznych nie zrobił. Analiza empiryczna (. . .) poprzedzona jest systematycznym omówieniem rodzajów błędów występujących w badaniach (. . .) w świetle najnowszych analiz, co jest oryginalnym przedsięwzięciem i stanowi niewątpliwie drugą zaletę tej pracy. Trzecią zaletą jest unaocznienie, w jakiej postaci błędy te występują w praktyce badawczej (. . .). Jeżeli chodzi o merytoryczny wkład do metodologii badań społecznych, to osiągnięciem autora jest pokazanie, na ile rzetelność pomiaru zależy od tematyki (objektu) badania (np. od rodzaju zmiennych: świadomościowe, czy pytania o fakty) i od struktury narzędzia pomiaru (pytania zamknięte, otwarte itd.).

Z recenzji prof. dr. hab. Henryka Domańskiego

ISBN 978-83-7683-037-7



9 788376 830377 >