

Uniwersytet im. Adama Mickiewicza w Poznaniu  
Wydział Socjologii

**Praktyki sondażowe na świecie i w Polsce: dokumentacja,  
reprezentatywność, preferencja techniki**

**Survey practices globally and locally: documentation, representativeness,  
mode preference**

Adam Rybak

Promotor: dr hab. Piotr Jabkowski, prof. UAM

Promotor pomocniczy: dr Piotr Cichocki

Poznań 2025



## Spis Treści

|                                                                                                  |     |
|--------------------------------------------------------------------------------------------------|-----|
| Wykaz prac naukowych wchodzących w skład cyklu                                                   | 5   |
| Streszczenie                                                                                     | 6   |
| Abstract                                                                                         | 7   |
| Autoreferat                                                                                      | 8   |
| I. Wstęp                                                                                         | 8   |
| II. Problemy i metody badawcze                                                                   | 10  |
| II.1. Cel pracy, problemy i pytania badawcze                                                     | 10  |
| II.2. Problemy badawcze na tle literatury przedmiotu                                             | 13  |
| II.2.1. Wpływ charakterystyk sondażu na wskaźnik realizacji próby oraz poziom reprezentatywności | 14  |
| II.2.1.1. Procedury terenowe                                                                     | 14  |
| II.2.1.2. Operaty losowania i typy prób badawczych                                               | 15  |
| II.2.1.3. Techniki sondażu                                                                       | 16  |
| II.2.2. Preferencja techniki sondażu w schematach <i>mixed-mode</i>                              | 16  |
| II.2.3. Miary jakości pomiaru sondażowego                                                        | 18  |
| II.2.3.1. Wskaźniki realizacji badania                                                           | 18  |
| II.2.3.2. Miary reprezentatywności próby                                                         | 19  |
| II.3. Metody zbierania i analizy danych                                                          | 20  |
| II.3.1. Metoda metaanalizy                                                                       | 21  |
| II.3.2. Badanie preferencji techniki sondażu wśród respondentów                                  | 22  |
| III. Dane z badań własnych i wyniki analiz                                                       | 22  |
| III.1. Inwentaryzacja charakterystyk sondaży z programów porównawczych                           | 22  |
| III.1.1. Zakres geograficzny danych objętych inwentaryzacją                                      | 24  |
| III.1.2. Dokumentacja                                                                            | 25  |
| III.1.3. Technika sondażu                                                                        | 27  |
| III.1.4. Typ operatu i schemat losowania                                                         | 31  |
| III.1.5. Technika sondażu a współczynniki realizacji                                             | 32  |
| III.1.6. Miary reprezentatywności – metaanalizy                                                  | 33  |
| III.1.7. Reprezentatywność próby dla poszczególnych programów                                    | 34  |
| III.2. Kluczowe wyniki z publikacji                                                              | 35  |
| III.2.1. Dokumentacja sondaży                                                                    | 35  |
| III.2.2. Reprezentatywność sondaży                                                               | 38  |
| III.2.3. Preferencja techniki sondażu                                                            | 40  |
| IV. Dyskusja                                                                                     | 42  |
| V. Podsumowanie i wnioski                                                                        | 46  |
| Literatura                                                                                       | 48  |
| Spis adresów repozytoriów wszystkich analizowanych programów                                     | 60  |
| Publikacje z cyklu                                                                               | 62  |
| Aneks                                                                                            | 143 |
| 1. Charakterystyki sondaży z bazy cd.                                                            | 143 |
| 1.1. Wiek włączający do próby                                                                    | 143 |

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| 1.2. Dobór wewnątrzgrupowy                                          | 143 |
| 1.3. Stosowanie doboru kwotowego i substytucji nierespondentów      | 144 |
| 1.4. Kwestie relacji z respondentem                                 | 144 |
| 1.5. Praca ankierów                                                 | 146 |
| 1.6. Kontrola wywiadów                                              | 147 |
| 1.7. Współczynniki realizacji próby                                 | 148 |
| 2. Metaanalizy cd.                                                  | 152 |
| 2.1. Metaanalizy w podgrupach cd.                                   | 152 |
| 2.2. Wpływ fieldwork efforts na reprezentatywność – metaanaliza     | 154 |
| 3. Zmienne z inwentaryzacji i ich rozkłady                          | 155 |
| Oświadczenie autora o wkładzie w przygotowanie publikacji naukowych | 160 |

Barring one or two exceptions, the problems that occur in surveys are opportunities for understanding once we take them seriously as facts of life.  
(Schuman 1982)

## **Wykaz prac naukowych wchodzących w skład cyklu**

**Publikacja 1:** Rybak, Adam. 2025. „Analiza jakości dokumentacji metodologicznej sondaży: Studium przypadku inwentaryzacji 30 programów badań porównawczych z całego świata”. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 87(1):287–309.

doi: 10.14746/rpeis.2025.87.1.15.

**Pkt. MN 2025: 70**

**Publikacja 2:** Rybak, Adam. 2023. „Survey Mode and Nonresponse Bias: A Meta-Analysis Based on the Data from the International Social Survey Programme Waves 1996–2018 and the European Social Survey Rounds 1 to 9”. *PLOS ONE* 18(3): e0283092.

doi: 10.1371/journal.pone.0283092.

**Pkt. MN 2023: 140; IF 2023: 2,9**

**Publikacja 3:** Rybak, Adam. 2019. „Komunikacja oparta na kanale wizualnym i audialnym a jakość pomiaru sondażowego. W stronę nowego ujęcia hipotezy preferencji techniki przy wykorzystaniu koncepcji pamięci roboczej”. *Przegląd Socjologiczny* 68(3):207–27.

doi: 10.26485/ps/2019/68.3/8.

**Pkt. MN 2019: 70**

**Publikacja 4:** Rybak, Adam. 2020. „Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference”. *Polish Sociological Review* 210(2):219–28.

doi: 10.26412/psr210.05.

**Pkt. MN 2020: 70; IF 2020: 0,28**

## Streszczenie

Badania sondażowe znalazły się w historycznie szczególnej sytuacji. Z jednej strony obserwujemy ich ilościowy rozwój, a kolejne programy obejmują coraz bardziej odległe zakątki świata. Z drugiej strony można dostrzec podwójny kryzys. Po pierwsze badania oparte na Big Data oferują analizy o coraz wyższej jakości, które mogą stanowić w pewnych obszarach alternatywę dla sondaży. Po drugie pandemia COVID-19 wywołała kryzys w przeprowadzaniu badań sondażowych, znacząco ograniczając stosowanie zdecydowanie dominującej w obszarze społeczno-naukowym techniki wywiadu kwestionariuszowego. W krótkim okresie uniemożliwiło to realizację wywiadów w domu respondentów, a w dłuższej perspektywie poważnie osłabiło funkcjonowanie sieci ankierskich. Ta sytuacja spowodowała przyspieszenie wdrażania alternatywnych technik – nawet instytucje silnie przywiązane do wywiadu kwestionariuszowego zmieniły swoje podejście. Zasygnalizowany problem sugeruje zasadność podjęcia refleksji metodologicznej i dokonania podsumowujących analiz, które mogą lepiej ukierunkować przyszły rozwój badań sondażowych.

Podstawą mojej pracy doktorskiej jest seria czterech publikacji. Dwie pierwsze były wynikiem inwentaryzacji praktyk sondażowych z programów porównawczych, które obejmowały wszystkie zamieszkane kontynenty i zostały przeprowadzone w ciągu ostatnich 65 lat. Omawiają one praktyki dokumentacyjne, a także prezentują analizy wpływu charakterystyk sondaży na reprezentatywność. Dwie kolejne prace dotyczą problematyki preferencji techniki sondażu, jedna z nich ma charakter teoretyczny, a druga bazuje na dwóch badaniach przeprowadzonych w Polsce. Choć dwie pierwsze prace nie omawiają *explicite* tematyki zróżnicowania praktyk sondażowych na świecie, to jednak ich inwentaryzacja była warunkiem *sine qua non* przeprowadzenia analiz, dlatego temat ten znalazł swoje miejsce w poprzedzającym publikację autoreferacie.

**Publikacja 1** łączy opis zmiennych dotyczących kompletności dokumentacji metodologicznej sondaży z programów porównawczych z jakościową oceną formy raportów i refleksją nad kierunkami ewentualnej standaryzacji tych dokumentów. **Publikacja 2** używa modeli metaregresji, by ocenić wpływ charakterystyk (operatu, techniki, programu, doświadczenia instytucjonalnego) badań na ich reprezentatywność mierzoną za pomocą kryterium wewnętrznego, opartego wyłącznie na strukturze próby. **Publikacja 3** ma charakter teoretyczny i poświęcona została próbie integracji hipotezy o preferencji techniki sondażu z kognitywistycznym modelem pamięci roboczej. **Publikacja 4** natomiast wykorzystuje model wielomianowej regresji logistycznej, bazującej na pochodzących od polskich respondentów danych na temat preferencji techniki do zbadania czynników kształtujących jej zróżnicowanie.

Najistotniejszymi wnioskami z całości analiz wydają się następujące kwestie: 1) Potrzeba wypracowania jednolitego standardu dokumentacji metodologicznej pozwalającego na automatyzację procesu zbierania danych metodologicznych. 2) Znaczenie dostępności operatu imiennego w sytuacji stopniowego odchodzenia od wywiadów kwestionariuszowych. 3) Potencjał sondaży *mixed-mode* do uzyskiwania prób o satysfakcjonującym poziomie reprezentatywności. 4) Zasadność badania źródeł i korelatów preferencji techniki sondażu w celu tworzenia modeli predykcyjnych pozwalających na planowanie przyszłych badań *mixed-mode* w schemacie równoległym w poszczególnych krajach.

**Słowa kluczowe:** metodologia badań sondażowych, dokumentacja sondażowa, reprezentatywność próby, preferencja techniki, sondaże mixed-mode

## Abstract

Survey research is facing a historically unprecedented situation. On the one hand, quantitative growth continues, and additional programs cover more and more distant corners of the world. On the other hand, a double crisis can be attested. Research based on Big Data offers analyses of increasingly better quality and can potentially push surveys out from some areas. The Covid-19 pandemic caused a crisis in conducting research in face-to-face mode, which is dominant in the social and academic sphere – both making it impossible in the short term and threatening the existence of interviewer networks in the long term. This has accelerated the implementation of alternative modes – even institutions strongly attached to face-to-face, such as the European Social Survey or the Centre for Public Opinion Research (PL), have changed their approach. This suggests the relevance of looking back and making summary analyses to better guide future developments.

My thesis is based on a series of four articles. The first two were the result of an inventory of survey practices from comparative programs carried out in the last 65 years. The data covers all inhabited continents. The first article uses the whole dataset to discuss documentation practices. The second deals with the impact of survey characteristics on the representativeness and uses a selected part of the dataset. Two other papers concern the issue of mode preference, the first in a theoretical way, and the other is based on two surveys conducted in Poland. These surveys are covered by the inventory. Papers do not explicitly discuss the subject of survey practices, but because their examination was a necessary condition for conducting analyses – this topic found its place in the self-report preceding the publications.

The **first publication** combines a quantitative analysis of the completeness of survey methodological reports with a qualitative assessment of their form and a reflection on the possible standardization of these documents. The **second publication** uses three-level cross-nested metaregression models to assess the impact of the survey characteristics (frame, mode, program, institutional experience) on their representativeness, measured by internal criteria based solely on the sample structure. The **third publication** is theoretical and concerns an attempt to integrate the hypothesis of mode preference with the cognitive model of working memory. The **fourth publication** contains a model of multinomial logistic regression based on declarative and behavioral data on mode preference of Polish respondents.

The most important conclusions from the entire analysis are the following issues. 1) The need to create a uniform standard of methodological documentation with fixed headings, which would enable the automation of collecting methodological data. 2) The key importance of the availability of an individual sampling frame in the situation of departure from face-to-face surveys. 3) High potential of mixed-mode surveys to obtain robust sample representativeness. 4) Importance of studying local and global predictors of mode preference to create predictive models for planning future parallel mixed-mode surveys.

**Keywords:** survey methodology, survey documentation, sample representativeness, mixed-mode surveys, survey mode preference

## Autoreferat

### I. Wstęp

Historia szeroko pojętych badań sondażowych jest przez niektórych autorów wyprowadzana od spisów powszechnych w Cesarstwie Rzymskim. Pierwszy zachowany „kwestionariusz” pochodzi z czasów Karola Wielkiego. Początki badań o charakterze bardziej naukowym mają zaś przypadać na wiek XVIII (Frankovic 2016; de Leeuw i Hox 2011). Z sondażami w rozumieniu wąskim, a więc z badaniami kwestionariuszowymi na próbie reprezentatywnej, mamy do czynienia jednak dopiero od pierwszej połowy XX wieku (Groves 2011; Neyman 1934). Początkowo nowa metoda badań wzbudzała na tyle duże zaniepokojenie, że uzyskanie wysokich wskaźników realizacji próby (czyli odsetka osób wylosowanych do badania, od których uzyskano informacje) nie sprawiało większego problemu. Ten stan rzeczy uległ jednak szybkiej zmianie i już 30 lat później badacze zwrócili większą uwagę na zagrożenie, jakie dla jakości pomiaru stanowią zmniejszające się odsetki kontaktów i kooperacji (Stoop 2005). Wraz z nastaniem kolejnych dekad zaczęto odchodzić od traktowania współczynnika realizacji próby (*response rate*) jako głównej miary jakości badania na rzecz analizy struktury samej próby (Bianchi i in. 2019; Hendra i Hill 2019; Roberts, Vandeplass i Herzing 2020; Schouten, Cobben, i Bethlehem 2009).

Badacze wyróżniają zazwyczaj około ośmiu podstawowych kategorii błędów sondażowych, składających się na tzw. całkowity błąd sondażowy (*Total Survey Error* (TSE) – jest to zarówno określenie błędu, jak i dominującego w metodologii sondażowej paradygmatu) (Biemer i Lyberg 2003; Groves i Lyberg 2010; Weisberg 2005). Zazwyczaj jednak to błąd wynikający z braków odpowiedzi oraz błąd pomiaru uznawane są za główne źródło całkowitego błędu sondażowego (Jabkowski 2015). Ponieważ badania sondażowe odgrywają wciąż istotną rolę w wielu dziedzinach, włączając socjologię (Davino i Fabbris 2013; Tepper i in. 2022), nawet pomimo rosnącego udziału analiz opartych o Big Data oferujących częściową alternatywę dla sondaży (Couper 2013), konieczny wydaje się wysiłek ukierunkowany na minimalizowanie wskazanych błędów. Zmagania z błędem braku odpowiedzi polegały w przeszłości głównie na podnoszeniu (lub bardziej realistycznie – utrzymywaniu na stałym poziomie) wskaźników realizacji – współcześnie odchodzi się coraz bardziej od tego jednowymiarowego podejścia. Optymalizacja struktury próby staje się głównym celem.

Przeprowadzane już w XXI wieku analizy dotyczące związku pomiędzy wskaźnikami realizacji prób badawczych oraz wielkością błędów mających swoje źródło w unikaniu przez respondentów udziału w badaniu nie przyniosły jednoznacznych wyników. Przeważać zdają się analizy ukazujące, że związek ten jest bardzo niewielki (Beullens i Loosveldt 2012; Groves 2006; Groves i Peytcheva 2008; Kohler 2007), można jednak spotkać się z wynikami sugerującymi istotną zależność (Cornesse i Bosnjak 2018; Jabkowski i Cichocki 2019). Poziom uzyskiwanego wskaźnika realizacji próby wyznacza przy tym górną granicę błędu, jaki może się pojawić w konsekwencji braków odpowiedzi. Poświęcanie mniejszej uwagi wskaźnikom realizacji jest istotne w przypadku badań przeprowadzanych innymi technikami niż tradycyjny wywiad kwestionariuszowy. Sondaże *mixed-mode*, które zazwyczaj notują niższy *response rate*, ale zdaniem części badaczy wykazują potencjał do uzyskiwania zbliżonej lub nawet większej reprezentatywności, są szczególnie interesującym przypadkiem (Cornesse i Bosnjak 2018; Fowler i in. 2002; Luiten i Schouten 2013; Olson, Smyth i Wood 2012). Konieczność prowadzenia analiz metodologicznych w tym obszarze jest dodatkowo wzmacniana przez fakt odchodzenia od wywiadów kwestionariuszowych z przyczyn finansowych – w krajach

rozwinętych wzrost kosztów tradycyjnych sondaży (Villard i Fitzgerald 2017) zmusza instytucje badawcze do szukania nowych rozwiązań. Dotyczy to nawet wielkich projektów badawczych o stabilnym finansowaniu (ESS ERIC 2024). Proces ten przyspieszyła także pandemia COVID-19, która doprowadziła do załamania profesjonalnych siatek ankierskich w niektórych krajach.

Wszelkie działania mające na celu ograniczenie negatywnych konsekwencji błędu wynikającego z braku odpowiedzi powinny opierać się na gruntownych analizach metodologicznych, ukierunkowanych na ocenę efektów wdrażania poszczególnych procedur sondażowych i pozwalających na zidentyfikowanie skutków wynikających z podejmowania przez badaczy konkretnych decyzji. Rezultaty takich badań mogą doprowadzić do opracowania planów służących (co najmniej) spowolnieniu negatywnych trendów lub zniwelowaniu ich skutków poprzez wprowadzanie różnych rozwiązań adaptacyjnych (Tourangeau i in. 2016).

Należy jednak mieć na uwadze, że wysiłki na rzecz zwiększania reprezentatywności badania poprzez dążenie do redukcji błędów braków odpowiedzi mogą wprowadzać inne zagrożenia. Warto na przykład postawić pytanie, czy respondenci, których trudno było przekonać do udziału w badaniu (np. konieczne były wielokrotne kontakty lub dodatkowe zachęty materialne), mogą być źródłem danych gorszej jakości. Przyczyn takiego ryzyka można upatrywać w ich niskiej motywacji do współpracy, słabych zdolnościach poznawczych czy też po prostu w chęci szybkiego pozbycia się „natrętnego” ankietera. Stosowano różne metody badania tego zjawiska, m.in. grupom respondentów przekazywano zachęty materialne o zróżnicowanej wartości, by ustalić wpływ takiej interwencji na jakość odpowiedzi, albo *post factum* porównywano odpowiedzi tych respondentów, do których dotarcie wymagało większego wysiłku. Można przywołać badania wskazujące, że zwiększona intensywność zachęt i ponagleń nie tylko nie zwiększyła błędów pomiaru, a wręcz go zmniejszyła (Grauenhorst, Blohm i Koch 2016; Peytchev, Peytcheva i Groves 2010; Willimack i in. 1995), ale zarazem na jednoznacznie gorszy pomiar wskazuje drugi zbiór badań (Biemer 2001; Cannell i Fowler 1963; Wagner i in. 2017). Natomiast wg zespołów Daverna (2003) czy Hoxa (Hox, de Leeuw i Chang 2012) związku takiego nie udało się ustalić. Jeszcze inna grupa badań (Kaminska, McCutcheon i Billiet 2010; Olson 2006) dostarczyła wyników wskazujących na wciąż liczne niejasności w obrębie tego zagadnienia. Bardziej pogłębiona analiza może opierać się na odwołaniu do psychologicznych teorii procesu odpowiadania na pytania sondażowe, ze szczególnym uwzględnieniem powiązanej z motywacją respondentów koncepcji *satisficingu* (Krosnick 1991).

W literaturze metodologicznej odnaleźć można przykłady wielu analiz podejmujących problematykę relacji pomiędzy różnymi charakterystykami badań sondażowych a jakością pomiaru. Opierają się one w większości na zbiorczej metaanalizie rezultatów badań opublikowanych w artykułach naukowych, analizie jakości jednego sondażu, porównaniu jakości wielu sondaży z różnych fal jednego projektu oraz porównaniu kilku sondaży z niewielkiej liczby rund różnych projektów. Niedawno opublikowany przegląd systematyczny metaanaliz w metodologii sondażowej (Cehovin, Bosnjak i Lozar Manfreda 2018) poszerzył wiedzę o tym, w jaki sposób badane są związki pomiędzy charakterystykami sondaży a różnymi rodzajami błędów. Część zawartych w cytowanym zestawieniu prac została uwzględniona w przeglądzie literatury zamieszczonym w dalszej części autoreferatu.

Choć w metodologii sondażowej dominujące były do niedawna głównie badania fragmentaryczne, to w ostatnich latach pojawiły się w Polsce projekty obejmujące zbieranie danych metodologicznych na dużą skalę, którymi praca ta jest częściowo inspirowana. Pierwszym z tych badań jest projekt zespołu Kazimierza M. Słomczyńskiego – *Survey Data Recycling* (SDR), którego efektem były liczne publikacje (Dubrow i in. 2016; Jenkins i Kwak

2022; Kwak i in. 2020; Słomczyński, Powalko i Krauze 2017). Drugim badaniem była natomiast inwentaryzacja projektów sondażowych, dokonana przez Piotra Jabkowskiego, której wynikiem były publikacje dotyczące jakości dokumentacji i praktyk terenowych w europejskich programach sondażowych (Jabkowski 2019, 2023; Jabkowski i Cichocki 2024; Jabkowski, Cichocki i Kołczyńska 2021; Jabkowski i Kołczyńska 2020).

Jednym ze znaczących osiągnięć obu projektów było stworzenie względnie dużych baz danych charakterystyk sondaży z programów międzykrajowych (odpowiednio 1721 i 1537 obserwacji). SDR inwentaryzował jednak jedynie niezbędne minimum informacji o charakterystykach badań. Ocenę jakości sondaży opierał zaś głównie na analizie jakości kodowania baz danych i ich zgodności z dokumentacją. Drugi projekt obejmował z kolei wyłącznie 5 największych sondażowych projektów europejskich w okresie od 1985 roku. Dwie pierwsze publikacje wchodzące w skład przedstawionego przeze mnie cyklu bazują na szerokim projekcie inwentaryzacji dokumentacji metodologicznej z 30 wybranych porównawczych międzykrajowych programów sondażowych, którego efektem było zebranie informacji o charakterystykach 10493 sondaży z całego świata. Projekt korzystał zarówno z dokumentacji badań, jak i baz danych wynikowych. Jego przeprowadzenie było możliwe dzięki wsparciu z grantu PRELUDIUM-17 (NC2019/33/N/HS6/00322).

Trzeba zaznaczyć, że aktualnie zainteresowanie tematem harmonizacji danych sondażowych zaczyna przeżywać rozkwit, o czym świadczą zwłaszcza niedawne prace polskich autorów, np. Słomczyńskiego (Słomczyński, Tomescu-Dubrow i Wysmulek 2022) i Wysmulek (Wysmulek, Tomescu-Dubrow i Kwak 2022) z zespołami, Kołczyńskiej (2022; Kołczyńska, Bürkner i in. 2024), Kiszkiela i współautorów (2024), a także ważne opracowanie zbiorowe poświęcone temu tematowi (Tomescu-Dubrow i in. 2024).

## **II. Problemy i metody badawcze**

### **II.1. Cel pracy, problemy i pytania badawcze**

Zagadnienia, które podejmuję w pracy doktorskiej, to zarówno jakość dokumentacji sondażowej, zróżnicowanie praktyk sondażowych na świecie, jak i wpływ tych praktyk na reprezentatywność wyników badań sondażowych. Choć badania sondażowe zwykle się uważa za względnie precyzyjną metodę pomiaru, to jednak różnice dotyczące staranności prowadzenia badań i kontroli ich jakości pomiędzy i w obrębie programów bywają bardzo znaczące.

Badacze nieuczestniczący w danym projekcie wiedzę o procedurach zastosowanych w danym sondażu czy niektórych wskaźnikach jakości jego realizacji czerpać mogą przede wszystkim z dokumentacji metodologicznej. Z tego powodu szczególnie istotne jest to, by zawierała ona odpowiedni i wyczerpujący zakres informacji.

Jeśli zaś chodzi o same praktyki sondażowe, to rosnące na świecie od dekad trudności z osiągnięciem satysfakcjonujących wartości wskaźników realizacji próby, które dodatkowo przyczyniają się do nieustannego zwiększania kosztów badania, są niewątpliwie motorem innowacji w metodologii badań sondażowych<sup>1</sup>. Jednym z ich skutków jest zróżnicowanie charakterystyk przeprowadzanych badań wśród poszczególnych ośrodków. W tym kontekście wydaje się celowe stworzenie konkretnego opracowania przedstawiającego jakie schematy planowania i realizacji niekomercyjnych badań surveyowych na świecie były i są stosowane, by następnie odpowiedzieć na pytanie, jak wpływają one na jakość pomiaru. W związku z powyższym moja praca stara się odpowiedzieć na trzy główne pytania badawcze:

---

<sup>1</sup> Jeden z badaczy holenderskich w prywatnej rozmowie przyznał, że źródłem imponującego rozwoju tej gałęzi nauki w jego kraju było właśnie wczesne doświadczenie dużych trudności z terenową realizacją badań.

**Jak wygląda rozkład praktyk, zwyczajów i standardów dotyczących dwóch procesów: dokumentowania (P1) i przeprowadzania (P2) sondaży na świecie, a także jaki jest wpływ opisanych w dokumentacji charakterystyk sondaży na jakość pomiaru (P3).**

Szczegółowe pytania badawcze, precyzujące tak zarysowany obszar analiz prezentują się następująco:

**P1a:** Jakie charakterystyki sondaży są najlepiej i najslabiej udokumentowane w raportach z projektów międzynarodowych?

**P1b:** Czy rozwój standardów dokumentowania sondaży jest adekwatny do poziomu możliwości technicznych?

**P1c:** Czy w analizowanym obszarze obserwujemy stabilny wzrost kompletności dokumentacji metodologicznej sondaży.

**P1d:** Jakie są międzynarodowe różnicowania kompletności dokumentacji?

**P1e:** Jakie są różnice w jakości dokumentacji pomiędzy programami sondażowymi?

**P2a:** Które sondażowe programy międzynarodowe zapewniają zróżnicowanie charakterystyk umożliwiające pożądane analizy metodologiczne?

**P2b:** Które programy międzynarodowe zapewniają najlepszą jakość pomiaru sondażowego?

**P3a:** Czy sondaże oparte o techniki ankiety i sondaże *mixed-mode* prowadzone bez obecności ankietera uzyskują niższe wskaźniki realizacji próby, niż wywiady kwestionariuszowe?

**P3b:** Czy sondaże przeprowadzone technikami ankiety i *mixed-mode* uzyskują porównywalną do wywiadów kwestionariuszowych reprezentatywność próby?

**P3c:** Jak typ dostępnego operatu moderuje związek między techniką sondażu, a reprezentatywnością próby?

**P3d:** Czy koncepcja *satisficingu* wyjaśnia problemy związane z preferencją techniki sondażu?

**P3e:** Czy kognitywistyczna koncepcja pamięci roboczej może tłumaczyć zmienność preferencji techniki sondażu wśród respondentów?

**P3f:** Jak wyglądają związki pomiędzy preferencją techniki polskich respondentów, a ich charakterystykami demograficznymi?

**P3g:** Czy pomiar preferencji techniki drogą pytań o deklaracje daje zbliżone wyniki do obserwacji realnych wyborów techniki udziału w badaniu w sondażu *mixed-mode*.

W tekście referatu kluczowe ustalenia dotyczące odpowiedzi na powyższe pytania szczegółowe oznaczone są odpowiednimi skrótami.

Odpowiedzi na pytania z pierwszego obszaru znajdują się w **Publikacji 1**, bazującej na projekcie inwentaryzacji sondaży z 30 międzynarodowych programów porównawczych. Efektem projektu jest baza zawierająca 10493 rekordy, opracowane w oparciu o publicznie dostępną dokumentację i bazy danych wynikowych. Artykuł omawia praktyki dokumentowania sondaży w analizowanych programach, przede wszystkim kompletność dokumentacji (**P1a**), ale także jej standaryzację, jednolitość (brak rozdrobnienia generującego sprzeczności) i formę. Weryfikuję hipotezy o: bardzo powolnym rozwijaniu się jakości dokumentacji w stosunku do postępujących możliwości technicznych (**P1b**); powolnym, ale zauważalnym wzroście kompletności dokumentacji w kolejnych dekadach (**P1c**) (Jabkowski 2019, 2023; Jabkowski i Kołczyńska 2020). Ponawiam także, w oparciu o nowy, szerszy zakres danych, analizy dotyczące globalnego zróżnicowania kompletności dokumentacji, osiągając wyniki odmienne, niż uzyskane wcześniej przez badaczy (**P1d**) (Kołczyńska i Schoene 2018). Porównuję stopień kompletności informacji o poszczególnych charakterystykach sondaży dostępnych w raportach wszystkich programów porównawczych, a także przeprowadzam bardziej szczegółową analizę

dla kilku wybranych projektów (**P1e**). Wnioski z moich analiz sugerują konieczność dążenia do większej standaryzacji raportów metodologicznych, optymalnie drogą konsensusu w środowisku badaczy. Wskazuję również na korzyści, jakie w dobie nowych technologii może przynieść raportowanie w formacie, który łatwo poddawałby się zautomatyzowanemu zbieraniu i przetwarzaniu danych. Umożliwiłoby ono znacznie szybsze i mniej obciążone błędami kodowania ręczne zbieranie danych metodologicznych do porównań.

Odpowiedź na pytanie dotyczące zróżnicowania praktyk związanych z realizacją badań sondażowych konieczna była w procesie tworzenia inwentaryzacji, będącej podstawą dla **Publikacji 1 i 2**. Problematyka ta jest w nich jednak opisana (z konieczności) skrótowo i wybiórczo, dlatego przedstawiam wyniki inwentaryzacji w rozszerzonej formie w autoreferacie, w części bazy empirycznej danych własnych (III.1) oraz w aneksie. Opisuję w tych miejscach zróżnicowanie w obrębie charakterystyk, które zostały użyte jako podstawa modeli analitycznych wykorzystanych w Publikacji 2 (a także, choć w skali jednego kraju i węższym zakresie, w Publikacji 4). Informacje o charakterystykach sondaży były również podstawą do obliczenia wskaźników kompletności dokumentacji zaprezentowanych w Publikacji 1. Omawiam zmianę trendów dotyczących stosowanych rozwiązań, np. wzrost popularności pewnych technik sondażu w czasie. Pozwala to zauważyć, że badania w schemacie mieszania technik stosowane były w zbadanych programach corocznie już od 1996 roku. Tym bardziej podkreśla to zasadność dokładnego omówienia podejścia opartego na technikach mieszanych, które przedstawiam w Publikacjach 2, 3 i 4. Kolejnym efektem przeprowadzonej inwentaryzacji jest konstatacja, że najciekawszego materiału (ze względu na zróżnicowanie metodologiczne) do porównań jakości sondaży przeprowadzanych w skali globalnej dostarczają programy Comparative Study of Electoral Systems (CSES) i International Social Survey Programme (ISSP), które jednocześnie są wyjątkowo dobrze udokumentowane, jak na programy o zasięgu wykraczającym poza Europę (**P2a**) (o czym piszę w Publikacji 1, a dane z programu ISSP wykorzystuję w Publikacji 2). Porównawcze programy międzykrajowe, które według założonych kryteriów osiągają najlepszą jakość pomiaru, to poza dwoma wspomnianymi także European Social Survey (ESS) (**P2b**) będący najkompletniej udokumentowanym programem porównawczym na świecie (wg analiz z Publikacji 1).

W celu sformułowania możliwie precyzyjnych wniosków na temat związków pomiędzy charakterystykami a miarami reprezentatywności stosuję w **Publikacji 2** metaanalizy oparte na danych z programów ESS i ISSP. Testuję hipotezę dotyczącą tego, że badania wykorzystujące technikę ankiety pocztowej i schematy mieszania technik sondażu, choć osiągają zazwyczaj niższe wskaźniki realizacji (**P3a**), nie muszą być z tego tylko powodu mniej reprezentatywne (Cornesse i Bosnjak 2018; Fowler i in. 2002; Luiten i Schouten 2013; Olson i in. 2012). W analizie uwzględniam jednocześnie takie charakterystyki, jak operat, fala, kraj, program i doświadczenie instytucjonalne. Wyniki wskazują, że sondaże używające wywiadu kwestionariuszowego osiągały najgorsze wskaźniki reprezentatywności (**P3b**), mimo istotnie wyższego *response rate*, jednak kluczowa była tu dostępność operatu imiennego (**P3c**), bez którego konieczne jest losowanie respondenta na poziomie gospodarstwa domowego. Proces doboru wewnątrzspółowego bez nadzoru ankietera stanowi tu zapewne największą przeszkodę – ze względu na skomplikowanie zadania i trudność w zakomunikowaniu jego ważności członkom gospodarstwa.

**Publikacje 3 i 4** dotyczą związku pomiędzy techniką sondażu a osiągniętymi wskaźnikami realizacji oraz błędami pomiaru. Analizuję tę zależność w odniesieniu do zjawiska „preferencji techniki sondażu” (*mode preference*). Publikacja 3 ma charakter teoretyczny, natomiast 4 – ze względu na bardziej szczegółowe ujęcie problematyki – ogranicza się do danych z sondaży przeprowadzonych w Polsce w ramach 7. i 8. rundy ESS, który poza

najwyższym współczynnikiem kompletności dokumentacji jest także jednym z trzech programów na świecie o najlepszych wskaźnikach reprezentatywności (patrz III.1.7.).

W Publikacjach 3 i 4 omawiam zjawiska *satisficingu* i preferencji techniki w oparciu o dostępną literaturę. W **Publikacji 3** przedstawiam związki między pojęciami *satisficingu* (Krosnick 1991) – a więc próby obniżenia przez respondenta nakładu wysiłku związanego z udziałem w badaniu, która może dotyczyć zarówno problemów z obszaru błędu braku odpowiedzi, jak i błędu pomiaru – a pojęciem *mode preference* (**P3d**) (Groves i Kahn 1979). Uzupełniam to także interdyscyplinarnym poszukiwaniem związków z kognitywistyczną koncepcją pamięci roboczej (**P3e**). Przedstawiona w obu publikacjach literatura sugeruje, że dopasowanie techniki do preferencji respondenta poprawia jakość pomiaru. Dlatego też w **Publikacji 4** pokazuję przykłady dwóch metod badania preferencji (jednej opartej na obserwacji deklaracji, a drugiej – realnych wyborów), uwzględniając ich zalety oraz wady, a także tworzę model mający ambicje odpowiedzieć na pytanie, w jaki sposób można łączyć metody badania preferencji techniki. Zaproponowany model opisuje zróżnicowanie preferencji techniki według cech społeczno-demograficznych respondentów w Polsce (**P3f**). Wnioskiem jest rekomendacja uprzedniego planowania takich sondaży *mixed-mode*, które w stopniu optymalnym wykorzystują użycie konkretnych technik sondażu dopasowanych do przewidywanych preferencji respondentów w celu zwiększenia jakości pomiaru. Przy czym analizy tego typu mogą być doskonalone przy użyciu nowoczesnych metod uczenia maszynowego (Jankowsky i Schroeders 2022; Küfner, Sakshaug i Zins 2022). Ponadto poddaję krytycznej analizie obecne w literaturze założenie, że pytanie respondentów o ich preferencję jest właściwym sposobem pozyskania informacji na temat ich rzeczywistych preferencji w tym obszarze (**P3g**) (Smyth, Olson i Millar 2014). Interesującym wnioskiem jest tu istotna niezgodność pomiędzy deklaracjami respondentów a obserwacjami ich wyborów, gdy zostają im zaproponowane alternatywne techniki udziału w badaniu.

## II.2. Problemy badawcze na tle literatury przedmiotu

Aby nadać strukturę poniższemu narracyjnemu przeglądowi literatury, podzieliłem opisywane zagadnienia na trzy kategorie łatwych do wyróżnienia grup charakterystyk sondaży: „wysiłku” włożonego w zbieranie danych (*fieldwork efforts*), doboru próby i techniki sondażu. Przegląd ten dotyczy wpływu charakterystyk zarówno na wskaźniki realizacji próby, jak i na (różnie mierzoną przez badaczy) reprezentatywność. Temat sondaży *mixed-mode* rozwijam w osobnym podrozdziale, gdyż jest on szczególnie istotny dla tematyki trzech publikacji włączonych do cyklu. Dalej natomiast przedstawiłem opis szczegółowy miar jakości pomiaru sondażowego.

Aby analizować występowanie różnych procedur badawczych w ramach sondaży i ich wpływ na jakość pomiaru należy posiadać o nich możliwie pełne i pewne informacje, w praktyce jednak ich dostępność jest mocno zróżnicowana (Mohler i in. 2010; Mohler, Pennell i Hubbard 2008; Vardigan, Granda i Hoelter 2016). Badania nad jakością dokumentacji sugerują ciągły (trwający od dekad) wzrost kompletności raportów metodologicznych w najważniejszych programach (Jabkowski 2019, 2023; Jabkowski i Kołczyńska 2020). Wniosek ten sformułowano jednak czysto ilościowo, co w konsekwencji nie oddaje pełni zróżnicowania form raportów, o których piszę w mojej pracy doktorskiej. Część badaczy zwraca także uwagę na konieczność położenia większego nacisku na kwestię wagi poszczególnych informacji zawartych w dokumentacji dla różnych, zdefiniowanych grup odbiorców (AAPOR/WAPOR 2021; Jedinger, Watteler i Förster 2018). Wcześniejsze badania sugerują, że w przypadku porównywania sondaży z programów międzynarodowych, najwyższą kompletnością charakteryzują się te przeprowadzane w Europie (Kołczyńska i Schoene 2018), gdzie sondaży

tych występuje najwięcej. Warto zasygnalizować, że wyniki przeprowadzonych przeze mnie badań pokazują odmienny obraz – wielka liczba europejskich sondaży nie przekłada się pozytywnie na lepsze (uśrednione) wskaźniki kompletności dokumentacji (Rybak 2025).

Przedstawiona poniżej literatura dotycząca związków pomiędzy charakterystykami sondażu, a jakością pomiaru, nie przynosi wprawdzie wyczerpującej odpowiedzi na postawione przeze mnie pytania badawcze (ze względu na częstą niekonkluzywność ustaleń), ale służy określeniu obszarów, które są przez badaczy uznawane za istotne w analizie takich relacji. Kierowałem się tą wiedzą, tworząc katalog zmiennych wykorzystanych w inwentaryzacji i modelach analitycznych. Z tego powodu wspomniany przegląd literatury odgrywa szczególnie istotną rolę w odniesieniu do Publikacji 1, gdyż stanowi uzasadnienie doboru zmiennych, których kompletność analizowałem, oraz Publikacji 2, ponieważ był punktem wyjścia dla doboru zmiennych do metaanalizy. W kolejnej części poruszone są także tematy dotyczące zjawiska *satisficingu* oraz preferencji respondentów wobec określonej techniki, co stanowi ramę teoretyczną dla analiz zaprezentowanych w Publikacjach 3 i 4.

## **II.2.1. Wpływ charakterystyk sondażu na wskaźnik realizacji próby oraz poziom reprezentatywności**

### **II.2.1.1. Procedury terenowe**

Pojęcie *fieldwork efforts* (Stoop i in. 2010:102–13) dotyczy starań i środków włożonych w proces przygotowania i realizacji fazy terenowej. W przypadku sondaży przeprowadzanych technikami wywiadu obejmuje to np. liczbę ankierów, sposób ich wynagradzania, kontrolę ich pracy czy objętość przeprowadzanych przez nich wywiadów. Charakterystyki, które poza wywiadami dotyczyć mogą również sondaży realizowanych technikami ankiety (pocztowej czy internetowej, np. w formie *push-to-web*) czy schematów mieszających różne techniki (*mixed-mode*), to rozmaite formy listów zapowiednich, zachęt materialnych czy przewidzianej w protokołach badawczych minimalnej liczby ponawianych kontaktów z potencjalnym respondentem.

Jak już wspomniano, dotychczasowe badania dotyczące wpływu charakterystyk sondaży na reprezentatywność były często fragmentaryczne i oparte na stosunkowo małych zbiorach danych. Nie zmienia to faktu, że ustalenia badaczy są ważne i formułowanie wartościowych konkluzji na ich podstawie jest możliwe, szczególnie w obszarze *fieldwork efforts*. Dla przykładu, powszechnie wskazuje się, że skuteczność listów zapowiednich w zwiększaniu wskaźników realizacji sondażu występuje niezależnie od zastosowanej techniki (de Leeuw i in. 2007; Porter i Whitcomb 2007; Yu i Cooper 1983). Niepозbawione znaczenia są także forma i nadawca tego listu (Tourangeau, Rips i Rasinski 2000:307). Uwzględnienie gratyfikacji dla respondenta również zdaje się zwiększać *response rates* we wszystkich technikach (Church 1993; Hopkins i Gullickson 1992; Singer i in. 1999).

Analiza posługująca się indeksem działań zwiększających *fieldwork efforts* (Stoop i in. 2010:102–13), oparta na danych ESS, sugeruje jednak negatywną zależność pomiędzy działaniami mającymi zwiększyć liczbę kontaktów i kooperację, a wskaźnikiem realizacji badania. Autorka tłumaczy to koniecznością podjęcia dodatkowego wysiłku przez instytuty badawcze w krajach, w których inne warunki (*survey climate*) zwiększają trudność przeprowadzenia badania. Dopiero uwzględnienie danych o wskaźnikach realizacji próby w poprzednich falach umożliwia uchwycenie tego, że wzrost *fieldwork efforts* prowadzi do ich zwiększenia.

Wprowadzenie dodatkowych zachęt materialnych i zwiększenie liczby kontaktów skutecznie zwiększa *response rate*, jeśli jest kierowane do odpowiednich grup respondentów w ramach procedur tzw. *adaptive* albo *responsive design* (Tourangeau i in. 2016). Jednocześnie

część dotychczasowych analiz wydaje się sugerować, że zwiększenie wysiłku i środków włożonych w realizację fazy terenowej badania, szczególnie powiększenie liczby prób nawiązania kontaktu z respondentami oraz wprowadzenie zachęt materialnych, skutkuje wzrostem *item-nonresponse*, czyli uchylaniem się od odpowiedzi na poszczególne pytania (Jäckle i Lynn 2007; Keeter i in. 2006). Może to być związane z faktem, że respondenci, którzy zostali skłonieni do wzięcia udziału w badaniu poprzez nachalne próby kontaktu albo wzbudzanie poczucia konieczności odwdzięczenia się (np. poprzez bezwarunkowe wynagrodzenia), udzielać będą odpowiedzi o niskiej jakości. W konsekwencji, dążenie do większego wskaźnika realizacji próby nie musi zawsze przekładać się na wyższą jakość badania.

Analizy wpływu wzrostu *fieldwork efforts* – poprzez używanie zachęt materialnych, listów zapowiednich czy zwiększanie minimalnej liczby prób kontaktu – na różne wskaźniki reprezentatywności sondażu, oparte o R-indicators (opisane w sekcji II.2.3.2.), zdają się nie prowadzić do jednoznacznych wniosków (Roberts, Vandenplas i Stahl 2014; Vogl i in. 2019; Wagner i in. 2017). Jeśli chodzi o ankietów, to wynagradzanie ich za liczbę wywiadów nie przynosi istotnego zwiększenia *response rates* (Rosen i in. 2011). Co do wpływu takiej praktyki na błąd braku odpowiedzi, to jest on potencjalnie niekorzystny, gdyż taki sposób wynagradzania może skłonić niedostatecznie nadzorowanych ankietów do skoncentrowania swoich wysiłków na tych potencjalnych respondentach, których łatwiej jest nakłonić do udziału w badaniu (Eckman i Koch 2019).

#### II.2.1.2. Operaty losowania i typy prób badawczych

Charakterystyki dotyczące losowania próby obejmują kwestie związane z operatem zawierającym wykaz jednostek populacji lub obiektów pozwalających na losowanie jednostek populacji w drodze doboru wielostopniowego (lub sposobem postępowania w przypadku braku operatu), definicją populacji, z której losowana jest próba, oraz schematem losowania, w tym metodami doboru wewnątrzspółowego w przypadku prób innych niż imienne.

Jeżeli chodzi o typy wykorzystanego operatu, to próby adresowe i próby gospodarstw domowych zdają się wpływać negatywnie na reprezentatywność zrealizowanego badania w porównaniu do prób imiennych (zwiększają także ryzyko trudnych do kontroli nadużyć ankietarskich) (Jabkowski 2018; Jabkowski i Cichocki 2019; Menold 2014). Losowanie przestrzenne może wywierać negatywny wpływ na reprezentatywność (Eckman i Koch 2019), a stosowanie metody doboru *random route*, nawet jeśli funkcje osoby losującej adres oraz ankietera są rozdzielone, wiąże się często z błędami doboru próby (Bauer 2014, 2016).

Wpływ na jakość badania ma także przyjęty sposób przeprowadzenia doboru wewnątrzspółowego. Opierająca się na tzw. siatce Kisha procedura selekcji zwiększa wskaźnik odmów udziału w badaniu i zmniejsza wskaźnik kooperacji w porównaniu do procedur quasi-losowych. Procedury te (wykorzystujące pytanie o ostatnie bądź najbliższe urodziny) skutkują z kolei zmniejszeniem reprezentatywności badania w stosunku do procedury Kisha (Jabkowski 2017). Co ciekawe, zaangażowanie ankietera we fragment procesu doboru próby może zmieniać kierunek związku pomiędzy wskaźnikami realizacji a reprezentatywnością badania. Jest to związane z faktem, że w przypadku prawidłowo skonstruowanej próby imiennej dobór respondentów opiera się na procedurze w pełni losowej, natomiast próba wymagająca udziału ankietera w losowaniu (np. użycie metody *random route*) może generować błędy wynikające z niedostatecznego przeszkolenia, lub nawet złej woli ankietera. Dlatego zaobserwowano, że w przypadku próby imiennej wzrost *response rates* skutkował wzrostem reprezentatywności, a w przypadku doboru metodą *random route* (czy, w

mniejszym stopniu, w przypadku stosowania doboru wewnątrzspółowego) wyższe wskaźniki realizacji były związane z większym błędem (Kołczyńska, Jabkowski i Eckman 2024).

### II.2.1.3. Techniki sondażu

Wybrana technika sondażu (lub schematy *mixed-mode*, polegające na równoczesnym albo sekwencyjnym wykorzystaniu wielu odmiennych technik sondażu) zdaje się mieć duży wpływ na różne aspekty pomiaru. Zgodnie z ustaleniami badaczy (Daikeler, Bosnjak i Manfreda 2020; Fox, Crask i Kim 1988; Goyder 1985; Hox i Leeuw 1994; de Leeuw 2008; Shih i Fan 2008) standaryzowane wywiady kwestionariuszowe osiągają zazwyczaj wyższe wskaźniki realizacji badania niż wywiady telefoniczne (choć konieczność stosowania quasi-losowej metody generowania numerów, wynikającej z rozkwitu telefonii komórkowej i zaniku stacjonarnej, stwarza problemy dla ich obliczania). Te z kolei osiągają tutaj wyniki wyższe niż ankiety pocztowe, choć niektórzy badacze zwracają uwagę na przykłady dobrze zaprojektowanych sondaży pocztowych, których *response rate* był równy lub wyższy w porównaniu do wywiadów osobistych w podobnym otoczeniu (Dillman, Smyth i Christian 2014:351–52). Najgorzej pod tym względem wypadają ankiety internetowe. Technika wywiadu standaryzowanego jest także (czy też była dotąd) zdecydowanie najpopularniejsza w sondażach o orientacji naukowej i społecznej, jednak problemem jest najwyższy spośród wszystkich technik (Villard i Fitzgerald 2017) i wciąż rosnący koszt jej stosowania, co z czasem będzie miało prawdopodobnie efekt zaporowy.

Jeśli zaś chodzi o reprezentatywność – metaanaliza de Leeuw (1992:45) wykazała, że w zestawieniu trzech technik (wywiad bezpośredni, telefoniczny i ankieta pocztowa) jedynie wywiad kwestionariuszowy generował istotne statystycznie zaburzenie proporcji płci w próbie względem populacji w stosunku do informacji z danych urzędowych. Gdy Laaksonen i Heiskanen (2014) porównali za pomocą R-indicatorów reprezentatywność równoległych sondaży realizowanych techniką kwestionariuszową, telefoniczną i internetową, okazało się, że wywiad kwestionariuszowy także wypadł pod tym względem najsłabiej. Z kolei Klausch, Hox i Schouten (2015) uzyskali wynik odwrotny – według eksperymentu porównującego sondaż *face-to-face*, CATI, ankietę pocztową i internetową to wywiad bezpośredni i ankieta internetowa osiągały najwyższą reprezentatywność, mierzoną za pomocą R-indicatorów.

Wpływ techniki na błąd pomiaru opiera się głównie na opozycji technik bazujących na kontakcie ankietera z respondentem (bezpośrednim lub telefonicznym) i technik potencjalnie zapewniających poczucie większej anonimowości (a więc kwestionariuszy do samodzielnego wypełniania) (Chang i Krosnick 2010). W analizie umożliwiającej walidację odpowiedzi respondentów w oparciu o dane urzędowe Sakshaug, Yan i Tourneau (2010) wykazali, że techniki bezosobowe zmniejszały zarówno błąd pomiaru, jak i wskaźnik realizacji w stosunku do CATI.

### II.2.2. Preferencja techniki sondażu w schematach *mixed-mode*

Ocena wpływu schematów mieszania technik na reprezentatywność badania jest trudna z powodu mnogości ich wariantów (de Leeuw i Nejc 2016). Mieszanie technik sondażu może być stosowane w schemacie równoległym lub sekwencyjnym. Ten pierwszy zakłada realizację fazy terenowej w różnych technikach jednocześnie, przy czym mogą być one odgórnie przypisane do respondentów lub jest im przedstawiony wybór z katalogu technik przewidzianych w protokole badawczym. Drugi natomiast polega na zmianie jednej techniki na inną na ustalonym etapie fazy terenowej. Schemat równoległy w sytuacji zaoferowania respondentom wyboru techniki zdaje się skutkować zmniejszeniem *response rate* (Medway i Fulton 2012) – przy czym powszechnie cytowana metaanaliza, na której opiera się ten wniosek,

zestawia jedynie badania korzystające z techniki pocztowej i internetowej. Celem badań *mixed-mode* w schemacie sekwencyjnym jest w pierwszej kolejności redukcja kosztów badania.

Należy wspomnieć w tym miejscu o hipotetycznym zjawisku preferencji techniki (*mode preference*). Ma to być względnie stała cecha respondenta, określająca jego przychylniejszy stosunek do udzielania odpowiedzi w sondażu realizowanym w konkretnej technice (lub technikach). Odpowiednie wykorzystanie dopasowania preferencji do użytej techniki ma skutkować zarówno zwiększeniem wskaźnika realizacji badania (Olson i in. 2012), jak i poprawą jakości udzielanych odpowiedzi (o czym mowa poniżej). Problemem jest tu z jednej strony stabilność w czasie, gdyż prawdopodobnie preferencja techniki zmienia się wraz z wiekiem (Al Baghal i Kelley 2016; Rybak 2018, 2019, 2020). Z drugiej strony kwestią problematyczną jest samo badanie tej preferencji – rzeczywiste wybory respondentów mogą nie zgadzać się z deklaracjami, zauważono ponadto silny wpływ techniki, w której zadano pytanie, na deklarację o preferencji (Groves i Kahn 1979; Tarnai i Paxson 2004).

Wyniki poniższych badań wydają się w dużej mierze wzajemnie sprzeczne. Z jednej strony wpływ użycia *mixed-mode* na reprezentatywność wydaje się korzystny. Porównanie sondażu *mixed-mode*, łączącego ankietę pocztową z CATI, z samą ankietą pocztową skutkowało stwierdzeniem większej reprezentatywności pierwszego rozwiązania (Fowler i in. 2002). Gdy Luiten i Schouten (2013) zestawili sondaż CATI z sondażem *mixed-mode*, łączącym CATI z ankietą pocztową i internetową, to okazało się, że miały one zbliżone *response rates* i koszt, ale użycie R-indicatorów ponownie wykazało większą reprezentatywność sondażu *mixed-mode*. Klausch, Hox i Schouten (2015) porównali z kolei sondaże telefoniczne, pocztowe i internetowe z ich odpowiednikami *mixed-mode*, w których dodano kolejną fazę w postaci wywiadu kwestionariuszowego. W większości przypadków błąd braku odpowiedzi generowany przez sondaże mieszające techniki, a mierzony przy użyciu R-indicatorów, był mniejszy. Cornesse i Bosnjak (2018), przeprowadzając metaanalizę 96 sondaży i wykorzystując R-indicatory, stwierdzili, że mieszanie technik prowadzi do zwiększenia reprezentatywności w porównaniu do sondaży stosujących tylko jedną technikę.

Z drugiej strony część badań wskazuje na porównywalne problemy z reprezentatywnością. Eksperyment niemiecki przeprowadzony w ramach European Values Study (EVS) wykazał zwiększenie wskaźnika realizacji, a zmniejszenie kosztu i długości trwania terenowej fazy badania starannie zaprojektowanego sondażu *mixed-mode* w porównaniu do analogicznego i równoczesnego badania stosującego wywiad kwestionariuszowy, problemy ze składem zrealizowanej próby w obu technikach były porównywalne, choć odmienne (Wolf i in. 2021). Szersze zestawienie równoległych eksperymentów z reszty krajów, które testowały schematy mieszania technik w ramach tej samej rundy EVS (EVS2017), wskazało na potencjał do zwiększenia wskaźnika odpowiedzi i zmniejszenia kosztów przy jednoczesnym wzroście ryzyka problemów z reprezentatywnością próby (Luijkx i in. 2021). Z kolei eksperyment polski (Sztabiński 2018), porównujący wywiady kwestionariuszowe w ramach ESS z równoległym sondażem sekwencyjno-równoległym (poczta/Internet→wywiad→poczta) *mixed-mode*, doprowadził do sytuacji, w której to drugie badanie osiągnęło jednocześnie wyższy koszt i niższy *response rate*, a ocena różnic w reprezentatywności była niejednoznaczna. Bianchi, Biffignandi i Lynn (2017) pokazali przykład panelu sondażowego, który po zmianie techniki w kolejnych falach – z wywiadu kwestionariuszowego na schemat mieszania technik – osiągnął zbliżony wskaźnik realizacji i strukturę próby (przy istotnej obniżce kosztów).

Wreszcie mamy też do czynienia z badaniami, które sugerują niekorzystny wpływ mieszania technik. Villard i Fitzgerald (2017) porównali wyniki sześciu innych eksperymentów sondażowych przeprowadzonych w ramach ESS i doszli do wniosku, że użycie *mixed-mode* w

porównaniu do CAPI/PAPI powoduje raczej zmniejszenie wskaźnika realizacji i zwiększenie błędu niż odwrotnie. Jak można zauważyć, sprzeczność wyników licznych badań (głównie eksperymentalnych) wydaje się sugerować, że metoda narracyjna nie udzieli w tym przypadku konkluzyjnych odpowiedzi. Z tego powodu we własnych badaniach zdecydowałem się zastosować metodę metaanalizy opartej na dużym zbiorze danych obserwacyjnych. Metoda ta omówiona jest w części II.3.1. autoreferatu i w Publikacji 2.

Na koniec należy wspomnieć o wpływie mieszania technik na kolejny istotny element całkowitego błędu sondażowego, jakim jest błąd pomiaru. Problematyką tą zajmuję się w Publikacji 3. Wpływ mieszania technik sondażu na błąd pomiaru może z jednej strony być korzystny ze względu na dopasowanie preferencji respondenta i techniki badania – wykazano możliwość zmniejszania się natężenia *satisficingu* w takich okolicznościach (Smyth, Olson i Kasabian 2014), z drugiej natomiast może pogarszać sytuację ze względu na problem odróżnienia efektu selekcji od efektu techniki (Vannieuwenhuyze 2013). Dopasowanie techniki do preferencji może także skutkować poprawą udzielanych odpowiedzi ze względu na związek z procesami poznawczymi jednostki (Rybak 2019).

### II.2.3. Miary jakości pomiaru sondażowego

Poniższe miary jakości pomiaru sondażowego odnoszą się do tego komponentu całkowitego błędu sondażowego (TSE), który niesie ze sobą największe ryzyko (Biemer i Lyberg 2003:116; Jabkowski 2015:61) generowania błędu systematycznego, tj. błędu braku odpowiedzi. Problematyka związana z błędem pomiaru (Jabkowski 2015:82; Sztabiński 2011:53–60), który z błędem braku odpowiedzi może konkurować jako źródło błędu całkowitego, została szerzej omówiona w Publikacji 3, nie była jednak podstawą dla tworzenia modeli analitycznych. Poniżej charakteryzuję wybrane rodzaje wskaźników błędu braku odpowiedzi. Pierwszy z nich (*response rate*) został użyty w Publikacji 2 oraz pośrednio w Publikacji 4. W Publikacji 1 badałem natomiast dostępność w dokumentacji informacji na jego temat. Drugi wskaźnik (kryterium zewnętrzne) wykorzystam w dalszej części autoreferatu i aneksie do omówienia jakości zebranych danych, natomiast trzeci (kryterium wewnętrzne) został użyty w Publikacji 2. Cztery (*R-indicators*) nie został przeze mnie wykorzystany, co uzasadniam.

#### II.2.3.1. Wskaźniki realizacji badania

Pierwszą miarą, którą opisuję ze względu na jej podstawowy charakter i powszechne stosowanie, są wskaźniki realizacji próby (dotyczące proporcji zrealizowanych wywiadów, kooperacji, odmów oraz nawiązania kontaktu z respondentem), zestandaryzowane zgodnie z wytycznymi American Association for Public Opinion Research (2023):

$$\begin{aligned} RR6 &= \frac{(I + P)}{(I + P) + (R + NC + O)} \\ COOP2 &= \frac{(I + P)}{(I + P) + R + O} \\ REF3 &= \frac{R}{(I + P) + (R + NC + O)} \\ CON3 &= \frac{(I + P) + R + O}{(I + P) + (R + NC + O)} \end{aligned}$$

Gdzie I to zrealizowany wywiad, P – wywiad częściowo zrealizowany, R – odmowa, NC – brak kontaktu, O – brak wywiadu z innych przyczyn (ale dana osoba stanowiła część

próby). Wybranie takich konkretnie definicji wskaźników (ze zbioru różnych propozycji przedstawionych przez AAPOR) pozwala uniknąć problemu ze zliczaniem kategorii respondentów, co do których nie ma pewności, jak ich zakodować – a jednoznaczna informacja o tej kategorii często nie jest dostępna. Dodatkowo nie ma ogólnie przyjętej definicji wywiadu wypełnionego częściowo, a także liczba takich wywiadów jest często sumowana z liczbą wywiadów kompletnych, więc dobrze, by wskaźnik uwzględniał je tak samo. Stąd decyzja o przyjęciu wskaźnika RR6 w Publikacji 1 i 2.

Same jednak wskaźniki realizacji pozwalają wyznaczyć jedynie maksymalną granicę błędu systematycznego wynikającego z braku odpowiedzi i są jedynie w stopniu niewielkim powiązane z wielkością błędu (Groves 2006; Groves i Peytcheva 2008). Dlatego konieczne jest opisanie innych miar reprezentatywności.

### II.2.3.2. Miary reprezentatywności próby

Zewnętrzne kryterium reprezentatywności, opisane przez Grovesa (2006), polega na porównaniu proporcji społeczno-demograficznych cech respondentów pomiędzy zrealizowaną próbą a populacją – o której informacje czerpiemy z publicznych statystyk. W przypadku poniższych analiz stosowane było porównanie proporcji udziału kobiet w wieku 25–64 lat (ze względu na różnice w definiowaniu populacji między sondażami) pomiędzy populacją respondentów a populacją kraju w roku badania – informacje o niej czerpane były z bazy danych ONZ (United Nations 2022):

$$ES_i = |\hat{p}_i - p_i|$$

$$V_i = p_i(1 - p_i)/n_i$$

Gdzie  $\hat{p}$  to proporcja płci w próbie w przyjętym zakresie wiekowym, a  $p$  to proporcja płci w populacji (wg danych oficjalnych) w tym samym zakresie wiekowym. Umożliwia to ocenę, w jakim stopniu zrealizowana próba odzwierciedla rzeczywisty skład populacji. Możliwą słabością tego kryterium jest różna jakość i aktualność danych urzędowych, szczególnie w krajach Trzeciego Świata.

Wewnętrzne kryterium reprezentatywności zostało zaproponowane przez Sodeura (1997) i zaaplikowane po raz pierwszy przez Kohlera (2007), a następnie w kolejnych badaniach (Eckman i Koch 2019, 2019; Jabkowski i Cichocki 2019; Jabkowski i in. 2021; Jabkowski, Cichocki i Kołczyńska 2024; Koch i in. 2014; Kołczyńska, Jabkowski i in. 2024; Menold 2014):

$$ES_i = |prop_{fem} - 0,5|$$

$$V_i = 0,25/n_i$$

Propozycja Sodeura polega na skupieniu uwagi na podpróbie respondentów zamieszkujących dwuosobowe gospodarstwa domowe heteroseksualnych partnerów, w których z definicji proporcja kobiet oraz mężczyzn wynosi dokładnie 50/50. W tej koncepcji każde odstępstwo od apriorycznej proporcji płci, jeśli przekracza błąd standardowy, świadczy o istotnych uchybieniach jakości pomiaru. Umożliwia to ocenę jakości zrealizowanej próby bez odwoływania się do danych urzędowych, które mogą być, szczególnie w krajach o słabo rozwiniętych instytucjach, bardzo niedokładne lub nieaktualne. Jednym z problemów związanych ze stosowaniem tego wskaźnika, który może z czasem się nasilać (Fischer 2016), jest kwestia gospodarstw domowych zamieszkiwanych przez partnerów tej samej płci. Tylko niewielki odsetek programów zawiera bowiem w kwestionariuszach pytanie o płeć partnera. Dotychczasowe szacunki (Jabkowski 2023; Jabkowski i in. 2021) wskazują, że aktualnie problem ten jest możliwy do zignorowania. Korelacja pomiędzy wskaźnikiem błędu

obliczonym z i bez uwzględnienia dwuosobowych gospodarstw partnerów tej samej płci wyniosła odpowiednio 0,99 dla ESS i 0,95 dla programu European Quality of Life Survey (EQLS).

W metaanalizach służących obliczaniu wskaźników reprezentatywności nie stosowałem wag. Tak zwane *design weights*, czyli wagi korygujące nierówne prawdopodobieństwa selekcji wynikające ze schematu doboru próby, były dostępne w bardzo niewielu przypadkach. Natomiast przy konstruowaniu zastosowanych wskaźników reprezentatywności stosowanie wag poststratyfikacyjnych, czy też wag w ogóle nieopisanych, konstruowanych w różny sposób w różnych krajach, latach i projektach jest niewskazane. Szersze omówienie tego problemu czytelnik znajdzie w pracy Jabkowskiego z zespołem (2021).

Innym popularnym wskaźnikiem reprezentatywności sondażu są zaproponowane przez Schoutena (Roberts i in. 2020; Schouten i in. 2009; Schouten, Shlomo i Skinner 2011) tak zwane *R-indicators*. Opierają się one na wariancji estymowanych prawdopodobieństw wzięcia udziału w badaniu obliczonych w oparciu o zmienne pomocnicze (najczęściej demograficzne). Zmienne te powinny być dostępne zarówno dla respondentów, jak i osób, które nie wzięły udziału w badaniu. Ich zastosowanie było niemożliwe ze względu na niedostępność ujednoliconych zmiennych pomocniczych dla wszystkich projektów w różnych krajach.

### **II.3. Metody zbierania i analizy danych**

W celu znalezienia odpowiedzi na założone problemy badawcze używałem różnych metod zbierania i analizowania danych. Cały proces, po selekcji sondażowych programów porównawczych, wykonanej w oparciu o obecnie niedostępną bazę przygotowaną przez GESIS (2019), musiał się rozpocząć od zbierania raportów metodologicznych i baz danych wynikowych dostępnych w repozytoriach internetowych. Spis repozytoriów znajduje się po spisie odwołań literaturowych.

Odpowiedź na pytanie dotyczące praktyk dokumentacyjnych ma charakter zarówno jakościowy, jak ilościowy. Ocena jakościowa opierała się na lekturze olbrzymiej liczby raportów, porównywaniu ich konstrukcji, standaryzacji, jednolitości oraz formy graficznej. Mogłem zaobserwować bardzo różne standardy konstrukcji dokumentacji, począwszy od przygotowywanych przez instytucje koordynujące zbiorczych raportów dotyczących całej fali (niekiedy uwzględniających formy graficzne ułatwiające ręczne kodowanie), poprzez kwestionariusze metodologiczne wypełniane przez poszczególne instytucje krajowe, aż po przypadki, kiedy jedna krajowa realizacja zaprezentowana była w kilku dokumentach, niekiedy w różnych formatach (dokumenty potrafiły sobie przeczyć).

Kolejnym krokiem było kodowanie charakterystyk sondaży z dokumentacji, by następnie ilościowo ocenić ich kompletność. Udzielenie odpowiedzi na pytanie badawcze dotyczące praktyk było więc etapem koniecznym przed pełną oceną dokumentacji. Dobór kodowanych charakterystyk wynikał z obserwacji realnych praktyk, z literatury, ale też z arbitralnych (z konieczności) decyzji. Różna struktura, nieprzyjazne formaty i brak wykorzystania w analizowanych programach pełnego potencjału architektury stworzonej w ramach Data Documentation Initiative (Rasmussen 2014; Van Wettere 2022) uniemożliwiały automatyzację zbierania danych. Charakterystyki były kodowane do bazy w programie Excel. Poprawność inwentaryzacji kontrolowali wykonawcy zatrudnieni w ramach grantu. Spis wszystkich zmiennych utworzonych w procesie inwentaryzacji znajduje się na końcu aneksu.

Aby odpowiedzieć na pytanie o kompletność raportów, stworzyłem 16 zmiennych kodujących obecność w dokumentacji metodologicznej informacji o poszczególnych charakterystykach. Umożliwiło to ocenę zmian kompletności informacji na przestrzeni lat w obrębie poszczególnych projektów, porównania pomiędzy projektami i okresami, jak i

całościową ocenę dostępności informacji o poszczególnych charakterystykach w całej inwentaryzacji.

Zebrana wiedza o charakterystykach sondaży umożliwiła także odpowiedź na pytanie na temat związku pomiędzy nimi a reprezentatywnością. Wzory dotyczące wykorzystanych miar reprezentatywności przedstawiłem powyżej.

Odpowiedź na pytanie dotyczące łącznych konsekwencji podstawowych charakterystyk, takich jak technika i operat, dla jakości pomiaru możliwa była dzięki zastosowaniu metaanalizy, o tej metodzie piszę poniżej. W Publikacji 2 przedstawiłem modele trzypoziomowej, krzyżowo zagnieżdżonej (Fernández-Castilla i in. 2019) metaregresji, które poza wpływem cech sondażu (osobno i w interakcji) pozwalały kontrolować (za pomocą tzw. *random intercepts*) także efekty związane z klimatem dla badań sondażowych (*survey climate*) w poszczególnych krajach czy różnicą w tematyce sondaży. Co prawda tematyka badań nie była kodowana, była ona jednak współdzielona w obrębie fal programów, w których obserwacje (krajowe sondaże) były zagnieżdżone. Dokonując metaanaliz, wykorzystałem pakiet metafor (Viechtbauer 2010) języka R (R Core Team 2024).

Aby porównywać porównywalne, z modeli wyłączyłem sondaże niemieszczące się w przyjętych kryteriach badania losowego, a więc te, które stosowały kwoty demograficzne i substytucje nierespondentów (*nonrespondents*), nie korzystały z losowych metod doboru wewnątrzspołowego czy schematów doboru próby spełniających wymogi losowości. Wyłączone zostały takie sondaże, w których dokumentacja dotycząca tych obszarów była niekompletna. Poniżej omawiam samą metodę metaanalizy.

### II.3.1. Metoda metaanalizy

Uzyskanie odpowiedzi na pytania badawcze dotyczące związków między charakterystykami sondażu i reprezentatywnością zrealizowanej próby możliwe było dzięki zastosowaniu metody metaanalizy do danych ze stworzonej bazy. Wartość metaanalizy opiera się na możliwości uniknięcia pułapek mniej zaawansowanych metod syntezy wyników wielu badań, których rezultaty mogą prowadzić do wrażenia subiektywności i niekonkluzywności rozstrzygnięć (Card 2012:7). Tradycyjne metody syntezy są nieprzydatne w przypadku konieczności wzięcia pod uwagę olbrzymiej ilości badań, jednocześnie porównywanie samych tylko wyników istotności poszczególnych testów w sytuacji nierównolicznych prób badawczych skutkuje fałszywym przekonaniem o mniejszym niż jest w istocie związku między zmiennymi. Inaczej jest w przypadku metaanalizy, która to umożliwia zbadanie istotności efektu dla wszystkich badań łącznie (Hunter i Schmidt 2004:3–32). Jedną z pierwszych decyzji w procesie planowania metaanalizy jest ustalenie miary wielkości efektu. „Wielkość efektu (*effect size*) to po prostu wielkość czegokolwiek, co badamy” (Cumming 2012:34), jeśli jest to mierzone w sposób zestandaryzowany dla wszystkich uwzględnianych badań lub doprowadzone do takiego stanu przez autorów metaanalizy. Na przykład w badaniach epidemiologicznych używaną wielkością efektu może być średnia lub proporcja w populacji (Barendregt i in. 2013). Card proponuje następującą roboczą definicję wielkości efektu: „indeks natężenia i kierunku związku pomiędzy dwoma zmiennymi” (Card 2012:87).

Kolejną ważną decyzją jest wybór pomiędzy modelami efektów stałych (*fixed-effects*) i modelami efektów losowych (*random-effects*) (Borenstein i in. 2009:61–86). Modele efektów stałych są uzasadnione wyłącznie w przypadku, kiedy badania, na których oparty jest model, są bardzo zbliżone pod względem konstrukcji, a badacz nie chce ich potraktować jako części szerszej reprezentacji (możliwych) badań. Modele *random-effects* są bardziej odporne na wpływ przypadków odstających i w dużo większym stopniu pozwalają na uogólnienie wyników. Najprostszy przykład modelu *random-effects* to:

$$\widehat{OES}^* = \frac{\sum_{i=1}^k w_i^* \times ES_i}{\sum_{i=1}^k w_i^*}$$

Gdzie  $k$  to liczba uwzględnionych badań, a  $ES_i$  to wielkość efektu jednego badania. Dla każdego badania wagi  $w_i^*$  są obliczane jako odwrotności wariancji  $\widehat{OES}^*$ , tj.  $V_i^* = V_i + \tau^2$ , gdzie  $V_i$  to wariancja w obrębie pojedynczego badania, a  $\tau^2$  to wariancja pomiędzy badaniami (Borenstein i in. 2009; Jabkowski i Cichocki 2019). We wszystkich przedstawionych metaanalizach używam metody „*restricted maximum likelihood*” do szacowania wielkości  $\tau^2$ , czyli „wariancji w rozkładzie prawdziwych wielkości efektów” (Borenstein i in. 2009:114–17). Dodatkowo, podczas obliczania modeli metaregresji stosuję metody zmniejszające ryzyko uzyskania fałszywych wyników, takie jak test Knapa-Hartunga (Knapp i Hartung 2003). Metoda metaanalizy była już stosowana w obszarze metodologii sondażowej, zarówno w oparciu o bardziej typową procedurę syntezy wyników badań z artykułów naukowych, jak i na podstawie samodzielnie obliczanych miar bazujących na danych wynikowych sondaży (Cehovin i in. 2018; Church 1993; Cornesse i Bosnjak 2018; Daikeler i in. 2020; Groves i Peytcheva 2008; Hox i Leeuw 1994; Jabkowski 2017; Manfreda i in. 2008; Medway i Fulton 2012; Shih i Fan 2007, 2008; Sturgis i in. 2017; Yammarino, Skinner i Childers 1991). Metodę metaanalizy stosuję w Publikacji 2, w autoreferacie i aneksie używam jej dodatkowo do opisu zebranych danych.

### II.3.2. Badanie preferencji techniki sondażu wśród respondentów

Pogłębiona odpowiedź na pytanie o wpływ techniki sondażu na jakość pomiaru była możliwa dzięki analizie wyników na poziomie poszczególnych respondentów. Jak zostało przedstawione w przeglądzie literatury, preferowanie konkretnych technik przez badane osoby może mieć wpływ zarówno na błąd braku odpowiedzi, jak i na błąd pomiaru.

Źródłem odpowiedzi na pytanie o kształt tego wpływu były zarówno studia literaturowe, przedstawione w Publikacji 3 (skupiającej się na błędzie pomiaru), jak i model wielomianowej regresji logistycznej (Best i Wolf 2015), zawarty w Publikacji 4 (omawiającej predyktory *mode preference*), obliczony za pomocą programu STATA. Do jego stworzenia użyłem danych z eksperymentu metodologicznego (Sztabiński 2018) przeprowadzonego w ramach 7. rundy ESS w Polsce, w którym użyto sekwencyjno-równoległego (poczta/Internet->wywiad->poczta) schematu *mixed-mode*, oraz z 8. rundy ESS w Polsce, w której pytano respondentów o preferowaną technikę. Połączyłem obie bazy, a w modelu poza zmiennymi społeczno-demograficznymi użyłem dodatkowej – określającej, czy obserwacja ma charakter deklaracyjny, czy behawioralny (wybór techniki). Umożliwiło to uchwycenie zarówno podobieństw, jak i różnic pomiędzy związkami charakterystyk respondenta z preferencją badaną na dwa sposoby.

## III. Dane z badań własnych i wyniki analiz

### III.1. Inwentaryzacja charakterystyk sondaży z programów porównawczych

Dane do publikacji (z wyjątkiem Publikacji 3, która ma charakter teoretyczny i czerpie ze studiów literaturowych) zakresowo zawierają się w inwentaryzacji, którą przeprowadziłem w ramach finansowanego przez NCN projektu *Metaanaliza wpływu cech konstrukcji badań sondażowych na jakość ich realizacji na podstawie danych z porównawczych projektów międzykrajowych*. Poniżej przedstawiam opis stworzonej bazy, uwzględniający charakterystyki analizowane w publikacjach. Spis pozostałych zmiennych, ich rozkłady i dodatkowe metaanalizy znajdują się w aneksie. Tabela 1 przedstawia liczebności realizacji krajowych w

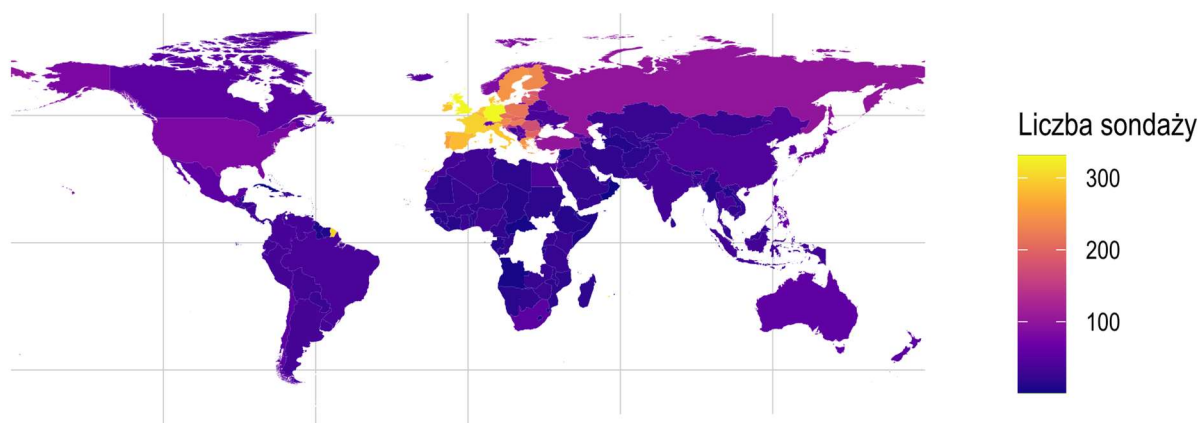
ramach poszczególnych programów porównawczych. Omówienie kryteriów selekcji programów do inwentaryzacji zostało zamieszczone w Publikacji 1.

*Tabela 1. Liczebności i zakres dat programów badawczych uwzględnionych w analizie (za Publikacją 1)*

| <b>Program</b>                                                              | <b>Liczba sondaży</b> | <b>Pierwszy i ostatni rok badań uwzględnionych w bazie</b> |
|-----------------------------------------------------------------------------|-----------------------|------------------------------------------------------------|
| AB – Afrobarometer                                                          | 205                   | 1999–2021                                                  |
| ABS – Asian Barometer Survey                                                | 69                    | 2001–2021                                                  |
| ARAB – Arab Barometer                                                       | 63                    | 2006–2022                                                  |
| ASES – Asia Europe Survey                                                   | 18                    | 2001                                                       |
| ASIAB – Asia Barometer                                                      | 34                    | 2005–2008                                                  |
| CCEB – Candidate Countries Eurobarometer                                    | 142                   | 2001–2004                                                  |
| CCS – Civic Culture Study                                                   | 5                     | 1959                                                       |
| CEEB – Central and Eastern Eurobarometer                                    | 120                   | 1990–1997                                                  |
| CNEP – Comparative National Elections Project                               | 61                    | 1990–2021                                                  |
| CONS – Consolidation of Democracy in Central and Eastern Europe             | 24                    | 1990–2001                                                  |
| CRRC – Caucasus Barometer                                                   | 30                    | 2006–2020                                                  |
| CSES – Comparative Study of Electoral Systems                               | 209                   | 1996–2020                                                  |
| EASS – East Asia Social Survey                                              | 24                    | 2006–2018                                                  |
| EB – Eurobarometer                                                          | 4604                  | 1974–2019                                                  |
| ECS – European Communities Study                                            | 20                    | 1970–1973                                                  |
| EES – European Election Studies                                             | 174                   | 1989–2019                                                  |
| EQLS – European Quality of Life Surveys                                     | 125                   | 2003–2016                                                  |
| ESS – European Social Survey                                                | 243                   | 2002–2022                                                  |
| EUNB – EU Neighbourhood Barometer                                           | 96                    | 2012–2014                                                  |
| EVS – European Values Study                                                 | 147                   | 1981–2020                                                  |
| GALLUP – Gallup World Poll                                                  | 2280                  | 2006–2021                                                  |
| INTUNE – Integrated and United                                              | 35                    | 2007–2009                                                  |
| ISSP – International Social Survey Programme                                | 961                   | 1985–2022                                                  |
| ISSP-SOR – Study of Religion of Additional 14 Countries, based on ISSP 2018 | 14                    | 2018                                                       |
| LAPOP – Latin American Public Opinion Project (Americas Barometer)          | 275                   | 1973–2021                                                  |
| LITS – Life in Transition Survey                                            | 98                    | 2006–2016                                                  |
| MEDW – Making Electoral Democracy Work                                      | 10                    | 2010–2016                                                  |
| NBB – New Baltic Barometer                                                  | 12                    | 1993–2000                                                  |
| NEB/NDB – New Europe Barometer/New Democracies Barometer                    | 94                    | 1991–2007                                                  |
| WVS – World Values Survey                                                   | 301                   | 1981–2021                                                  |
| <b>Razem</b>                                                                | <b>10493</b>          | <b>1959–2022</b>                                           |

### III.1.1. Zakres geograficzny danych objętych inwentaryzacją

Zakres krajów realizacji sondaży zawartych w inwentaryzacji przedstawia wykres 1. Posługując się podziałem używanym przez ONZ (Statistics Division of the United Nations Secretariat 1999); (Cypr, Gruzja i Armenia są zaklasyfikowane jako Azja) można stwierdzić, że najliczniej reprezentowanym regionem była Europa Północna (2091), a najsłabiej Afryka Środkowa (80). Najwięcej sondaży przeprowadzono w Niemczech (331, plus 3 w Niemieckiej Republice Demokratycznej) i Wielkiej Brytanii (329). W Afryce najwięcej sondaży zrealizowano w Republice Południowej Afryki (55) i Egipcie (42). W Azji – na Cyprze (176), w Turcji (102) i Japonii (76). W Ameryce Łacińskiej – w Meksyku (57) i Chile (53). W Ameryce Północnej w USA (81). Wśród krajów Oceanii największą liczbę sondaży odnotowano w Australii (58).



Wykres 1. Liczba sondaży w bazie wg kraju

Najdłużej trwające wśród uwzględnionych programy porównawcze to LAPOP (od 1973) i EB (od 1974 lub licząc ECS – od 1970). Najkrócej trwały CCS (1959), ASES (2001) i ISSP-SOR (2018) – wszystkie ok. roku. Historia programów wykraczających poza obszar zwyczajowego Zachodu rozpoczyna się już w pierwszym roku uwzględnionym w bazie. W pierwszej i jedynej rundzie CCS, którą literatura uznaje za początek porównawczych międzykrajowych badań sondażowych (Almond i Verba 1963; Lagos 2008), poza krajami zimnowojennego pierwszego świata, brał udział Meksyk. Dekadę później ('70) kraj ten uczestniczył również w pierwszej rundzie prowadzonego przez Uniwersytet Vanderbilta programu LAPOP. Wreszcie, po kolejnych dziesięciu latach ('80), w pierwszej rundzie WVS obecne były, poza krajami Zachodu, Japonia, Korea Południowa, RPA i raz jeszcze Meksyk. W kolejnej dekadzie ('90) ISSP rozpoczyna programy międzykontynentalne, a dalej powstają badania skupione już wyłącznie na obszarach takich jak Afryka, Bliski Wschód czy Azja Wschodnia.

Najwięcej sondaży przeprowadzono w roku 2014 (694), kolejno plasują się lata 2012 (556) i 2011 (551). Pierwsze badanie w Afryce miało miejsce w roku 1982 (RPA, WVS), ale coroczne badania rozpoczynają się dopiero w 1999. W Azji po raz pierwszy badanie przeprowadzono w roku 1979 (Izrael, LAPOP), zaś sondaże coroczne rozpoczynają się w 1989. Pierwsze badanie w ramach programów porównawczych, które obejmowało Europę, Amerykę Północną i Łacińską, to wspomniany CCS (1959). W Europie sondaże porównawcze, uwzględnione w bazie, odbywają się corocznie już od 1970, w Ameryce Łacińskiej od 1983, a w Północnej od 1985 (USA). W Oceanii pierwsze uwzględnione badanie pochodzi z 1981 (Australia, WVS), natomiast badania coroczne również są realizowane od 1985.

Programy można podzielić na ograniczające się do jednego kontynentu i te, które obejmowały ich wiele. Dziesięć programów, które uwzględniały tylko jeden kontynent, to AB, ABS, ASIAB, CONS, CRRC, EASS, INTUNE, MEDW, NBB, NEB/NBD. Kraje z dwóch kontynentów były badane przez EB, EES, EQLS, ESS i EVS. LAPOP badał z kolei kraje z trzech kontynentów, a sondaże na wszystkich sześciu kontynentach przeprowadzono w ramach CSES, GALLUP, ISSP i WVS.

### III.1.2. Dokumentacja

Tabela 2 przedstawia opis i podsumowanie kompletności zakodowanych zmiennych metodologicznych dla wszystkich analizowanych sondaży (**P1a**). Co ciekawe, technika sondażu była najkompletniej udokumentowaną charakterystyką. Kolejne miejsca zajmowały informacje dotyczące doboru próby, takie jak operat, zakres wieku badanej populacji, zastosowanie substytucji oraz kwoty demograficzne. Te charakterystyki były obecne w ponad 90% raportów. Natomiast bardziej szczegółowe, choć równie istotne informacje o metodach doboru wewnątrzspołowego uwzględniono w niecałych 25% raportów. Podczas gdy dane o samej liczbie przeprowadzonych wywiadów pojawiały się w około 90% raportów, współczynniki realizacji, które są często uznawane za podstawowy wskaźnik jakości badania, znalazły się tylko w około 15% dokumentów.

Dane dotyczące kwestii często analizowanych przez metodologów, takich jak użycie listów zapowiednich i zachęt materialnych dla respondentów, pojawiły się jedynie w niecałych 15% raportów. Można jednak argumentować, że autorzy dokumentacji mogli nie czuć się zobowiązani do wspominania o tych charakterystykach sondażu, które nie były obecne w ich badaniach. Kwestia ta została szerzej omówiona w Publikacji 1. Nie można jednak w ten sposób wyjaśnić braków w odniesieniu do informacji dotyczących liczby i sposobu wynagradzania ankieterów w badaniach prowadzonych metodą wywiadu kwestionariuszowego, które mogą wpływać na jakość zebranych danych. Te informacje pojawiły się odpowiednio w około 5% i 10% analizowanych raportów.

*Tabela 2. Opis zmiennych dotyczących kompletności dokumentacji i podsumowanie ich dostępności w bazie (za Publikacją 1)*

| Nazwa         | Dostępność informacji o:         | Liczebności | % sondaży z dostępną informacją |
|---------------|----------------------------------|-------------|---------------------------------|
| mod_info      | technice                         | Nie: 137    | 0,987                           |
|               |                                  | Tak: 10356  |                                 |
| loage_info    | minimalnym wieku                 | Nie: 432    | 0,959                           |
|               |                                  | Tak: 10061  |                                 |
| upage_info    | maksymalnym wieku                | Nie: 450    | 0,957                           |
|               |                                  | Tak: 10043  |                                 |
| fra_info      | operacie                         | Nie: 630    | 0,94                            |
|               |                                  | Tak: 9863   |                                 |
| int_info      | liczbie wywiadów                 | Nie: 997    | 0,905                           |
|               |                                  | Tak: 9496   |                                 |
| subquota_info | substytucjach/kwotach            | Nie: 1631   | 0,845                           |
|               |                                  | Tak: 8862   |                                 |
| cont_info     | minimalnej liczbie prób kontaktu | Nie: 6362   | 0,394                           |
|               |                                  | Tak: 4131   |                                 |
| long_info     | długości wywiadu                 | Nie: 7209   | 0,313                           |
|               |                                  | Tak: 3284   |                                 |

|                  |                                |           |       |
|------------------|--------------------------------|-----------|-------|
| hhselection_info | dobrze wewnątrzspółowym        | Nie: 7584 | 0,277 |
|                  |                                | Tak: 2909 |       |
| check_info       | back checkach                  | Nie: 8535 | 0,187 |
|                  |                                | Tak: 1958 |       |
| outcome_info     | umożliwiających obliczenie RR6 | Nie: 8843 | 0,157 |
|                  |                                | Tak: 1650 |       |
| rr_info          | dowolnym RR                    | Nie: 9014 | 0,141 |
|                  |                                | Tak: 1479 |       |
| lett_info        | listach zapowiednich           | Nie: 9056 | 0,137 |
|                  |                                | Tak: 1437 |       |
| inc_info         | zachętach materialnych         | Nie: 9165 | 0,127 |
|                  |                                | Tak: 1328 |       |
| paid_info        | sposobie wynagradzania ankiet. | Nie: 9410 | 0,103 |
|                  |                                | Tak: 1083 |       |
| work_info        | liczbie ankierów               | Nie: 9887 | 0,058 |
|                  |                                | Tak: 606  |       |

Warto porównać opisany przeze mnie zestaw zmiennych dotyczących kompletności informacji w raportach metodologicznych ze wskazaniem międzynarodowego raportu dotyczącego (w części) stanu obecnego i wskazań na przyszłość w dziedzinie dokumentacji metodologicznej sondaży (AAPOR/WAPOR 2021). Porównanie wykazanych w raporcie obszarów, które powinny być poruszone w dokumentacji z wybranymi przeze mnie zmiennymi przedstawia tabela 3. Z tego punktu widzenia w inwentaryzacji brakuje uwzględnienia informacji instytucjonalnych (finansowanie, instytucja krajowa), o temacie badania, oraz tych dotyczących kwestionariusza, jego tłumaczenia czy testowania.

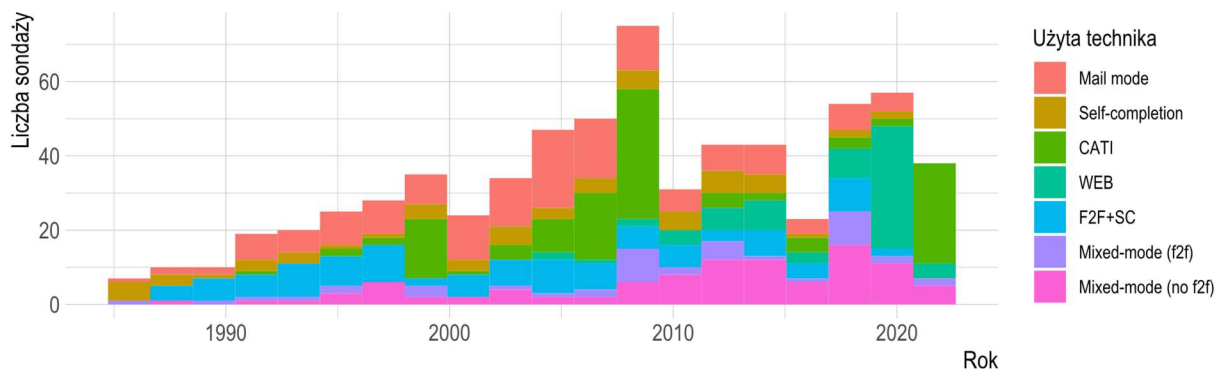
*Tabela 3. Porównanie zaleceń zawartych w Załączniku 6 raportu (AAPOR/WAPOR, 2021) odnośnie obszarów, o których powinna informować dokumentacja ze zmiennymi zawartymi w inwentaryzacji. + pełna zgodność; 0 zgodność tematyczna; - brak zgodności (za Publikacją 1)*

| <b>Zalecenia raportu:</b>  | <b>Informacje w bazie o:</b>                                               | <b>Zgodność</b> |
|----------------------------|----------------------------------------------------------------------------|-----------------|
| Finansowanie               |                                                                            | -               |
| Instytucja badawcza        |                                                                            | -               |
| Cele badania               |                                                                            | -               |
| Przekrojowe/podłużne       | (Zbierane dla niektórych programów, nie uwzględnione w bazie dokumentacji) | -               |
| Populacja badana           | minimalnym wieku                                                           | +               |
|                            | maksymalnym wieku                                                          | +               |
| Jednostka obserwacji       | (We wszystkich tylko jednostka)                                            | +               |
| Dobór próby                | operacie                                                                   | +               |
|                            | dobrze wewnątrzspółowym                                                    | 0               |
|                            | substytucjach/kwotach                                                      | +               |
| Zachęty materialne         | zachętach materialnych                                                     | +               |
|                            | listach zapowiednich                                                       | 0               |
| Technika                   | technice                                                                   | +               |
| Kwestionariusz             | długości wywiadu                                                           | 0               |
| Testowanie kwestionariusza |                                                                            | -               |

|                             |                                                                       |             |
|-----------------------------|-----------------------------------------------------------------------|-------------|
| Tłumaczenie kwestionariusza |                                                                       | -           |
| Ankieterzy                  | sposobie wynagradzania ankieterów<br>liczbie ankieterów               | 0<br>+      |
| Procedury terenowe          | minimalnej liczbie prób kontaktu<br>back checkach                     | 0<br>0      |
| Response rate               | liczbie wywiadów<br>dowolnym RR<br>umożliwiająca obliczenie RR6 AAPOR | 0<br>0<br>+ |
| Wagi                        | (Nie uwzględniono celowo)                                             | -           |
| Czas i miejsce              | (Dostępne da wszystkich)                                              | +           |

### III.1.3. Technika sondażu

Wśród zastosowanych w uwzględnionych badaniach technik zdecydowanie najpopularniejszy był standaryzowany wywiad kwestionariuszowy (9003 – PAPI i CAPI kodowane łącznie), a najmniej popularne – *ex aequo* – sondaże zrealizowane przy użyciu techniki zakładającej, że respondent dostaje od ankietera kwestionariusz do samodzielnego wypełnienia (*self-completion*) (62), oraz sondaż *mixed-mode*, gdzie wywiad kwestionariuszowy połączony był z inną techniką niż wspomniane *self-completion* (62). Sondaże przeprowadzonych przy pomocy ankiety pocztowej jest łącznie 155, a CATI – 792, ale zdecydowana większość z nich została zrealizowana w ramach programu GALLUP, gdyż nie jest to technika popularna w badaniach niekomercyjnych. Nie licząc tego programu, sondaże wykonanych techniką wywiadu telefonicznego jest 130. Badania wykorzystujące ankietę internetową znalazły się w bazie 71 razy.



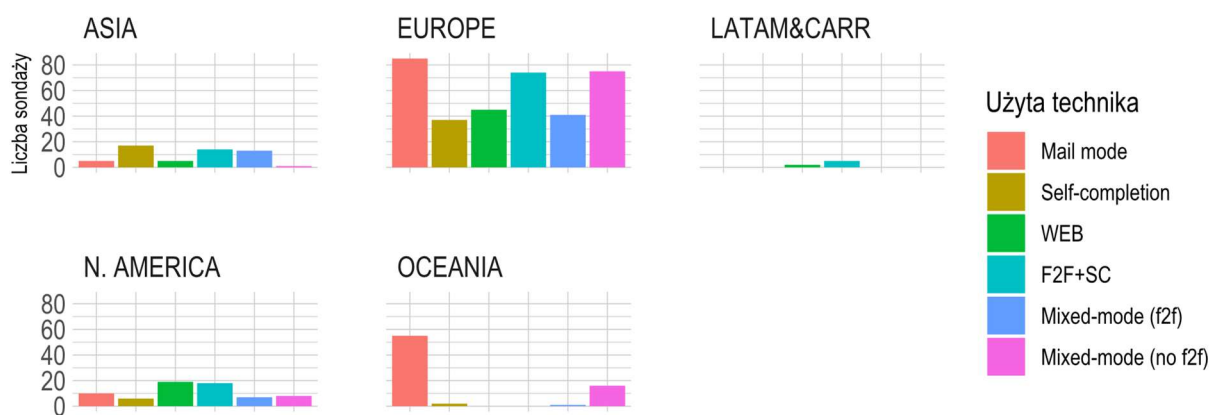
Wykres 2. Techniki inne niż wywiad kwestionariuszowy wg roku realizacji badania (z pominięciem programu GALLUP)

Jeśli chodzi o badania w schemacie mieszania technik, to – poza tą wspomnianą powyżej – technika łącząca wywiad standaryzowany z kwestionariuszem do samodzielnego wypełnienia (najczęściej jako dołączony moduł – bardzo popularne rozwiązanie w ISSP) była wykorzystana 111 razy, a łączenie technik innych niż wywiad standaryzowany (czyli np. pocztowej i internetowej) zastosowano 100 razy. Zmiany popularności technik w czasie prezentuje wykres 2.

Historia wąsko rozumianego sondażu rozpoczyna się od techniki standaryzowanego wywiadu kwestionariuszowego. Początkowo technika ta była optymalna ze względu na bardzo wysoki współczynnik realizacji, osiągany za sprawą nowego rodzaju badania, który budził zainteresowanie respondentów. Dlatego też uwzględnione w bazie sondaże przez początkowy

okres wykonywane były wyłącznie w tej technice. Według danych zebranych w bazie ankieta pocztowa po raz pierwszy pojawia się w Niemczech Zachodnich, w pierwszej rundzie ISSP (1985), a od roku 1990 jest stosowana niemal corocznie. Szczyt jej popularności przypada na rok 2012, kiedy obserwujemy 12 badań wykonanych tą techniką, a później następuje pewne osłabienie zainteresowania nią. Technika wykorzystująca wyłącznie kwestionariusz do samodzielnego wypełnienia, który jest dostarczany respondentowi osobiście (a więc nie drogą pocztową), pojawiła się w roku 1985 jednocześnie w USA oraz w Wielkiej Brytanii (ISSP) i od tego czasu jest stosowana niemal nieprzerwanie. W przypadku tej techniki spotykamy pewne problemy z nazewnictwem, o czym mowa będzie dalej. Pierwsze uwzględnione badanie CATI przeprowadzono w 1992 w USA (CNEP), a coroczne stosowanie (będące głównie zasługą programu GALLUP) rozpoczyna się w 1996. Pierwsze wykorzystanie ankiety internetowej, równocześnie w ramach CNEP (USA) i EES (Holandia), przypada na rok 2004, a o corocznym stosowaniu możemy mówić od 2010.

Badanie łączące wywiad kwestionariuszowy z modułem do samodzielnego wypełnienia pierwszy raz pojawiło się w 1987 w USA (ISSP) i od tego czasu jest stosowane corocznie. Badania *mixed-mode*, łączące wywiad kwestionariuszowy z techniką inną niż *self-completion*, po raz pierwszy użyto w 1985 w Australii – wywiad stosowano tam wśród mieszkańców miast, a kwestionariusz pocztowy trafiał do mieszkańców wsi. Coroczne stosowanie tego rozwiązania baza notuje od 2006. Pierwszy uwzględniony sondaż, który nie wykorzystuje wywiadu bezpośredniego w schemacie badania *mixed-mode*, przeprowadzono w Australii w 1987 – łączył on kwestionariusz do samodzielnego wypełnienia z wywiadem telefonicznym. Technika ta jest stosowana corocznie od 1996, gdy w Szwecji i Nowej Zelandii kwestionariusz pocztowy połączono z CATI lub kwestionariuszem dostarczonym przez ankietę. Szczyt jej popularności przypada na okres od 2005 (8) do 2018 (8).



Wykres 3. Techniki wg kontynentów – bez wywiadu kwestionariuszowego i telefonicznego (i z pominięciem Afryki, gdyż tam tylko te techniki były stosowane)

Wśród technik sondażu można zaobserwować regionalne zróżnicowanie popularności. O ile wywiad kwestionariuszowy jest uniwersalny w ujęciu globalnym, to reszta technik już nie, co częściowo pokazuje wykres 3. Ankieta pocztowa jest niezwykle popularna w Australii i Nowej Zelandii (łącznie 55, w ramach ISSP, WVS i CSES). Niewiele rzadziej była stosowana w państwach nordyckich (Szwecji, Norwegii i Finlandii, łącznie 49 razy, niemal wyłącznie w ramach ISSP). Badania takie najczęściej używają operatu imiennego, który pozwala nie tylko na personalizację przesyłki z kwestionariuszem, ale przede wszystkim nie obarcza członków gospodarstwa domowego koniecznością przeprowadzenia losowania ostatecznego respondenta. Interesujący wyjątek możemy zaobserwować we Francji, gdzie 18 razy (ISSP)

zastosowano technikę pocztową, wykorzystując operat adresowy. Jak można było się spodziewać, respondenci mieli problem z prawidłowym przeprowadzeniem losowania i miary reprezentatywności wykazały w tym przypadku bardzo słabe wyniki – temat ten porusza Publikacja 2.

Kwestionariusz do samodzielnego wypełniania najpopularniejszy był w Japonii, gdzie został użyty 16 razy w ramach ISSP. Dokumentacja nie wspomina, że kwestionariusz dołączony był do innego badania, więc można przypuszczać, że jest to związane z lokalną specyfiką kulturową, np. podejrzeniem silnego efektu ankieterskiego. Z drugiej strony warto odnotować, że w ramach EASS w tym kraju wykonano liczne badania łączące wywiad z modułem do samodzielnego wypełniania, więc możliwe, że raporty ISSP zawierają błędy w tym względzie. Problem rozstrzyga jednak prawdopodobnie opis w ISSP Study Description Form dla fali z 2017 roku w Japonii, który przedstawia *fieldwork method* jako: „Self-completion, arrives with interviewer, drops-off, pick up later”. Opis tej techniki sprawia problemy, gdyż w niektórych przypadkach *self-completion* mogło oznaczać ankietę pocztową, moduł do samodzielnego wypełniania dołączony do innej techniki lub technikę samodzielną, jak w przypadku japońskim. Za każdym razem rozstrzygnięcia poszukiwano poprzez analizę kontekstu zawartego w dokumentacji. Znacząca popularność tej techniki w Europie (37, ISSP) z dużym prawdopodobieństwem jest związana z nieścisłościami w dokumentacji, a realnie istotna część badań była tu wykonana jako moduł dołączony do wywiadu. W związku z tym w przypadku modelowania związku techniki z innymi zmiennymi warto łączyć te dwie kategorie (*self-completion* i schemat łączący tę technikę z wywiadem kwestionariuszowym). Dlatego można od razu zauważyć, że moduł dołączony do wywiadu najpopularniejszy jest w Europie (74, ISSP), Ameryce Północnej (18, ISSP) i Azji Wschodniej (14, Japonia – EASS i WVS i Korea, ISSP).

Zróznicowanie popularności wywiadu telefonicznego (CATI) staje się szczególnie widoczne, jeśli zestawimy badania komercyjne i badania mające charakter społeczny, często naukowy. O ile w pierwszym przypadku jest on niezwykle popularny ze względu na szybkość realizacji i niski koszt, o tyle w drugiej sytuacji występuje dosyć rzadko. Do przyczyn tego faktu należy zaliczyć głównie długość kwestionariusza w badaniach niekomercyjnych – godzinne badanie ciężko jest przeprowadzić przez telefon, a także problemy z reprezentatywnością zarówno losowania metodą *random digit dialing*, jak i z wykorzystaniem zakupionych od podmiotów komercyjnych operatorów abonentów telefonicznych (Kalsbeek i Agans 2008). Badanie metodą wywiadu telefonicznego było w uwzględnionych w bazie badaniach niekomercyjnych stosowane rzadziej niż ankietę pocztową (patrz wyżej). Jednocześnie w mającym charakter komercyjny programie GALLUP zastosowano je aż 1597 razy na 2280 sondaży. GALLUP korzysta z tej techniki we wszystkich krajach, w których penetracja usług telefonicznych wynosi przynajmniej 80%, oraz w tych, gdzie tradycyjnie wykorzystuje się tę technikę (GALLUP 2019), mowa więc głównie o krajach wysokorozwiniętych. Niekomercyjne sondaże telefoniczne przeprowadzano głównie w krajach europejskich (bez Europy Wschodniej). Jednocześnie z powodu ograniczeń związanych z pandemią COVID-19 w Ameryce Łacińskiej całą falę LAPOP 2021 zrealizowano w ten sposób. Po wyłączeniu danych GALLUPa najwięcej sondaży telefonicznych odnotować można w Niemczech (9, CSES, EES, INTUNE, WVS).

Ankieta internetowa stosowana jest głównie w krajach o szerokim rozpowszechnieniu Internetu, choć możliwe jest wykorzystanie metody *web-push* (gdy kontakt z respondentem przeprowadzony jest inną techniką niż właściwe zbieranie danych), w której zachęcony respondent mógłby udać się np. do biblioteki w celu wypełnienia kwestionariusza. Problemem jest także brak operatu adresów e-mail, jednak niektóre kraje stopniowo wprowadzają oficjalne

cyfrowe skrzynki pocztowe dla obywateli i z takiego operatu skorzystano w rundach 2018, 2019 i 2020 w ISSP w Danii. Niekiedy w programach porównawczych wykorzystywane są panele internetowe, np. w LAPOP w USA i w Kanadzie, a także w większości krajów uczestniczących w rundzie 2019 EES w Europie. W Ameryce Północnej technika ankiety internetowej była stosowana najczęściej – 12 razy w USA (CNEP, LAPOP, WVS) i 7 razy w Kanadzie (LAPOP, WVS).

Badania sondażowe w schemacie mieszania technik, w którym łączono wywiad kwestionariuszowy z inną techniką (ale z wyłączeniem kwestionariuszy do samodzielnego wypełnienia – patrz wyżej) były najmniej popularne we wszystkich analizowanych programach. Najczęściej stosowano tę technikę w Chinach – 12 razy, łącząc wywiad kwestionariuszowy z wywiadem telefonicznym (GALLUP). W Islandii (EVS, ISSP) i USA (CSES, ISSP) schemat ten – w połączeniu z różnymi technikami – pojawił się po 5 razy.

Schemat *mixed-mode* łączący techniki inne niż wywiad kwestionariuszowy zdecydowanie najpopularniejszy był w krajach nordyckich. Tego typu badania przeprowadzono 16 razy w Danii (ISSP), po 11 razy w Norwegii i Finlandii (oba ISSP), 8 razy w Szwecji (CSES, ISSP) i 6 razy na Islandii (ISSP). Popularne były także w Oceanii – 10 zrealizowano w Nowej Zelandii (CSES, ISSP, WVS) i 6 w Australii (ABS, CSES, ISSP, WVS). Przeprowadzono je również 6 razy w Kanadzie (CSES). 89% z tych badań wykorzystywało wśród technik ankietę pocztową. W zależności od tego, jak sformułowany był opis dostępny w dokumentacji, w tej kategorii mogły się znaleźć sondaże internetowe *web-push*, dlatego warto te kategorie analizować łącznie.

Tabela 4. Techniki sondażu wg wybranych programów. Skala kolorów w kolumnach

| Program | Brak info. | F2F  | Mail | Self-c. | CATI | Web | F2F +s-c | Mixed-mode (z f2f) | Mixed-mode (bez f2f) |
|---------|------------|------|------|---------|------|-----|----------|--------------------|----------------------|
| ISSP    | 38         | 541  | 137  | 61      |      | 4   | 96       | 10                 | 74                   |
| CSES    | 7          | 131  | 6    | 1       | 24   | 1   | 6        | 11                 | 22                   |
| WVS     | 27         | 252  | 9    |         | 2    | 4   | 1        | 3                  | 3                    |
| ABS     |            | 68   |      |         |      |     |          |                    | 1                    |
| EES     | 1          | 90   | 2    |         | 44   | 27  |          | 10                 |                      |
| EVS     | 1          | 136  | 1    |         |      | 1   |          | 8                  |                      |
| GALLUP  | 3          | 1597 |      |         | 662  |     |          | 18                 |                      |
| CNEP    | 2          | 40   |      |         | 4    | 13  | 1        | 1                  |                      |
| LITS    |            | 97   |      |         |      |     |          | 1                  |                      |
| LAPOP   | 7          | 230  |      |         | 24   | 14  |          |                    |                      |
| INTUNE  |            | 10   |      |         | 25   |     |          |                    |                      |

Jak pokazuje tabela 4, zdecydowanie największe zróżnicowanie technik sondażu występowało w programach ISSP i CSES, dlatego to one w największym stopniu mogą być wykorzystywane do porównywania efektów stosowania technik sondażu w modelach wykorzystujących dane obserwacyjne (P2a). W przypadku WVS znajdujemy przykłady stosowania niemal wszystkich technik, ale każda z tych kategorii jest bardzo nieliczna.

Badania wykorzystujące technikę pocztową i kwestionariusze do samodzielnego wypełnienia to domena ISSP. Wywiady telefoniczne, poza GALLUPem, były obecne w EES – stanowiły np. niemal całą rundę 2009. Runda 2021 programu LAPOP w Ameryce Łacińskiej została także z powodu pandemii w całości zrealizowana tą techniką. Wywiady telefoniczne w

ramach CSES były przeprowadzane niemal wyłącznie w Europie – począwszy od 1999 roku. W obu rundach INTUNE, 2007 i 2009, również dominowała ta technika.

Badania techniką ankiety internetowej najpopularniejsze były, jak już wspomniano, w rundzie 2019 EES, a także w badaniach LAPOP przeprowadzanych od 2008 roku w USA i w Kanadzie. Największe zróżnicowanie geograficzne sondaży internetowych występuje w CNEP – w programie tym przeprowadzono je w Azji, Ameryce i Europie.

Jeśli chodzi o sondaże *mixed-mode*, to jako nowość wprowadzono je w rundzie 2017 EVS, w większości łączyły tam wywiad kwestionariuszowy z ankietą pocztową i internetową jednocześnie. Były one także konsekwentnie obecne w CSES i ISSP.

#### III.1.4. Typ operatu i schemat losowania

Kierując się stopniem kompletności informacji zawartych w dokumentacji, za następną charakterystykę omawianych sondaży należy uznać operat losowania. Baza zawiera informacje o badaniach ze wszystkich kontynentów, które przeprowadzane były w dużej mierze w krajach rozwijających się. W większości przypadków w krajach takich nie jest nawet dostępny wiarygodny operat adresowy, a ograniczona penetracja usług telefonicznych jest argumentem przeciw zastosowaniu tam losowania numerów telefonu. W takiej sytuacji sondaż musi opierać się na losowaniu przestrzennym, które często ma charakter losowania fragmentu siatki na mapie (Harter i in. 2010). Sondaże tego typu zdecydowanie dominowały w bazie, stanowiąc 7243 przypadki, a więc dziesięciokrotność opisanych operatów imiennych czy adresowych. Należy zauważyć, że bardzo często ten typ próby był stosowany także przez EB, który zgodnie z nazwą jest realizowany w krajach europejskich, w większości wysokorozwiniętych.



Wykres 4. Stosowanie operatów imiennych i adresowych wg kontynentu

Porównanie częstości stosowania operatów imiennych i adresowych / gospodarstw domowych na poszczególnych kontynentach przedstawia wykres 4. Badania wykorzystujące ten pierwszy przeważały jedynie w Europie i Oceanii, gdzie indziej więcej było badań opierających się na operacie adresowym. Operat imienny bardzo dobrze nadaje się do badań metodą ankiety pocztowej, ze względu na możliwość personalizacji, która z jednej strony sprzyja zwiększeniu odsetka odpowiedzi, a z drugiej usuwa trudność samodzielnego losowania respondenta przez członków gospodarstwa domowego. Jak pokazuje podrozdział dotyczący technik sondażu, z możliwości tej korzystano często na obu kontynentach, ale szczególnie w Oceanii. Oceania jest też obszarem, na którym badania oparte na operacie imiennym przeważają liczebnie nad wszystkimi innymi łącznie. W Azji operat imienny stosowano głównie w Japonii, Korei, w Hongkongu i na Tajwanie. W Europie był wykorzystywany szczególnie często w krajach skandynawskich – w Danii 65 razy, w Szwecji 46 razy, a w

Norwegii 45. Kontynentami, na których operat adresowy był stosowany częściej niż losowanie przestrzenne, są Oceania i Ameryka Północna.

Operat imienny najczęściej wykorzystywany był w ramach ISSP (352) i ESS (117). W tym drugim programie operat imienny, co wyjątkowe, stosowano częściej niż adresowy (107). Operat adresowy i imienny były stosowane częściej niż losowanie przestrzenne w programach CSES, ESS i ISSP, natomiast EVS korzystał częściej z operatu adresowego niż z losowania przestrzennego. Losowanie z udziałem metody *random digit dialing*, zazwyczaj związane z techniką wywiadu telefonicznego, rzadko używane w programach o bardziej naukowej orientacji, szczególnie częste było w programie GALLUP (661). Względnie chętnie korzystano z niego także w EES (46), CSES (24), LAPOP (24) i INTUNE (18).

Jeśli chodzi o schematy losowania, to jedynym operatem, dla którego możliwe jest zastosowanie losowania jednostopniowego, jest operat imienny. Losowanie jednostopniowe (najczęściej prosta próba losowa) było stosowane istotnie często w przypadku programów ISSP (147), ESS (50) i CSES (37). Wykorzystywano je głównie w Europie: Dania, Finlandia, Szwecja – po 31 razy i Norwegia – 30 razy, a także w Oceanii: Australia – 25 razy, a Nowa Zelandia – 21. Losowanie jednostopniowe pojawiało się częściej niż wielostopniowe w badaniach wykorzystujących technikę ankiety pocztowej oraz schemat mieszania technik bez wywiadu kwestionariuszowego.

### **III.1.5. Technika sondażu a współczynniki realizacji**

Najniższy średni *response rate* notujemy w przypadku sondaży przeprowadzonych techniką wywiadu telefonicznego. Technika ta, odbierana przez część respondentów jako natarczywa i nieprzyjemna, wiąże się z bardzo wysokim odsetkiem odmów. Pozytywnym aspektem jest tu natomiast niezwykle niski koszt jednostkowego badania, co w połączeniu z często bardzo dużą próbą pozwala względnie łatwo osiągnąć założoną liczbę wywiadów.

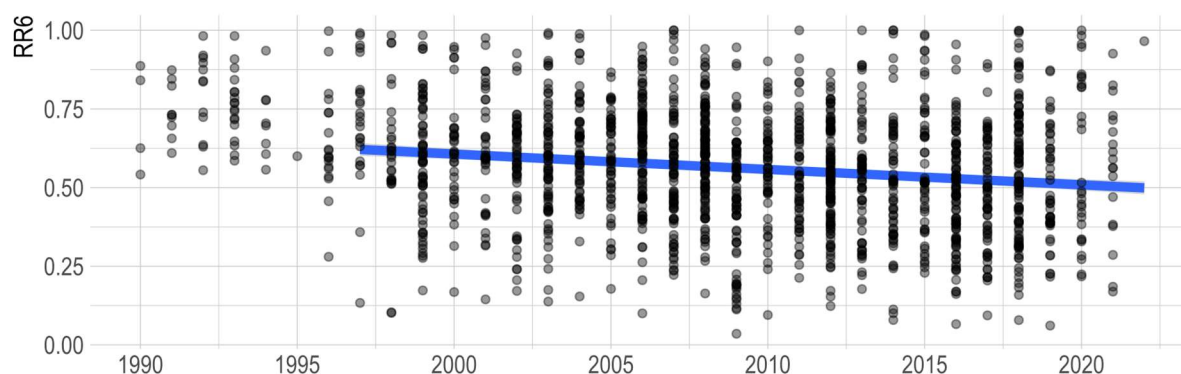
Jako drugie od końca pod omawianym względem plasują się sondaże stosujące schemat mieszania technik bez udziału wywiadu kwestionariuszowego. Taki *mixed-mode* jest wykorzystywany, jak wspomniano wcześniej, głównie w krajach nordyckich oraz w Oceanii i w Kanadzie, a więc na obszarach bardzo wysoko rozwiniętych, gdzie zawsze należy spodziewać się niższego *response rate*; przy czym w przypadku tej techniki jest to często rekompensowane lepszą reprezentatywnością (Rybak 2023). Najwyższy średni wskaźnik realizacji pojawia się w przypadku badań techniką wywiadu kwestionariuszowego lub kwestionariusza do samodzielnego wypełnienia, który jest dostarczony i odebrany przez ankietera (**P3a**). W obu tych przypadkach potencjalny respondent może być pod wpływem bezpośrednich, doskonalonych przez dekady badań sondażowych technik motywowania i konwersji osób odmawiających. Jest to niewątpliwa zaleta badań określanych w języku angielskim jako *face-to-face*. Średnie wartości wskaźników realizacji próby dla poszczególnych technik przedstawia tabela 5.

Ta prosta analiza nie wystarcza jednak do adekwatnego opisanego tego zjawiska, dlatego model opisujący związki między techniką sondażu, a *response rate* uwzględniający także inne czynniki przedstawiłem w Publikacji 2 i za nią w części III.2.2. autoreferatu.

Tabela 5. Średnie wartości wskaźników realizacji sondaży, dla których dało się je samodzielnie obliczyć, oraz liczba takich sondaży w podziale na techniki. Skala kolorów w kolumnach

| Technika                         | RR6  | COOP2 | REF3 | CON3 | Liczba |
|----------------------------------|------|-------|------|------|--------|
| F2F                              | 0,60 | 0,67  | 0,23 | 0,88 | 1192   |
| Mail                             | 0,49 | 0,79  | 0,14 | 0,65 | 130    |
| Self-c.                          | 0,60 | 0,67  | 0,22 | 0,89 | 46     |
| CATI                             | 0,37 | 0,44  | 0,43 | 0,85 | 54     |
| Web                              | 0,57 | 0,75  | 0,18 | 0,75 | 8      |
| F2F +s-c                         | 0,56 | 0,59  | 0,30 | 0,93 | 86     |
| Mixed-mode (z f2f), poza f2f+s-c | 0,49 | 0,61  | 0,30 | 0,83 | 33     |
| Mixed-mode (bez f2f)             | 0,46 | 0,73  | 0,13 | 0,68 | 93     |

Dodatkowo warto zwrócić uwagę na to, że literatura zgodnie dokumentuje postępujący od dekad spadek *response rates* osiąganych przez badania sondażowe na świecie (de Leeuw, Hox i Luiten 2018; Luiten, Hox i de Leeuw 2020). Ten trend potwierdza wykres 5, który przedstawia zmianę średniego wskaźnika realizacji w czasie, w oparciu o sondaże uwzględnione w bazie, dla których informacje były dostępne.



Wykres 5. Obliczony samodzielnie response rate (RR6) wg roku rozpoczęcia fieldworku (w tle regresja liniowa dla lat 2000–2022)

### III.1.6. Miary reprezentatywności – metaanalizy

Aby przedstawić zakres zebranych danych dotyczących wewnętrznego i zewnętrznego wskaźnika reprezentatywności, zamiast prezentacji standardowych tabel wybrałem wykonanie metaanaliz, stosując wewnętrzne i zewnętrzne kryterium reprezentatywności jako *effect sizes*. Opisana procedura dotyczy prostego porównania programów w dalszej części tej sekcji i metaanaliz w aneksie. Bardziej rozbudowane modele (ale oparte na wybranych programach) przedstawiam w Publikacji 2 i za nią w części III.2.2. autoreferatu.

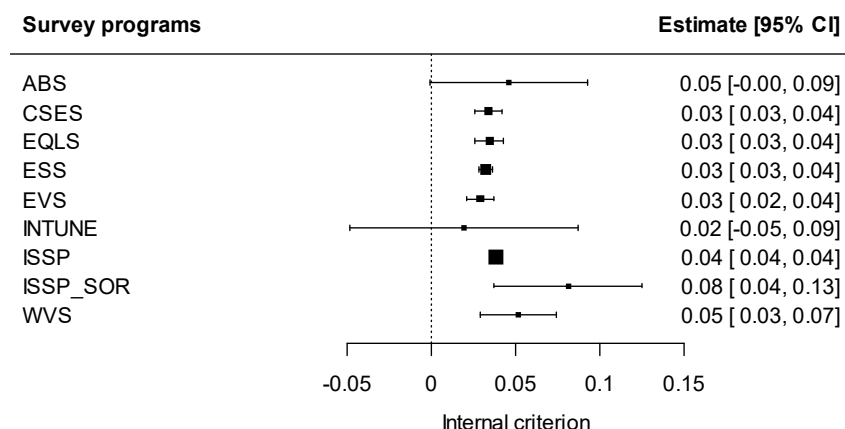
Korelacja Pearsona pomiędzy wewnętrznym i zewnętrznym wskaźnikiem reprezentatywności obliczonym dla całej bazy jest bardzo słaba ( $r(2758)=0,09$ ;  $p<0,000$ ). Jeśli ograniczymy się do badań, które według deklaracji zawartej w dokumentacji spełniają warunki doboru losowego, to wzrasta ona nieznacznie do ( $r(1060)=0,14$ ;  $p<0,000$ ). Tak niewielki związek pomiędzy obydwoma wskaźnikami reprezentatywności wymaga dalszych badań, szczególnie w odniesieniu do możliwości ich wymiennego stosowania. Korelacje pomiędzy obiema miarami a obliczonym *response rate* nie są istotne statystycznie.

Poniższe analizy są przeprowadzane na podzbiorze sondaży określonych jako zachowujące standardy doboru losowego. Jest ich 1134, jednak liczebności w poszczególnych modelach mogą się wahać ze względu na braki danych dotyczących innych zmiennych. Standardowa, dwupoziomowa metaanaliza z efektami losowymi (*random-effects*) i poprawką Knappa-Hartunga, wykorzystująca metodę REML do estymacji  $\tau^2$  i wewnętrzne kryterium reprezentatywności jako wielkość efektu, zidentyfikowała jego wartość na poziomie 0,036 (95% CI 0,033-0,038). Test heterogeniczności jest istotny statystycznie ( $Q(df=1042)=1775,9871$ ;  $p<0,000$ ).  $I^2$  („procent zmienności między wielkościami efektów, która występuje między badaniami w stosunku do całkowitej zmienności między wielkościami efektów”) (Card 2012:188–89) wynosi 42,67% i wartość tę można określić jako średnią (Higgins i in. 2003); oznacza ona, że heterogeniczność pomiędzy sondażami jest wystarczająca, żeby prowadzić bardziej pogłębione analizy.

Standardowa, dwupoziomowa metaanaliza z efektami losowymi (*random-effects*) i poprawką Knappa-Hartunga, w której stosuję metodę REML do estymacji  $\tau^2$ , używając zewnętrznego kryterium reprezentatywności jako wielkości efektu, pozwoliła określić jego średnią wartość na poziomie 0,043 (95% CI 0,041-0,044). Test heterogeniczności jest istotny statystycznie ( $Q(df=1080)=4493,6728$ ;  $p<0,000$ ).  $I^2$  wynosi 75.9%, a więc osiąga wartość wysoką, tak więc dalsze analizy należy uznać za wskazane.

### III.1.7. Reprezentatywność próby dla poszczególnych programów

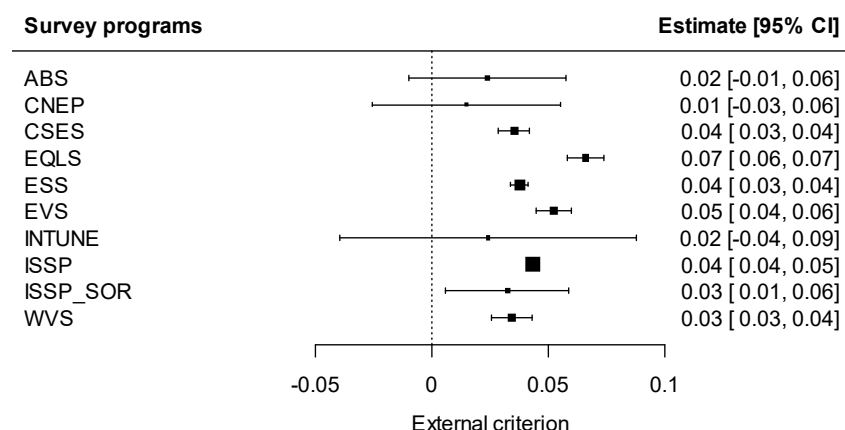
Metaanalizy w podgrupach wg programów przedstawione zostały na dwóch poniższych wykresach. Wykres 6 prezentuje porównanie wartości wewnętrznego wskaźnika reprezentatywności pomiędzy poszczególnymi programami sondażowymi. Górna granica przedziału ufności ma najmniejszą wartość w przypadku sondaży CSES, EQLS, ESS, EVS i ISSP – dla wszystkich wynosi 0,4. Im mniejsza wartość wskaźnika, tym spodziewana mniejsza wartość błędu wynikającego z braku odpowiedzi.



Wykres 6. Wewnętrzne kryt. reprezentatywności wg programu – wyniki metaanalizy w podgrupach

Wykres 7 przedstawia metaanalizę w grupach wg programów i szacuje wartość wskaźnika zewnętrznego, bazującego na danych populacyjnych. W tym przypadku najniższe wartości górnej granicy przedziału ufności prezentują programy CSES, ESS, WVS – dla wszystkich jest to 0,4; i ew. ISSP, dla którego wartość to 0,5. EVS i EQLS zostały ocenione dużo gorzej przez wskaźnik zewnętrzny niż wewnętrzny, a w przypadku WVS było odwrotnie. Zatem programami, które według oceny zewnętrznym i wewnętrznym wskaźnikiem reprezentatywności wypadają najlepiej, są CSES, ESS i ISSP (Dane z ESS i ISSP zostały

wykorzystane w Publikacji 2) (**P2b**). Jest to pozytywny wniosek dla dalszych badań, gdyż CSES i ISSP dają możliwość analizy wpływu dużo większego bogactwa rozwiązań metodologicznych niż EVS, EQLS i WVS. Programy, które w ocenie dokonanej przy użyciu obu wskaźników wypadły bardzo słabo lub całkiem niejednoznacznie, należałoby uwzględnić w analizach metodologicznych dopiero w dalszej kolejności.



Wykres 7. Zewnętrzne kryt. reprezentatywności wg programu – wyniki metaanalizy w podgrupach

### III.2. Kluczowe wyniki z publikacji

Poniżej przedstawiam najważniejsze rezultaty otrzymane w analizach przeprowadzonych w ramach prezentowanego cyklu publikacji (zbiorcza tabela kompletności zmiennych i druga, dotycząca porównania zakresu inwentaryzacji z zewnętrznymi zaleceniami, obie pochodzące z Publikacji 1, zostały już przedstawiona w sekcji III.1.2 autoreferatu).

#### III.2.1. Dokumentacja sondaży

Tabela 6 prezentuje jeden z nielicznych pozytywnych trendów dotyczących dokumentacji wśród opisanych w Publikacji 1. O ile ogromne różnice w jakości raportów metodologicznych, wybiórcze informowanie o części tematów czy bardzo powolne uaktualnianie formatu publikacji do współczesnych standardów technologicznych należy ocenić jednoznacznie negatywnie (**P1b**), o tyle widoczny jest w całej bazie postępujący z dekady na dekadę wzrost wskaźników kompletności dokumentacji (**P1c**). Taką prawidłowość potwierdzają także ustalenia innych badaczy (Jabkowski 2023; Kołczyńska i Schoene 2018). Wzrost ten jest szczególnie silny w przypadku takich charakterystyk jak użycie substytucji nieodpowiadających i kwot demograficznych, minimalnej liczby podejmowanych prób kontaktów z respondentem czy długości trwania wywiadu. Jediną informacją, której dostępność zmalała, jest liczba przeprowadzonych wywiadów. Po początkowym wzroście nastąpił także istotny spadek dostępności danych na temat liczebności kategorii nieodpowiadających.

*Tabela 6. Kompletność dokumentacji metodologicznej w podziale na dekady. Fragment bazy 1980–2019; skala kolorów w wierszach (za Publikacją 1)*

| Kompletność      | 1980 – 1989 | 1990 – 1999 | 2000 – 2009 | 2010 – 2019 |
|------------------|-------------|-------------|-------------|-------------|
| mod_info         | 0,96        | 0,96        | 0,98        | 1           |
| loage_info       | 0,93        | 0,9         | 0,96        | 0,98        |
| upage_info       | 0,93        | 0,89        | 0,96        | 0,97        |
| fra_info         | 0,83        | 0,81        | 0,95        | 0,97        |
| int_info         | 0,93        | 0,92        | 0,88        | 0,91        |
| sub quota_info   | 0,5         | 0,74        | 0,83        | 0,92        |
| cont_info        | 0           | 0,08        | 0,39        | 0,47        |
| long_info        | 0,07        | 0,06        | 0,24        | 0,41        |
| hhselection_info | 0,04        | 0,24        | 0,32        | 0,28        |
| check_info       | 0           | 0,15        | 0,22        | 0,18        |
| outcome_info     | 0           | 0,15        | 0,22        | 0,14        |
| rr_info          | 0,01        | 0,12        | 0,15        | 0,15        |
| lett_info        | 0           | 0,03        | 0,19        | 0,15        |
| inc_info         | 0           | 0,03        | 0,16        | 0,14        |
| paid_info        | 0           | 0,04        | 0,15        | 0,1         |
| work_info        | 0           | 0,03        | 0,06        | 0,07        |

W przeciwieństwie do dotychczasowych (opartych mniejszych zbiorach danych) ustaleń (Kołczyńska i Schoene 2018) należy odnotować, że sondaże europejskie cechują się najmniejszą średnią kompletnością dokumentacji (**P1d**), co widać w tabeli 7. Najwięcej informacji o kodowanych zmiennych znajdowało się w raportach z sondaży przeprowadzanych w Oceanii i Ameryce Północnej.

Na niski wynik sondaży europejskich w badaniu korzystającym z tak dużej bazy wpływ ma zdecydowanie największa liczba sondaży realizowanych na tym kontynencie (patrz sekcja III.1. autoreferatu). Duża ich część to programy charakteryzujące się niewielką kompletnością dokumentacji, takie jak np. Eurobarometr. Z kolei badania z Bliskiego Wschodu czy Ameryki Łacińskiej to z reguły efekt dużych programów globalnych (WVS czy ISSP), cechujących się akceptowalną kompletnością raportów i programów lokalnych (np. ARAB albo LAPOP), które trzymają pewien bazowy poziom. Okres analizy zdaje się nie odgrywać tutaj dużej roli, nawet przy uwzględnieniu jedynie danych od 2015 roku europejska dokumentacja cechuje się średnio najniższą kompletnością. Wyniki te mogą nasuwać wniosek o zasadności większej integracji wysiłków badawczych na naszym kontynencie, by nie trwonić czasu i środków na bardzo liczne badania o potencjalnie słabszej jakości.

Tabela 7. Kompletność dokumentacji metodologicznej wg kontynentu. Dane 1990+; skala kolorów w wierszach (za Publikacją 1)

| Kompletność      | Afryka | Azja | Europa | A. Łacińska i Karaiby | A. Północna | Oceania |
|------------------|--------|------|--------|-----------------------|-------------|---------|
| mod_info         | 0,99   | 0,99 | 0,99   | 0,99                  | 0,95        | 0,99    |
| loage_info       | 0,99   | 0,93 | 0,98   | 0,84                  | 0,79        | 0,91    |
| upage_info       | 0,99   | 0,93 | 0,98   | 0,84                  | 0,79        | 0,91    |
| fra_info         | 0,99   | 0,93 | 0,94   | 0,96                  | 0,88        | 0,98    |
| int_info         | 0,83   | 0,89 | 0,91   | 0,98                  | 0,97        | 0,93    |
| sub quota_info   | 0,84   | 0,77 | 0,89   | 0,82                  | 0,83        | 0,87    |
| cont_info        | 0,76   | 0,68 | 0,27   | 0,63                  | 0,6         | 0,8     |
| long_info        | 0,92   | 0,59 | 0,17   | 0,45                  | 0,28        | 0,32    |
| hhselection_info | 0,22   | 0,32 | 0,27   | 0,38                  | 0,59        | 0,68    |
| check_info       | 0,25   | 0,25 | 0,16   | 0,24                  | 0,41        | 0,48    |
| outcome_info     | 0,04   | 0,18 | 0,17   | 0,13                  | 0,4         | 0,56    |
| rr_info          | 0,17   | 0,16 | 0,14   | 0,09                  | 0,29        | 0,28    |
| lett_info        | 0,11   | 0,14 | 0,14   | 0,11                  | 0,33        | 0,47    |
| inc_info         | 0,11   | 0,13 | 0,13   | 0,14                  | 0,34        | 0,44    |
| paid_info        | 0,05   | 0,14 | 0,1    | 0,12                  | 0,18        | 0,13    |
| work_info        | 0,01   | 0,04 | 0,07   | 0,03                  | 0,02        | 0,15    |
| <b>Średnia</b>   | 0,52   | 0,50 | 0,46   | 0,48                  | 0,54        | 0,62    |

Tabela 8 przedstawia średnie kompletności poszczególnych zmiennych dotyczące dokumentacji metodologicznej ze wszystkich 30 programów sondażowych analizowanych w Publikacji 1. Najkompletniej udokumentowane programy to ESS, EQLS i EVS. Mają one jednak zasięg ograniczający się niemal wyłącznie do Europy. Najlepiej udokumentowane programy obejmujące po kilka kontynentów to CSES i ISSP (oraz ISSP-SOR) (**P1e**). Dane z sondaży ESS i ISSP uwzględnione są w poniższych metaanalizach.

Tabela 8. Kompletność dokumentacji metodologicznej wg programu (Za Publikacją 1)

| Program  | mod | loage | upage | fra | int | sub quota | cont | long | hhselect | check | outcome | rr  | lett | inc | paid | work | Średnia |
|----------|-----|-------|-------|-----|-----|-----------|------|------|----------|-------|---------|-----|------|-----|------|------|---------|
| AB       | 1   | 1     | 1     | 1   | 0,4 | 0,7       | 0,3  | 1    | 0,3      | 0,7   | 0       | 0,6 | 0,3  | 0,3 | 0    | 0    | 0,55    |
| ABS      | 1   | 0,2   | 0,2   | 0,3 | 0,3 | 0,2       | 0,1  | 0,9  | 0,2      | 1     | 0       | 0,1 | 0    | 0   | 0    | 0,2  | 0,29    |
| ARAB     | 1   | 1     | 1     | 1   | 1   | 0         | 0    | 1    | 0,9      | 1     | 0       | 0,5 | 0    | 0   | 0    | 0    | 0,52    |
| ASES     | 1   | 1     | 1     | 1   | 0   | 0,1       | 0,2  | 0    | 0,9      | 0,1   | 0       | 0   | 0    | 0   | 0    | 0    | 0,33    |
| ASIAB    | 1   | 1     | 1     | 1   | 1   | 0,1       | 0,1  | 0    | 0,7      | 0,1   | 0,1     | 0   | 0    | 0   | 0    | 0    | 0,38    |
| CCEB     | 1   | 1     | 1     | 1   | 1   | 0         | 0    | 0    | 0        | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,31    |
| CCS      | 1   | 1     | 1     | 1   | 1   | 1         | 0,2  | 0    | 1        | 0     | 0,8     | 1   | 0    | 0   | 0    | 0    | 0,56    |
| CEEB     | 1   | 1     | 1     | 0   | 1   | 1         | 0    | 0    | 0,1      | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,32    |
| CNEP     | 1   | 0,2   | 0,2   | 0,8 | 1   | 0,1       | 0    | 0    | 0,6      | 0     | 0       | 0,8 | 0    | 0   | 0    | 0,1  | 0,30    |
| CONS     | 1   | 1     | 0     | 0   | 1   | 0         | 0    | 0    | 1        | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,25    |
| CRRC     | 1   | 0,8   | 0,8   | 1   | 1   | 0,1       | 0,7  | 0,4  | 1        | 0,6   | 0,1     | 0,8 | 0,2  | 0   | 0    | 0,3  | 0,54    |
| CSES     | 1   | 0,9   | 0,9   | 0,9 | 0,9 | 0,7       | 0,8  | 0    | 0,9      | 0,6   | 0,7     | 0,7 | 0,9  | 0,9 | 0    | 0,3  | 0,69    |
| EASS     | 1   | 0,8   | 0,8   | 0,8 | 1   | 0         | 0    | 0    | 0        | 0     | 0,5     | 1   | 0    | 0   | 0    | 0    | 0,36    |
| EB       | 1   | 1     | 1     | 1   | 0,9 | 1         | 0    | 0    | 0,1      | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,37    |
| ECS      | 1   | 1     | 1     | 0,5 | 1   | 0,3       | 0    | 0    | 0,2      | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,31    |
| EES      | 1   | 1     | 1     | 0,3 | 0,9 | 0,1       | 0,6  | 0,4  | 0,4      | 0,2   | 0,2     | 0,4 | 0    | 0   | 0    | 0,6  | 0,45    |
| EQLS     | 1   | 1     | 1     | 0,8 | 1   | 1         | 1    | 0,8  | 0,8      | 0,8   | 1       | 0,7 | 1    | 0,3 | 0    | 1    | 0,82    |
| ESS      | 1   | 1     | 1     | 1   | 1   | 1         | 1    | 1    | 1        | 1     | 1       | 1   | 1    | 1   | 1    | 1    | 0,99    |
| EUNB     | 1   | 1     | 1     | 1   | 0   | 0         | 0    | 1    | 1        | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,38    |
| EVS      | 1   | 1     | 1     | 0,9 | 1   | 0,7       | 0,7  | 1    | 0,6      | 0,7   | 0,7     | 0,7 | 0,5  | 0,5 | 0,7  | 0    | 0,74    |
| GALLUP   | 1   | 1     | 1     | 1   | 1   | 1         | 1    | 1    | 0        | 0     | 0       | 0   | 0    | 0   | 0    | 0    | 0,50    |
| INTUNE   | 1   | 1     | 1     | 1   | 1   | 0,1       | 0,8  | 0    | 0,9      | 0     | 0       | 0,9 | 0    | 0   | 0    | 0    | 0,49    |
| ISSP     | 1   | 0,9   | 0,9   | 0,9 | 0,9 | 0,8       | 0,7  | 0    | 0,9      | 0,8   | 0,8     | 0,4 | 0,7  | 0,6 | 0,6  | 0    | 0,68    |
| ISSP-SOR | 0,9 | 0,9   | 0,9   | 0,9 | 0,9 | 0,9       | 0,6  | 0    | 0,9      | 0,9   | 0,9     | 0   | 0,9  | 0,9 | 0,9  | 0    | 0,72    |
| LAPOP    | 1   | 0,6   | 0,6   | 0,9 | 1   | 0,6       | 0,1  | 0    | 0,6      | 0,3   | 0       | 0,1 | 0    | 0,1 | 0    | 0    | 0,36    |
| LITS     | 1   | 1     | 1     | 1   | 1   | 0,4       | 1    | 0    | 1        | 0     | 0,3     | 0,7 | 0,4  | 0,4 | 0    | 0    | 0,56    |
| MEDW     | 0,7 | 0,7   | 0,3   | 0   | 0,7 | 0         | 0,7  | 0    | 0,7      | 0     | 0,7     | 0,7 | 0,7  | 0,7 | 0    | 0    | 0,41    |
| NBB      | 1   | 1     | 1     | 1   | 1   | 0,2       | 0,4  | 0    | 1        | 0,3   | 0,5     | 0,9 | 0    | 0   | 0    | 0    | 0,52    |
| NEB/NDB  | 0,5 | 0,4   | 0,5   | 0,5 | 0,5 | 0         | 0    | 0    | 0        | 0,4   | 0       | 0   | 0    | 0   | 0    | 0    | 0,17    |
| WVS      | 0,9 | 0,9   | 0,9   | 0,8 | 0,9 | 0,7       | 0,6  | 0,1  | 0,8      | 0,6   | 0,4     | 0,2 | 0,1  | 0   | 0,5  | 0,2  | 0,53    |

### III.2.2. Reprezentatywność sondaży

Publikacja 2 szczegółowo opisuje relacje pomiędzy technikami sondażu, a *response rates* i reprezentatywnością próby. Procedury obliczenia poniższych modeli opisane są w tej publikacji. Model metaregresji, prezentujący relacje pomiędzy charakterystykami sondażu a współczynnikiem realizacji (RR6), zamieściłem w tabeli 9. W modelu obserwacje są krzyżowo

zagnieżdżone (Fernández-Castilla i in. 2019) w falach badania i krajach realizacji, co pozwala częściowo kontrolować (za pomocą tzw. *random intercepts*) efekt wspólnej dla fali tematyki kwestionariusza i *survey climate* w danym kraju (niestety bez uwzględnienia zmian w czasie). Stosuje on uznany za optymalny w procedurze testowania (opisanej w Publikacji 2) zestaw zmiennych niezależnych, obejmujący cztery techniki sondażu, operat dzielony binarnie na imienny i resztę losowych, dwa programy porównawcze i liczbę łącznych historycznych realizacji sondażu w danym kraju i w danym programie (doświadczenie instytucjonalne). Zawarte są w nim także interakcje pomiędzy techniką a operatem – w celu uchwycenia moderującego efektu tego drugiego na relację między techniką sondażu a *response rate* (a w następnych modelach reprezentatywności próby). Technika opisana jako sondaż *mixed-mode* z udziałem wywiadów kwestionariuszowych w zdecydowanej większości dotyczy badań z kwestionariuszem do samodzielnego wypełnienia dołączonym do wywiadu, dlatego druga kategoria *mixed-mode* jest w tym przypadku istotniejsza.

W przypadku stosowania operatu imiennego schemat mieszania technik bez wywiadu kwestionariuszowego skutkuje istotnie niższym wskaźnikiem realizacji, niż wywiad *face-to-face*. W sytuacji użycia operatu innego niż imienny technika ankiety pocztowej wiąże się z istotnie niższym *response rate*, niż wywiad bezpośredni (**P3a**). Kolejne modele pokażą, że związek technik z reprezentatywnością może mieć inny kierunek, niż by na to wskazywały wskaźniki realizacji. ESS charakteryzuje się wyższym wskaźnikiem realizacji niż ISSP.

*Tabela 9. Krzyżowo zagnieżdżona metaregresja szacująca związek pomiędzy charakterystykami badania a response rate (RR6). \* $p < 0,05$ ; \*\* $p < 0,01$ ; \*\*\* $p < 0,001$ .*

*N=795 sondaży, 2 434 078 jednostek do obliczenia wariancji wielkości efektu (Suplement z Publikacji 2)*

| Effect size = RR6                             | coef   | se    | tval   | df  | pval  | ci.lb  | ci.ub  |     |
|-----------------------------------------------|--------|-------|--------|-----|-------|--------|--------|-----|
| Intercept                                     | 0,584  | 0,026 | 22,237 | 22  | <,001 | 0,529  | 0,638  | *** |
| Mixed-mode (with f2f)                         | 0,013  | 0,021 | 0,602  | 785 | 0,548 | -0,029 | 0,054  |     |
| Mail mode                                     | 0,024  | 0,021 | 1,154  | 785 | 0,249 | -0,017 | 0,065  |     |
| Mixed-mode (no f2f)                           | -0,048 | 0,02  | -2,339 | 785 | 0,02  | -0,088 | -0,008 | *   |
| Nonindividual frame                           | -0,015 | 0,015 | -1     | 785 | 0,318 | -0,045 | 0,015  |     |
| ESS                                           | 0,063  | 0,02  | 3,089  | 22  | 0,005 | 0,021  | 0,105  | **  |
| Repetitions                                   | -0,001 | 0,001 | -0,958 | 785 | 0,338 | -0,003 | 0,001  |     |
| Mixed-mode (with f2f)*<br>Nonindividual frame | 0,035  | 0,027 | 1,292  | 785 | 0,197 | -0,018 | 0,089  |     |
| Mail mode*<br>Nonindividual frame             | -0,165 | 0,03  | -5,564 | 785 | <,001 | -0,223 | -0,107 | *** |
| Mixed-mode (no f2f)*<br>Nonindividual frame   | 0,039  | 0,059 | 0,66   | 785 | 0,509 | -0,077 | 0,155  |     |

Dwa najważniejsze spośród modeli metaregresji zaprezentowanych w Publikacji 2 zamieszczono w tabeli 10. Miarą użytą jako *effect size* jest ich przypadku wskaźnik oparty o wewnętrzne kryterium reprezentatywności. Pierwszy z nich stosuje zestaw zmiennych niezależnych taki, jak w poprzednim modelu z tabeli 9. Drugi model, ukierunkowany na estymację efektów cząstkowych techniki przy stałym poziomie zmiennej *response rate*, zawiera dodatkowo zmienną z obliczoną wartością współczynnika RR6. Oba modele są krzyżowo zagnieżdżone w falach i krajach, jak opisano powyżej. Modele są ogólnie zgodne ze sobą w wynikach, dlatego opisuję je łącznie.

Oba modele szacują, że w przypadku operatu imiennego wszystkie techniki osiągają istotnie niższy wskaźnik błędu braku odpowiedzi niż wywiad kwestionariuszowy (**P3b**). Oznacza to, że w takiej sytuacji schemat *mixed-mode* bez udziału ankietera wiąże się z niższym wskaźnikiem realizacji, ale wyższą reprezentatywnością.

Jeśli chodzi o operat losowy inny niż imienny (**P3c**), jedyną techniką osiągającą wynik istotnie gorszy od wywiadu kwestionariuszowego, jest ankieta pocztowa. Poprzedni model pokazał, że w takiej sytuacji osiąga ona także niższy *response rate* niż *face-to-face*. Jest to spowodowane głównie przez wyjątkowo nieudane ankiety pocztowe z ISSP we Francji, stanowiące 16 z 28 obserwacji w tej kategorii. W celu zilustrowania tego zagadnienia zamieszczam poniżej instrukcję samodzielnego losowania respondenta z ISSP2017. Zgodnie z oczekiwaniami ESS osiąga istotnie niższy wskaźnik błędu niż ISSP, a każda kolejna realizacja krajowa w obrębie programu istotnie podnosi reprezentatywność.

**Merci de transmettre au plus vite le questionnaire à la personne qui doit répondre : c'est la personne majeure de votre foyer, dont l'anniversaire est le plus tôt dans l'année, à partir du 1er janvier 2017.**

- Si vous êtes né-e en **août**, votre conjoint en **mars**, votre fille en **février**, c'est bien votre fille, si elle est majeure, qui doit répondre (*même si elle ne réside pas en permanence à votre domicile*).
- Si elle est mineure, c'est votre conjoint qui doit répondre.

*Rysunek 1. Instrukcja losowania wewnątrzgrupowego w ISSP 2017 we Francji, przeprowadzonego techniką ankiety pocztowej (ISSP Research Group 2019)*

*Tabela 10. Krzyżowo zagnieżdżona metaregresja szacująca związek pomiędzy charakterystykami badania a wewnętrznym wskaźnikiem reprezentatywności. \* $p < 0,05$ ; \*\* $p < 0,01$ ; \*\*\* $p < 0,001$ .  $N = 791$  sondaży, 301 406 respondentów w subpopulacjach do obliczenia wariacji wielkości efektu (za Publikacją 2)*

|                                               | Model II (tutaj jako pierwszy) |        |        |     | Model III (tutaj jako drugi) |        |        |     |
|-----------------------------------------------|--------------------------------|--------|--------|-----|------------------------------|--------|--------|-----|
| Effect size = Int. Crit.                      | coef                           | ci.lb  | ci.ub  |     | coef                         | ci.lb  | ci.ub  |     |
| Intercept                                     | 0,051                          | 0,041  | 0,060  | *** | 0,066                        | 0,051  | 0,081  | *** |
| Mixed-mode (with f2f)                         | -0,010                         | -0,021 | 0,000  | *   | -0,010                       | -0,021 | 0,000  | *   |
| Mail mode                                     | -0,019                         | -0,028 | -0,009 | *** | -0,019                       | -0,029 | -0,010 | *** |
| Mixed-mode (no f2f)                           | -0,017                         | -0,027 | -0,008 | *** | -0,020                       | -0,029 | -0,010 | *** |
| Nonindividual frame                           | 0,001                          | -0,005 | 0,008  |     | 0,002                        | -0,005 | 0,008  |     |
| Mixed-mode (with f2f)*<br>Nonindividual frame | 0,008                          | -0,006 | 0,022  |     | 0,009                        | -0,005 | 0,023  |     |
| Mail mode*<br>Nonindividual frame             | 0,070                          | 0,056  | 0,083  | *** | 0,065                        | 0,051  | 0,079  | *** |
| Mixed-mode (no f2f)*<br>Nonindividual frame   | 0,009                          | -0,015 | 0,033  |     | 0,009                        | -0,014 | 0,033  |     |
| ESS                                           | -0,020                         | -0,028 | -0,012 | *** | -0,020                       | -0,027 | -0,012 | *** |
| Repetitions                                   | -0,001                         | -0,001 | 0,000  | *** | -0,001                       | -0,001 | 0,000  | *** |
| Response rate (RR6)                           |                                |        |        |     | -0,025                       | -0,043 | -0,008 | **  |
| Pseudo R2                                     | 0,67                           |        |        |     | 0,69                         |        |        |     |

### III.2.3. Preferencja techniki sondażu

Publikacja 3 wskazuje, że możliwe są różne działania korzystające z preferowania przez respondenta konkretnych technik, aby podnieść jakość pomiaru sondażowego. Droga prostego

oferowania respondentowi wyboru nie wydaje się jednak dobrym rozwiązaniem. Po pierwsze prawdopodobnie zmniejsza to *response rate* (Medway i Fulton 2012). Po drugie już sam kontakt z respondentem odbywa się w jakiejś technice i np. nieufna osoba nie otworzy drzwi ankieterowi, który mógłby jej zaoferować kwestionariusz do samodzielnego wypełnienia albo kontakt do ankietera CATI. Osoba, która nie znosi papierowej korespondencji, nie doczyta w liście, że kwestionariusz można też wypełnić w Internecie.

W sytuacji ponowienia kontaktu z respondentami z wcześniejszego sondażu, jeśli zadano im pytanie o preferowaną technikę, można przydzielić ich do technik według deklaracji. Zauważono jednak wpływ techniki pytania na treść deklaracji na ten temat (Groves i Kahn 1979; Tarnai i Paxson 2004) – może to być skutek efektu selekcji (osoby silnie preferujące inną technikę nie wzięły udziału w badaniu) albo niekiedy uprzejmości względem ankietera (respondent może chcieć uniknąć sugestii, że rozmowa z ankieterem jest nieprzyjemna).

Optymalnym rozwiązaniem wydaje się stworzenie modeli predykcyjnych opartych na deklaracjach (Smyth, Olson i Millar 2014) i obserwacjach wyborów dokonywanych przez respondentów (Haan, Ongena i Aarts 2014), których położenie społeczno-demograficzne znamy. Należałoby także uwzględnić wiedzę z dziedziny nauk o poznaniu, dotyczącą łatwości przetwarzania informacji w trybie audialnym i wizualnym, opisując to zagadnienie w Publikacji 3, opierając się na psychologicznych teoriach udziału w sondażu (w tym do koncepcji *satisficingu*) (**P3d**) (Krosnick 1991; Tourangeau i in. 2000) oraz kognitywistycznej teorii pamięci roboczej (**P3e**) (Clearman, Klinger i Szűcs 2017; Constantinidou i Baker 2002; Kumar i Priyadarshi 2013; Miller i Bichsel 2004; Olive, Kellogg i Piolat 2008; Pham i Hasson 2014; Tahmasbipour i Ahmady 2011).

Przeprowadzane w Publikacji 4 analizy odniosły się do wspomnianych problemów poprzez zestawienie i porównanie preferencji mierzonych za pomocą deklaracji oraz zachowań (tu określonych jako behawioralne dane o preferencji, w literaturze niekiedy określanych jako *mode choice*) w ramach jednego modelu wielomianowej regresji logistycznej.

Tabela 11 przedstawia model, który – korzystając z danych behawioralnych z eksperymentu metodologicznego w ramach ESS7 i pytania o deklarację z ESS 8 – szacuje prawdopodobieństwa preferowania techniki ankiety pocztowej i internetowej wśród polskich respondentów (**P3f**). W przypadku kobiet bardziej prawdopodobne jest preferowanie techniki pocztowej. Ankieta pocztowa jest także istotnie częściej preferowana wraz ze wzrostem wielkości miejsca zamieszkania respondenta; podobnie jest w przypadku ankiety internetowej. Interesująca jest tu krzywoliniowa zależność od wieku – największe prawdopodobieństwo preferowania techniki pocztowej występuje wśród osób w wieku 15–19 i 35–64 lat. Z kolei upodobania osób w wieku 20–24 i 65–74 lat nie różnią się tu istotnie od preferencji respondentów z kategorii wiekowej 75+.

Ustalenia dotyczące tej starszej grupy wiekowej zgodne są z wynikami eksperymentów przedstawionych w Publikacji 3, sugerujących, „że osoby „bardzo stare” [70+] (...) będą chętniej i lepiej uczestniczyły w sondażach prowadzonych techniką operującą na sygnałach słuchowych, natomiast osoby w wieku średnim i we „wczesnej starości” łatwiej poradzą sobie w badaniach wykorzystujących wizualny kanał komunikacji.” (**P3e**). Ze względu na odwołanie do *satisficingu* traktuję trudność zadania udziału w sondażu w danej technice jako predyktor preferencji.

W przypadku ankiety internetowej zależność ma prostszy kształt, największe prawdopodobieństwo jej preferowania dotyczy osób w wieku 15–24 lat, natomiast osoby w przedziale wiekowym 55–74 lat nie różnią się istotnie w tym względzie od jeszcze starszych. Przy czym pochodzące z obserwacji zachowań informacje na temat preferencji różnią się istotnie od danych uzyskanych w wyniku zbierania deklaracji (**P3g**). W rzeczywistości szanse

na wybranie ankiety pocztowej były prawie dziesięciokrotnie większe, niż wynikałoby to z deklaracji.

*Tabela 11. Wielomianowa regresja logistyczna: współczynniki ryzyka względnego preferencji techniki pocztowej i internetowej w stosunku do face-to-face w zależności od charakterystyk respondenta i rodzaju pomiaru preferencji. N=1918 \*\*  $p < 0,01$ ; \*  $p < 0,05$ ; '  $p < 0,051$  (za Publikacją 4)*

| <b>Model (base=F2F)</b>                  | <b>Mail RRR</b> |    | <b>Web RRR</b> |    |
|------------------------------------------|-----------------|----|----------------|----|
| Intercept                                | 0,011           | ** | 0,008          | ** |
| <b>Female</b>                            | 2,149           | ** | 1,196          |    |
| <b>Age (base = 75+)</b>                  |                 |    |                |    |
| 15-19                                    | 3,012           | *  | 21,56          | ** |
| 20-24                                    | 2,429           |    | 27,99          | ** |
| 25-34                                    | 2,669           | *  | 14,92          | ** |
| 35-44                                    | 2,907           | *  | 12,12          | ** |
| 45-54                                    | 3,758           | ** | 6,944          | *  |
| 55-64                                    | 2,903           | ** | 4,095          |    |
| 65-74                                    | 2,198           |    | 1,374          |    |
| <b>Urbanicity (base = village)</b>       |                 |    |                |    |
| city <10k                                | 1,330           |    | 0,969          |    |
| city 10-19k                              | 1,859           | *  | 2,069          | *  |
| city 20-49k                              | 1,950           | ** | 1,479          |    |
| city 50-99k                              | 2,456           | ** | 3,517          | ** |
| city 100-199k                            | 1,806           | *  | 2,744          | ** |
| city 200-499k                            | 3,186           | ** | 3,363          | ** |
| city 500k+                               | 3,547           | ** | 7,164          | ** |
| <b>Type of data (base = declarative)</b> |                 |    |                |    |
| Behavioral data                          | 9,302           | ** | 0,620          | '  |

#### IV. Dyskusja

Praktyki sondażowe jako materiał badań mogą niekiedy dostarczać informacji bardziej interesujących niż same wyniki sondaży. Z jednej strony pozwalają przyjrzeć się temu, jak instytucje badawcze próbują dostosować się do lokalnych warunków, by możliwie starannie, jednak w ramach własnych możliwości materialnych zdobyć wiedzę o życiu interesujących ich populacji. Z drugiej strony możemy obserwować reakcje ludzi na różne metody i techniki uzyskiwania od nich interesujących badacza informacji.

Powyższy autoreferat i poniższe publikacje służą poszukiwaniu odpowiedzi na pytania o charakter historycznych i obecnych praktyk dokumentacyjnych w sondażowych programach międzykrajowych, oraz o wpływ opisanych w dokumentacji charakterystyk sondażu na jakość pomiaru. Udzielenie odpowiedzi wymagało uprzedniego zinventaryzowania samych praktyk sondażowych, co zostało uwypuklone w autoreferacie. Przedstawione analizy mają jednak liczne ograniczenia, których opis poprzedza wnioski płynące z serii publikacji.

Ocena praktyk dokumentacyjnych oparta była na danych zebranych na potrzeby opisu ogólnych praktyk sondażowych w celu stworzenia modeli badających wpływ charakterystyk sondażu na jakość pomiaru. Dobór zmiennych wykorzystanych do obliczenia wskaźników kompletności dokumentacji podporządkowany był zatem właśnie temu celowi analitycznemu. Z jednej strony opierałem się w tym procesie na literaturze metodologicznej, uwzględniając te

charakterystyki, które badacze określili jako istotne dla oceny reprezentatywności badania. Z drugiej strony brałem pod uwagę zazwyczaj te zmienne, które pojawiały się w dokumentacji więcej niż jednego programu.

Analiza ilościowa wymagała w tym przypadku przyjęcia arbitralnych założeń dotyczących (braku) różnicy w wadze pewnych informacji. Różne charakterystyki bowiem mają różną wagę dla odmiennych użytkowników danych. Badacz oceniający reprezentatywność zwróci uwagę na inne cechy badania niż dziennikarz, ustawodawca czy analityk substancjalnych wyników. Starając się, mimo wymienionych ograniczeń, zapewnić możliwie szerokiemu gronu odbiorców możliwość wykorzystania moich analiz, nie różnicowałem wagi poszczególnych charakterystyk (i udostępniłem całość bazy do własnych porównań).

Należy zwrócić uwagę na problemy, które wymykają się możliwości prostego ilościowego podsumowania. Analizowana dokumentacja metodologiczna wykazywała diametralne zróżnicowanie pod wieloma względami. W przypadku oceniania takich cech raportowania jak jednolitość (tzn. czy jest jeden, lub więcej raportów wspólnych dla całej fali, czy nawet realizacji krajowej) pojawia się pytanie, jak inaczej niż jakościowo podsumować różnice między projektami typu ESS czy ISSP, których główny raport metodologiczny zazwyczaj przygotowany jest przez instytucję koordynującą, z CSES, który udostępnia wypełnione (niekiedy odrębnie) kwestionariusze metodologiczne dla każdego kraju i osobne niestandardyzowane omówienie doboru próby w poszczególnych krajach.

Szczególnie problematyczne są przypadki, kiedy niejednolita dokumentacja wzajemnie sobie przeczy (np. liczebność zrealizowanej próby w EVS99 dla Austrii i Chorwacji czy EVS08 dla Francji). Czy traktować to jako brak informacji, czy przyjąć bardziej negatywną ocenę? A może należy spróbować skontaktować się z instytucją i wyjaśnić przyczynę różnic? Jak ustosunkować się do sytuacji, kiedy informacja jest podana, ale wyraźnie błędna – przykład CSES3 i 4 w Tajlandii, gdzie podane liczebności kategorii nieodpowiadających przeczą obliczonemu *response rate*? W Publikacji 1, ze względu na to, że proces zbierania danych był podporządkowany opisowi praktyk, stosowałem zasadę pewnej naiwności i kodowałem informację (lub jedną z informacji w sytuacji sprzeczności między dokumentami), która była dostępna.

Narzuca to jako podstawowy wniosek potrzebę dążenia do większej standaryzacji, optymalnie wypracowanej drogą kompromisu w środowisku badaczy. Pożądanym stanem docelowym, po opracowaniu i wdrożeniu jednego standardu raportu, byłoby zawarcie informacji o całej fali w jednym dokumencie, podzielonym na rozdziały dotyczące kolejnych realizacji krajowych (na wzór ESS), w formacie umożliwiającym automatyzację zbierania danych. W sytuacji, kiedy pełne spektrum możliwości architektury .xml, zaproponowanej przez Data Documentation Initiative (Van Wettère 2022), nie zostało wykorzystane przez żaden z projektów, już samo publikowanie raportów w skoroszytach .csv ogromnie ułatwiłoby automatyzację. Istotne byłoby tu powszechne przyjęcie jednolitych nagłówek poprzedzających poszczególne informacje o charakterystykach. W obecnych warunkach ręczne kodowanie, abstrahując od jego niezwyklej czasochłonności, skutkuje licznymi błędami, które trzeba wyłapywać w kolejnych fazach kontroli. Proces ten jest ważny nie tylko w kontekście opisu dokumentacji czy praktyk, ale przede wszystkim dla dokonywania metodologicznych analiz *ex post*, dotyczących związku pomiędzy charakterystykami badania a różnymi miarami jakości pomiaru. Wskazane mogłoby być opracowanie nowych raportów dla starszych badań – ale dopiero po pełnej implementacji jednolitego standardu – na wzór dosyć powszechnie przyjętych kategorii nierespondentów wyszczególnionych przez AAPOR (The American Association for Public Opinion Research 2023) i stosowanych w większości programów.

Wśród pozytywnych wniosków należy wymienić fakt, że zbiorcza kompletność raportów metodologicznych zwiększa się z dekady na dekadę. Obserwację tę potwierdzają także ustalenia innych badaczy (Jabkowski 2023; Kołczyńska i Schoene 2018). Dodatkowo należy wyróżnić te programy, które charakteryzują się największą kompletnością informacji w raportach. Są to ESS, EQLS i EVS w Europie oraz CSES i ISSP na świecie. Te programy, charakteryzujące się również najwyższą reprezentatywnością wyników, powinny być traktowane priorytetowo zarówno w analizach metodologicznych, jak i substancjalnych. Warto wspomnieć, że moja analiza, oparta na kilkukrotnie większym zbiorze danych, przyniosła odmienny niż wcześniejsza (Kołczyńska i Schoene 2018) wynik w zakresie globalnego zróżnicowania kompletności raportów. Cytowani badacze wskazali sondaże europejskie jako najlepiej udokumentowane, natomiast według moich badań dokumentacja europejska wykazywała najwięcej braków, a najlepsze pod tym względem okazały się Oceania i Ameryka Północna. Jest to związane z występującą na naszym kontynencie inflacją badań, wśród których istotna część (jak np. EB) spełnia wyłącznie bardzo podstawowe wymagania. Może to sugerować korzyści płynące z potencjalnej integracji wysiłków badaczy, by badać mniej, ale dokładniej.

Choć ekstensywne omówienie praktyk sondażowych (terenowych) na świecie nie znajduje się *explicite* w żadnej z publikacji, to zebranie informacji na ten temat było warunkiem *sine qua non* ich opisanie i wyniki tego procesu zostały przedstawione w autoreferacie oraz w aneksie. Wartym odnotowania wnioskiem jest tu wskazanie Azji jako kontynentu o zaskakująco dużej dostępności operatorów imiennych. W połączeniu ze znacznym zróżnicowaniem stosowanych tam technik sondażowych, jak i metod motywowania respondentów (drugi po Europie kontynent z największą liczbą sondaży z zachętami materialnymi) sugeruje to, że rozwój badań sondażowych w tej części świata należy obserwować z rosnącą uwagą, gdyż może to dostarczyć wiele nowej wiedzy metodologicznej.

Przechodząc do wniosków z badania związków pomiędzy charakterystykami sondażu a reprezentatywnością, należy podkreślić, że metaanalizy przedstawione w Publikacji 2 (i autoreferacie) opierają się na danych obserwacyjnych. Nie jest to niczym niezwykłym, jednak fakt, że charakterystyki nie są przypisywane do analizowanych sondaży w sposób zrandomizowany, a wynikają częściowo z zastanych uwarunkowań, wpływa ograniczająco na możliwości uogólniania wyników. Ograniczenie to jest łagodzone przez fakt, że w poszczególnych programach sondaże w tych samych latach i krajach były często przeprowadzane w różny sposób. Stosowanie danych obserwacyjnych do takich analiz jest także przyjętą praktyką. W przeglądzie systematycznym metaanaliz w metodologii sondażowej (Cehovin i in. 2018) wśród 54 uwzględnionych badań 28, a więc ponad połowa, opiera się właśnie na danych obserwacyjnych. Kolejnym ograniczeniem jest tylko częściowe uwzględnienie lokalnej specyfiki nastawienia populacji do udziału w sondażach (*survey climate*) (Wagner i Stoop 2018). W modelach trzypoziomowych obserwacje były zagnieżdżone w falach i krajach jednocześnie, jednak nie pozwoliło to na uchwycenia zmian w czasie, jakie mogły pod tym względem zachodzić w danych krajach.

Wyniki prezentowane w Publikacji 2 mogą napawać optymizmem. Chociaż istnieje względnie rozpowszechnione przekonanie na temat techniki sondażu, że „im drożej, tym lepiej”, prezentowane modele wykazały, że choć zazwyczaj najdroższe sondaże *face-to-face* związane były z uzyskiwaniem najwyższego *response rate*, to już nie osiągały najwyższego wskaźnika reprezentatywności uzyskanej próby. Wręcz odwrotnie, w przypadku stosowania próby imiennej zarówno sondaże pocztowe, jak i *mixed mode* miały niższe wskaźniki błędu braku odpowiedzi, a jedyną techniką, która uzyskała niższą reprezentatywność niż *face-to-face*, był sondaż pocztowy w sytuacji próby adresowej. Płyną z tego dwa wnioski. Po pierwsze

alternatywne techniki sondażu mogą dostarczać danych bardzo dobrej jakości i warto z nich korzystać. Po drugie jednak wykorzystanie tej szansy często wymaga dostępności operatu imiennego w danym kraju, więc środowisko badaczy w krajach, gdzie taki operat nie jest dostępny, powinno lobbować na rzecz jego stworzenia lub udostępnienia do badań o dużej doniosłości naukowej (jeśli potencjalny operat istnieje, ale instytucje państwa się nim nie dzielą). Upowszechnienie przedstawionych wyników może także ułatwić pracę badaczom, którzy mierzą się z trudnościami związanymi z rosnącymi stawkami ankietatorów czy wręcz upadkiem sieci ankieterskich po pandemii COVID-19.

Przechodząc na niższy poziom analizy, Publikacje 3 i 4 dotyczą problemu relacji pomiędzy konkretną charakterystyką badania (techniką) a jakością pomiaru, ujmowanej w odniesieniu do hipotetycznego zjawiska preferencji techniki sondażu. Choć zwrócono na nie uwagę już w latach 70. XX w., to badania poświęcone temu zjawisku są stosunkowo nowe i nieliczne, zatem kwestia ta tym bardziej wydawała się interesującym zagadnieniem do zgłębienia. Preferencja ma mieć pozytywny wpływ zarówno na chęć uczestnictwa w badaniu (Olson i in. 2012), jak i jakość udzielanych odpowiedzi (Smyth, Olson i Kasabian 2014). Podstawowym ograniczeniem w zakresie formułowania wniosków na ten temat jest wątpliwość, na ile *mode preference* jest zjawiskiem realnym i istotnym. Do problematycznych kwestii dotyczących tego zagadnienia można zaliczyć ograniczone oddziaływanie preferencji mierzonej za pomocą deklaracji (Millar, O'Neill i Dillman 2009), wpływ techniki, w której pytanie o preferencję jest zadawane (Groves i Kahn 1979; Tarnai i Paxson 2004), a także stabilność preferencji w czasie (Al Baghal i Kelley 2016).

Dodatkowym ograniczeniem moich analiz jest to, że część behawioralna bazy danych użytej w Publikacji 4 opiera się na sondażu *mixed-mode* o schemacie sekwencyjno-równoległym (Sztabiński 2018), a optymalny byłby schemat wyłącznie równoległy; mimo to jego konstrukcja nie wpływa w dużym stopniu na wnioski, które wyciągam z badania.

Nawiązując do teorii psychologicznych (teorii udziału w sondażu i koncepcji *satisficingu*) oraz kognitywistycznych, sformułowałem tezę, że można postawić znak równości pomiędzy różnymi ujęciami interakcji między trudnością zadania i zdolnościami (oraz motywacją) respondenta a relacją pomiędzy obciążeniem poznawczym i możliwościami poznawczymi w teorii pamięci roboczej. Na podstawie studiów literaturowych prezentowanych w Publikacji 3 przedstawiłem hipotezę, że ze względu na (między innymi) nierównomierny spadek wydajności podsystemów pamięci roboczej osoby mające więcej niż 70 lat będą częściej niż młodsze wołały odpowiadać na pytania sondażowe w technikach wywiadu (a więc w kanale audialnym), a osoby między 50. a 70. rokiem życia będą wykazywać większą skłonność niż osoby starsze i młodsze do preferowania techniki ankiety (a więc komunikacji w kanale wizualnym). Hipoteza ta znalazła częściowe poparcie w behawioralnych i deklaracyjnych danych empirycznych przedstawionych w Publikacji 4. Najsilniejsze upodobanie względem techniki pocztowej wśród wszystkich grup wiekowych przejawiały osoby z kategorii 45–54 lat. (W Publikacji 3 poruszam także, co zrozumiałe, biorąc pod uwagę analizowane grupy wiekowe, temat potencjalnego wpływu rozpowszechnienia wad wzroku i słuchu na preferencję techniki)

Należy zaznaczyć, że możliwość uogólniania wniosków z Publikacji 4 jest oczywiście ograniczona, gdyż wpływ na preferencję techniki mogą mieć rozmaite czynniki kulturowe i instytucjonalne, a dane empiryczne użyte w modelu pochodziły wyłącznie z Polski. Najciekawszym wynikiem analizy wydaje się jednak to, jak bardzo silny może być efekt rodzaju pomiaru preferencji, gdyż deklaracje diametralnie odstawały od realnych wyborów. W odpowiedzi na pytanie ankietera respondenci jako preferowaną formę sondażu dwa razy częściej wskazywali ankietę internetową niż pocztową. Natomiast gdy dochodziło do realnego wyboru, prawie 7 razy częściej wypełniali ankietę pocztową niż internetową. Przy czym

między okresami zbierania danych minęło tylko 1,5 roku, więc przyczyną różnicy nie mógł być stopień rozpowszechnienia Internetu. Większego wpływu nie miał także schemat badania *mixed-mode*, ponieważ w pierwszej fazie respondenci dostawali równoległy wybór między ankietami. Nie miał też znaczenia efekt wskazywania na technikę udzielania odpowiedzi, gdyż pytanie było zadane w trakcie wywiadu.

Nasuują się dwa prawdopodobne wytłumaczenia takiego stanu rzeczy. Pytanie o preferencję techniki mogło okazać się kwestią drażliwą. Być może respondenci chcieli zaprezentować się przed ankieterem jako nowocześniejsi i bardziej zaznajomieni z technologiami współczesnego świata niż byli w rzeczywistości<sup>2</sup>. Mogło być też tak, że ten temat był na tyle słabo urefleksyjniony, że po długim badaniu (pytanie znajdowało się na końcu kwestionariusza) respondenci udzielali już mocno nieprzemyślanych odpowiedzi. W obu przypadkach należy wyciągnąć wniosek, że opieranie się wyłącznie na deklaracjach dotyczących preferencji nie jest dobrą strategią. Zamiast tego wskazane byłoby tworzenie modeli predykcyjnych opartych na danych z różnych źródeł – deklaracji, wyborów, uzupełnionych o wiedzę z analiz kognitywistycznych. Można wykorzystać w tym celu także nowoczesne, bardziej rozbudowane techniki uczenia maszynowego (Jankowsky i Schroeders 2022; Küfner i in. 2022).

## V. Podsumowanie i wnioski

Cztery publikacje składające się na przedstawiony cykl opisują kolejne elementy metodologicznej analizy praktyk sondażowych. Pierwsza obejmuje problematykę kompletności informacji zawartych w dokumentacji sondaży z programów porównawczych, a także opis jednolitości, spójności i formatu zastanych raportów. Jej wyniki potwierdzają dotychczasowe ustalenia badaczy mówiące o stabilnym wzroście kompletności dokumentacji metodologicznej w czasie, natomiast w przypadku zróżnicowania geograficznego tej zmiennej sugerują wniosek przeciwny do wyników wcześniejszych badań. Dotyczy on najniższej wśród wszystkich kontynentów kompletności raportów z sondaży europejskich, mogącej mieć związek z inflacją badań na tym obszarze. Z tego powodu treść publikacji zaleca większą koordynację w zakresie ustalania norm dokumentacyjnych (i nie tylko) pomiędzy ośrodkami badawczymi. Za wzór stawia w tym przypadku Europejski Sondaż Społeczny, który przewyższa wszystkie inne programy porównawcze pod względem jakości dokumentacji.

Druga publikacja dotyczy kwestii związku pomiędzy charakterystykami sondażu – przede wszystkim techniką i operatem – a reprezentatywnością badania (a także, choć pobocznie, wskaźnikami realizacji próby). Posługuje się zaproponowanym przez Wolfganga Sodeura wewnętrznym kryterium reprezentatywności, stosowanym jako miara wielkości efektu w modelach metaanalizy. Pokazuje, że mniejsze wskaźniki realizacji kojarzone z sondażami stosującymi techniki inne niż wywiad kwestionariuszowy (ze szczególnym uwzględnieniem nowoczesnych sondaży *mixed-mode*, przeprowadzanych bez bezpośredniego kontaktu respondenta z ankieterem) nie wiążą się z niższą jakością struktury próby. Wręcz przeciwnie, w sprzyjających warunkach (obejmujących dostępność operatu imiennego) badania realizowane tymi technikami osiągają często reprezentatywność wyższą, niż badania *face-to-face*. Wynik ten jest krokiem w stronę bardziej jednoznacznej (pozytywnej) oceny wpływu stosowania schematów mieszania technik sondażu na jakość pomiaru, dotychczasowe analizy bowiem prezentują zauważalnie sprzeczne wnioski.

Publikacja trzecia, o charakterze teoretycznym, stawia sobie za cel wskazanie związków i podobieństw pomiędzy ujęciami procesu udziału respondenta w badaniach sondażowych

---

<sup>2</sup> Za tę interpretację dziękuję prof. de Leeuw.

przedstawionymi przez takich badaczy jak Roger Tourangeau (ujęcie psychologiczne) i Jon Krosnick (*satisficing*), a hipotezą preferencji techniki sondażu i kognitywistyczną koncepcją pamięci roboczej. Zwraca uwagę na wyniki eksperymentalnych analiz dotyczących wpływu starzenia się na nierównomierny spadek poziomu funkcjonowania podsystemów tej pamięci. Pełni ona także rolę rozbudowanego wstępu teoretycznego do publikacji czwartej, która skupia się na empirycznej analizie związków pomiędzy cechami społeczno-demograficznymi polskich respondentów, a prawdopodobieństwem preferowania przez nich określonej techniki sondażu. Wyniki tych analiz są zgodne z opartymi na teoretycznych rozważaniach przypuszczeniami, że osoby w wieku średnim (45-54 lat) będą charakteryzowały się największą skłonnością do udziału w badaniu w technikach ankiety (w tym wypadku pocztowej). Najciekawszą obserwacją prezentowaną przez ostatnią z publikacji jest jednak nie sam kształt preferencji, ale diametralna różnica pomiędzy deklarowanymi preferencjami, a realnymi wyborami techniki udziału w badaniu. Poddaje ona w wątpliwość możliwość opierania przyszłych badań koncentrujących się na zjawisku *mode preference* na samych tylko deklaracjach respondentów.

Za najistotniejsze wnioski płynące z całości przeprowadzonych analiz należy uznać:

1) Konieczność uzgodnienia w środowisku jednolitego standardu dokumentacji metodologicznej o stałych nagłówkach, który umożliwi automatyzację procesu zbierania danych metodologicznych; 2) Kluczową istotność dostępności operatu imiennego w sytuacji stopniowego odchodzenia od przeprowadzania sondaży techniką wywiadu kwestionariuszowego; 3) Wysoki potencjał sondaży *mixed-mode* do uzyskiwania dużej reprezentatywności próby; 4) Zasadność badania źródeł i korelatów preferencji techniki sondażu w celu tworzenia modeli predykcyjnych służących do planowania przyszłych badań *mixed-mode* w schemacie równoległym w poszczególnych krajach.

Implementacja wskazanych zaleceń może przyczynić się do utrzymania istotnej roli badań sondażowych w sytuacji wzrastającego poziomu wykorzystania Big Data jako alternatywnego źródła danych i zwiększających się trudności z terenową realizacją badań.

## Literatura (wg American Sociological Association Style Guide 7th edition)

- AAPOR/WAPOR. 2021. *AAPOR/WAPOR Task Force Report on Quality in Comparative Surveys*.
- Afrobarometer. 2011. „Round 5 Survey Manual”.
- Al Baghal, Tarek i Jennifer Kelley. 2016. „The Stability of Mode Preferences: Implications for Tailoring in Longitudinal Surveys”. *Methods, Data, Analyses* 10(102):143–66. doi: 10.12758/mda.2016.012.
- Almond, Gabriel Abraham i Sidney Verba. 1963. *The Civic Culture*. Princeton: Princeton University Press.
- Barendregt, Jan J., Suhail A. Doi, Yong Yi Lee, Rosana E. Norman i Theo Vos. 2013. „Meta-analysis of prevalence”. *Journal of Epidemiology and Community Health* 67(11):974–78. doi: 10.1136/jech-2013-203104.
- Bauer, Johannes J. 2014. „Selection Errors of Random Route Samples”. *Sociological Methods & Research* 43(3):519–44. doi: 10.1177/0049124114521150.
- Bauer, Johannes J. 2016. „Biases in Random Route Surveys”. *Journal of Survey Statistics and Methodology* 4(2):263–87. doi: 10.1093/jssam/smw012.
- Best, Henning i Christof Wolf, red. 2015. *The Sage Handbook of Regression Analysis and Causal Inference*. Los Angeles: SAGE.
- Beullens, Koen i Geert Loosveldt. 2012. „Should high response rates really be a primary objective?” *Survey Practice* 5(3):1–5. doi: 10.29115/sp-2012-0019.
- Bianchi, Annamaria, Silvia Biffignandi i Peter Lynn. 2017. „Web-face-to-face mixed-mode design in a longitudinal survey: Effects on participation rates, sample composition, and costs”. *Journal of Official Statistics* 33(2):385–408. doi: 10.1515/jos-2017-0019.
- Bianchi, Annamaria, Natalie Shlomo, Barry Schouten, Damião N. da Silva i Chris Skinner. 2019. „Estimation of response propensities and indicators of representative response using population-level information”. *Survey Methodology* 45(2):231–64.
- Biemer, Paul P. 2001. „Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing”. *Journal of Official Statistics* 17(2):295–320.
- Biemer, Paul P. i Lars Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins i Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester: Wiley.
- Cannell, Charles F. i Floyd J. Fowler. 1963. „Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study”. *Public Opinion Quarterly* 27(2):250. doi: 10.1086/267165.
- Card, Noel A. 2012. *Applied Meta-Analysis for Social Science Research*. New York, London: Guilford Press.
- Castorena, Oscar. 2021. *Methodological Note #007. Survey Weights in Americas Barometer Data*. Vanderbilt University.

- Cehovin, Gregor, Michael Bosnjak i Katja Lozar Manfreda. 2018. „Meta-Analyses in Survey Methodology: A Systematic Review”. *Public Opinion Quarterly* 82(4):641–60. doi: 10.1093/poq/nfy042.
- Chang, Linchiat i Jon A. Krosnick. 2010. „Comparing oral interviewing with self-administered computerized questionnaires: An experiment”. *Public Opinion Quarterly* 74(1):154–67. doi: 10.1093/poq/nfp090.
- Church, Allan H. 1993. „Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis”. *Public Opinion Quarterly* 57(1):62. doi: 10.1086/269355.
- Clearman, Jack, Vojtěch Klinger i Dénes Szűcs. 2017. „Visuospatial and verbal memory in mental arithmetic”. *Quarterly Journal of Experimental Psychology* 70(9):1837–55. doi: 10.1080/17470218.2016.1209534.
- Constantinidou, Fofi i Susan Baker. 2002. „Stimulus modality and verbal learning performance in normal aging”. *Brain and Language* 82(3):296–311. doi: 10.1016/S0093-934X(02)00018-4.
- Cornesse, Carina i Michael Bosnjak. 2018. „Is there an association between survey characteristics and representativeness? A meta-analysis”. *Survey Research Methods* 12(1):1–13. doi: 10.18148/srm/2018.v12i1.7205.
- Couper, Mick. 2013. „Is the sky falling? New technology, changing media, and the future of surveys”. *Survey Research Methods* 7(3):145–56.
- Cumming, Geoff. 2012. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York, London: Routledge.
- Daikeler, Jessica, Michael Bosnjak i Katja Lozar Manfreda. 2020. „Web versus other survey modes: An updated and extended meta-analysis comparing response rates”. *Journal of Survey Statistics and Methodology* 8(3):513–39. doi: 10.1093/jssam/smz008.
- Davern, Michael, Todd H. Rockwood, Randy Sherrod i Stephen Campbell. 2003. „Prepaid Monetary Incentives and Data Quality in Face-to-Face Interviews”. *Public Opinion Quarterly* 67(1):139–47. doi: 10.1086/346012.
- Davino, Cristina i Luigi Fabbri. 2013. „Preface”. w *Survey Data Collection and Integration*, zredagowane przez C. Davino i L. Fabbri. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dillman, Don A., Jolene D. Smyth i Leah Melani Christian. 2014. *Internet, phone, mail, and mixed mode surveys: The tailored design method (4th ed.)*. Hoboken: John Wiley & Sons Inc.
- Dubrow, Joshua Kjerulf, J. Craig Jenkins, Marta Kołczyńska, Olena Oleksienko, Przemek Powalko, Kazimierz Maciej Słomczyński, Irina Tomescu-Dubrow, Ilona Wysmułek i Marcin W. Zieliński, red. 2016. *Democratic Values and Protest Behavior. Harmonization of Data from International Survey Projects*. Warsaw: IFIS PAN Publishers.
- Eckman, Stephanie i Achim Koch. 2019. „Interviewer Involvement in Sample Selection Shapes the Relationship between Response Rates and Data Quality”. *Public Opinion Quarterly* 83(2):313–37. doi: 10.1093/poq/nfz012.

- ESS ERIC. 2024. *The Road to A Sustainable Self-Completion Future European Social Survey Eric Strategic Plan 2024-2029*. European Social Survey. European Research Infrastructure Consortium.
- Fernández-Castilla, Belén, Marlies Maes, Lies Declercq, Laleh Jamshidi, S. Natasha Beretvas, Patrick Onghena i Wim Van den Noortgate. 2019. „A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis”. *Behavior Research Methods* 51(3):1286–1304. doi: 10.3758/s13428-018-1063-2.
- Fischer, Mirjam M. 2016. „Identifying same-sex couples in cross-national survey data: A comparison of same-sex couples’ demographic and socio-economic traits in six European countries”. *DANS symposium publications* 8:50–79.
- Fowler, Floyd Jackson Jr, Patricia M. Gallagher, Vickie L. Stringfellow, Alan M. Zaslavsky, Joseph W. Thompson i Paul D. Cleary. 2002. „Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members”. *Medical care* 40(3):190–200. doi: 10.1097/00005650-200203000-00003.
- Fox, Richard J., Melvin R. Crask i Jonghoon Kim. 1988. „Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response”. *Public Opinion Quarterly* 52:467–91. doi: 10.1086/269125.
- Frankovic, Kathleen A. 2016. „Observations on the Historical Development of Polling”. S. 87–102 w *The SAGE Handbook of Survey Methodology*, zredagowane przez C. Wolf, D. Joye, T. Smith i Y. Fu. London: SAGE Publications Ltd.
- GALLUP. 2019. „How Does the Gallup World Poll Work? Measures the Attitudes and Behaviors of the World’s Residents”. Pobrano 1 styczeń 2022 (<https://www.gallup.com/178667/gallup-world-poll-work.aspx>).
- Gaziano, C. 2005. „Comparative Analysis of Within-Household Respondent Selection Techniques”. *Public Opinion Quarterly* 69(1):124–57. doi: 10.1093/poq/nfi006.
- GESIS. 2019. „Overview of Comparative Surveys Worldwide”. Pobrano 1 czerwiec 2019 (<http://www.gesis.org/ComparativeSurveyOverview>).
- Goyder, John. 1985. „Face-to-face interviews and mailed questionnaires: The net difference in response rate”. *Public Opinion Quarterly* 49(2):234–52. doi: 10.1086/268917.
- Grauenhorst, Thomas, Michael Blohm i Achim Koch. 2016. „Respondent Incentives in a National Face-to-Face Survey: Do They Affect Response Quality?” *Field Methods* 28(3):266–83. doi: 10.1177/1525822X15612710.
- Groves, Robert. 2006. „Nonresponse rates and nonresponse bias in household surveys: What Do We Know about the Linkage between Nonresponse Rates and Nonresponse Bias?” *Public Opinion Quarterly* 70(5):646–75. doi: 10.1093/poq/nfl033.
- Groves, Robert M. 2011. „Three eras of survey research”. *Public Opinion Quarterly* 75(5 spec. issue):861–71. doi: 10.1093/poq/nfr057.
- Groves, Robert M. i Robert L. Kahn. 1979. *Surveys by telephone : a national comparison with personal interviews*. New York: Academic Press.
- Groves, Robert M. i Lars Lyberg. 2010. „Total survey error: Past, present, and future”. *Public Opinion Quarterly* 74(5):849–79. doi: 10.1093/poq/nfq065.

- Groves, Robert M. i Emilia Peytcheva. 2008. „The impact of nonresponse rates on nonresponse bias: A meta-analysis”. *Public Opinion Quarterly* 72(2):167–89. doi: 10.1093/poq/nfn011.
- Haan, Marieke, Yfke P. Ongena i Kees Aarts. 2014. „Reaching hard-to-survey populations: Mode choice and mode preference”. *Journal of Official Statistics* 30(2):355–79. doi: 10.2478/jos-2014-0021i.
- Harter, Rachel, Stephanie Eckman, Ned English i Colm O’Muircheartaigh. 2010. „Applied Sampling for Large-Scale Multistage Area Probability Designs”. w *Handbook of survey research*, zredagowane przez P. V. Marsden i J. D. Wright. Bingley: Emerald Group Publ.
- Hendra, Richard i Aaron Hill. 2019. „Rethinking Response Rates: New Evidence of Little Relationship Between Survey Response Rates and Nonresponse Bias”. *Evaluation Review* 43(5):307–30. doi: 10.1177/0193841X18807719.
- Higgins, Julian P. T., Simon G. Thompson, Jonathan J. Deeks i Douglas G. Altman. 2003. „Measuring inconsistency in meta-analyses”. *BMJ* 327(7414):557–60. doi: 10.1136/bmj.327.7414.557.
- Hopkins, Kenneth D. i Arlen R. Gullickson. 1992. „Response Rates in Survey Research: A Meta-Analysis of the Effects of Monetary Gratuities”. *The Journal of Experimental Education* 61(1):52–62. doi: 10.1080/00220973.1992.9943849.
- Hox, Joop J. i Edith D. de Leeuw. 1994. „A comparison of nonresponse in mail, telephone, and face-to-face surveys. Applying multilevel modeling to meta-analysis”. *Quality & Quantity* 28:329–44.
- Hox, Joop J., Edith D. de Leeuw i Hsuan Tzu Chang. 2012. „Nonresponse versus Measurement Error: Are Reluctant Respondents Worth Pursuing?” *Bulletin de Methodologie Sociologique* 113(1):5–19. doi: 10.1177/0759106311426987.
- Hunter, John E. i Frank L. Schmidt. 2004. *Methods of meta-analysis : correcting error and bias in research findings*. Thousand Oaks: Sage.
- ISSP Research Group. 2019. „International Social Survey Programme: Social Networks and Social Resources - ISSP 2017”. *GESIS Data Archive, Cologne. ZA6980 Data file Version 2.0.0*. doi: 10.4232/1.13322.
- Jabkowski, Piotr. 2015. *Reprezentatywność badań reprezentatywnych: analiza wybranych problemów metodologicznych oraz praktycznych w paradygmacie całkowitego błędu pomiaru*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.
- Jabkowski, Piotr. 2017. „A Meta-analysis of Within-Household Selection Impact on Survey Outcome Rates, Demographic Representation and Sample Quality in the European Social Survey”. *Ask. Research & Methods* 26(1):31–60. doi: 10.18061/1811/81932.
- Jabkowski, Piotr. 2018. „The Effect of Using Different Types of Survey Sample on Sample Execution in the European Social Survey”. *Polish Sociological Review* 2(202):265–78. doi: 10.26412/psr202.07.
- Jabkowski, Piotr. 2019. „Zróżnicowanie praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie archiwizacji 1537 surveyów zrealizowanych w krajach europejskich w latach 1981–2017”. *Przegląd Socjologiczny* 68(1):9–38. doi: 10.26485/PS/2019/68.1/1.

- Jabkowski, Piotr. 2023. „Increase in the Quality of Methodological Documentation of Cross-National Pan-European Multi-Wave Surveys over the Last 40 Years – a Research Note”. *International Journal of Social Research Methodology* 26(6):817–24. doi: 10.1080/13645579.2022.2097394.
- Jabkowski, Piotr i Piotr Cichocki. 2019. „Within-household selection of target-respondents impairs demographic representativeness of probabilistic samples: Evidence from seven rounds of the European social survey”. *Survey Research Methods* 13(2):167–80. doi: 10.18148/srm/2019.v13i2.7383.
- Jabkowski, Piotr i Piotr Cichocki. 2024. „Survey Response Rates in European Comparative Surveys: A 20-Year Decline Irrespective of Sampling Frames or Survey Modes”. *Quality & Quantity*. doi: 10.1007/s11135-024-01993-9.
- Jabkowski, Piotr, Piotr Cichocki i Marta Kołczyńska. 2021. „Multi-Project Assessments of Sample Quality in Cross-National Surveys: The Role of Weights in Applying External and Internal Measures of Sample Bias”. *Journal of Survey Statistics and Methodology* 11(2):1–24. doi: 10.1093/jssam/smab027.
- Jabkowski, Piotr, Piotr Cichocki i Marta Kołczyńska. 2024. „Kish Grid vs Birthday Procedures: Survey Refusal Rates by the Type of within-Household Selection Procedure in the European Social Survey”. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 163(1):77–87. doi: 10.1177/07591063241258087.
- Jabkowski, Piotr i Marta Kołczyńska. 2020. „Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981-2017”. *Methodology* 16(3):186–207. doi: 10.5964/meth.2795.
- Jäckle, Annette i Peter Lynn. 2007. *Respondent incentives in a multi-mode panel survey: cumulative effects on nonresponse and bias*. Colchester: ISER.
- Jankowsky, Kristin i Ulrich Schroeders. 2022. „Validation and Generalizability of Machine Learning Prediction Models on Attrition in Longitudinal Studies”. *International Journal of Behavioral Development* 46(2):169–76. doi: 10.1177/016502542211075034.
- Jedinger, Alexander, Oliver Watteler i André Förster. 2018. „Improving the quality of survey data documentation: A total survey error perspective”. *Data* 3(4):1–10. doi: 10.3390/data3040045.
- Jenkins, J. Craig i Joonghyun Kwak. 2022. „How to Capture Global Protest Trends: Using Survey Data Recycling Data to Construct Cross-National Trends in Protest”. *American Behavioral Scientist* 66(4):459–88. doi: 10.1177/00027642211021627.
- Kalsbeek, William D. i Robert P. Agans. 2008. „Sampling and Weighting in Household Telephone Surveys”. w *Advances in Telephone Survey Methodology*, zredagowane przez J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. D. Leeuw, L. Japac, P. J. Lavrakas, M. W. Link i R. L. Sangster. Wiley.
- Kaminska, Olena, Allan L. McCutcheon i Jaak Billiet. 2010. „Satisficing among reluctant respondents in a cross-national context”. *Public Opinion Quarterly* 74(5):956–84. doi: 10.1093/poq/nfq062.
- Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best i Peyton Craighill. 2006. „Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey”. *Public Opinion Quarterly* 70(5):759–79. doi: 10.1093/poq/nfl035.

- Kish, Leslie. 1949. „A Procedure for Objective Respondent Selection within the Household”. *Journal of the American Statistical Association* 44(247):380–87. doi: 10.1080/01621459.1949.10483314.
- Kiszkziel, Lukasz, Piotr Paweł Laskowski, David Voas, Rachel J. Bacon, Wesley J. Wildman, Ivan Puga-Gonzalez, F. LeRon Shults i Konrad Talmont-Kaminski. 2024. „Dataset of Integrated Measures of Religion (DIM-R). Harmonization of Religiosity Data from Selected International Multiwave Surveys”. *Religion, Brain & Behavior* 1–41. doi: 10.1080/2153599X.2024.2305447.
- Klausch, Thomas, Joop Hox i Barry Schouten. 2015. „Selection error in single- and mixed mode surveys of the Dutch general population”. *Journal of the Royal Statistical Society. Series A: Statistics in Society* 178(4):945–61. doi: 10.1111/rssa.12102.
- Knapp, Guido i Joachim Hartung. 2003. „Improved tests for a random effects meta-regression with a single covariate”. *Statistics in Medicine* 22(17):2693–2710. doi: 10.1002/sim.1482.
- Koch, Achim, Verena Halbherr, Ineke A. L. Stoop i Joost W. S. Kappelhof. 2014. *Assessing ESS sample quality by using external and internal criteria*. Mannheim: European Social Survey, GESIS.
- Kohler, Ulrich. 2007. „Surveys from inside: An assessment of unit nonresponse bias with internal criteria”. *Survey Research Methods* 1(2):55–67.
- Kończyńska, Marta. 2022. „Combining Multiple Survey Sources: A Reproducible Workflow and Toolbox for Survey Data Harmonization”. *Methodological Innovations* 15(1):62–72. doi: 10.1177/20597991221077923.
- Kończyńska, Marta, Paul-Christian Bürkner, Lauren Kennedy i Aki Vehtari. 2024. „Modeling Public Opinion Over Time and Space: Trust in State Institutions in Europe, 1989–2019”. *Survey Research Methods* 18(1):1–19. doi: 10.18148/SRM/2024.V18I1.8119.
- Kończyńska, Marta, Piotr Jabkowski i Stephanie Eckman. 2024. „Interviewer Involvement in Respondent Selection Moderates the Relationship Between Response Rates and Sample Bias in Cross-National Survey Projects in Europe”. *Journal of Survey Statistics and Methodology* 12(1):1–13. doi: 10.1093/jssam/smad013.
- Kończyńska, Marta i Matthew Schoene. 2018. „Survey Data Harmonization and the Quality of Data Documentation in Cross-national Surveys”. S. 963–84 w *Advances in Comparative Survey Methods*, zredagowane przez T. P. Johnson, B. Pennell, I. A. L. Stoop i B. Dorer. Wiley.
- Krosnick, Jon. 1991. „Response Strategies for Coping with the Cognitive Demands of Attitude Measure in Surveys”. *Applied Cognitive Psychology* 5:213–36.
- Küfner, Benjamin, Joseph W. Sakshaug i Stefan Zins. 2022. „Analysing Establishment Survey Non-Response Using Administrative Data and Machine Learning”. *Journal of the Royal Statistical Society Series A: Statistics in Society* 185(Supplement 2):310–42. doi: 10.1111/rssa.12942.
- Kumar, Navnit i Brajesh Priyadarshi. 2013. „Differential effect of aging on verbal and visuo-spatial working memory.” *Aging and disease* 4(4):170–77.

- Kwak, Joonghyun, Irina Tomescu-Dubrow, Kazimierz M. Slomczynski i Joshua K. Dubrow. 2020. „Youth, Institutional Trust, and Democratic Backsliding”. *American Behavioral Scientist* 64(9):1366–90. doi: 10.1177/0002764220941222.
- Laaksonen, Seppo i Markku Heiskanen. 2014. „Comparison of three modes for a crime victimization survey”. *Journal of Survey Statistics and Methodology* 2(4):459–83. doi: 10.1093/jssam/smu018.
- Lagos, Marta. 2008. „International Comparative Surveys: Their Purpose, Content and Methodological Challenges”. S. 580–93 w *The SAGE Handbook of Public Opinion Research*, zredagowane przez W. Donsbach i M. Traugott. London: SAGE Publications Ltd.
- de Leeuw, E. D., M. Callegaro, J. Hox, E. Korendijk i G. Lensvelt-Mulders. 2007. „The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis”. *Public Opinion Quarterly* 71(3):413–43. doi: 10.1093/poq/nfm014.
- de Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone and Face to Face surveys*. Amsterdam: TT-Publikaties.
- de Leeuw, Edith D. 2008. „Choosing the Method of Data Collection”. S. 113–35 w *International Handbook of Survey Methodology*, zredagowane przez E. D. de Leeuw, J. J. Hox i D. A. Dillman. New York: L. Erlbaum Associates.
- de Leeuw, Edith D. i Joop. J. Hox. 2011. „Internet surveys as part of a mixed-mode design”. S. 45–76 w *Social and Behavioral Research and the Internet*. London: Taylor & Francis.
- de Leeuw, Edith D., Joop J. Hox i Annemieke Luiten. 2018. „International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data”. *Survey Methods: Insights from the Field* (1981):1–11. doi: 10.13094/SMIF-2018-00008.
- de Leeuw, Edith D. i Berzelak Nejc. 2016. „Survey Mode or Survey Modes?” S. 142–56 w *The SAGE Handbook of Survey*, zredagowane przez C. Wolf, D. Joye, T. W. Smith i Y. Fu. Thousand Oaks: Sage.
- Luijkx, Ruud, Guðbjörg Andrea Jónsdóttir, Tobias Gummer, Michèle Ernst Stähli, Morten Frederiksen, Kimmo Ketola, Tim Reeskens, Evelyn Brislinger, Pablo Christmann, Stefán Þór Gunnarsson, Árni Bragi Hjaltason, Dominique Joye, Vera Lomazzi, Angelica M. Maineri, Patricia Milbert, Michael Ochsner, Alexandre Pollien, Marlène Sapin, Ivet Solanes, Sascha Verhoeven i Christof Wolf. 2021. „The European Values Study 2017: On the Way to the Future Using Mixed-Modes”. *European Sociological Review* 37(2):330–46. doi: 10.1093/esr/jcaa049.
- Luiten, Annemieke, Joop Hox i Edith D. de Leeuw. 2020. „Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys”. *Journal of Official Statistics* 36(3):469–87. doi: 10.2478/jos-2020-0025.
- Luiten, Annemieke i Barry Schouten. 2013. „Tailored fieldwork design to increase representative household survey response: An experiment in the Survey of Consumer Satisfaction”. *Journal of the Royal Statistical Society. Series A: Statistics in Society* 176(1):169–89. doi: 10.1111/j.1467-985X.2012.01080.x.
- Manfreda, Katja Lozar, Michael Bosnjak, Jernej Berzelak, Iris Haas, Vasja Vehovar, Michael Bosnjak i Iris Haas. 2008. „Web surveys versus other survey modes: A meta-analysis

- comparing response rates". *International Journal of Market Research* 50(1):79–104. doi: 10.1177/147078530805000107.
- Medway, Rebecca L. i Jenna Fulton. 2012. „When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates". *Public Opinion Quarterly* 76(4):733–46. doi: 10.1093/poq/nfs047.
- Menold, Natalja. 2014. „The influence of sampling method and interviewers on sample realization in the european social survey". *Survey Methodology* 40(1):105–23.
- Millar, Morgan M., Allison C. O'Neill i Don A. Dillman. 2009. *Are Mode Preferences Real?* Pullman W.A.: Social and Economic Sciences Research Center, Washington State Universit.
- Miller, Heather i Jacqueline Bichsel. 2004. „Anxiety, working memory, gender, and math performance". *Personality and Individual Differences* 37(3):591–606. doi: 10.1016/j.paid.2003.09.029.
- Mohler, Peter Ph., Sue Ellen Hansen, Beth-Ellen Pennell, Wendy Thomas, Joachim Wackerow i Frost Hubbard. 2010. „A Survey Process Quality Perspective on Documentation". S. 299–314 w *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, zredagowane przez J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B.-E. Pennell i T. W. Smith. Hoboken: John Wiley & Sons.
- Mohler, Peter Ph., Beth-Ellen Pennell i Frost Hubbard. 2008. „Survey Documentation: Toward Professional Knowledge Management in Sample Surveys". S. 403–20 w *International Handbook of Survey Methodology*, zredagowane przez E. D. de Leeuw, J. J. Hox i D. A. Dillman. New York: L. Erlbaum Associates.
- Neyman, Jerzy. 1934. „On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection". *Journal of the Royal Statistical Society* 97(4):558–625. doi: 10.2307/2342192.
- Olive, Thierry, Ronald Kellogg i Annie Piolat. 2008. „Verbal, visual, and spatial working memory demands during text composition". *Applied Psycholinguistics* 29(4):669–87. doi: <http://dx.doi.org/10.1017/S0142716408080284>.
- Olson, Kristen M. 2006. „Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias". *Public Opinion Quarterly* 70(5):737–58. doi: 10.1093/poq/nfl038.
- Olson, Kristen, Jolene D. Smyth i Heather M. Wood. 2012. „Does giving people their preferred survey mode actually increase survey participation rates? an experimental examination". *Public Opinion Quarterly* 76(4):611–35. doi: 10.1093/poq/nfs024.
- Peytchev, Andy, Emilia Peytcheva i Robert M. Groves. 2010. „Measurement Error, Unit Nonresponse, and Self-Reports of Abortion Experiences". *Public Opinion Quarterly* 74(2):319–27. doi: 10.1093/poq/nfq002.
- Pham, Andy V. i Ramzi M. Hasson. 2014. „Verbal and visuospatial working memory as predictors of children's reading ability". *Archives of Clinical Neuropsychology* 29(5):467–77. doi: 10.1093/arclin/acu024.
- Porter, Stephen R. i Michael E. Whitcomb. 2007. „Mixed-mode contacts in web surveys". *Public Opinion Quarterly* 71(4):635–48. doi: 10.1093/poq/nfm038.

- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rasmussen, Karsten Boye. 2014. „Social Science Metadata and the Foundations of the DDI”. *IASSIST Quarterly* 37(1–4):28. doi: 10.29173/iq499.
- Roberts, Caroline, Caroline Vandenplas i Jessica M. E. Herzing. 2020. „A Validation of R-Indicators as a Measure of the Risk of Bias Using Data from a Nonresponse Follow-Up Survey”. *Journal of Official Statistics* 36(3):675–701. doi: 10.2478/jos-2020-0034.
- Roberts, Caroline, Caroline Vandenplas i Michele Ernst Stahli. 2014. „Evaluating the Impact of Response Enhancement Methods on the Risk of Nonresponse Bias and Survey costs”. *Survey Research Methods* 8(2):67–80.
- Rosen, Jeffrey, Joe Murphy, Andy Peytchev, Sarah Riley i Mark Lindblad. 2011. „The Effects of Differential Interviewer Incentives on a Field Data Collection Effort”. *Field Methods* 23(1):24–36. doi: 10.1177/1525822X10383390.
- Rybak, Adam. 2018. „The Selection of Survey Modes on the Basis of Respondents’ Preferences Analysis”. *Ask: Research and Methods* 27(1):23–39. doi: 10.18061/ask.v27i1.0002.
- Rybak, Adam. 2019. „The effects of visual and aural communication on the quality of survey measurement. The proposition of a new theoretical approach to survey mode preference on the basis of the working memory concept”. *Przegląd Socjologiczny* 68(3):207–27. doi: 10.26485/ps/2019/68.3/8.
- Rybak, Adam. 2020. „Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference”. *Polish Sociological Review* 210(2):219–28. doi: 10.26412/psr210.05.
- Rybak, Adam. 2023. „Survey Mode and Nonresponse Bias: A Meta-Analysis Based on the Data from the International Social Survey Programme Waves 1996–2018 and the European Social Survey Rounds 1 to 9”. *PLOS ONE* 18(3). doi: 10.1371/journal.pone.0283092.
- Rybak, Adam. 2025. „Analiza jakości dokumentacji metodologicznej sondaży: Studium przypadku inwentaryzacji 30 programów badań porównawczych z całego świata”. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 87(1):287–309. doi: 10.14746/rpeis.2025.87.1.15.
- Sakshaug, Joseph W., Ting Yan i Roger Tourangeau. 2010. „Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items”. *Public Opinion Quarterly* 74(5):907–33. doi: 10.1093/poq/nfq057.
- Salmon, Charles T. i John Spicer Nichols. 1983. „The Next-Birthday Method of Respondent Selection”. *Public Opinion Quarterly* 47(2):270–76. doi: 10.1086/268785.
- Schouten, Barry, Fannie Cobben i Jelke Bethlehem. 2009. „Indicators for the representativeness of survey response”. *Survey Methodology* 35(1):101–13.
- Schouten, Barry, Natalie Shlomo i Chris Skinner. 2011. „Indicators for monitoring and improving representativeness of response”. *Journal of Official Statistics* 27(2):231–53.
- Schuman, Howard. 1982. „Artifacts Are in the Mind of the Beholder”. *The American Sociologist* 17(1):21–28.

- Shih, Tse-Hua i Xitao Fan. 2007. „Response rates and mode preferences in web-mail mixed-mode surveys: a meta-analysis”. *International Journal of Internet Science* 2(1):59–82.
- Shih, Tse-Hua i Xitao Fan. 2008. „Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis”. *Field Methods* 20(3):249–71. doi: 10.1177/1525822X08317085.
- Singer, Eleanor, John Van Hoewyk, Nancy Gebler, Trivellore Raghunathan i Katherine Mcgonagle. 1999. „The effect of incentives on response rates in interviewer-mediated surveys”. *Journal of Official Statistics* 15(2):217–30.
- Slomczynski, Kazimierz M., Irina Tomescu-Dubrow i Ilona Wysmulek. 2022. „Survey Data Quality in Analyzing Harmonized Indicators of Protest Behavior: A Survey Data Recycling Approach”. *American Behavioral Scientist* 66(4):412–33. doi: 10.1177/00027642211021623.
- Slomczynski, Kazimierz Maciek, Przemek Powalko i Tadeusz Krauze. 2017. „Non-Unique Records in International Survey Projects: The Need for Extending Data Quality Control”. *Survey Research Methods* Vol 11:1–16. doi: 10.18148/SRM/2017.V11I1.6557.
- Smyth, Jolene D., Kristen Olson i Alian Kasabian. 2014. „The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement”. *Survey Research Methods* 8(3):137–52.
- Smyth, Jolene D., Kristen Olson i Morgan M. Millar. 2014. „Identifying predictors of survey mode preference”. *Social Science Research* 48:135–44. doi: 10.1016/j.ssresearch.2014.06.002.
- Sodeur, Wolfgang. 1997. „Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen”. *ZA-Information / Zentralarchiv für Empirische Sozialforschung* (41):58–82.
- Statistics Division of the United Nations Secretariat. 1999. „Standard Country or Area Codes for Statistical Use (M49)”. Pobrano 11 luty 2021 (<https://unstats.un.org/unsd/methodology/m49/>).
- Stoop, Ineke. 2005. *The Hunt for the Last Respondent*. Den Haag: Social and Cultural Planning Office.
- Stoop, Ineke, Jaak Billiet, Achim Koch i Rory Fitzgerald. 2010. *Improving Survey Response: Lessons Learned from the European Social Survey*. Chichester: John Wiley & Sons, Ltd.
- Sturgis, Patrick, Joel Williams, Ian Brunton-Smith i Jamie Moore. 2017. „Fieldwork Effort, Response Rate, and the Distribution of Survey Outcomes: A Multilevel Meta-analysis”. *Public Opinion Quarterly* 81(2):523–42. doi: 10.1093/poq/nfw055.
- Sztabiński, Franciszek. 2011. *Ocena jakości danych w badaniach surveyowych*. Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński, Paweł B. 2018. „Does mixed mode data collection improve the achieved sample? A comparison of the ESS 7 and a mixed mode experiment”. w *New Uncertainties and Anxieties in Europe: Seven Waves of the European Social Survey*, zredagowane przez F. Sztabiński, H. Domański i P. B. Sztabiński. Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang.

- Tahmasbipour, Najaf i K. Ahmady. 2011. „The investigation of prevalence rate and diagnosis of exam anxiety risk factors among Shahid Rajaee University students (SRTTU)”. *Procedia - Social and Behavioral Sciences* 30:717–21. doi: 10.1016/j.sbspro.2011.10.140.
- Tarnai, John i M. Chris Paxson. 2004. „Survey Mode Preferences of Business Respondents”. w *59th Annual Conference, American Association for Public Opinion Research*.
- Tepper, Stephanie .. J., Mikaela K. Spruill, Bharathy Premachandra i Neil A. Lewis. 2022. „Surveys as Conversations between Makers and Takers: A Conversational Framework for Assessing and Responding to Community Needs”. *Analyses of Social Issues and Public Policy* 22(3):857–75. doi: 10.1111/asap.12326.
- The American Association for Public Opinion Research. 2023. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 10th edition. AAPOR.
- Tomescu-Dubrow, Irina, Christof Wolf, Kazimierz M. Slomczynski i J. Craig Jenkins, red. 2024. *Survey Data Harmonization in the Social Sciences*. Hoboken, New Jersey: Wiley.
- Tourangeau, Roger, J. Michael Brick, Sharon Lohr i Jane Li. 2016. „Adaptive and responsive survey designs: a review and assessment”. *Journal of the Royal Statistical Society. Series A: Statistics in Society*. doi: 10.1111/rssa.12186.
- Tourangeau, Roger, Lance J. Rips i Kenneth A. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- United Nations. 2022. „Demographic indicators by region, subregion and country 1950-2021”.
- Van Wettene, Niek. 2022. „How to Implement the Metadata Standard «Data Documentation Initiative (DDI)»?”
- Vannieuwenhuyze, Jorre T. a. 2013. „On the Relative Advantage of Mixed-Mode versus Single-Mode Surveys”. *Survey Research Methods* 8(1):31–42. doi: 10.18148/srm/2014.v8i1.5485.
- Vardigan, Mary, Peter Granda i Lynette Hoelter. 2016. „Documenting Survey Data Across the Life Cycle”. S. 443–59 w *The SAGE Handbook of Survey Methodology*, zredagowane przez C. Wolf, D. Joye, T. Smith i Y. Fu. London: SAGE Publications Ltd.
- Viechtbauer, Wolfgang. 2010. „Conducting Meta-Analyses in R with the metafor Package”. *Journal of Statistical Software* 36(3):1–48. doi: 10.18637/jss.v036.i03.
- Villard, Ana i Rory Fitzgerald. 2017. „Using mixed modes in survey research: Evidence from six experiments in the ESS”. S. 273–310 w *Values and identities in Europe: Evidence from the European Social Survey*, zredagowane przez M. J. Breen. Abingdon: Routledge.
- Vogl, Susanne, Jennifer A. Parsons, Linda K. Owens i Paul J. Lavrakas. 2019. „Experiments on the Effects of Advance Letters in Surveys”. S. 89–110 w *Experimental Methods in Survey Research*, zredagowane przez P. Lavrakas, M. Traugott, C. Kennedy, A. Holbrook, E. de Leeuw i B. West. Hoboken: Wiley.
- Wagner, James i Ineke A. L. Stoop. 2018. „Comparing Nonresponse and Nonresponse Biases in Multinational, Multiregional, and Multicultural Contexts”. S. 807–33 w *Advances*

- in Comparative Survey Methods*, zredagowane przez T. P. Johnson, B.-E. Pennell, I. A. L. Stoop i B. Dorer. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Wagner, James, Brady T. West, Heidi Guyer, Paul Burton, Jennifer Kelley, Mick P. Couper i William D. Mosher. 2017. „The Effects of a Mid-Data Collection Change in Financial Incentives on Total Survey Error in the National Survey of Family Growth”. S. 155–77 w *Total Survey Error in Practice*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Weisberg, Herbert. 2005. *The Total Survey Error Approach*. Chicago: The University of Chicago Press.
- Willimack, Diane K., Howard Schuman, Beth-Ellen Pennell i James M. Lepkowski. 1995. „Effects of a Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-Face Survey”. *Public Opinion Quarterly* 59(1):78. doi: 10.1086/269459.
- Wolf, Christof, Pablo Christmann, Tobias Gummer, Christian Schnaudt i Sascha Verhoeven. 2021. „Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys”. *Public Opinion Quarterly* 85(2):623–48. doi: 10.1093/poq/nfab039.
- Wysmulek, Ilona, Irina Tomescu-Dubrow i Joonghyun Kwak. 2022. „Ex-Post Harmonization of Cross-National Survey Data: Advances in Methodological and Substantive Inquiries”. *Quality & Quantity* 56(3):1701–8. doi: 10.1007/s11135-021-01187-7.
- Yammarino, Francis J., Steven J. Skinner i Terry L. Childers. 1991. „Understanding mail survey response behavior a meta-analysis”. *Public Opinion Quarterly* 55(4):613–39. doi: 10.1086/269284.
- Yu, Julie i Harris Cooper. 1983. „A Quantitative Review of Research Design Effects on Response Rates to Questionnaires”. *Journal of Marketing Research* 20(1):36–44. doi: 10.2307/3151410.

## **Spis adresów repozytoriów wszystkich analizowanych programów**

**AB:** [www.afrobarometer.org](http://www.afrobarometer.org)

**ABS:** [www.asianbarometer.org](http://www.asianbarometer.org)

**ARAB:** [www.arabbarometer.org](http://www.arabbarometer.org)

**ASES:** [www.icpsr.umich.edu/web/ICPSR/studies/22324](http://www.icpsr.umich.edu/web/ICPSR/studies/22324)

**ASIAB:**

05 – [www.openicpsr.org/openicpsr/project/149101](http://www.openicpsr.org/openicpsr/project/149101)

06 – [www.openicpsr.org/openicpsr/project/163441](http://www.openicpsr.org/openicpsr/project/163441)

07 – [www.openicpsr.org/openicpsr/project/163461](http://www.openicpsr.org/openicpsr/project/163461)

08 – [www.openicpsr.org/openicpsr/project/163481](http://www.openicpsr.org/openicpsr/project/163481)

**CCEB:** [www.gesis.org/en/eurobarometer-data-service/survey-series/candidate-countries-eb](http://www.gesis.org/en/eurobarometer-data-service/survey-series/candidate-countries-eb)

**CCS:** [search.gesis.org/research\\_data/ZA0624](http://search.gesis.org/research_data/ZA0624)

**CEEB:** [www.gesis.org/eurobarometer-data-service/survey-series/central-eastern-eb](http://www.gesis.org/eurobarometer-data-service/survey-series/central-eastern-eb)

**CNEP:** [u.osu.edu/cnep/surveys/surveys-through-2012](http://u.osu.edu/cnep/surveys/surveys-through-2012)

**CONS:** [search.gesis.org/research\\_data/ZA4054](http://search.gesis.org/research_data/ZA4054)

**CRRC:** [caucasusbarometer.org/en/downloads](http://caucasusbarometer.org/en/downloads)

**CSES:** [cses.org/data-download/download-data-documentation](http://cses.org/data-download/download-data-documentation)

**EASS:** [www.icpsr.umich.edu/web/ICPSR/series/486](http://www.icpsr.umich.edu/web/ICPSR/series/486)

**EB, ECS:** [www.gesis.org/en/eurobarometer-data-service](http://www.gesis.org/en/eurobarometer-data-service)

**EES:** [www.gesis.org/en/services/finding-and-accessing-data/international-survey-programs/european-election-studies](http://www.gesis.org/en/services/finding-and-accessing-data/international-survey-programs/european-election-studies)

**EQLS:** [doi.org/10.5255/UKDA-Series-200013](http://doi.org/10.5255/UKDA-Series-200013)

**ESS:** [www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)

**EUNB:** [www.gesis.org/en/services/finding-and-accessing-data/international-survey-programs/eu-neighbourhood-barometer](http://www.gesis.org/en/services/finding-and-accessing-data/international-survey-programs/eu-neighbourhood-barometer)

**EVS:** [europeanvaluesstudy.eu](http://europeanvaluesstudy.eu)

**GALLUP:**

[www.gallup.com/services/177797/country-data-set-details.aspx](http://www.gallup.com/services/177797/country-data-set-details.aspx)

[www.gallup.com/analytics/318875/global-research.aspx](http://www.gallup.com/analytics/318875/global-research.aspx)

**INTUNE:**

I – [search.gesis.org/research\\_data/ZA5695](http://search.gesis.org/research_data/ZA5695)

II – [search.gesis.org/research\\_data/ZA5696](http://search.gesis.org/research_data/ZA5696)

**ISSP:** [www.gesis.org/en/issp/home](http://www.gesis.org/en/issp/home)

**ISSP-SOR:** [search.gesis.org/research\\_data/ZA7630](http://search.gesis.org/research_data/ZA7630)

**LAPOP:** [www.vanderbilt.edu/lapop](http://www.vanderbilt.edu/lapop)

**LITS:** [www.ebrd.com/what-we-do/economic-research-and-data/data/lits.html](http://www.ebrd.com/what-we-do/economic-research-and-data/data/lits.html)

**MEDW:** [dataverse.harvard.edu/dataverse/MEDW](http://dataverse.harvard.edu/dataverse/MEDW)

**NBB:**

93 – [doi.org/10.5255/UKDA-SN-6489-1](https://doi.org/10.5255/UKDA-SN-6489-1)

95 – [doi.org/10.5255/UKDA-SN-6490-1](https://doi.org/10.5255/UKDA-SN-6490-1)

96 – [doi.org/10.5255/UKDA-SN-6491-1](https://doi.org/10.5255/UKDA-SN-6491-1)

00 – [doi.org/10.5255/UKDA-SN-6492-1](https://doi.org/10.5255/UKDA-SN-6492-1)

**NEB/NDB:**

I – [doi.org/10.5255/UKDA-SN-6452-1](https://doi.org/10.5255/UKDA-SN-6452-1)

II – [doi.org/10.5255/UKDA-SN-6453-1](https://doi.org/10.5255/UKDA-SN-6453-1)

III – [doi.org/10.5255/UKDA-SN-6454-1](https://doi.org/10.5255/UKDA-SN-6454-1)

IV – [doi.org/10.5255/UKDA-SN-6455-1](https://doi.org/10.5255/UKDA-SN-6455-1)

V – [doi.org/10.5255/UKDA-SN-6456-1](https://doi.org/10.5255/UKDA-SN-6456-1)

VI – [doi.org/10.5255/UKDA-SN-5242-1](https://doi.org/10.5255/UKDA-SN-5242-1)

VII – [doi.org/10.5255/UKDA-SN-5243-1](https://doi.org/10.5255/UKDA-SN-5243-1)

I-XIII – [doi.org/10.5255/UKDA-SN-5241-1](https://doi.org/10.5255/UKDA-SN-5241-1)

**WVS:** [www.worldvaluessurvey.org](http://www.worldvaluessurvey.org)

## **Publikacje z cyklu**

ADAM RYBAK<sup>a</sup>

**ANALIZA JAKOŚCI DOKUMENTACJI  
METODOLOGICZNEJ SONDAŻY:  
STUDIUM PRZYPADKU INWENTARYZACJI  
30 PROGRAMÓW BADAŃ PORÓWNAWCZYCH  
Z CAŁEGO ŚWIATA**

**SURVEY DOCUMENTATION QUALITY: A CASE STUDY  
OF THE METHODOLOGICAL INVENTORY COVERING  
CHARACTERISTICS OF SURVEYS FROM 30 COMPARATIVE  
PROGRAMMES FROM AROUND THE GLOBE**

Methodological documentation serves as the primary source of information about sampling schemes, fieldwork procedures and related sources of potential measurement error in survey research. The quality of such documentation, including its completeness and standardization, is crucial for conducting methodological analyses based on inventories of multiple comparative survey programmes. Individual examples and quantitative summaries of historical development, as well as current methodological reports, serve as a foundation for reflecting on potential future analyses based on large methodological inventories. Therefore, the aim of this paper was to thoroughly analyse the gathered data, identify critical issues, and prepare proposals for a standardized report format suitable for automatic processing. The implementation of these proposals could enhance the capacity to conduct extensive methodological analyses. Given the increasing costs and labour intensity of survey research, both substantial and methodological data should be thoroughly utilized. The analysis is based on a methodological inventory of 10,493 surveys from 30 comparative programmes, spanning all inhabited continents. The paper demonstrates the challenges encountered during data collection and presents completeness indicators for both selected programmes and the complete dataset. The findings reveal that low levels of standardization between analysed methodological reports necessitate a critical evaluation of current practices. This situation impedes the development of future methodological inventories. The paper identifies critical barriers to creating quantitative comparisons of documentation quality and proposes solutions. Additionally, it presents a standardized report format addressing existing issues and potentially facilitating harmonization efforts.

**Keywords:** methodological documentation; comparative surveys; survey methodology; presenting research results

---

<sup>a</sup> Adam Mickiewicz University, Poznań, Poland /  
Uniwersytet im. Adama Mickiewicza w Poznaniu, Polska  
[adam.rybak@amu.edu.pl](mailto:adam.rybak@amu.edu.pl), <https://orcid.org/0000-0003-3981-1081>

Podstawowym źródłem informacji o schematach losowania próby i procedurach terenowych badań sondażowych oraz związanych z nimi potencjalnych źródłach błędów pomiaru jest dokumentacja metodologiczna. Jej szeroko pojęta jakość, której istotny element stanowią kompletność i standaryzacja, jest czynnikiem determinującym możliwość prowadzenia badań metodologicznych na dużych zbiorach danych pochodzących z międzykrajowych programów porównawczych. Za punkt wyjścia do poprawy jakości dokumentacji metodologicznej powinno służyć dokładne rozpoznanie dotychczasowego stanu rzeczy. Stanowi to pierwszy cel poniższego badania. Kolejnym celem była analiza problemów obecnych w tym obszarze, która umożliwiła sformułowanie zaleceń dążących do zwiększenia standaryzacji tworzonych raportów i tym samym możliwości lepszego wykorzystania danych pochodzących z badań sondażowych. Analiza zawarta w artykule opiera się na inwentaryzacji danych metodologicznych z 30 międzykrajowych porównawczych programów sondażowych z całego świata. Podsumowuję całość zbioru opartego na 10 493 realizacjach krajowych, poszczególne programy, a także przedstawiam problematyczne praktyki na podstawie konkretnych raportów metodologicznych. Przykłady i zestawienia ukazują rozwój i obecny stan jakości opracowanej dokumentacji. Efektem badania jest krytyczna ocena obecnych praktyk dokumentacyjnych, które ze względu na niewielki poziom standaryzacji pomiędzy projektami bardzo utrudniają praktyczną inwentaryzację informacji na temat przyjętych procedur badawczych. Przedstawiam zestawienia problematycznych praktyk oraz propozycję standaryzacji formy raportów, która rozwiązałaby większość trudności pojawiających się obecnie w procesie harmonizacji danych metodologicznych.

Słowa kluczowe: dokumentacja metodologiczna; międzykrajowe badania porównawcze; metodologia sondażowa; komunikowanie wyników badań

---

## I. WPROWADZENIE

Podstawowym źródłem, z którego korzysta się w przypadku analiz praktyk sondażowych, jest dokumentacja metodologiczna. Badacz ma w pierwszym rzędzie dostęp do tych informacji, które osoby przygotowujące raporty postanowiły w nich zawrzeć. Jeśli nie podejmie dodatkowych wysiłków, jak próba osobistego kontaktu z koordynatorami krajowymi programów sondażowych, przeczesywanie surowych baz danych wynikowych w poszukiwaniu zmiennych metodologicznych czy studia monografii opisujących historię badań – będzie widział tyle, ile chciano mu pokazać. Artykuł ten jest efektem inwentaryzacji praktyk dokumentacyjnych z 30 porównawczych międzykrajowych programów sondażowych z całego świata.

Inwentaryzacja kompletności informacji zawartej w dokumentacji była w przypadku niniejszego badania pochodną inwentaryzacji praktyk sondażowych, co miało szereg konsekwencji. Jedną z nich było to, że dobór zmiennych opierał się na zestawie zadeklarowanych obszarów, które – opierając się na wynikach badań innych metodologów – mogłem określić jako mające wpływ na jakość pomiaru. Z tego powodu zakres danych, które znalazły się w stworzonej przeze mnie bazie kompletności dokumentacji sondaży, jest dosyć szeroki, ale podporządkowany konkretnemu typowi odbiorcy. Nie uwzględnia perspektywy innych grup zainteresowanych użytkowaniem danych, jak dziennikarze,

decydenci polityczni czy badacze zainteresowani np. paradata. Obszary, które zostały wyznaczone jako potencjalnie doniosłe, to liczba i motywowanie ankietów, kontakt z respondentami i ich motywowanie, różne aspekty doboru próby, oraz techniki sondażu.

Poza kilkoma arbitralnie pojętymi decyzjami o sięgnięciu głębiej, zbieranie danych ograniczyłem do informacji umieszczonych w repozytoriach i na stronach internetowych projektów. Przykładami takich podjętych przeze mnie decyzji mogą być: udany kontakt z zespołem Europejskiego Sondażu Społecznego w celu uzgodnienia przelicznika kategorii wywiadów niezrealizowanych na szerzej przyjęty standard (The American Association for Public Opinion Research, 2023); upewnienie się, jak często w Americas Barometer używano tzw. *frequency matching* (Castorena, 2021); czy incydentalne użycie informacji o technice badania zawartych w surowych danych wynikowych w celu ułatwienia kodowania. Nie znaczy to jednak, że inwentaryzacja danych była monotonna, informacje były przedstawione w najróżniejszy sposób, o czym mowa będzie dalej. W przeszłości analizy procedur dokumentacji badań sondażowych skupiały się w dużej mierze na wskazywaniu dobrych praktyk (Mohler i in., 2008, 2010), jednak absorpcja tych rozwiązań przez instytucje prowadzące badania wydaje się powolna.

W niniejszym artykule przyjmuję odwrotne podejście, skupiając się na krytycznej analizie stanu obecnego. Z opisu tego, jak rzeczywiście wyglądają raporty metodologiczne, wyprowadzam propozycje refleksji i ewentualnych ulepszeń. Efektem jest zestawienie zarówno konkretnych przykładów, jak i zbiorczych podsumowań dotyczących jakości dokumentacji w sondażowych programach porównawczych. Przedstawiam także zbiór problemów dotyczących możliwości ilościowych porównań jakości dokumentacji. Brak odpowiedzi na nie może hamować rozwój tego typu analiz.

## 1. Dostępne rozwiązanie problemu standaryzacji raportów metodologicznych

Dla dużej części problemów zawartych w tej pracy rozwiązanie teoretyczne już istnieje. To, dlaczego dotąd jest implementowane w bardzo ograniczonym zakresie, wymaga jednak zupełnie innego rodzaju studiów. Prawdziwa dostępność, a więc możliwość łatwego gromadzenia, przetwarzania, agregowania danych metodologicznych, wymaga, by nie miały one różnych formatów i form prezentacji graficznej, ale były zestandaryzowane. Idea takiej standaryzacji nie jest niczym nowym – projekt Data Documentation Initiative (DDI), utworzony w 1995 r. (Rasmussen, 2014; Vardigan, 2014; Vardigan i in., 2008), oferuje rozwiązanie, które jest otwarte i rozwinięte do stabilnej wersji 3.3 (Van Wette, 2022). Opiera się ono na architekturze .xml, co umożliwia zautomatyzowany odczyt maszynowy.

Data Documentation Initiative umożliwia także udoskonalenie kontroli jakości w całym procesie przygotowania i realizacji badania, ale dla rozważań w niniejszym artykule interesujący jest jedynie jego potencjał do wy-

tworzania łatwej w obróbce dokumentacji o stałych, zestandaryzowanych nagłówkach. W praktyce DDI nie stało się dotąd alternatywą dla dotychczasowych form dokumentacji metodologicznej. Tym samym nie można wykorzystać jego potencjału, prowadząc badania skupiające się na inwentaryzacji dużych zbiorów danych sondażowych ukierunkowanych na analizy metodologiczne. Nie potrafię wskazać przyczyny tego stanu rzeczy, w innych analizach wskazuje się np. mentalność izolacji (*silo mentality*) i strach przed zmianą (Fonteneau, 2011).

## 2. Przykłady obecnych w literaturze sposobów analizowania dokumentacji

Sposób oceny dokumentacji metodologicznej, który przedstawiam w tym artykule, nie jest ani jedynym, ani najlepszym z możliwych podejść. Jego niewątpliwym ograniczeniem jest fakt, że był podporządkowany procesowi inwentaryzacji praktyk badawczych, był więc środkiem, a nie celem.

Zanim omówię dalej własne podejście, przedstawię przykłady alternatyw. Najbardziej odmienne sposoby oceny jakości dokumentacji prezentują badania będące wynikiem projektu Survey Data Recycling ([SDR]; Dubrow i in., 2016; Slomeczynski i in., 2017, 2022; Slomeczynski i Tomescu-Dubrow, 2018). Badania te koncentrują się nie na samej zawartości dokumentacji, ale ocenie zgodności pomiędzy dokumentacją a danymi wynikowymi (np. zgodność między opisem zmiennych w kluczach kodowych i danych wynikowych), jak również jakością samych danych wynikowych (np. obecnością duplikatów obserwacji). W tym przypadku jest to związane z celem użycia tych danych ponownie, w formie zharmonizowanej, wzbogaconej o odpowiednie wskaźniki i flagi opracowane na podstawie znalezionych niezgodności i błędów powstałych przy opracowywaniu wyników.

Przykład propozycji sposobu analizowania dokumentacji, który częściowo pokrywa się z przedstawionym przeze mnie, ale jednocześnie proponuje analizę szerszej problematyki, przedstawił Jedinger z zespołem (2018). Wymieniony przez nich zestaw kluczowych pytań, jakie należy sobie zadać przy ocenie dokumentacji metodologicznej, zawiera, poza kwestiami obecnymi w moim badaniu, poszukiwanie w nich odpowiedzi na takie pytania jak: Jaki był cel zbierania danych? Jakie zbierano informacje? Jak dane były edytowane, kodowane i ważone? Jakie podjęto środki w celu ochrony danych osobowych? Takie podejście pozwala na bardziej całościową ocenę przydatności danych metodologicznych i staranności opracowywania wyników badań. Umożliwia także większe skupienie się na aspekcie dokumentacji jako sposobu komunikowania się zbierającego dane z ich potencjalnym użytkownikiem (Tepper i in., 2022). Nie zastosowałem go, gdyż byłoby nadmiarowe dla osiągnięcia celu opisu praktyk i stworzenia modeli metaanaliz reprezentatywności (Ry-

bak, 2023), które były głównym celem projektu będącego źródłem danych do niniejszego artykułu.

Wreszcie najbardziej zbliżony sposób działania zaprezentowano w analizach Kołczyńskiej (2014; Kołczyńska i Schoene, 2018), która jako pierwsza zaprezentowała koncepcję ilościowej oceny kompletności dokumentacji badań sondażowych, opierając się na danych z krajów Europy Południowo-Wschodniej. Badaczka uczestniczyła następnie zarówno w badaniach związanych ze wspomnianym projektem SDR, jak i w inspirowanych jej dokonaniach analizach zespołu skupionego wokół ośrodka poznańskiego (Jabkowski, 2019, 2023; Jabkowski i Kołczyńska, 2020; Jabkowski i in., 2021). Podstawową różnicą między podejściem cytowanych badaczy a moim jest fakt, że autorzy ci podjęli się prób porównania kompletności pomiędzy modułami zmiennych w obrębie programu oraz położyli silny nacisk na porównania pomiędzy programami, ograniczając krytykę dostępnej dokumentacji wyłącznie do aspektu jej kompletności. To podejście wydaje się częściowo problematyczne ze względu na dramatyczną różnicę w formie i standaryzacji dokumentów pomiędzy programami, jak również na różną wagę poszczególnych informacji z dokumentacji dla odbiorcy końcowego. Te fragmenty niniejszego artykułu, które opierają się na podsumowaniach kompletności zmiennych dotyczących informacji zawartych w dokumentacji, kontynuują podejście zaproponowane przez wymienionych badaczy.

Istotnym dla omawianej problematyki wydarzeniem jest przedstawienie przez łączoną komisję American Association for Public Opinion Research [AAPOR] i World Association of Public Opinion Research [WAPOR] raportu dotyczącego jakości sondażowych badań porównawczych (AAPOR/WAPOR, 2021). Przedstawiono w nim rozdział dotyczący dokumentacji, który porusza między innymi kwestie dotyczące zakresu tematów badanych w tym artykule. Jego zakres jest jednak szerszy. Przykładowo, poniższe analizy planowo pomijają kwestie przygotowania, testowania i tłumaczenia kwestionariuszy, które są istotne dla wielu badaczy. W wytycznych raportu wskazano także problemy dostępności różnych rodzajów para- i metadata przydatnych do wtórnych analiz. Z punktu widzenia tego artykułu interesujące są wytyczne dotyczące *study-level-metadata*, czyli informacji o konkretnych sondażach krajowych w ramach projektu.

Tabela 1 zawiera porównanie elementów, których obecność sprawdzałem w swojej analizie z wytycznymi z raportu. Warto zaznaczyć, że informacje o tym, czy sondaż był częścią panelu, były pomocniczo kodowane tylko dla tej niewielkiej części programów, w których panowała bardzo duża dowolność metodologiczna i panele rzeczywiście występowały, nie posiadam więc wiedzy o obecności tej informacji dla całej bazy. Pomiąłem także w całości temat wag statystycznych, gdyż planowane metaanalizy dotyczące reprezentatywności nie miały posługiwać się nimi ze względu na dotychczasowe ustalenia badaczy (Jabkowski i in., 2021).

**Tabela 1**

Porównanie zaleceń zawartych w Załączniku 6 raportu (AAPOR/WAPOR, 2021)  
ze zmiennymi zawartymi w inwentaryzacji

| Zalecenia AAPOR/WAPOR       | Informacje w bazie o:                                                        | Zgodność |
|-----------------------------|------------------------------------------------------------------------------|----------|
| Finansowanie                |                                                                              | –        |
| Instytucja badawcza         |                                                                              | –        |
| Cele badania                |                                                                              | –        |
| Przekrojowe/podłużne        | (zbierane dla niektórych programów,<br>nieuwzględnione w bazie dokumentacji) | –        |
| Populacja badana            | minimalnym wieku                                                             | +        |
|                             | maksymalnym wieku                                                            | +        |
| Jednostka obserwacji        | (we wszystkich tylko jednostka)                                              | +        |
| Dobór próby                 | operacje                                                                     | +        |
|                             | doborze wewnątrzgrupowym                                                     | 0        |
|                             | substytucjach/kwotach                                                        | +        |
| Zachęty materialne          | zachętach materialnych                                                       | +        |
|                             | listach zapowiednich                                                         | 0        |
| Technika                    | technice                                                                     | +        |
| Kwestionariusz              | długości wywiadu                                                             | 0        |
| Testowanie kwestionariusza  |                                                                              | –        |
| Tłumaczenie kwestionariusza |                                                                              | –        |
| Ankieterzy                  | sposobie wynagradzania ankieterów                                            | 0        |
|                             | liczbie ankieterów                                                           | +        |
| Procedury terenowe          | minimalnej liczbie prób kontaktu                                             | 0        |
|                             | backcheckach                                                                 | 0        |
| Response rate               | liczbie wywiadów                                                             | 0        |
|                             | dowolnym RR                                                                  | 0        |
|                             | umożliwiająca obliczenie RR6 AAPOR                                           | +        |
| Wagi                        | (nieuwzględnione celowo)                                                     | –        |
| Czas i miejsce              | (dostępne da wszystkich)                                                     | +        |

+ pełna zgodność, 0 – zgodność tematyczna, – brak zgodności

Źródło: opracowanie własne.

## II. DANE

Uwzględnione sondaże pochodzą ze 179 krajów ze wszystkich zamieszkałych kontynentów (z uwzględnieniem krajów współcześnie nieistniejących, jak Jugosławia, Związek Sowiecki, Czechosłowacja czy NRD, oraz nieuznawanych, jak Tajwan). Zakres danych podsumowuje tabela 2.

Punktem wyjścia selekcji badań do analizy była lista porównawczych sondaży międzykrajowych wykonana i udostępniona w Internecie przez GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS, 2019). Lista ta nie jest już udostępniana przez tę instytucję.

**Tabela 2**

Liczebności i zakres dat programów badawczych uwzględnionych w analizie

| <b>Program</b>                                                              | <b>Liczba sondaży</b> | <b>Pierwszy i ostatni rok badań uwzględnionych w bazie</b> |
|-----------------------------------------------------------------------------|-----------------------|------------------------------------------------------------|
| AB – Afrobarometer                                                          | 205                   | 1999–2021                                                  |
| ABS – Asian Barometer Survey                                                | 69                    | 2001–2021                                                  |
| ARAB – Arab Barometer                                                       | 63                    | 2006–2022                                                  |
| ASES – Asia Europe Survey                                                   | 18                    | 2001                                                       |
| ASIAB – Asia Barometer                                                      | 34                    | 2005–2008                                                  |
| CCEB – Candidate Countries Eurobarometer                                    | 142                   | 2001–2004                                                  |
| CCS – Civic Culture Study                                                   | 5                     | 1959                                                       |
| CEEB – Central and Eastern Eurobarometer                                    | 120                   | 1990–1997                                                  |
| CNEP – Comparative National Elections Project                               | 61                    | 1990–2021                                                  |
| CONS – Consolidation of Democracy in Central and Eastern Europe             | 24                    | 1990–2001                                                  |
| CRRC – Caucasus Barometer                                                   | 30                    | 2006–2020                                                  |
| CSES – Comparative Study of Electoral Systems                               | 209                   | 1996–2020                                                  |
| EASS – East Asia Social Survey                                              | 24                    | 2006–2018                                                  |
| EB – Eurobarometer                                                          | 4 604                 | 1974–2019                                                  |
| ECS – European Communities Study                                            | 20                    | 1970–1973                                                  |
| EES – European Election Studies                                             | 174                   | 1989–2019                                                  |
| EQLS – European Quality of Life Surveys                                     | 125                   | 2003–2016                                                  |
| ESS – European Social Survey                                                | 243                   | 2002–2022                                                  |
| EUNB – EU Neighbourhood Barometer                                           | 96                    | 2012–2014                                                  |
| EVS – European Values Study                                                 | 147                   | 1981–2020                                                  |
| GALLUP – Gallup World Poll                                                  | 2 280                 | 2006–2021                                                  |
| INTUNE – Integrated and United                                              | 35                    | 2007–2009                                                  |
| ISSP – International Social Survey Programme                                | 961                   | 1985–2022                                                  |
| ISSP_SOR – Study of Religion of Additional 14 Countries, based on ISSP 2018 | 14                    | 2018                                                       |
| LAPOP – Latin American Public Opinion Project (Americas Barometer)          | 275                   | 1973–2021                                                  |
| LITS – Life in Transition Survey                                            | 98                    | 2006–2016                                                  |
| MEDW – Making Electoral Democracy Work                                      | 10                    | 2010–2016                                                  |
| NBB – New Baltic Barometer                                                  | 12                    | 1993–2000                                                  |
| NEB/NDB – New Europe Barometer/New Democracies Barometer                    | 94                    | 1991–2007                                                  |
| WVS – World Values Survey                                                   | 301                   | 1981–2021                                                  |
| <b>Razem</b>                                                                | <b>10 493</b>         | <b>1959–2022</b>                                           |

Źródło: opracowanie własne.

Kryteriami, które decydowały o obecności programu badawczego w stworzonej bazie, były:

1) obecność w zestawieniu programów porównawczych (GESIS, 2019) – dostępne w archiwum autora i w narzędziu Internet Archive Wayback Machine (<https://web.archive.org/>);

2) anglojęzyczna dokumentacja w wolnym dostępie internetowym (z tego powodu wykluczyłem duży program Latinobarometro, który swoją dokumentację publikuje konsekwentnie wyłącznie po hiszpańsku);

3) niekomercyjny charakter (poza GALLUP);

4) bazy danych wynikowych w wolnym dostępie internetowym (poza GALLUP i ASES);

5) włączyłem tylko projekty badające (niemal) całą populację kraju, to jest nie uwzględniłem tych obejmujących wyłącznie osoby starsze, młodzież, miasta albo wybrany region kraju. (Kilka uwzględnionych i dodatkowo oznaczonych przypadków badań obejmujących tylko wybrany obszar kraju zostało w większości wychwyconych na etapie kontrolnym. Te, w których mniejszościowy obszar kraju był wyłączony z badania na skutek wojny, powstania, albo klęki żywiołowej, kodowałem normalnie);

6) uwzględniłem tylko badania, które dotyczyły (co najmniej w części kwestionariusza) jednostki, a nie, przykładowo, gospodarstwa domowego jako całości;

7) nie uwzględniłem „programów” będących tylko harmonizacją *post factum* wcześniejszych badań.

### III. METODOLOGICZNA INWENTARYZACJA SONDAŻY: KRYTYCZNA ANALIZA FORMUŁY RAPORTÓW

#### 1. Jakość samej formy publikacji metodologicznych – przykłady

Dokumentacja metodologiczna sondaży rodzi się wtedy, kiedy one same – w pierwszej połowie XX w. (Mohler i in., 2008), a więc w czasach maszynopisu. Z punktu widzenia jakości samego zapisu powinniśmy się spodziewać wyłącznie postępu – do typografii cyfrowej. Jednak jest wprost przeciwnie, część zgromadzonych przeze mnie kwestionariuszy metodologicznych wypełniona jest na maszynie, ale wiele nadal odręcznie długopisem (CSES1 w 1996 r. w Holandii, na Tajwanie, w Czechach, Norwegii, Ukrainie; CSES2 w 2001 r. Norwegii – rys. S1 [Rybak, 2024b]<sup>1</sup>; WVS2 z 1990 w Chinach – rys. S2 [Rybak, 2024b] – ma postać wyrwanej z żółtego notatnika kartki zapisanej ołówkiem, z tej samej fali dla Chile ma postać krzywego wydruku z drukarki igłowej).

Kolejne przykłady trudnych w odczytaniu raportów to np. raport WVS4 z Iraku pokazujący tabelkę, która nie mieści się na kartce i jej część pozostaje dla odbiorcy nieznana. Kwestionariusz metodologiczny z WVS4 w Ugandzie także wypełniony jest długopisem. Raport LITS1 ma tabele w postaci obrazków tak skompresowanych, że liczby są nieczytelne (zob. rys. S3; Rybak, 2024b). Większość dokumentacji ISSP przedstawiona jest w formie centralnie

<sup>1</sup> Wszystkie ryciny, do których odwołania znajdują się w tekście, dostępne są w Suplemencie zdeponowanym w repozytorium (Rybak, 2024b), dlatego odwołania do nich poprzedzone są literą S.

zaplanowanego raportu dla całej fali, ale już na przykład CSES i WVS mają zazwyczaj osobny kwestionariusz metodologiczny dla każdego kraju, który potrafi być wypełniony, jeśli nie odręcznie, to różną czcionką w różnym kolorze, stosując różne symbole zaznaczenia konkretnej opcji („iksy”, „ptaszki”, „krzyże”, wypełnienie całego pola jednym kolorem, wypełnienie gradientem, zostawienie pola pustego, a zakreślenie na żółto tekstu, który stoi przy polu (tak np. Kirgistan w WVS6, tutaj także adnotacje cyrylicą w anglojęzycznym kwestionariuszu, zob. rys. S4; Rybak, 2024b). Wspomniany ISSP w fali 2020 zrezygnował z jednolitego raportu i także wprowadził pełne kwestionariusze metodologiczne dla krajów – ustawił tam jednak pola tekstowe stałej wielkości, dlatego tekst je wypełniający jest ucinany.

Bardzo wiele dokumentów ma postać skanów: dokumentacja EES99 posiada po bokach ślady otworów po dziurkaczu, wszystkie fale CEEB to krzywo zeskanowane kartki. Nawet kiedy dokument ma formę cyfrowej typografii, to potrafi mieć różny format, od najpopularniejszego PDF, przez doc (np. CNEP10 z Dominikany), arkuszy Excel (lista wszystkich EB albo większości CNEP), po txt (całościowa dokumentacja fal CSES, w odróżnieniu od osobnych kwestionariuszy dla krajów). Na rycinach w Suplemencie (Rybak, 2024b) przedstawiam reprodukcje kilku przykładów.

## 2. Ujednolicenie treści publikacji metodologicznych – przykłady

Istotne dla jakości dokumentacji są jej kompletność i jednolitość, ściśle ze sobą powiązane. Rozdzielenie dokumentacji sondażu na kilka plików znacznie utrudnia ich zinwentaryzowanie, a jednocześnie może wprowadzać niezgodności w opisie pomiędzy dokumentami. Przykładowo, większość fal EB posiada jeden raport w repozytorium GESIS, drugi, różny, w repozytorium Komisji Europejskiej i jeszcze inny dokument w formacie Excel (GESIS Eurobarometer Data Service, 2023), który zawiera daty prac fieldworku czy wielkości próby wszystkich badań ECS, EB i CEEB. Co więcej, informacje potrafią się różnić między tymi dokumentami.

Wspomniany CSES (Comparative Study of Electoral Systems, 2023) jest częściowo omówiony metodologicznie w plikach PDF „design” dla krajów, a częściowo w pliku txt dla całej fali. Część badań CNEP znajduje swoje omówienie we wspomnianym skoroszycie Excel (The Ohio State University, 2024), reszta jest natomiast opisana w osobnych dokumentach na tej samej stronie. W dokumentacji EVS nie zgadzają się liczebności zrealizowanej próby pomiędzy raportami metodologicznymi w repozytorium (European Values Study, 2023) a umieszczonym tam również plikiem Excel, zatytułowanym „EVS 1981–2017 Method Overview”, np. w fali 2008 dla Francji, w fali 1999 np. dla Austrii czy Chorwacji. Repozytorium WVS (World Values Survey, 2023) dla wielu krajów zawiera pliki, które nazywają się raportami technicznymi, kwestionariuszami metodologicznymi, *study designs* itd., których liczba dla każdego sondażu jest zmienna, a zawartość częściowo się pokrywa, a częściowo uzupełnia.

Wreszcie problem jednolitości i kompletności łączy się z problemem trwałej archiwizacji. Strona i jednocześnie repozytorium AB do lipca 2022 r. udostępniały manuale – a więc jedynie deklaracje odnośnie do charakterystyk ba-

dania, fiszki techniczne dla danych krajów, które zawierały tylko podstawowe charakterystyki, oraz (w osobnym pliku) szczegółowe wskaźniki kompletności, odmów, kontaktów i kooperacji, które jednak nie dawały się przeliczyć na liczby naturalne, gdyż były przybliżeniami ułamków. Po przebudowie strony duża część tych dokumentów zniknęła z repozytorium.

Te wyliczenia nie mają za zadanie przedstawienia kompletnego zbioru wszystkich błędów i nieścisłości, ale pokazanie możliwie realistycznie, w jakim stanie znajduje się dokumentacja najważniejszych światowych projektów i jak wygląda praca naukowa z jej wykorzystaniem.

### 3. Pytania przed analizą ilościową

Należy się także zastanowić, czy w opisanej sytuacji „pomiar” jakości dokumentacji jest w ogóle możliwy, a jeśli tak, to jakie problemy stawia przed nami takie przedsięwzięcie. Pojawiają się następujące pytania:

a) Jak traktować różnice pomiędzy projektami, które dopuszczają pewną różnorodność, np. w kwestii techniki sondażu? Czy dokument metodologiczny, który stwierdza, że w każdym badanym kraju przeprowadzono wywiad kwestionariuszowy, jest istotnie lepszym źródłem informacji niż taki, który opisuje stosowane w różnych krajach różne techniki, ale koordynator w jednym z tych krajów zaniedbał podanie odpowiedniej informacji?

Uwagi: O ile w obecnej sytuacji informacje podawane zbiorczo (nie tylko w sytuacji jednej techniki dla programu, ale też zbiorczych tabel, jak w ISSP) ułatwiają kodowanie manualne, o tyle w sytuacji upowszechnienia standardu łatwego do kodowania automatycznego zalecana byłaby osobna tabela dla każdej realizacji krajowej.

b) Czy program sondażowy, który w tzw. manualu opisuje intencję zastosowania jakiegoś rozwiązania należy traktować tak samo, jak program, który w raporcie z pracy terenowej zaświadcza, że to rozwiązanie zastosował?

Uwagi: Literatura zaleca (AAPOR/WAPOR, 2021) prowadzenie dokumentacji na bieżąco, odbiorcę końcowego interesuje zazwyczaj stan finalny i on przede wszystkim powinien być opisany.

c) Uwzględnienie których charakterystyk w dokumentacji jest konieczne? Przykładowo, czy jeśli cały program sondażowy nie stosuje zachęt materialnych dla respondentów i z racji tego nie wspomina o tym dokumentacja, to jest ona gorszej jakości? Czy konieczne jest udokumentowanie liczby pakietów kwestionariuszy przesłanych respondentom pocztą (dostępne np. w części raportów ISSP)? Czy liczba ankietów pracujących w terenie jest niezbędną informacją?

Uwagi: Możliwa jest inspiracja z innej dyscypliny, sugerująca uwzględnianie wszystkich względnie często spotykanych rozwiązań. Przy pisaniu wyników diagnozy medycznej sugeruje się zawrzeć informacje o wszystkich obszarach, które składały się na badanie, nie tylko o tym, co jest anormalne (Raffle i Gray, 2007). Użycie schematów (*templates*) diagnozy, które uzupełnia

się w trakcie badania, poprawia jakość dokumentacji (Mohammad i in., 2021; Rose i in., 2001). Na tej samej zasadzie dokumentacja sondażowa powinna zawierać komplet informacji o przeprowadzonym badaniu, nie tylko o metodach użytych, lecz także tych, których nie użyto (zachęty materialne, listy, substytucje itd.).

d) Jak traktować różnice pomiędzy programami, które opisują falę w jednej publikacji, takimi, które mają osobną publikację dla każdego kraju, a takimi, które mają kilka publikacji zbiorczych, z których każda zawiera tylko część informacji o interesujących nas zmiennych?

Uwagi: Ze względu zarówno na archiwizację, dokumentowanie źródeł, jak i potencjał do stosowania algorytmów automatycznego zbierania danych wskazane jest stosowanie jednej publikacji dla jednej fali.

e) Jak rozróżniać stopień czytelności dokumentów metodologicznych, jeśli wiemy, że część z nich jest wypełniona długopisem, zawiera rozpikselowane tabele albo wstawki w innych językach?

Uwagi: Czytelność pisma odręcznego i grafik w bardzo niskiej rozdzielczości jest subiektywna.

f) Jak traktować różnicę pomiędzy programem, którego dokumentacja nie zawiera jakiegokolwiek informacji, a takim, która zawiera, ale w dwóch sprzecznych wersjach? Albo jest podana, ale w oczywisty sposób błędnie (casus CSES 3 i 4 w Tajlandii, z której dokumentacji wynikało jednoznacznie, że wskaźnik realizacji został nieprawidłowo obliczony)?

Uwagi: Przy braku innych przesłanek w badaniach praktyk stosowałem zapisywanie losowej wartości. Przy czystej analizie dokumentacji należałoby uznać to za brak danych.

g) Po ustaleniu katalogu niezbędnych do wyszczególnienia charakterystyk należy przejść do przypisania każdej z nich odpowiedniej wagi, wszak wskaźnik realizacji jest zazwyczaj bardziej istotną informacją o jakości badania niż średnia długość wypełniania kwestionariusza.

Uwagi: W dotychczasowej analizie (Rybak, 2023) związków pomiędzy charakterystykami badania a reprezentatywnością stosowałem podstawowy zestaw wymaganych zmiennych: technika, operat, dobór wewnętrzny, kwoty/substytucje, *response rate*. Inni odbiorcy mogą definiować swoje wymagania minimalne zupełnie inaczej, producent dokumentacji powinien wziąć pod uwagę różnych użytkowników i zdefiniować ich potrzeby. To samo podejście można zastosować przy analizie.

#### 4. Analizowane zmienne i ich kompletność

Tabela 3 zawiera opis i podsumowanie kompletności zakodowanych zmiennych metodologicznych dla wszystkich sondaży łącznie. Co zaskakujące, najkompletniej udokumentowaną charakterystyką sondażu była technika. Po

niej następne w kolejności były podstawowe informacje o doborze próby, jak operat, zakres wieku badanej populacji czy użycie substytucji i kwot demograficznych. Bardziej szczegółowa, ale nie mniej istotna informacja o sposobie doboru wewnątrzgrupowego była obecna nie w około 90%, jak poprzednie, ale w niecałej 1/4 raportów. Informacja dosyć niewielkiej wartości o samej tylko liczbie wywiadów była obecna także w około 90% raportów, ale już uważane zazwyczaj za najbardziej podstawowy wskaźnik jakości badania współczynniki realizacji zostały zawarte w około 15% dokumentów.

Informacja o często analizowanych przez metodologów kwestiach użycia listów zapowiednich i zachętach materialnych przeznaczonych dla badanych znajduje się w niecałych 15% raportów. W tym wypadku jednak krytyczna analiza może spotkać się z odpowiedzią, że być może autorzy dokumentacji nie czuli się zobowiązani do wspomnienia tych charakterystyk sondażu, które nie były obecne w ich badaniu. Poruszyłem ten temat w sekcji III.3 artykułu. Nie można w ten sposób podsumować jednak wspólnych dla wszystkich badań przeprowadzonych techniką wywiadu kwestionariuszowego problemów liczby i sposobu wynagradzania ankietowanych, które mogą mieć wpływ na jakość zebranych danych, a były opisane w kolejno około 5% i 10% analizowanych raportów.

Nie przedstawiam obliczeń kompletności dla kategorii tematycznych złożonych z kilku zmiennych. Wynika to z faktu, że, przykładowo, dla badacza chcącego ocenić reprezentatywność badań informacja o zastosowaniu doboru kwotowego na poziomie gospodarstwa domowego będzie bardziej istotna niż to, czy operat był imienny czy adresowy. Natomiast dla osoby chcącej skorzystać z uzyskanych danych substancjalnych zakres wieku populacji, z której losowana jest próba, może być ważniejszy niż wszystkie inne charakterystyki procesu doboru. Problemem jest więc nadanie poszczególnym zmiennym w obrębie kategorii wag.

**Tabela 3**

Opis zmiennych uwzględnionych w analizie i podsumowanie ich dostępności w bazie

| Nazwa      | Opis: Czy jest dostępna informacja o... | Liczebności                | Procent sondaży w bazie z dostępnymi informacjami |
|------------|-----------------------------------------|----------------------------|---------------------------------------------------|
| mod_info   | technice                                | Nie: 137<br>Tak: 10<br>356 | 0,987                                             |
| loage_info | minimalnym wieku                        | Nie: 432<br>Tak: 10<br>061 | 0,959                                             |
| upage_info | maksymalnym wieku                       | Nie: 450<br>Tak: 10<br>043 | 0,957                                             |
| fra_info   | operacie                                | Nie: 630<br>Tak: 9 863     | 0,94                                              |

Tab. 3 *cd.*

| Nazwa            | Opis: Czy jest dostępna informacja o... | Liczebności             | Procent sondaży w bazie z dostępnymi informacjami |
|------------------|-----------------------------------------|-------------------------|---------------------------------------------------|
| int_info         | liczbie wywiadów                        | Nie: 997<br>Tak: 9 496  | 0,905                                             |
| subquota_info    | substytucjach/kwotach                   | Nie: 1631<br>Tak: 8 862 | 0,845                                             |
| cont_info        | minimalnej liczbie prób kontaktu        | Nie: 6362<br>Tak: 4 131 | 0,394                                             |
| long_info        | długości wywiadu                        | Nie: 7209<br>Tak: 3 284 | 0,313                                             |
| hhselection_info | doborze wewnątrzgrupowym                | Nie: 7584<br>Tak: 2 909 | 0,277                                             |
| check_info       | backcheckach                            | Nie: 8535<br>Tak: 1 958 | 0,187                                             |
| outcome_info     | umożliwiająca obliczenie RR6 wg AAPOR   | Nie: 8843<br>Tak: 1 650 | 0,157                                             |
| rr_info          | dowolnym RR                             | Nie: 9014<br>Tak: 1 479 | 0,141                                             |
| lett_info        | listach zapowiednich                    | Nie: 9056<br>Tak: 1 437 | 0,137                                             |
| inc_info         | zachętach materialnych                  | Nie: 9165<br>Tak: 1 328 | 0,127                                             |
| paid_info        | sposobie wynagradzania ankierów         | Nie: 9410<br>Tak: 1 083 | 0,103                                             |
| work_info        | liczbie ankierów                        | Nie: 9887<br>Tak: 606   | 0,058                                             |

Źródło: opracowanie własne.

W tabeli 4 przedstawiam średnie wartości kompletności dla zmiennych w poszczególnych obszarach – z wyraźnym zastrzeżeniem, że dane te należy potraktować krytycznie. Sekcje III.1 i III.2 artykułu pokazują, że różnice w sposobach dokumentowania badań w różnych programach sondażowych są często tak wielkie, że ciężko przyłożyć do nich jedną miarę. Dodatkowo różne programy były realizowane w różnych latach. Okres, w którym była tworzona dokumentacja, ma istotny wpływ na jej kształt.

Można zauważyć, że z programów o zasięgu globalnym najkompletniejszą (pod względem założonych kryteriów) dokumentację posiadają CSES, ISSP i (będący jedną falą) ISSP\_SOR. Wśród projektów obejmujących, z niewielkimi wyjątkami, kontynent europejski za najlepiej udokumentowane należy uznać EQLS, ESS i EVS. Dodatkowo ESS jest najlepiej udokumentowanym programem ze wszystkich zbadanych.

Tabela 4

Kompletność dokumentacji metodologicznej według programu

| Programy | mod | loage | upage | fra | int | sub   quota | cont | long | hhselect | check | outcome | rr  | lett | inc | paid | work |
|----------|-----|-------|-------|-----|-----|-------------|------|------|----------|-------|---------|-----|------|-----|------|------|
| AB       | 1   | 1     | 1     | 1   | 0,4 | 0,7         | 0,3  | 1    | 0,3      | 0,7   | 0       | 0,6 | 0,3  | 0,3 | 0    | 0    |
| ABS      | 1   | 0,2   | 0,2   | 0,3 | 0,3 | 0,2         | 0,1  | 0,9  | 0,2      | 1     | 0       | 0,1 | 0    | 0   | 0    | 0,2  |
| ARAB     | 1   | 1     | 1     | 1   | 1   | 0           | 0    | 1    | 0,9      | 1     | 0       | 0,5 | 0    | 0   | 0    | 0    |
| ASES     | 1   | 1     | 1     | 1   | 0   | 0,1         | 0,2  | 0    | 0,9      | 0,1   | 0       | 0   | 0    | 0   | 0    | 0    |
| ASIAB    | 1   | 1     | 1     | 1   | 1   | 0,1         | 0,1  | 0    | 0,7      | 0,1   | 0,1     | 0   | 0    | 0   | 0    | 0    |
| CCEB     | 1   | 1     | 1     | 1   | 1   | 0           | 0    | 0    | 0        | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| CCS      | 1   | 1     | 1     | 1   | 1   | 1           | 0,2  | 0    | 1        | 0     | 0,8     | 1   | 0    | 0   | 0    | 0    |
| CEEB     | 1   | 1     | 1     | 0   | 1   | 1           | 0    | 0    | 0,1      | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| CNEP     | 1   | 0,2   | 0,2   | 0,8 | 1   | 0,1         | 0    | 0    | 0,6      | 0     | 0       | 0,8 | 0    | 0   | 0    | 0,1  |
| CONS     | 1   | 1     | 0     | 0   | 1   | 0           | 0    | 0    | 1        | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| CRRC     | 1   | 0,8   | 0,8   | 1   | 1   | 0,1         | 0,7  | 0,4  | 1        | 0,6   | 0,1     | 0,8 | 0,2  | 0   | 0    | 0,3  |
| CSES     | 1   | 0,9   | 0,9   | 0,9 | 0,9 | 0,7         | 0,8  | 0    | 0,9      | 0,6   | 0,7     | 0,7 | 0,9  | 0,9 | 0    | 0,3  |
| EASS     | 1   | 0,8   | 0,8   | 0,8 | 1   | 0           | 0    | 0    | 0        | 0     | 0,5     | 1   | 0    | 0   | 0    | 0    |
| EB       | 1   | 1     | 1     | 1   | 0,9 | 1           | 0    | 0    | 0,1      | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| ECS      | 1   | 1     | 1     | 0,5 | 1   | 0,3         | 0    | 0    | 0,2      | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| EES      | 1   | 1     | 1     | 0,3 | 0,9 | 0,1         | 0,6  | 0,4  | 0,4      | 0,2   | 0,2     | 0,4 | 0    | 0   | 0    | 0,6  |
| EQLS     | 1   | 1     | 1     | 0,8 | 1   | 1           | 1    | 0,8  | 0,8      | 0,8   | 1       | 0,7 | 1    | 0,3 | 0    | 1    |
| ESS      | 1   | 1     | 1     | 1   | 1   | 1           | 1    | 1    | 1        | 1     | 1       | 1   | 1    | 1   | 1    | 1    |
| EUNB     | 1   | 1     | 1     | 1   | 0   | 0           | 0    | 1    | 1        | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| EVS      | 1   | 1     | 1     | 0,9 | 1   | 0,7         | 0,7  | 1    | 0,6      | 0,7   | 0,7     | 0,7 | 0,5  | 0,5 | 0,7  | 0    |
| GALLUP   | 1   | 1     | 1     | 1   | 1   | 1           | 1    | 1    | 0        | 0     | 0       | 0   | 0    | 0   | 0    | 0    |
| INTUNE   | 1   | 1     | 1     | 1   | 1   | 0,1         | 0,8  | 0    | 0,9      | 0     | 0       | 0,9 | 0    | 0   | 0    | 0    |
| ISSP     | 1   | 0,9   | 0,9   | 0,9 | 0,9 | 0,8         | 0,7  | 0    | 0,9      | 0,8   | 0,8     | 0,4 | 0,7  | 0,6 | 0,6  | 0    |
| ISSP_SOR | 0,9 | 0,9   | 0,9   | 0,9 | 0,9 | 0,9         | 0,6  | 0    | 0,9      | 0,9   | 0,9     | 0   | 0,9  | 0,9 | 0,9  | 0    |
| LAPOP    | 1   | 0,6   | 0,6   | 0,9 | 1   | 0,6         | 0,1  | 0    | 0,6      | 0,3   | 0       | 0,1 | 0    | 0,1 | 0    | 0    |
| LITS     | 1   | 1     | 1     | 1   | 1   | 0,4         | 1    | 0    | 1        | 0     | 0,3     | 0,7 | 0,4  | 0,4 | 0    | 0    |
| MEDW     | 0,7 | 0,7   | 0,3   | 0   | 0,7 | 0           | 0,7  | 0    | 0,7      | 0     | 0,7     | 0,7 | 0,7  | 0,7 | 0    | 0    |
| NBB      | 1   | 1     | 1     | 1   | 1   | 0,2         | 0,4  | 0    | 1        | 0,3   | 0,5     | 0,9 | 0    | 0   | 0    | 0    |
| NEB/NDB  | 0,5 | 0,4   | 0,5   | 0,5 | 0,5 | 0           | 0    | 0    | 0        | 0,4   | 0       | 0   | 0    | 0   | 0    | 0    |
| WVS      | 0,9 | 0,9   | 0,9   | 0,8 | 0,9 | 0,7         | 0,6  | 0,1  | 0,8      | 0,6   | 0,4     | 0,2 | 0,1  | 0   | 0,5  | 0,2  |

Źródło: opracowanie własne.

Tabela 5 zawiera wskaźniki średniej kompletności dokumentacji sondaży z czterech wybranych dekad. Dane te przynoszą wniosek umiarkowanie pozytywny: wskaźniki kompletności rosną w kolejnych okresach. Jest to zgodne z wcześniejszymi ustaleniami badaczy (Jabkowski, 2023; Kołczyńska i Schone, 2018). Największy wzrost dostępnych informacji dotyczy opisu operatu losowania i informacji o substytucjach i kwotach, a więc ogólnie tematu doboru próby. Jest to ważne, gdyż taka informacja umożliwia wstępną ocenę reprezentatywności danych wynikowych dla badanej populacji.

Udostępniam zbiór wszystkich 10 493 sondaży z przypisanymi im zmiennymi zawierającymi informację o kompletności wszystkich raportów w poszczególnych obszarach (Rybak, 2024a). Można go wykorzystać do dowolnych porównań pomiędzy programami bądź okresami, należy jednak pamiętać o wymienionych w niniejszym artykule problemach z tym związanych. Badacze mogą także użyć tej bazy w celu selekcji do własnych analiz sondaży o dokumentacji wysokiej jakości. Należy jednak pamiętać, że kompletność raportów metodologicznych nie musi się wiązać z reprezentatywnością badania (Jabkowski, 2023).

**Tabela 5**

Kompletność dokumentacji metodologicznej według dekady i kontynentu  
(dla kontynentu dane od roku 1990–)

| Zmienne        | Dekada        |               |               |               | Kontynent |      |        |                           |                     |         |
|----------------|---------------|---------------|---------------|---------------|-----------|------|--------|---------------------------|---------------------|---------|
|                | 1980–<br>1989 | 1990–<br>1999 | 2000–<br>2009 | 2010–<br>2019 | Afryka    | Azja | Europa | Amer.<br>Łac.–<br>Karaiby | Ameryka<br>Północna | Oceania |
| mod            | 0,96          | 0,96          | 0,98          | 1             | 0,99      | 0,99 | 0,99   | 0,99                      | 0,95                | 0,99    |
| loage          | 0,93          | 0,9           | 0,96          | 0,98          | 0,99      | 0,93 | 0,98   | 0,84                      | 0,79                | 0,91    |
| upage          | 0,93          | 0,89          | 0,96          | 0,97          | 0,99      | 0,93 | 0,98   | 0,84                      | 0,79                | 0,91    |
| fra            | 0,83          | 0,81          | 0,95          | 0,97          | 0,99      | 0,93 | 0,94   | 0,96                      | 0,88                | 0,98    |
| int            | 0,93          | 0,92          | 0,88          | 0,91          | 0,83      | 0,89 | 0,91   | 0,98                      | 0,97                | 0,93    |
| sub  <br>quota | 0,5           | 0,74          | 0,83          | 0,92          | 0,84      | 0,77 | 0,89   | 0,82                      | 0,83                | 0,87    |
| cont           | 0             | 0,08          | 0,39          | 0,47          | 0,76      | 0,68 | 0,27   | 0,63                      | 0,6                 | 0,8     |
| long           | 0,07          | 0,06          | 0,24          | 0,41          | 0,92      | 0,59 | 0,17   | 0,45                      | 0,28                | 0,32    |
| hhselect       | 0,04          | 0,24          | 0,32          | 0,28          | 0,22      | 0,32 | 0,27   | 0,38                      | 0,59                | 0,68    |
| check          | 0             | 0,15          | 0,22          | 0,18          | 0,25      | 0,25 | 0,16   | 0,24                      | 0,41                | 0,48    |
| outcome        | 0             | 0,15          | 0,22          | 0,14          | 0,04      | 0,18 | 0,17   | 0,13                      | 0,4                 | 0,56    |
| rr             | 0,01          | 0,12          | 0,15          | 0,15          | 0,17      | 0,16 | 0,14   | 0,09                      | 0,29                | 0,28    |
| lett           | 0             | 0,03          | 0,19          | 0,15          | 0,11      | 0,14 | 0,14   | 0,11                      | 0,33                | 0,47    |
| inc            | 0             | 0,03          | 0,16          | 0,14          | 0,11      | 0,13 | 0,13   | 0,14                      | 0,34                | 0,44    |
| paid           | 0             | 0,04          | 0,15          | 0,1           | 0,05      | 0,14 | 0,1    | 0,12                      | 0,18                | 0,13    |
| work           | 0             | 0,03          | 0,06          | 0,07          | 0,01      | 0,04 | 0,07   | 0,03                      | 0,02                | 0,15    |

Źródło: opracowanie własne.

Kołczyńska i Schoene (2018), analizując dokumentację 1721 sondaży z programów porównawczych, doszli do wniosku, że najlepiej udokumentowane badania pochodzą z Europy, a najgorzej z Ameryki Łacińskiej i Bliskiego Wschodu. Dane przedstawione w tabeli 5, ograniczone do lat 1990+ nie potwierdzają tej diagnozy (dane dla całej bazy dają identyczne wnioski). Dokumentacja w sondażach europejskich wykazuje najmniejszą kompletność, biorąc pod uwagę założone kryteria. Najlepiej udokumentowane sondaże były przeprowadzane w Oceanii i Ameryce Północnej. Zły wynik sondaży europejskich może wynikać z tego, że przeprowadza się tam najwięcej sondaży, z których duża część (jak np. bardzo liczny EB) posiada dokumentację niskiej jakości. Należy jednak raz jeszcze przypomnieć o tym, że dla odbiorcy końcowego nie wszystkie informacje mają jednakową wagę.

## 5. Analiza dokumentacji na poziomie programu

Niniejsze rozważania można uzupełnić o ilościowe i jakościowe spojrzenie na cztery projekty: ISSP, AB, CSES i ESS. W dwóch pierwszych widać bowiem wyraźny spadek jakości dokumentacji w ostatnich falach, który nie jest aż tak łatwy do uchwycenia na podstawie samych podsumowań ilościowych, natomiast jest od razu widoczny po spojrzeniu na konkretne raporty metodologiczne. Z kolei dwa kolejne to przykłady o wiele stabilniejszej tendencji w konstruowaniu dokumentacji, choć w swojej formie zupełnie różne.

To, co można zauważyć w przypadku dokumentacji ISSP (tab. 6), to fakt, że od roku 2009 raporty metodologiczne zaczynają uwzględniać *response rate* oraz że ogólna kompletność raportów poprawia się od 2012. Duży spadek kompletności w okolicach 2009 r. wynika z tego, że kraje, które zgodnie z opisem z repozytorium brały udział w fali, nie były uwzględnione w głównym raporcie. Nie znaczy to, że ich raporty nie mogą być ukryte w ramach innej publikacji, dociekliwość koderów jest jednak ograniczona. W dwóch ostatnich falach widzimy jedynie zniknięcie informacji o substytucjach, kwotach i *response rate*. Kiedy ogląda się same raporty, wrażenie przez nie wywierane jest zdecydowanie inne.

Zupełnie nieprzydatny do zautomatyzowanej inwentaryzacji, ale bardzo czytelny i przejrzysty format raportu, który stosowany był konsekwentnie do roku 2017, w kolejnej fali ustępuje miejsca tabelom o podobnym układzie, ale pozbawionym wszelkich udogodnień wizualnych. Dramatycznie zwiększa to potencjał błędów przy ręcznym tworzeniu inwentaryzacji. W dwóch ostatnich falach natomiast otrzymujemy, wspomniany wcześniej, a popularny wśród innych projektów format surowego kwestionariusza metodologicznego wypełnianego przez pracownika krajowej instytucji badawczej. Kodowanie charakterystyk z takiego formatu zajmuje zdecydowanie więcej czasu, a ze względu na brak realnej standaryzacji odpowiedzi na pytania kwestionariuszowe zrzuca obowiązek ich uzgodnienia na badacza. W tym przypadku występuje też wspomniany wcześniej problem techniczny: okna odpowiedzi udzielonych w aplikacji komputerowej przy tworzeniu wydruku do pliku PDF nie rozszerzyły się do rozmiaru tekstu. Tym samym część odpowiedzi jest ucięta. (Poszczególne formy raportów są przedstawione na rysunkach S5, S6 i S7; Rybak, 2024b).

Tabela 6

Kompletność dokumentacji fal według programu (ISSP tylko 2005–2020)

| Programy<br>i ich fale | mod  | loage | upage | tra  | int  | sub <br>quota | cont | long | hselect | check | outcome | rr   | lett | inc  | paid | work |
|------------------------|------|-------|-------|------|------|---------------|------|------|---------|-------|---------|------|------|------|------|------|
| ISSP05                 | 0,97 | 0,97  | 0,97  | 0,97 | 0,97 | 0,97          | 0,94 | 0    | 0,97    | 0,97  | 0,97    | 0    | 0,97 | 0,97 | 0,69 | 0    |
| 06                     | 1    | 0,97  | 0,97  | 1    | 1    | 0,97          | 0,97 | 0    | 1       | 0,97  | 1       | 0    | 0,97 | 0,97 | 0,71 | 0    |
| 07                     | 1    | 0,94  | 0,94  | 1    | 1    | 0,94          | 0,89 | 0    | 1       | 0,94  | 0,97    | 0    | 0,94 | 0,94 | 0,78 | 0    |
| 08                     | 1    | 0,95  | 0,95  | 0,98 | 1    | 0,9           | 0,83 | 0    | 0,95    | 0,93  | 1       | 0    | 0,85 | 0,93 | 0,75 | 0    |
| 09                     | 0,76 | 0,76  | 0,76  | 0,76 | 0,78 | 0,76          | 0,76 | 0    | 0,73    | 0,76  | 0,76    | 0,76 | 0,76 | 0,76 | 0,63 | 0    |
| 10                     | 0,91 | 0,88  | 0,88  | 0,94 | 0,94 | 0,85          | 0,82 | 0    | 0,94    | 0,76  | 0,94    | 0,76 | 0,79 | 0,76 | 0,61 | 0    |
| 11                     | 0,88 | 0,84  | 0,84  | 0,84 | 0,88 | 0,81          | 0,81 | 0    | 0,84    | 0,81  | 0,88    | 0,78 | 0,81 | 0,81 | 0,59 | 0    |
| 12                     | 1    | 1     | 1     | 1    | 1    | 1             | 0,95 | 0    | 1       | 1     | 1       | 1    | 1    | 1    | 0,76 | 0    |
| 13                     | 1    | 1     | 1     | 1    | 1    | 1             | 0,97 | 0    | 1       | 1     | 1       | 1    | 0,97 | 1    | 0,79 | 0    |
| 14                     | 1    | 1     | 1     | 1    | 1    | 0,97          | 1    | 0    | 0,97    | 1     | 0,97    | 1    | 1    | 1    | 0,74 | 0    |
| 15                     | 1    | 1     | 1     | 1    | 1    | 1             | 1    | 0    | 0,97    | 1     | 0,92    | 0,92 | 1    | 1    | 0,78 | 0    |
| 16                     | 1    | 1     | 1     | 1    | 1    | 1             | 1    | 0    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 0    |
| 17                     | 1    | 1     | 1     | 1    | 1    | 1             | 1    | 0    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 0    |
| 18                     | 1    | 1     | 1     | 1    | 1    | 1             | 0,97 | 0    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 0    |
| 19                     | 1    | 1     | 1     | 1    | 1    | 0,07          | 1    | 0    | 1       | 1     | 1       | 0    | 0,93 | 0,97 | 0,93 | 0    |
| 20                     | 1    | 1     | 1     | 1    | 1    | 0,07          | 1    | 0    | 1       | 0,93  | 1       | 0    | 1    | 0,93 | 0,93 | 0    |
| AB1                    | 1    | 1     | 1     | 1    | 0    | 1             | 0    | 1    | 0       | 0     | 0       | 0    | 0    | 0    | 0    | 0    |
| 2                      | 1    | 1     | 1     | 1    | 0    | 1             | 0    | 1    | 0       | 0     | 0       | 0,94 | 0    | 0    | 0    | 0    |
| 3                      | 1    | 1     | 1     | 1    | 0    | 1             | 0    | 1    | 0       | 0     | 0       | 0,94 | 0    | 0    | 0    | 0    |
| 4                      | 1    | 1     | 1     | 1    | 1    | 1             | 0    | 1    | 0       | 0     | 0       | 1    | 0    | 0    | 0    | 0    |

Tab. 6 *cd.*

| Programy<br>i ich fale | mod  | loage | upage | tra  | int  | sub qu-<br>ota | cont | long | hselect | check | outcome | rr   | lett | inc  | paid | work |
|------------------------|------|-------|-------|------|------|----------------|------|------|---------|-------|---------|------|------|------|------|------|
| 5                      | 1    | 1     | 1     | 1    | 0,97 | 1              | 1    | 1    | 1       | 1     | 0       | 1    | 1    | 1    | 0    | 0    |
| 6                      | 1    | 1     | 1     | 1    | 0,97 | 1              | 1    | 1    | 1       | 1     | 0       | 0,97 | 1    | 1    | 0    | 0    |
| 7                      | 1    | 1     | 1     | 1    | 0    | 0              | 0    | 1    | 0       | 1     | 0       | 0    | 0    | 0    | 0    | 0    |
| 8                      | 1    | 1     | 1     | 1    | 0    | 0              | 0    | 1    | 0       | 1     | 0       | 0    | 0    | 0    | 0    | 0    |
| CSES1                  | 0,9  | 0,78  | 0,8   | 0,76 | 0,73 | 0              | 0,71 | 0    | 0,76    | 0,02  | 0,54    | 0,61 | 0,83 | 0,83 | 0    | 0,15 |
| 2                      | 0,98 | 0,9   | 0,9   | 0,98 | 0,88 | 0,29           | 0,83 | 0    | 0,98    | 0,05  | 0,66    | 0,76 | 0,9  | 0,9  | 0    | 0,22 |
| 3                      | 0,96 | 0,94  | 0,94  | 0,94 | 0,88 | 0,96           | 0,86 | 0    | 0,94    | 0,9   | 0,78    | 0,62 | 0,94 | 0,94 | 0    | 0,34 |
| 4                      | 1    | 1     | 1     | 0,98 | 0,85 | 1              | 0,89 | 0    | 0,98    | 0,91  | 0,81    | 0,68 | 1    | 1    | 0    | 0,36 |
| 5                      | 1    | 1     | 1     | 1    | 0,93 | 0,97           | 0,8  | 0    | 1       | 0,93  | 0,9     | 0,7  | 1    | 1    | 0    | 0,3  |
| ESS1                   | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 0,95 | 1    | 1    |
| 2                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 3                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 4                      | 1    | 1     | 1     | 1    | 1    | 1              | 0,97 | 1    | 1       | 0,97  | 1       | 1    | 1    | 1    | 1    | 1    |
| 5                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 6                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 7                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 0,95    | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 8                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 9                      | 1    | 1     | 1     | 1    | 1    | 1              | 1    | 1    | 1       | 1     | 1       | 1    | 1    | 1    | 1    | 1    |
| 10                     | 1    | 1     | 1     | 1    | 1    | 0              | 1    | 1    | 1       | 1     | 0       | 0,9  | 1    | 1    | 1    | 1    |

Źródło: opracowanie własne.

W przypadku AB sytuacja jest jeszcze bardziej skomplikowana, ponieważ w trakcie inwentaryzacji doszło do przebudowy strony internetowej tego projektu, w trakcie której zniknął dostęp do części dokumentacji. Udało się ją jednak wcześniej zarchiwizować, dlatego w tabeli 6 można zauważyć, że fale 5 i 6 były zdecydowanie lepiej udokumentowane metodologicznie niż wcześniejsze i późniejsze.

W falach 5 i 6 stosowano zestandaryzowany format raportu, który choć nie bardzo rozbudowany, pozwala łatwo wydobyć jednolity zestaw informacji o realizacjach w poszczególnych krajach. Pokazuje to rysunek S8 (Rybak, 2024b). Dodatkowe informacje dla tych fal znajdowały się w manualach (jednak dokument taki powstaje przed realizacją jako zestaw zaleceń, wątpliwe jest więc traktowanie go jako opisu realnych działań terenowych) oraz w tabelach umieszczonych w skoroszycie Excel, zawierających obliczone już wskaźniki realizacji.

W przypadku CSES można zauważyć, przedstawiony w tabeli 6, stabilny wzrost dokładności opisu badania w kolejnych falach. Wynika on przede wszystkim z rozbudowy kwestionariuszy metodologicznych (rys. S9; Rybak, 2024b) wypełnianych przez instytucje z krajów uczestniczących w programie. Taki format był też stosowany np. w niemal całym WVS. Ma on pewne przewagi nad innymi formatami: po pierwsze – pozwala na dosyć dokładny opis charakterystyk sondażu przez instytucję krajową, a po drugie – wymaga najmniej pracy od koordynatora programu, który musi jedynie przygotować kwestionariusz. Z drugiej strony brak zaangażowania koordynatora w opracowanie zebranych danych metodologicznych powoduje, że cierpi standaryzacja danych. Mówiąc wprost, osoby wypełniające kwestionariusze, po pierwsze, różnie rozumieją pytania, a po drugie, mają dużą tendencję do opuszczania niektórych z nich. Z tej i innych wspomnianych powyżej (jak graficzne różnice w udzielaniu odpowiedzi) przyczyn raczej nie ma tu możliwości zastosowania żadnych procedur zautomatyzowanego gromadzenia danych metodologicznych.

Ostatnim przykładem jest ESS – program, który bez wahania można uznać za modelowy dla sondażowych badań porównawczych. Dokumentacja metodologiczna jest tu niezwykle szczegółowa, jak pokazuje tabela 6 (zob. też rys. S10; Rybak, 2024b). W zestawieniu ilościowym praktycznie nie widać żadnych istotnych różnic w jakości dokumentacji. Pewne braki w fali 10 wynikają z tego, że w momencie tworzenia inwentaryzacji dostępna była jedynie część danych. Nie oznacza to jednak, że różnice nie występują – widoczne są jednak na poziomie tak szczegółowym, że niemożliwym do uchwycenia przez stosowany schemat inwentaryzacji. W fali 9 zmienia się także sposób kodowania kategorii osób biorących (niebiorących) udział w badaniu. Nowy podział kategorii utrudniał jasne przełożenie ich na kategorie AAPOR, konieczny był więc kontakt z instytucjami ESS. Można tu zauważyć pewną niekonsekwencję, gdyż w przypadku inwentaryzacji innego programu taka sytuacja mogłaby hipotetycznie spowodować zakodowanie tego jako braku danych. Moją motywacją była tu jednak duża ilość bardzo dobrej jakości danych metodologicznych, których dostarcza ESS. Sposób konstrukcji dokumentacji ESS, gdyby tylko zastosowano format bardziej przyjazny niż pliki PDF, byłby potencjalnie najbardziej przystosowanym do łatwej automatyzacji zbierania danych metodologicznych.

#### IV. WNIOSKI

Podstawowym ograniczeniem niniejszej analizy jest fakt, że była ona środkiem do osiągnięcia innego celu. Sposób doboru kodowanych zmiennych przynajmniej częściowo podporządkowałem zamiarowi opisu konkretnych charakterystyk i stworzenia statystycznych modeli badających związki pomiędzy nimi a miarami reprezentatywności. Warto poszerzyć analizowane zmienne, przykładowo, o kwestie związane z odbiorem informacji przez jej użytkowników (a więc takie jak cele badania i związek pomiędzy nimi a jego charakterystykami).

Obecnie gromadzenie danych metodologicznych wymaga ręcznej pracy wielu koderów nieuchronnie podatnych na wprowadzanie błędów do tworzonych przez siebie inwentaryzacji, które to błędy nakładają się na te powstałe w trakcie tworzenia samej dokumentacji. Wprowadzenie takiego jak DDI lub podobnego rozwiązania uczyniłoby proces gromadzenia danych metodologicznych o wiele szybszym, mniej pracochłonnym i bardziej odpornym na błędy. Jeśli jednak DDI sprawia szczegółowe techniczne problemy w wypadku jakiegoś projektu, to już wprowadzenie jednolitego, zestandaryzowanego dla wszystkich fal skoroszytu, w którym z łatwością można rozpoznawać tekst, tabele, sumować dane i używać rozmaitych makr, dałoby takiej dokumentacji istotną przewagę nad plikami PDF, które na takie operacje nie pozwalają albo pozwalają w niewielkim tylko stopniu.

Potencjalnie owocnym podejściem mogłoby być przygotowanie jednego, czytelnego i atrakcyjnego wizualnie standardu tabeli o stałych, zestandaryzowanych nagłówkach, która zastąpiłaby obecne często dosyć dowolnie konstruowane kwestionariusze metodologiczne. Wzorem może tutaj być tabela kategorii osób, które nie uczestniczyły w badaniu, przygotowana przez AAPOR (2023), tabela ta została dosyć powszechnie przyjęta, nawet jeśli poszczególne projekty stosują równoległe inne klasyfikacje. Najlepiej byłoby jednak, gdyby taka tabela była powielana w formie lepiej nadającej się do automatycznej obróbki niż pliki PDF, na przykład w arkuszu kalkulacyjnym. W przypadku upowszechnienia dokumentacji w takiej formie należałoby spodziewać się także stworzenia łatwych w obsłudze dla wszystkich kategorii użytkowników końcowych aplikacji i skryptów open-source do ich gromadzenia i obróbki.

Pozytywne wyniki dotyczące wzrostu kompletności dokumentacji metodologicznej badań sondażowych potwierdzają wcześniejsze ustalenia badaczy. Wzrost ten następuje jednak bardzo powoli. W obecnym stanie rzeczy ilościowe porównywanie dokumentacji pomiędzy programami sondażowymi wciąż wydają się problematyczne ze względu na drastyczne różnice w ich formie i czytelności, jak też liberalne podejście do wypracowanych dotąd standardów. Nie zmienia to faktu, że krytyczny namysł nad stanem dokumentacji badawczej jest istotny dla procesu dalszego jej udoskonalania. Motywowanie zespołów badawczych do poprawy jakości ich własnych raportów, a także koordynowania tych wysiłków z całym środowiskiem byłoby najcenniejszym osiągnięciem tego artykułu.

- Author contributions / Indywidualny wkład autora (CRediT):** Adam Rybak – 100% (Conceptualization / Konceptualizacja; Data Curation / Zarządzanie danymi; Formal analysis / Formalna analiza; Funding acquisition / Pozyskanie funduszy; Investigation / Przeprowadzenie badań; Methodology / Metodologia; Project administration / Administracja projektu; Validation / Walidacja; Visualization / Wizualizacja; Writing – original draft / Pisanie – pierwszy szkic; Writing – review & editing / Pisanie – recenzja i edycja).
- Conflict of interest / Konflikt interesów:** The author declares no conflict of interest. / Autor nie zgłosił konfliktu interesów.
- Funding / Finansowanie:** National Science Centre (Poland), grant Preludium 17 no. 2019/33/N/HS6/00322. / Narodowe Centrum Nauki, Preludium 17, umowa nr 2019/33/N/HS6/00322.
- The use of AI tools / Wykorzystanie narzędzi AI:** The author declares no use of AI tools. / Autor oświadczył, że nie korzystał z narzędzi AI.
- Data availability / Dostępność danych:** A full data set of the completeness of the information contained in the documentation for each national implementation is available in the OSF repository. / Pełna baza danych zawierająca informacje o kompletności zawartych w dokumentacji informacji dla poszczególnych realizacji krajowych dostępna w repozytorium OSF. <https://doi.org/10.17605/OSF.IO/85CK9>
- Acknowledgement / Podziękowania:** I would like to thank Prof. Piotr Jabkowski for his support in conducting my research. / Dziękuję prof. Piotrowi Jabkowskiemu za wsparcie w realizacji moich badań.
- Supplementary material / Dodatkowe materiały:** Rybak, A. (2024a), Baza danych – Kompletność dokumentacji sondaży [Baza Danych], OSF, <https://doi.org/10.17605/OSF.IO/A5WBN>. Wszystkie ryciny, do których odwołania znajdują się w tekście: Rybak, A. (2024b), *Ryciny do artykułu* [plik PDF], OSF, <https://doi.org/10.17605/OSF.IO/9VPDM>. Spis wszystkich repozytoriów i stron internetowych, z których pobrane zostały informacje na temat charakterystyk sondaży: Rybak, A. (2024c), *Spis odnośników do baz danych* [plik PDF], OSF, <https://doi.org/10.17605/OSF.IO/DG2VF>

## References / Bibliografia

- AAPOR/WAPOR. (2021). *AAPOR/WAPOR task force report on quality in comparative surveys*. American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (wyd. 10). AAPOR.
- Castorena, O. (2021). *Methodological note #007: Survey weights in AmericasBarometer data*. Vanderbilt University.
- Comparative Study of Electoral Systems. (2023). *CSES download data and documentation*. <https://cses.org/data-download/download-data-documentation/>
- Dubrow, J. K., Jenkins, J. C., Kołczyńska, M., Oleksiyenko, O., Powalko, P., Słomczyński, K. M., Tomescu-Dubrow, I., Wysmulek, I., i Zieliński, M. W. (red.). (2016). *Democratic values and protest behavior: Harmonization of data from international survey projects*. IFIS PAN Publishers.
- European Values Study. (2023). *European values study: Methodology, data, documentation*. <https://europeanvaluesstudy.eu/methodology-data-documentation/>
- Fonteneau, F. (2011). *Experiences and benefits of implementing the DDI in data management*. UNESCAP Expert Group Meeting.
- GESIS. (2019). *Overview of comparative surveys worldwide*. <https://web.archive.org/web/20191211072538/https://www.gesis.org/angebot/daten-analysieren/weitere-sekundaer-daten/uebersichten/overview-of-comparative-surveys-worldwide>

- GESIS Eurobarometer Data Service. (2023). *Eurobarometer fieldwork dates and institutes, sample sizes, and geographic units by participating countries*. [https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/eurobarometer/eb\\_standard/eb\\_countries-over-time.xlsx](https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/eurobarometer/eb_standard/eb_countries-over-time.xlsx)
- Jabkowski, P. (2019). Zróżnicowanie praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie archiwizacji 1537 surveyów zrealizowanych w krajach europejskich w latach 1981–2017. *Przegląd Socjologiczny*, 68(1), 9–38. <https://doi.org/10.26485/PS/2019/68.1/1>
- Jabkowski, P. (2023). Increase in the quality of methodological documentation of cross-national pan-European multi-wave surveys over the last 40 years – a research note. *International Journal of Social Research Methodology*, 26(6), 817–824. <https://doi.org/10.1080/13645579.2022.2097394>
- Jabkowski, P., Cichocki, P., i Kołczyńska, M. (2021). Multi-Project Assessments of sample quality in cross-national surveys: The role of weights in applying external and internal measures of sample bias. *Journal of Survey Statistics and Methodology*, 11(2), 1–24. <https://doi.org/10.1093/jssam/smab027>
- Jabkowski, P., i Kołczyńska, M. (2020). Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981–2017. *Methodology*, 16(3), 186–207. <https://doi.org/10.5964/meth.2795>
- Jedinger, A., Watteler, O., i Förster, A. (2018). Improving the quality of survey data documentation: A total survey error perspective. *Data*, 3(4), 1–10. <https://doi.org/10.3390/data3040045>
- Kołczyńska, M. (2014). Representation of Southeast European countries in international survey projects: Assessing data quality. *Ask: Research and Methods*, 23(1), 57–78.
- Kołczyńska, M., i Schoene, M. (2018). Survey data harmonization and the quality of data documentation in cross-national surveys. W: *Advances in comparative survey methods* (s. 963–984). <https://doi.org/10.1002/9781118884997.ch44>
- Mohammad, L. M., Vidalis, B. M., Fatemi, L., Coffman, R. R., Qeadan, F., i Kimmell, K. T. (2021). Use of standardized history and physical examination for neurosurgical patients improves clinical documentation and reimbursement. *World Neurosurgery*, 148, e667–e673. <https://doi.org/10.1016/j.wneu.2021.01.056>
- Mohler, P. Ph., Hansen, S. E., Pennell, B.-E., Thomas, W., Wackerow, J., i Hubbard, F. (2010). A survey process quality perspective on documentation. W: J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B.-E. Pennell, i T. W. Smith (red.), *Survey methods in multinational, multiregional, and multicultural contexts* (s. 299–314). John Wiley & Sons.
- Mohler, P. Ph., Pennell, B.-E., i Hubbard, F. (2008). Survey documentation: Toward professional knowledge management in sample surveys. W: E. D. de Leeuw, J. J. Hox, i D. A. Dillman (red.), *International handbook of survey methodology* (s. 403–420). L. Erlbaum Associates.
- Ohio State University. (2024). *Comparative national elections project: Publicly available surveys*. <https://u.osu.edu/cnep/surveys/surveys-through-2012/>
- Raffle, A. E., i Gray, J. A. M. (2007). *Screening: Evidence and practice*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199214495.001.0001>
- Rasmussen, K. B. (2014). Social science metadata and the foundations of the DDI. *IASSIST Quarterly*, 37(1–4), 28. <https://doi.org/10.29173/iq499>
- Rose, E. A., Deshikachar, A. M., Schwartz, K. L., i Severson, R. K. (2001). Use of a template to improve documentation and coding. *Family Medicine*, 33(7), 516–521.
- Rybak, A. (2023). Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996–2018 and the European social survey rounds 1 to 9. *PLOS ONE*, 18(3). <https://doi.org/10.1371/journal.pone.0283092>
- Rybak, A. (2024a). Baza danych – Kompletność dokumentacji sondaży [Baza Danych]. OSF. <https://doi.org/10.17605/OSF.IO/A5WBN>
- Rybak, A. (2024b). Ryciny do artykułu [plik PDF]. OSF. <https://doi.org/10.17605/OSF.IO/9VPDM>
- Rybak, A. (2024c). Spis odnośników do baz danych [plik PDF]. OSF. <https://doi.org/10.17605/OSF.IO/DG2VF>

- Slomczynski, K. M., Powalko, P., i Krauze, T. (2017). Non-unique records in international survey projects: The need for extending data quality control. *Survey Research Methods*, 11(1), 1–16. <https://doi.org/10.18148/srm/2017.v11i1.6557>
- Slomczynski, K. M., i Tomescu-Dubrow, I. (2018). Basic principles of survey data recycling. W: *Advances in Comparative Survey Methods* (s. 937–962). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118884997.ch43>
- Slomczynski, K. M., Tomescu-Dubrow, I., i Wismulek, I. (2022). Survey data quality in analyzing harmonized indicators of protest behavior: A survey data recycling approach. *American Behavioral Scientist*, 66(4), 412–433. <https://doi.org/10.1177/00027642211021623>
- Tepper, S. J., Spruill, M. K., Premachandra, B., i Lewis, N. A. (2022). Surveys as conversations between makers and takers: A conversational framework for assessing and responding to community needs. *Analyses of Social Issues and Public Policy*, 22(3), 857–875. <https://doi.org/10.1111/asap.12326>
- Van Wettere, N. (2022, 12 stycznia). *How to implement the metadata standard „Data Documentation Initiative (DDI)”?* <https://doi.org/10.5281/zenodo.5841664>
- Vardigan, M. (2014). DDI Timeline. *IASSIST Quarterly*, 37(1–4), 51. <https://doi.org/10.29173/iq502>
- Vardigan, M., Heus, P., i Thomas, W. (2008). Data documentation initiative: Toward a standard for the social sciences. *International Journal of Digital Curation*, 3(1), 107–113. <https://doi.org/10.2218/ijdc.v3i1.45>
- World Values Survey. (2023). *World values survey: Documentation for download*. <https://www.worldvaluessurvey.org/WVSContents.jsp>

## Suplement do artykułu – Ryciny

Rybak, Adam. 2025. „Analiza jakości dokumentacji metodologicznej sondaży: Studium przypadku inwentaryzacji 30 programów badań porównawczych z całego świata”. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 87(1):287–309. doi: 10.14746/rpeis.2025.87.1.15.

1.

c) Refusal Conversion

Was an effort made to persuade respondents who were reluctant to be interviewed? No ☐ Yes ☐ (Please Describe): THE INTERVIEWERS ARE INSTRUCTED TO LISTEN TO THE RESPONDENTS CONCERNS AND REASONS FOR NOT PARTICIPATING, THEN DELIVER AN ADEQUATE ANSWER AS QUICKLY AS POSSIBLE

Were respondents who were reluctant to be interviewed sent a letter persuading them to take part? No ☐ Yes ☒ (Please Describe/ Include with Deposit): A SPECIAL LETTER EXPLAINING THE IMPORTANCE OF THE STUDY INCLUDING THE NAME OF THE INTERVIEWER WHO TOOK THE CB-CALL

Rysunek S1 Fragment raportu metodologicznego z CSES2 Norwegia („CSES Module 2: Norway 2001 Design Report”, 2004)

2.

Sampling China

Method

stratified multiple-stage random sampling

statu acc: region levels a ~~area~~ level

37% 1. Ea level region - coastal prov + 11 provinces + cities

Rysunek S2 Fragment raportu metodologicznego z WVS2 Chiny (Mingming, 1990)

3.

Table 9: No interviews, all cases by PSU

| Country    | Total | Max | No | 1  | 2  | 3  | 4  | 5  | 6 | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|------------|-------|-----|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Albania    | 226   | 15  | 6  | 13 | 8  | 3  | 4  | 4  | 6 | 11 | 6  | 14 | 6  | 6  | 3  | 7  | 4  | 1  | 10 | 5  | 14 | 4  |
| Armenia    | 228   | 12  | 6  | 10 | 6  | 3  | 8  | 5  | 2 | 10 | 5  | 11 | 4  | 11 | 2  | 12 | 4  | 1  | 2  | 10 | 6  | 1  |
| Azerbaijan | 374   | 36  | 11 | 30 | 36 | 26 | 34 | 11 | 3 | 21 | 14 | 18 | 9  | 0  | 13 | 26 | 5  | 3  | 4  | 12 | 7  | 1  |
| Belarus    | 296   | 16  | 3  | 11 | 8  | 4  | 6  | 0  | 2 | 5  | 9  | 4  | 6  | 3  | 12 | 8  | 1  | 0  | 3  | 0  | 8  | 11 |
| Bosnia     | 378   | 28  | 13 | 18 | 16 | 13 | 8  | 8  | 8 | 8  | 0  | 0  | 5  | 6  | 2  | 8  | 11 | 15 | 10 | 2  | 11 | 1  |
| Bulgaria   | 533   | 43  | 7  | 12 | 0  | 0  | 14 | 2  | 8 | 37 | 36 | 7  | 18 | 10 | 4  | 15 | 4  | 13 | 14 | 17 | 9  | 10 |

Rysunek S3 Fragment raportu metodologicznego z LITS I (Synovate, 2006)

4.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>IF STRATIFICATION FACTORS USED</b></p> <p>23. What stratification factors were used, and at what stage(s) of selection?<br/>PLEASE WRITE IN:</p> <p>Были определены 83 первичные единицы выборки, в каждой опрашивались в среднем 24 респондента. Каждая первичная единица выборки отражает избирательный участок, поскольку в каждом избирательном участке проживает равное количество человек. Избирательные участки для опроса выбирались случайным образом, с учетом количества респондентов в отдельных населенных пунктах и областях, репрезентативно отражающих</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Rysunek S4 Fragment raportu metodologicznego z WVS6 Kirgistan (Moldogaziev i Karabchuk, 2011)

5.

|                                          | BG | CDN | CH | CY | CZ |
|------------------------------------------|----|-----|----|----|----|
| Were a minimum number of calls required? |    |     |    |    |    |
| Yes                                      | X  |     | X  |    | X  |
| No                                       |    | X   |    | X  |    |
| Not answered                             |    |     |    |    |    |
| Not applicable                           |    |     |    |    |    |

Rysunek S5 ISSP od końca lat 90tych do 2017 – przykład formatu dokumentacji (Harkness i in., 1999)

6.

|                                                          | NO | PH | RU | SK |
|----------------------------------------------------------|----|----|----|----|
| Did you use any variables for stratification?            |    |    |    |    |
| Yes                                                      | X  |    |    |    |
| No                                                       |    | X  | X  | X  |
| What selection method was used to identify a respondent? |    |    |    |    |
| No selection needed                                      | X  |    |    |    |
| Kish grid                                                |    | X  |    |    |
| Last or next birthday                                    |    |    | X  | X  |
| Other                                                    |    |    |    |    |
| How many stages does your sampling                       |    |    |    |    |

*Rysunek S6 ISSP 2018-19 – przykład formatu dokumentacji (Sapin i in., 2020)*

7.

S10. At the different stages, the sampling method and the sampled units were:

S10a. at stage 1:

In the first stage, locality strata have been created and the localities were chosen from these strata with the help of random sampling.

All the regions (7) are involved into the sample. About 70 localities are selected.

S10b. at stage 2 (if any):

After creating the locality sample, the streets are defined with the help of a random technique using a database containing all the street names in the selected localities. The selected streets serve as starting points for the interviewers.

*Rysunek S7 ISSP 2020-2021 – przykład formatu dokumentacji (ISSP Research Group, 2020)*

## 8.

|                              |                                                                                                                                                                                           |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dates of Fieldwork:</b>   | 22 May to 4 June 2013                                                                                                                                                                     |
| <b>Sample size:</b>          | 1200                                                                                                                                                                                      |
| <b>Sampling frame:</b>       | 2007 Swaziland Population and Housing Census, with 2013 population projections, provided by the Central Statistics Office                                                                 |
| <b>Sample universe:</b>      | Citizens of Swaziland who are 18 years and older                                                                                                                                          |
| <b>Sample design:</b>        | Nationally representative, random, clustered, stratified, multi-stage area probability sample                                                                                             |
| <b>Stratification:</b>       | Regions and urban-rural location                                                                                                                                                          |
| <b>Stages:</b>               | Primary Sampling Units (PSUs) (from strata), start points, households, respondents                                                                                                        |
| <b>PSU selection:</b>        | Probability proportionate to population size (PPPS)                                                                                                                                       |
| <b>Cluster size:</b>         | 8 households per PSU                                                                                                                                                                      |
| <b>Household selection:</b>  | Randomly selected start points, followed by walk pattern using 5/10 interval                                                                                                              |
| <b>Respondent selection:</b> | Gender quota filled by alternating interviews between men and women; respondents of appropriate gender listed, after which household member draws a numbered card to select an individual |
| <b>Weighting:</b>            | Weighted to account for individual selection probabilities                                                                                                                                |
| <b>Margin of error:</b>      | +/- 3% at 95% confidence level                                                                                                                                                            |
| <b>Fieldwork by:</b>         | ActivQuest                                                                                                                                                                                |
| <b>Survey Languages:</b>     | English & siSwati                                                                                                                                                                         |
| <b>Main researchers:</b>     | Edward Okoth (NI), Sipho Kunene (Co-NI), Vimbai Tsododo (Co-NI)                                                                                                                           |
| <b>Outcome rates:</b>        | Contact rate: 84.0%<br>Cooperation rate: 70.5%<br>Refusal rate: 2.7%<br>Response rate: 59.3%                                                                                              |
| <b>EA Substitution Rate:</b> | 6.7% (10 PSUs)                                                                                                                                                                            |

*Rysunek S8 Fragment raportu z AB5 Eswatini (Park, 2015)*

## 9.

### Refusal Conversion

29a. Were efforts made to persuade respondents who were reluctant to be interviewed?

- ☒ Yes  
☐ No

Please describe: **Respondents who were reluctant were contacted again with attempt to persuade the respondent to be interviewed. The work to re-contact and persuade respondents are done by the most experienced interviewers. They use specially made arguments about the importance to participate, in purpose to persuade and motivate the respondent. If they refuse they will be offered to answer a shorter questionnaire.**

29b. Were respondents who were reluctant to be interviewed sent a letter persuading them to take part?

- ☐ Yes  
☒ No  
(If yes, please provide a copy of the letter or letters.)

If yes, please describe:

*Rysunek S9 Fragment raportu z CSES3 Szwecja (The Comparative Study Of Electoral Systems, 2010)*

## 10.

### 43 Fieldwork procedures

#### 43.1 Interviewer selection

|                                       |    |
|---------------------------------------|----|
| Total number of interviewers:         | 79 |
| Number of experienced interviewers:   | 75 |
| Number of inexperienced interviewers: | 4  |

#### 43.2 Briefing of interviewers

|                                                                        |         |
|------------------------------------------------------------------------|---------|
| How many of the interviewers received ESS specific personal briefing?: | 79      |
| Total length of ESS specific personal briefing(s) per interviewer:     | ½-1 day |
| Written ESS specific instructions:                                     | Yes     |
| Training in refusal conversion:                                        | Yes     |
| Training on how to fill in contact forms:                              | Yes     |
| Training on how to fill in observable and dwelling information:        | Yes     |
| <b>Materials on observable and dwelling information:</b>               |         |
| Photos                                                                 | No      |
| Audio-tape recording                                                   | No      |
| Other                                                                  | No      |

*Rysunek S10 Fragment raportu z ESS5 Chorwacja (ESS ERIC, 2018)*

### Odwołania:

CSES Module 2: Norway 2001 Design Report. (2004). *The Comparative Study Of Electoral Systems*. <https://doi.org/10.7804/CSES.MODULE2.2015-12-15>

ESS ERIC. (2018). ESS5—2010 Documentation Report Edition 4.2. *Sikt - Norwegian Agency for Shared Services in Education and Research*. <https://doi.org/10.21338/NSD-ESS5-2010>

Harkness, J. A., Langfeldt, B., i Scholz, E. (1999). Study Monitoring Report 1997. *International Social Survey Programme: Work Orientations II*. <https://doi.org/10.4232/1.3090>

ISSP Research Group. (2020). Hungary Technical Report 2020. *International Social Survey Programme: Environment IV*. <https://doi.org/10.4232/1.14153>

Mingming, S. (1990). *World Values Survey Wave 2 China Sampling & Methodology*. China Statistical Information & Consultancy Service Centre. [www.worldvaluessurvey.org/WVSDocumentationWV2.jsp](http://www.worldvaluessurvey.org/WVSDocumentationWV2.jsp)

Moldogaziev, T., i Karabchuk, T. (2011). *World Values Survey Wave 6 Kyrgyzstan Methodology Questionnaire*. Центр Изучения Общественного мнения и Прогнозирования «Эл-Пикир». [www.worldvaluessurvey.org/WVSDocumentationWV6.jsp](http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp)

Park, C. (2015). *Data codebook for a Round 5 Afrobarometer survey in 34 African countries*. Afrobarometer, Michigan State University, University of Cape Town, Center for Democratic Development.

Sapin, M., Joye, D., Nisple, K., Reveilhac, M., i Steinmetz, S. (2020). Study Monitoring Report 2018. *International Social Survey Programme ISSP 2018 - Religion IV*.  
<https://doi.org/10.4232/1.13629>

Synovate. (2006). *Life in Transition Survey (LITS) 2006: A brief report on observations, experiences and methodology from the survey*. The European Bank for Reconstruction and Development.

The Comparative Study Of Electoral Systems. (2010). CSES Module 3: Sweden 2006 Design Report. *The Comparative Study Of Electoral Systems*.  
<https://doi.org/10.7804/CSES.MODULE3.2015-12-15>

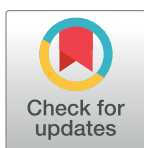
RESEARCH ARTICLE

# Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996–2018 and the European social survey rounds 1 to 9

Adam Rybak \*

Faculty of Sociology, Adam Mickiewicz University, Poznań, Poland

\* [adam.rybak@amu.edu.pl](mailto:adam.rybak@amu.edu.pl)



## OPEN ACCESS

**Citation:** Rybak A (2023) Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996–2018 and the European social survey rounds 1 to 9. PLoS ONE 18(3): e0283092. <https://doi.org/10.1371/journal.pone.0283092>

**Editor:** Olga Scrivner, Rose-Hulman Institute of Technology / Indiana University, UNITED STATES

**Received:** April 14, 2022

**Accepted:** March 1, 2023

**Published:** March 16, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0283092>

**Copyright:** © 2023 Adam Rybak. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are available at OSF repository: Dataset [10.17605/OSF.IO/VTWN7](https://doi.org/10.17605/OSF.IO/VTWN7) R script [10.17605/OSF.IO/QM5RU](https://doi.org/10.17605/OSF.IO/QM5RU) Data

## Abstract

The constant increase in survey nonresponse and fieldwork costs are the reality of survey research. Together with other unpredictable events occurring in the world today, this increase poses a challenge: the necessity to accelerate a switch from face-to-face data collection to different modes, that have usually been considered to result in lower response rates. However, recent research has established that the simple response rate is a feeble measure of study quality. Therefore, this article aims to analyze the effect of survey characteristics, especially the survey mode, on the nonresponse bias. The bias measure used is the internal criteria first proposed by Sodeur and first applied by Kohler. The analysis is based on the survey documentation and results from the International Social Survey Programme waves 1996–2018 and the European Social Survey rounds 1 to 9. Random-effects three-level meta-regression models, based on data from countries from each inhabited continent, were created in order to estimate the impact of the survey mode or modes, sampling design, fieldwork experience, year of data collection, and response rate on the nonresponse bias indicator. Several ways of nesting observations within clusters were also proposed. The results suggest that using mail and some types of mixed-mode surveys were connected to lower nonresponse bias than using face-to-face mode surveys.

## Introduction

The constant increase [1, 2] in survey nonresponses and, thus, possible nonresponse errors create difficulties with the implementation of the survey method. While in some countries this issue is addressed by increasing so-called “fieldwork effort” [3 p101–113], resulting in differing outcomes, in many others the skyrocketing costs and problems associated with gaining access to specific populations result in switching from “traditional” face-to-face surveys to other modes. The consequences of the COVID-19 pandemic amplified this changing approach. In many countries that still deemed standard face-to-face collection to be optimal, the law

citations [10.17605/OSF.IO/D6PZG](https://doi.org/10.17605/OSF.IO/D6PZG) SPSS syntaxes for NBI calculation [10.17605/OSF.IO/JKE5H](https://doi.org/10.17605/OSF.IO/JKE5H)  
Software citations [10.17605/OSF.IO/RX7Q5](https://doi.org/10.17605/OSF.IO/RX7Q5).

**Funding:** The author gratefully acknowledges funding from the National Science Centre, Poland (AR: PRELUDIUM-17 grant no. 2019/33/N/HS6/00322 <https://ncn.gov.pl/en>). The funder had no role in study design, data collection, analysis, decision to publish, or manuscript preparation.

**Competing interests:** The author has declared that no competing interests exist.

necessitated a change, which raises the question: Should we treat this situation as a risk or an opportunity? To reduce uncertainty in this subject, the following study aims at estimating the impact of using different modes of data collection with control of the variables, such as the sampling design, survey program, institutional experience, and response rate (RR), on survey representativeness. I treat these variables as controls because, for example, sampling design is usually predetermined by accessible sampling frames, while the survey mode used is (with limitations) a matter of diligent choice. As the data are hierarchical, differences between countries and program waves are also accounted for.

There is already quite a substantial body of survey methodology literature concerning relations between the survey mode and sampling on the one hand and nonresponse on the other. Unfortunately, there are far fewer studies that consider nonresponse bias. Additionally, most methodological analyses are based on more fragmented sources, such as single surveys or a few journal article meta-analyses. Exceptions [e.g., 2] usually lack the implementation of net sample quality indicators other than nonresponse rates. In most cases, they also stick only to European or US data, while there is a substantial tradition of doing non-face-to-face surveys in countries like Canada, Australia, and New Zealand. For this reason, I have based this study on the reports and data from both the European Social Survey (ESS) rounds 1 to 9 and the International Social Survey Programme (ISSP) waves 1996–2018. The reason is that ISSP surveys present a very high variability of survey modes compared to other international programs. Yet the limitation of this study is the usage of observational data; modes are not assigned randomly to surveys. To mitigate this issue a little, fully face-to-face ESS surveys have been added to the database to provide parallel surveys conducted in the same countries in the same timeframe, at least for a portion of cases. While some studies [4, 5] suggest that the quality of ISSP documentation is inferior to that of the ESS, it is still one of the better documented international survey programs today.

## Nonresponse bias

Unit nonresponse is a major component of total survey error, which itself is currently the default framework for conducting survey methodology research [6–8]. The other types of error are not a subject of this study, nor will I delve into other issues connected to cross-national comparisons. Measurement error or problems with survey research costs will only be mentioned. As for nonresponse itself, if not connected to systematic error, it will pose a threat only to the precision of the estimates. On the other hand, “nonresponse bias results from the contrast between those responding and not responding to the surveys” [9]. Nonresponse bias threatens the representativeness of a survey. For a formal presentation: [e.g., 3 p29–32].

There are many possible ways of estimating the size of nonresponse bias. The most common in survey practice—but highly criticized [10–13] for establishing only a higher bound of the possible error and having a limited relation with other measures—is the RR. As for the other most feasible, possible measures of survey representativeness, they can be divided into two categories. The first category requires individual-level auxiliary data about a whole drawn sample. The most popular members of this category are so-called R-indicators [14–16]. Unfortunately, such required auxiliary data are not available for this study. The second category, which does not require such data, can be divided into the two most popular approaches. The first one, sometimes called “external criteria of representativeness,” requires access to reliable population statistics [9, 17–19]; they can be obtained both from official statistics and other survey data. This approach would not be optimal for an assessment concerning survey data covering such a wide geographical area as the ISSP. Official data or other survey data can and do strongly differ in quality between countries (not to speak of continents). Therefore,

differentiation between the effects of nonresponse bias and bias coming from faulty population statistics would be problematic.

As for the second approach, Sodeur [20] proposed a way of estimating the quality of a net sample based on the so-called “internal criteria of representativeness.” It was then developed by Kohler [21] and used in several subsequent studies [e.g., 17, 22–24]. This method is based on finding the proportion in the sample, which is known *ex definitione*. Souder’s example is heterosexual couples living together in two-person households; the gender balance in this situation should be 50–50. “Internal criteria assessments probe the integrity of the sample by exploiting this a priori knowledge and point to potential irregularities in the survey process” [25]. Kohler [21] argues that this measure would not be affected by coverage and observational errors nor by item-nonresponse (unfortunately only for face-to-face surveys, where gender is in most cases assessed by the interviewer), making it a good candidate for the role of assessing survey representativeness. Ideally, it would be possible to control for same-sex couples, yet this is rarely the case (surely not for the ISSP data, crucial for this study). However, the working assumption could be made that there are no more or no fewer gay couples than lesbian couples in the population. For ESS rounds 1 to 8 and European Quality of Life Survey rounds 1 to 4, where controlling for same-sex couples is possible, the correlation between indicators using strict (with control) and lenient (as in this paper) approaches to assess nonresponse bias was 0.96 and 0.93, respectively [25]. The impossibility of using a strict approach is an important limitation of this study.

Additionally, counting only official marriages would make a big difference between countries with different legal landscapes regarding this matter, so both formal and informal partnerships should be included. In the following parts of this paper, I will refer to Sodeur’s internal criteria of nonresponse bias assessment as a nonresponse bias indicator (NBI).

For an extensive discussion on the usability of the different nonresponse bias measures and indicators in similar research, see the work of Jabkowski, Cichocki, and Kołczyńska [25]. Additionally, it must be stressed that, in reality, nonresponse bias is a property of a specific variable. It can have different sizes for different survey questions. For this reason, the measure used here must be treated only as a proxy. There are some additional possibilities for assessing the quality of the survey in terms of the nonresponse bias. One could use the gross sample information if it consists of individuals with known characteristics, but this is possible only in the case of individual samples [26]. Additionally, nonrespondents could be targeted by very short follow-up questionnaires, as in the ESS, but such follow-up may also be vulnerable to nonresponse. A similar vulnerability is shared by any method that compare early and late respondents, or individuals who needed the conversion procedures with the rest of the respondents [19]. Such assessment methods are only possible for the secondary data analysis, when required information is easily accessible to external researchers, which is rarely the case.

## Survey modes

The face-to-face mode has usually, especially in the past, been considered the best mode for achieving high RRs [27, 28] or considered the “gold standard,” which is reflected in the mode most commonly used in the most important comparative survey programs. For example, the Eurobarometer, ESS, and European Quality of Life Survey waves conducted in pre-COVID-19 times (i.e., before 2020) were based solely on the face-to-face mode. Moreover, in the case of the World Values Survey and the European Values Study, this mode was clearly dominant. The rationale for considering this mode the best may be the possibility it provides for the interviewer to use many motivational and conversion procedures and to gather paradata. It can also be driven by the mode preference of potential respondents. Due to very limited research,

preference for face-to-face surveys among respondents may be the most widespread, at least in some countries [29, 30]; but apparently, declared preference also depends on the mode of the question administration if the preference is studied through declarative, not behavioral, data [first observed by 31]. I will not discuss the possible causes of these preferences, whether are they psychological [32], cognitive [33], or of another type. The widespread usage of this mode in scientific surveys may also come from the lack of the sampling frame necessary to use any other mode, but this is mostly the case in low- and middle-income countries. On the other hand, the main downside of this survey mode is its cost—usually the highest of all other single-mode surveys [34]. Moreover, the high RRs it generates may stem, if proper supervision and backchecking is not implemented, from fraudulent activity on the part of the interviewer.

Mail surveys have commonly been considered as generating lower RR than face-to-face surveys [28, 35–37]. On the other hand, they are treated as a way to overcome the social desirability bias present mainly in the face-to-face context [38, 39]; however, this type of bias is not a focus of this study. The downside of the mail mode is that it becomes more problematic if no individual frame is available. Random within-household selection made by household members may be of lower quality than one made by a skilled interviewer. However, the RR of mail surveys may also be higher than those of current web surveys [40, 41]. Nevertheless, there are multiple examples of well-designed mail surveys that achieved similar or even higher RRs than one would expect from face-to-face surveys in a similar context [42 p351–352].

While mixed-mode surveys in a very broad sense were part of survey practice from the early years (when self-administered questionnaires started to be used for asking sensitive questions in face-to-face surveys) and started to be more widely discussed in the 1960s, they became very popular only in the 21st century [43]. The idea of mixing different modes of data collection is motivated mainly by the desire to reduce cost, social desirability bias, or nonresponse (but not all at once). It may also involve modes of contact and a follow-up phase. Mixed-mode surveys could possibly be cheaper than face-to-face surveys while achieving higher RR if carefully designed [44]. I will not discuss the complex differences between various types of mixed-mode [42 p398–449, 45–47] implementation: sequential and parallel design, mode choice [48], mode preference [29, 49–51], adaptive or responsive design [52]. It is because the mixed-mode surveys I will analyze in the forthcoming parts of this paper are heterogeneous in implementation specifics, as are single-mode surveys. In a study consisting of 795 cases, it would be impossible to address these differences substantially.

As stated previously, a body of literature concerning relations between survey mode and nonresponse bias exists but is not very large; nevertheless, any endeavor aimed at finding clear conclusions in it may be fruitless. This paper is a systematic review, not a narrative one. Therefore, the list of findings is more of an illustration. I will start with the single-mode comparisons. De Leeuw [28] in her experiment found that only face-to-face surveys (in comparison to mail and computer-assisted telephone interviewing, CATI) generated significant differences between gender proportion in the sample and the register data. Laaksonen and Heiskanen [53] found in face-to-face, phone, and web parallel surveys in Finland that the face-to face mode was the least representative as measured by R-indicators. Klausch, Hox, and Schouten [54] found in their mode experiment comparing face-to-face, CATI, web, and mail modes that face-to-face and web modes achieved the best representativeness measured by R-indicators. There are also some comparisons between single-mode and mixed-mode surveys. Fowler et al. [55] found results from a mail and CATI mixed-mode survey more representative than from a single-mode mail survey. Dillman et al. [56] showed that using a mixed-mode survey design can potentially raise RR compared to a single-mode survey, but only non-face-to-face modes were included in this study; the authors did not find any significant differences in representativeness between single-mode and mixed-mode approaches. Luiten and Schouten [57]

compared a single-mode CATI survey to a mixed-mode web, mail, and CATI survey; while both had similar RR and cost, the mixed-mode approach was more representative as measured by R-indicators. Klausch, Hox, and Schouten [54] compared single-mode CATI, mail, and web surveys with their respective mixed-mode approaches with added face-to-face follow-up; in most cases, the bias measured by R-indicators for mixed-mode surveys was lower. Bianchi et al. [58] found that switching the mode in the panel survey from a face-to-face to a mixed-mode approach resulted in similar RR and sample composition while lowering the cost. Villard and Fitzgerald [34], after analyzing a series of mode experiments conducted under the auspices of the ESS, concluded that “mixed-mode data collection designs are unlikely to increase RR or reduce nonresponse bias in survey estimates compared to data collected face to face. In fact, overall, they suggest that RR would decrease and nonresponse bias would probably increase.” Cornesse and Bosnjak [59], after conducting a meta-analysis of 96 surveys, found that using mixed-mode surveys resulted in nonresponse bias reduction measured by R-indicators compared to single-mode surveys. Sztabiński [60], in an ESS mode experiment in Poland, found that using a mail, web, and face-to-face mixed-mode survey resulted in lower RR, higher costs, and a similar sample composition as a parallel single-mode survey. Wolf et al. [44], after an analysis of the 2017 round of the European Values Study, found that face-to-face surveys were more representative than some mixed-mode surveys but that the latter were much cheaper and had higher RR.

The preceding section offered an extensive overview of the current state of knowledge about the relationship between survey modes and various propositions of measures of survey representativeness. While an important body of research exists in this field, the results can be difficult to reconcile due to the limitations of traditional forms of narrative review or even such approaches as simple vote counting of significant results. These limitations can lead readers to gain a false impression of conflicting results in many fields [61 p20]. To address these issues, this study endeavors to shed more light on this subject by harnessing the potential of meta-analysis, which is a “statistically defensible approach to synthesizing empirical findings” [62 p8]. Therefore, this study aims to answer the following research questions: What is the relation between survey mode and survey representativeness, and how is this relation affected by other survey characteristics and broader context?

## Data and methods

The data for this study come from methodological reports and data from ISSP waves 1996–2018 and ESS rounds 1 to 9. All ISSP documentation and datasets are stored and publicly available in the GESIS–Leibniz Institute for the Social Sciences repository [63]. The ISSP is a cross-national survey research collaboration between institutions from 57 countries from all inhabited continents, beginning in 1984 [64]. While being decentralized and differing highly in the survey modes used and fieldwork procedures, it maintains high standards of unified reporting and quality control [65]. ESS documentation and datasets are available on the official ESS site [66], while the Norwegian Centre for Research Data provides hosting. The ESS is one of the most sophisticated methodologically among big cross-national comparative social survey programs [67]. It started in 2001, inspired by the growing discontent of the academic community for previous comparative surveys [68]. All nine pre-COVID-19 rounds of the ESS were guided by a set of ambitious principles: “non-duplication, longitudinality, cross-national equivalence, potential for multilevel analysis, ongoing refinement of methodology, simultaneous flexibility and stability, easy and inexpensive data access and clear definitions of both population and sampling” [69]. Moreover, as stated previously, all the pre-COVID-19 ESS surveys were done in a face-to-face mode, either as PAPI or CAPI. Data references can be found in the separate file [70].

Reports and datasets were used to code each survey's characteristics as part of a funded research project. For this paper, the most important information was about the program, year and country, sampling design (including within-household selection), use of quotas or substitution of nonresponders, mode (or modes) of the survey, and outcome numbers used to calculate RR. RRs were calculated from the outcome numbers even if the separate value was present in (some of) the reports. Of the six ways of RR calculation proposed by the American Association for Public Opinion Research, I selected RR6 [71], because of some inconsistencies in the reporting of the "unknown eligibility" category in the ISSP.

$$RR6 = \frac{(I + P)}{(I + P) + (R + NC + O)}$$

where I and P represent complete and partial interviews; R represents refusals and break-offs; NC represents noncontacts; and O represents other (eligible) nonresponses. Because ESS outcome reporting is constructed in a different way, personal communication was needed to establish how to calculate RR6. Result datasets were used to code the NBI. The idea of this measure was described previously. NBI, being in this context a single study effect size, was calculated as:

$$ES_i = |prop_{fem} - 0.5|$$

where  $prop_{fem}$  is the proportion of females in a subsample of couples living together in two-person households and  $n_i$  is the size of a subsample of couples living together in two-person households [21]. As the proportion on which the  $ES_i$  is based is only an estimate and the true proportion of females in the defined subpopulation is already known to equal 0.5, the within-study variance may be calculated as:

$$V_i = 0.25/n_i$$

where  $n_i$  is the subsample size of the single study [21, 23, 24]. Information about the "survey experience" (how many times the survey was done before in one country in one program) was calculated afterwards by the author. While obviously not capturing all information about possible fieldwork institution change between waves and the degree of shared experience or field networks between such institutions, it is a proxy for how much information about the local surveying climate and context was available to people conducting such a survey.

Initially, there were 233 coded ESS surveys from rounds 1 to 9 and 760 coded ISSP surveys from waves 1996–2018. Exclusion criteria, meant to compare only surveys using entirely random sampling, were: using nonrespondents substitution, quotas, or nonrandom within-household selection, not providing complete information about sampling, survey mode, outcome numbers, and information needed to compute NBI. After the exclusion, the database consisted of 795 surveys (223 ESS and 572 ISSP). The created dataset is available at [72], and the SPSS syntaxes used for NBI calculations are available at [73]. The exclusion process is described in Fig 1.

Because the analysis was conducted on a database created by the author [based on 75], who used all the methodological reports and datasets provided by the repositories formerly listed to create it, and because of the traditions and specific nature of meta-analysis in the survey methodology context, it is not possible nor necessary to use the entire PRISMA standards [76] for reporting, as is done in clinical trial meta-analyses.

As stated previously, all ESS surveys are done in face-to-face mode and in European countries. ISSP surveys are done in multiple modes in countries from all inhabited continents. Table 1 shows the distribution of modes by the United Nations world regions.

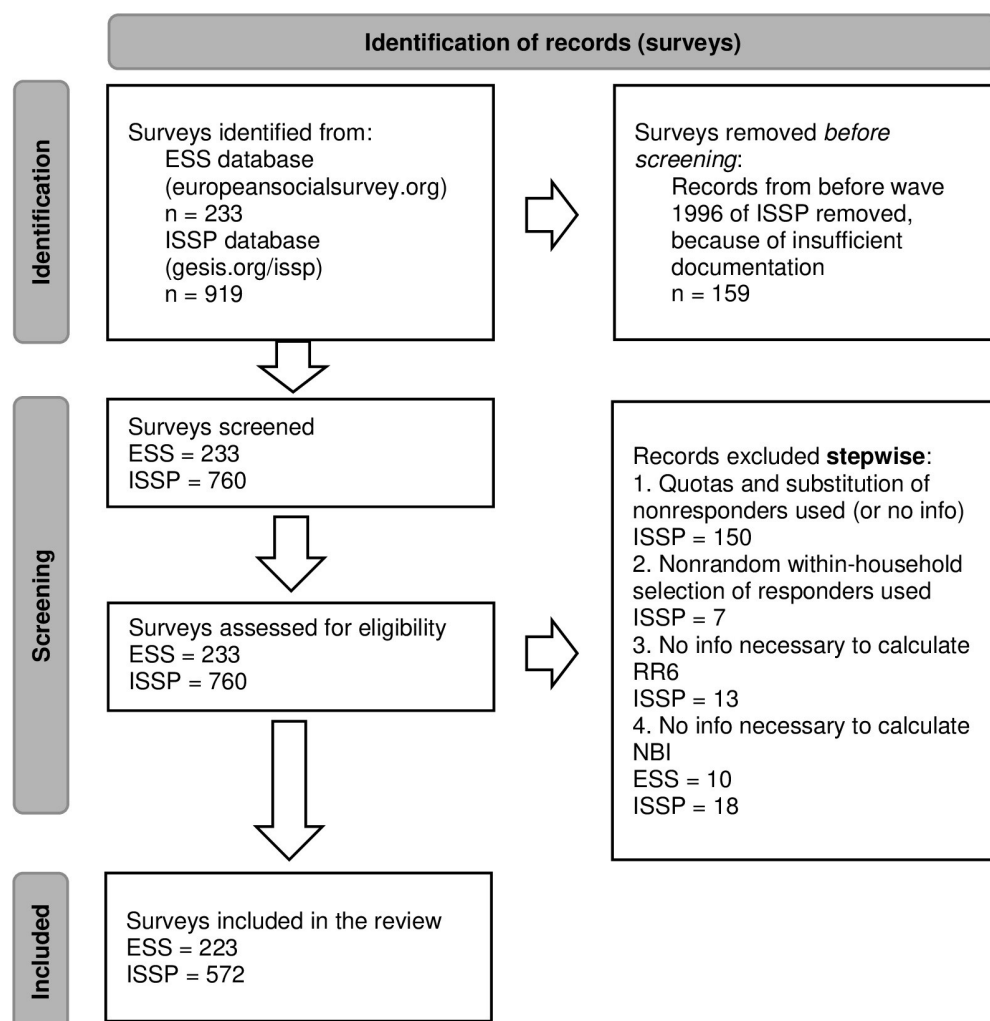


Fig 1. PRISMA flow diagram. From [74].

<https://doi.org/10.1371/journal.pone.0283092.g001>

As for the categorization of the modes, face-to-face and mail-mode categories included single-mode surveys conducted in the stated mode. ISSP questionnaires are quite often fielded as a self-administered supplement to the face-to-face main survey. Apart from that, sometimes ISSP surveys were reported as being fielded in single-mode self-completion (especially popular in ISSP Japan, 15 surveys); it indicates a self-administered questionnaire delivered and collected by an interviewer. Both of these situations fulfil the criteria for being called a mixed-mode survey [45, 46]. From all 96 ISSP mixed-mode surveys present in the final dataset, where face-to-face was one of the modes, only three contained any survey mode more than face-to-face and self-completion. It was in the year 2000 in Switzerland, where some additional contacts after these two modes were made by telephone and mail, and in the 2012 and 2013 ISSP in Iceland, where face-to-face nonresponders could fill out an online questionnaire. For this reason, all these surveys were included in one category: mixed-mode (with face-to-face). As for the last category—mixed-mode surveys (without face-to-face)—it included (ISSP) surveys fielded in different combinations of mail, web, and CATI survey modes. This type of survey was especially popular in Nordic countries: 15 surveys in Denmark, 10 in Norway, 8 in

Table 1. Distribution of survey modes in analyzed dataset by the United Nations world regions.

|                           | Face-to-face (of which ESS) | Mixed-mode (with f2f) | Mail mode | Mixed-mode (no f2f) | Total |
|---------------------------|-----------------------------|-----------------------|-----------|---------------------|-------|
| Australia and New Zealand | 0                           | 1                     | 27        | 3                   | 31    |
| Caribbean                 | 4                           | 0                     | 0         | 0                   | 4     |
| Central America           | 10                          | 0                     | 0         | 0                   | 10    |
| Eastern Asia              | 32                          | 20                    | 0         | 0                   | 52    |
| Eastern Europe            | 115 (47)                    | 9                     | 0         | 0                   | 124   |
| Northern America          | 15                          | 8                     | 9         | 0                   | 32    |
| Northern Europe           | 100 (72)                    | 26                    | 39        | 42                  | 207   |
| South America             | 20                          | 2                     | 0         | 0                   | 22    |
| Southeastern Asia         | 16                          | 0                     | 0         | 0                   | 16    |
| Southern Africa           | 14                          | 0                     | 0         | 0                   | 14    |
| Southern Asia             | 6                           | 0                     | 0         | 0                   | 6     |
| Southern Europe           | 93 (38)                     | 0                     | 0         | 0                   | 93    |
| Western Asia              | 40 (13)                     | 0                     | 0         | 0                   | 40    |
| Western Europe            | 81 (53)                     | 30                    | 26        | 7                   | 144   |
| Total                     | 546                         | 96                    | 101       | 52                  | 795   |

<https://doi.org/10.1371/journal.pone.0283092.t001>

Finland, and 5 in Iceland. As stated previously and subsequently, ESS face-to-face surveys were included in the model to alleviate issues originating from this study's observational, non-experimental character.

## Meta-analysis

Most commonly, meta-analysis is used to create proper statistical synthesis of results taken from many articles or clinical trials. In that situation, some typically used dependent variables, called effect sizes or treatment effects [77 p17], are usually computed to either statistically standardize the different measures used by different authors or comply with standard settings of experimental studies in medicine. Therefore, readers probably encountered two types of effect sizes: “those based on differences of two means and those based on differences of two proportions” [78 p13]. However, these examples do not exhaust the usefulness of meta-analysis. On the contrary, “effect size is simply the size of anything that may be of interest” [79 p34], whether it is measured in exactly the same way for every study or recalculated to make it so. For example, in epidemiological studies, the true population mean, or proportion, may be of interest [80]. Card proposed that the working definition of effect size could be “an index of the direction and magnitude of association between two variables” [62 p87], a definition that would include the measure used in this paper, if interpreted *sensu largo*.

An important question while doing a meta-analysis is whether to use a fixed- or random-effects model [77 p61–86]. Fixed-effects models are justified only when the studies included are nearly identical and one is not interested in treating them as part of a bigger population. Therefore, random-effects models, being more robust and allowing for some generalizations, should be used in most cases. The simplest meta-analytical random-effects model looks like the equation below:

$$\widehat{OES}^* = \frac{\sum_{i=1}^k w_i^* \times ES_i}{\sum_{i=1}^k w_i^*}$$

“Parameter k is the number of included studies and estimator  $ES_i$  stands for single-study effect size, i.e., the difference between the observed and the true proportion of females in two-

person heterosexual households. For each study, weights  $w_i^*$  are computed as reciprocals of the variance of  $\widehat{OES}^*$ , i.e.,  $V_i^* = V_i + \tau^2$ , where  $V_i$  is within-study variance and  $\tau^2$  denotes between-study variance" [23, 77].

When enough heterogeneity between studies is found, meta-regression is the usual method of choice to analyze it [81]. To avoid spurious findings while conducting meta-regression, methods such as using the so-called Knapp-Hartung test [82] or permutation test were proposed [83–85]. Nevertheless, these typical two-level models, where the “hidden” first levels are usually individuals and the second levels are study summaries are sometimes not enough [86–89]. When one could expect that studies are nested in some higher-level cluster, one should use the three-level meta-regression model to deal with such dependencies. For discussion on why using more common hierarchical models is not enough, please see [90]. Studies can and often are nested in more than one set of clusters at a time—that is, random factors may be crossed; to address that situation, cross-classified models can be used [91]. Many types of meta-analytical models have already been used to conduct research in the area of survey methodology, all of which either synthesize results from different papers or are based on self-calculated measures for selected survey programs [e.g., 10, 37, 40, 41, 48, 59, 92–98].

The effect size I used for the meta-analysis was, as stated before, NBI. The choice of moderators for the models was based on both state-of-the-art and statistical evaluation. Survey mode and frame constitute probably the most basic set of survey characteristics. The subject of mode was extensively described in the introductory parts of this paper. As for the sampling frame, the frame of individuals is usually considered better in terms of representativeness than other types of frames, primarily because of the burden connected to the process of within-household selection of the respondent at the last sampling stage [22, 99, 100]. Therefore, I used the binary variable individual/nonindividual (household/address, area probability sampling). To not overcomplicate the model, I ignored the possible impact of the within-household selection method on nonresponse bias [96]. Exploratory subgroup analysis and previous knowledge about possible issues connected with the aforementioned respondent selection not being supervised by the interviewer (as in, for example, mail mode) led to the conclusion that the interaction between those two moderators may also be an important factor in explaining differences in NBI values (Readers can conduct all analyses mentioned in the article, and others, using the R script available at [101].) Therefore, mode, frame, and interaction between them were selected for the simplest model specification. The analyzed dataset consists of studies from two survey programs that differ significantly in terms of funding and organization [102, 103], hence the inclusion of the binary variable ESS/ISSP in the extended model. Institutional experience can influence outcome rates of a survey [3 p102–113] and thus at least the upper boundary of possible nonresponse bias. I chose to include a moderator that for each observation simply counts how many surveys were done previously in this country and program. This study, as stated previously, is based on observational data—like many meta-analyses in survey methodology [98]. If one, however, would like to apply the categories of variables usually included in the causal inference models, then it should be noted that this study is focused chiefly on the effect of survey mode on nonresponse bias and that every candidate for the meta-regression moderator listed above should be treated as a confounder. The sampling frame, institutional experience, and survey program can usually have an impact on the choice of mode(s) and separately on the nonresponse bias. As for the last candidate for a moderator—the RR (RR6)—its inclusion can be used to estimate the partial effects of survey modes holding RR constant. In the causal inference model, it would be a mediator. RR6 was included only in the last of the models presented in the results.

After testing if the inclusion of stated moderators is not only theoretically but also statistically sound (addressed in the results section), identification and exclusion of outliers were

performed based on the influence diagnostics of outliers [104] and on quantile-quantile plots [105, 106]. The four identified highly problematic observations included ESS surveys in Slovakia conducted in rounds 4, 5, and 6 and an ESS survey in Lithuania from round 9. All of them were face-to-face surveys with unusually high value of NBI. Lithuanian cases were already identified by Jabkowski, Cichocki, and Kołczyńska [25]; the sampling design switched from one based on the individual register frame in round 3 to an area sample with house enumeration in round 4. “This change dramatically increased interviewer discretion in regard to respondent selection,” which resulted in oversampling women. As for the case of Lithuania in round 9, it is a bit more problematic. Between rounds 8 and 9, the data collector changed, the mode switched from PAPI to CAPI, and the Kish grid replaced the last birthday method. Moreover, the DEFFp rose from 1.34 to 1.64 and the number of completed interviews dropped from 2,122 to 1,835 (References to documentation are available at [70].) Nevertheless, none of these, unfortunately, explains such a rise of NBI. The final models after the exclusion of outliers are based on 791 observations.

To answer the research questions, I created random-effects meta-regression (alternatively called mixed-effects) models using the restricted maximum likelihood method to estimate  $\tau^2$  [77 p114–117]. Every model was created with the R package metafor [107]. Software citations are available at [108] and the R script for replication at [101]. To address the issue of surveys probably being nested in some third-level clusters, which may result in observation dependency, I introduced cross-classified models [91], where surveys were nested simultaneously in program waves and countries. Surveys conducted on one country’s population at different points in time can be treated as repeated observations by some, therefore the need for nesting. Countries can also differ in survey climate, but it is hard to measure, and such measure was not included. Instead, the random intercepts for countries may, to some degree, serve the role of estimating it. The author considered including a variable measuring cultural trust at this level, but the most important, comprehensive source of such were the World Value Survey results, and the idea of basing survey evaluation on data from a methodologically less “diligent” survey was rejected. Surveys being part of specific comparative program waves (counted here separately for both programs) can share many characteristics, such as timeframe, methodological standards, and especially questionnaire topics. Topic saliency can have significant impact on unit nonresponse [109–111], but it is not uniform for the whole population. Therefore, it was accounted for only in the random intercept of waves. The moderator that explained part of wave variance only is the survey program because it does not differ inside the cluster. It is usually considered that there should be at least 20 third-level clusters [112 p46–48]. However, it is a matter of discussion [113] and is not always followed by researchers [86, 114]. In this case, surveys were nested in 32 waves and 53 countries, which is a clearly sufficient number. As the ICC calculation and the checking of  $I^2$  distribution between model levels [115] are extremely problematic in cross-classified models, such values were gathered for surveys nested in countries and in waves separately. Cross-nested models with three sets of moderators are presented in the results part. The final two-level model and the three-level model with RR6 as an effect size can be consulted in the S1 Table.

The raw, nontransformed proportions [for transformation proposal, see, e.g., 80] were used to calculate NBI. It is because the standard error of the NBI is based on the known “true” proportion of 0.5; therefore, there was no identified risk for the variance calculation. Moreover, subsamples and proportions were calculated without using weights. As weights are not standardized between countries and in the ISSP most of them are poststratification weights, it is better for comparisons to use raw data [For extensive discussion on this subject, please see 25.] If design weights were available for all analyzed cases, it would be the best choice to use them in calculations.

## Results

Standard, two-level random-effects meta-analysis of surveys using NBI as an effect size identified the mean value of effect (thus, NBI) as 0.0366 (95% CI 0.0341–0.0391). The test of heterogeneity is significant ( $Q(df = 794) = 1532.43$ ;  $p < 0.0001$ ).  $I^2$  (“the percentage of variability among effect sizes that exists between studies relative to the total variability among effect sizes” [62 p188–189]) has a value of 49.15%. After the exclusion of four outliers, the only major change concerned  $I^2$ , which dropped to 31.71%, a value that could be interpreted as a low-to-moderate level [116], a still sufficient heterogeneity between studies to conduct a meta-regression.

As we can see in Fig 2, simple random-effects meta-analysis in subgroups shows that NBI values for mixed-mode surveys with face-to-face (mostly face-to-face with additional self-administered questionnaires) and for the mail mode clearly overlap with the estimate for single-mode face-to-face surveys. At this level, it is not possible to tell whether values of bias estimate for other mixed-mode surveys are smaller than for face-to-face. For context, we can see in Fig 3 that the values of RR6 for modes using face-to-face surveys in any configuration clearly outperform other ones, which is in accordance with common expectations concerning this subject. Based on the presented literature, it was expected that the connection between realization and bias would be weak at most.

## Meta-regression

Meta-regression based on the full set of observations, with the basic set of moderators (mode\*frame), was compared with models including additional variables: program, repetitions, and a set of both. The likelihood ratio test suggests that, statistically, the full set is the best:  $\chi^2(2) =$

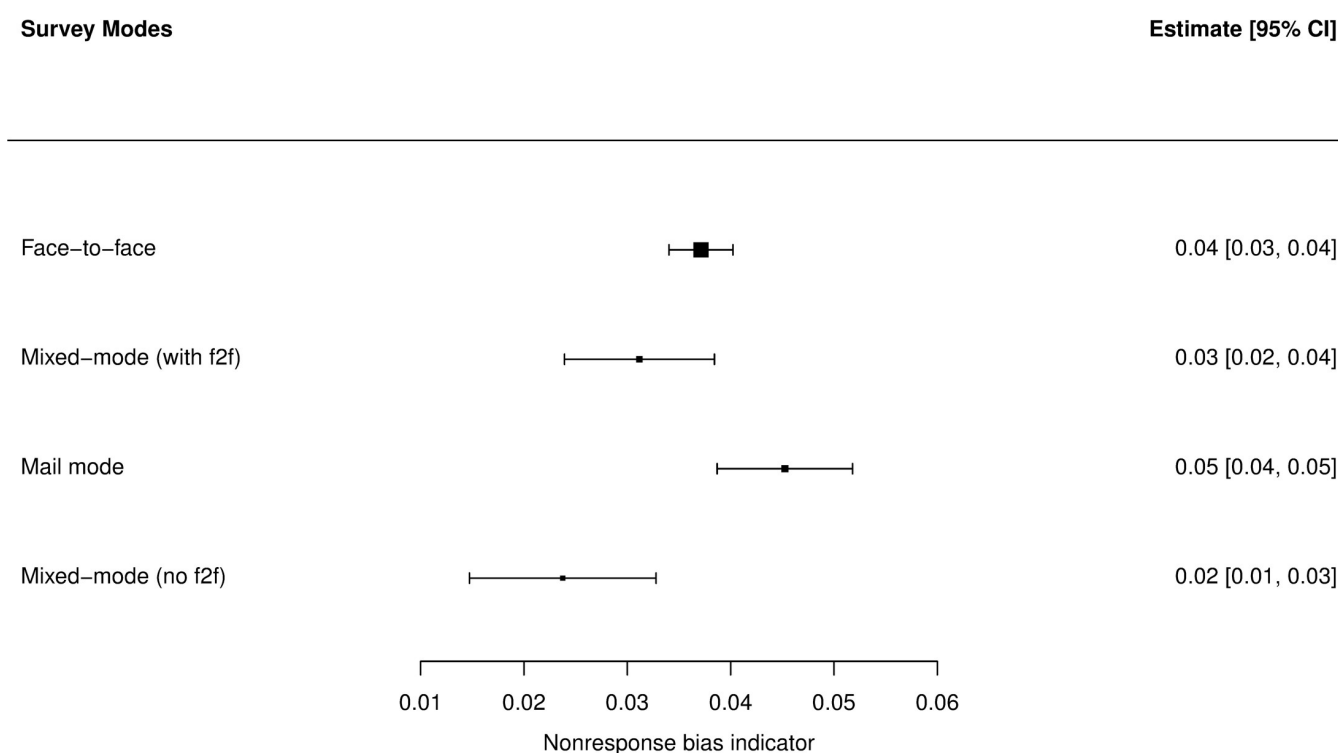
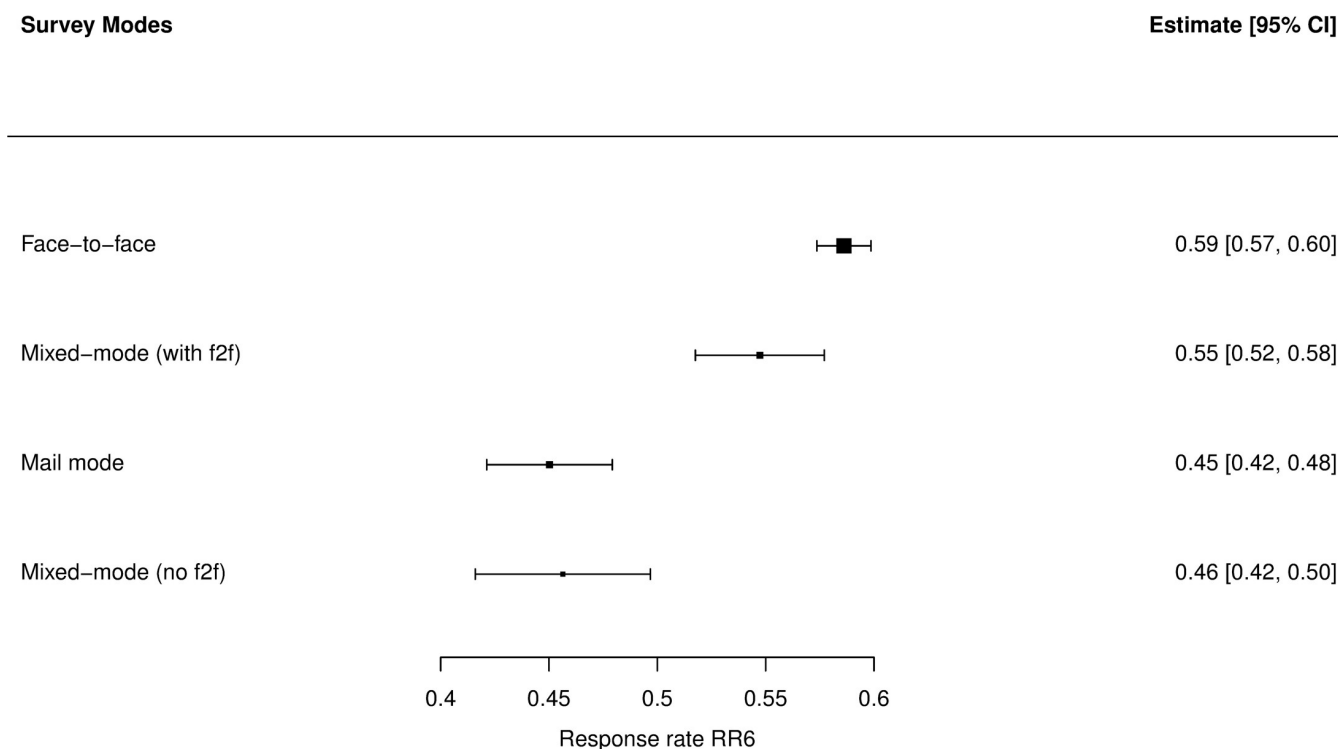


Fig 2. Distribution of NBI values of surveys in the dataset by survey modes—forest plot.  $N = 795$ .

<https://doi.org/10.1371/journal.pone.0283092.g002>



**Fig 3. Distribution of RR6 values of surveys in the dataset by survey modes—forest plot.** N = 795.

<https://doi.org/10.1371/journal.pone.0283092.g003>

13.09;  $p = 0.001$ . The model using this set of moderators, restricted maximum likelihood, and the Knapp-Hartung method, with  $p$ -values adjusted according to the permutation method, found mail mode, mixed-mode (no face-to-face), frame, program, repetitions, and mail-non-individual frame interaction moderators to be significant. This model was used for quantile-quantile plot and influence diagnostics, which both have resulted in flagging four observations as problematic outliers. These outliers were already described in the data and methods section and were excluded from any further analysis, which lowered the number of observations to 791. The two-level model designed the same as in the previous step, but with outliers excluded, differs in the frame moderator not being statistically significant and the mixed-mode (with face-to-face) starting to be statistically significant, having a pseudo- $R^2$  of 77.52%, and a significant test of moderators:  $F(df1 = 9, df2 = 781) = 29.0204$ ;  $p = 0.001$ . Multimodel inference [117] confirmed that the final set of moderators is optimal. This model is available in S1 Table.

As calculating the ICC and  $I^2$  distribution is problematic for cross-nested models, I calculated them for models with country and wave as the third level separately. Every three-level model in this paper uses  $t$ - and  $F$ -distributions for making inference, and uses “an improved method for approximating the degrees of freedom” [118 p128] for these distributions, as the full Knapp-Hartung method is unavailable, and the models also exclude the four outliers. According to a likelihood ratio test (intercept only), the model with surveys nested in waves (and no moderators) is statistically better than the two-level model:  $\chi^2(1) = 5.2$ ;  $p = 0.02$  and has an ICC value of only 0.1, 28.6% of the total variance at level 2, and only 3.28% at level 3. The model with surveys nested in countries (and no moderators) is statistically better than the two-level model:  $\chi^2(1) = 97.64$ ;  $p < 0.0001$  and has an ICC value of 0.65, 13.1% of the total variance at level 2, and 24.1% at level 3. The cross-classified model (surveys nested in countries

and waves) has all three variance components identifiable according to likelihood profile plots [118 p171]. **Table 2** presents a set of three cross-classified models with different sets of moderators. The first model uses the most basic set—mode, frame, and their interaction—and has a pseudo- $R^2$  of 58%. The second model uses the optimal set of available moderators with added program and repetition (“institutional experience”) variables and has a pseudo- $R^2$  of 67%. The third model, aimed at estimating the partial effect of the mode holding outcome rates constant, includes an additional moderator: RR (RR6). It has a pseudo- $R^2$  of 69%. For context, readers can find a similar cross-classified model using RR6 as an effect size in [S1 Table](#).

The three presented models agree that when the frame is held at the “individual” level, mail mode and mixed-mode without face-to-face in the mix have significantly lower NBI than single-mode face-to-face. According to Model II, when the frame is held at the “individual” level, every mode is better in terms of nonresponse bias than pure face-to-face. When one holds the frame at the “nonindividual” level, the only mode bearing significantly different results than face-to-face is the mail mode; that mode used with a nonindividual frame is connected to a massive and significant increase in the NBI (This is the strongest effect observed in this study.) This problematic combination of mail and address frame in the survey design was most used in France (16 out of 28 cases). One should bear in mind that there are far fewer cases of mail and mixed-mode (no face-to-face) surveys connected to the household/address frame than to the frame of individuals: 28 to 73 for mail and 3 to 49 for mixed-mode. The ESS has significantly lower NBI than the ISSP and also more experience possibly gathered by the institution results in slightly lower bias. Model III, estimating partial effects of the mode, bears very similar results. Additionally, in this model, a higher RR is connected to lower bias.

## Discussion and conclusion

This study primarily aims to assess the relationship between survey mode and nonresponse bias. The choice of mode is one of the most crucial decisions when designing survey research. Whereas the existing opportunities usually impose the type of frame used and the sampling design, the mode choice is the compromise between methodological stance and possible budget. The most widespread belief is that the more expensive the mode, the better; and for cross-cut surveys, the face-to-face mode is usually the most expensive (It may not be true for poorly designed mixed-mode surveys.) While having the highest RRs from the analyzed modes, face-to-face does not have the lowest bias. The models presented here suggest that at least mail mode and mixed-mode surveys (without face-to-face) resulted in significantly smaller nonresponse bias than single mode face-to-face surveys when the frame of individuals was used. For mail surveys, this result was probably obtained—according to estimates present in the literature—with a much lower cost than face-to-face surveys. For mixed-mode surveys (without face-to-face), if they were designed properly, this remarkable result was also achieved with a lower cost than face-to-face surveys. These forms of mixed-mode surveys using different combinations of mail, web, and CATI were applied mainly in Nordic countries, such as Denmark (in 15 ISSP waves), Norway (10), Finland (8), and Iceland (5). Models II and III found that self-administered surveys, solo or mixed with face-to-face, also resulted in significantly lower NBI when the frame of individuals was used. In a scenario when sampling design was based on other types of frames, only mail mode was connected to higher NBI than face-to-face. It all suggests that the switch from face-to-face to other modes should be considered a great opportunity to raise research quality, even considering possible risks, mainly some kinds of mode effects [119].

All these results might mean that face-to-face surveys may not fully deserve the reverence they are held in based mostly upon the high RRs achieved, especially if we consider that this

Table 2. Cross-classified meta-regression models estimating the relationship between survey characteristics and NBI.

|                               | Model I |       |        |     |        |        | Model II |      |        |       |        |      | Model III |        |        |     |        |       |        |       |        |        |        |     |
|-------------------------------|---------|-------|--------|-----|--------|--------|----------|------|--------|-------|--------|------|-----------|--------|--------|-----|--------|-------|--------|-------|--------|--------|--------|-----|
|                               | coef    | se    | tval   | df  | pval   | ci.lb  | ci.ub    | coef | se     | tval  | df     | pval | ci.lb     | ci.ub  | coef   | se  | tval   | df    | pval   | ci.lb | ci.ub  |        |        |     |
| Intercept                     | 0.034   | 0.003 | 10.843 | 24  | < .001 | 0.028  | 0.041    | ***  | 0.051  | 0.005 | 11.131 | 22   | < .001    | 0.041  | 0.060  | *** | 0.066  | 0.007 | 9.302  | 21    | < .001 | 0.051  | 0.081  | *** |
| Mixed-mode (with f2f)         | -0.005  | 0.005 | -0.944 | 783 | 0.346  | -0.014 | 0.005    |      | -0.010 | 0.005 | -1.996 | 781  | 0.046     | -0.021 | 0.000  | *   | -0.010 | 0.005 | -2.013 | 780   | 0.044  | -0.021 | 0.000  | *   |
| Mail mode                     | -0.011  | 0.004 | -2.475 | 783 | 0.014  | -0.020 | -0.002   | *    | -0.019 | 0.005 | -3.938 | 781  | < .001    | -0.028 | -0.009 | *** | -0.019 | 0.005 | -4.086 | 780   | < .001 | -0.029 | -0.010 | *** |
| Mixed-mode (no f2f)           | -0.012  | 0.005 | -2.525 | 783 | 0.012  | -0.021 | -0.003   | *    | -0.017 | 0.005 | -3.607 | 781  | < .001    | -0.027 | -0.008 | *** | -0.020 | 0.005 | -4.006 | 780   | < .001 | -0.029 | -0.010 | *** |
| Nonindividual frame           | 0.005   | 0.003 | 1.426  | 783 | 0.154  | -0.002 | 0.012    |      | 0.001  | 0.003 | 0.367  | 781  | 0.714     | -0.005 | 0.008  |     | 0.002  | 0.003 | 0.502  | 780   | 0.616  | -0.005 | 0.008  |     |
| Mixed-mode (with f2f)*        | 0.005   | 0.007 | 0.763  | 783 | 0.446  | -0.009 | 0.019    |      | 0.008  | 0.007 | 1.147  | 781  | 0.252     | -0.006 | 0.022  |     | 0.009  | 0.007 | 1.243  | 780   | 0.214  | -0.005 | 0.023  |     |
| Nonindividual frame           |         |       |        |     |        |        |          |      |        |       |        |      |           |        |        |     |        |       |        |       |        |        |        |     |
| Mail mode*Nonindividual frame | 0.067   | 0.007 | 9.601  | 783 | < .001 | 0.053  | 0.081    | ***  | 0.070  | 0.007 | 10.053 | 781  | < .001    | 0.056  | 0.083  | *** | 0.065  | 0.007 | 9.196  | 780   | < .001 | 0.051  | 0.079  | *** |
| Mixed-mode (no f2f)*          | 0.011   | 0.012 | 0.866  | 783 | 0.387  | -0.013 | 0.035    |      | 0.009  | 0.012 | 0.744  | 781  | 0.457     | -0.015 | 0.033  |     | 0.009  | 0.012 | 0.760  | 780   | 0.447  | -0.014 | 0.033  |     |
| Nonindividual frame           |         |       |        |     |        |        |          |      |        |       |        |      |           |        |        |     |        |       |        |       |        |        |        |     |
| ESS                           |         |       |        |     |        |        |          |      | -0.020 | 0.004 | -5.310 | 22   | < .001    | -0.028 | -0.012 | *** | -0.020 | 0.004 | -5.377 | 21    | < .001 | -0.027 | -0.012 | *** |
| Repetitions                   |         |       |        |     |        |        |          |      | -0.001 | 0.000 | -3.328 | 781  | 0.001     | -0.001 | 0.000  | *** | -0.001 | 0.000 | -3.818 | 780   | < .001 | -0.001 | 0.000  | *** |
| Response rate (RR6)           |         |       |        |     |        |        |          |      |        |       |        |      |           |        |        |     | -0.025 | 0.009 | -2.823 | 780   | 0.005  | -0.043 | -0.008 | **  |
| Pseudo-R <sup>2</sup>         | 0.58    |       |        |     |        |        |          |      | 0.67   |       |        |      |           |        |        |     | 0.69   |       |        |       |        |        |        |     |

\* p&lt;0.05

\*\* p&lt;0.01

\*\*\* p&lt;0.001

<https://doi.org/10.1371/journal.pone.0283092.t002>

mode is also a victim of possible measurement issues, due mainly to the interviewer effects [120]. This finding is particularly optimistic in the context of limitations imposed by the COVID-19 pandemic, beginning in 2020, which made doing face-to-face research extremely hard. While the health concerns are in most cases lower at the time of writing this paper, changes in people's attitudes may be more long-lasting.

On the other hand, it must be stated that, compared especially to mail and CATI modes, direct interviewer involvement enables the collection of a lot more paradata, which is highly important in survey methodology research. This is an undeniable loss produced by the mode switch. Another disadvantage of using non-face-to-face modes concerns the interaction between the mail mode and the nonindividual frame (household, address, or area probability sampling). It resulted in a significant rise of NBI, probably because the within-household selection of the final respondent needed to be done by household members. They may not have understood the written instructions or might not have found these instructions that important when no interviewer was present to guide and motivate them to do it properly. This problematic situation was most common in the ISSP in France (16 out of 28 cases in the database). Lobbying by the research community to establish individual frames in countries where it is not present is highly advisable. Another factor resulting in lower bias was the (proxy of) institutional experience; doing more surveys in one country and one program resulted in lower NBI, which means that people do learn and improve their work over time. In addition, the ESS, according to the univocal assessment of the survey methodologist, being more methodologically "diligent" than the ISSP, had smaller NBI values.

This study's main limitation is that the survey modes in the analyzed programs were not assigned randomly to country and wave. Therefore, there is a possibility that behind all the models, there is one important hidden variable: the survey climate of the country [121 p815–17]. It was at least partially accounted for by using random intercepts for every country, but this solution is only viable if the survey climate is constant over time. This statement is probably false at least in some countries. The best way of overcoming this issue is through experimental research, yet it is very costly and impossible to conduct on a scale such as was used in this study. Of 54 meta-analyses in survey methodology listed by Čehovin, Bošnjak, and Lozar Manfreda, 28, more than half, used observational data [98]. Therefore, this limitation is real but very common in the field. The method of alleviating this limitation at least a little, used here, was the inclusion of ESS data. While ESS surveys used the single face-to-face mode only, they provided the possibility of comparing the results of surveys using this mode with ISSP surveys using another mode, but done in the same country and the same period. Unfortunately, it could only be done for European countries, thereby excluding examples of continuous use of the mail mode in New Zealand (14), Australia (13), and Canada (9). Another limitation comes from the impossibility, in the context of the ISSP data, of using the previously described "strict method" of NBI calculation—that is, controlling for the sex of the respondent's partner. A previous study suggests that this approach does not carry much risk [25]. The last limitation worth mentioning here is the fact that this paper covers modes and approaches used mostly in large comparative survey programs, which are relatively well funded, based on international collaboration, and usually academically and/or methodologically focused. For researchers dealing with smaller surveys, usually local, or especially commercially focused, and who are not used to pursuing full statistical representativeness, the data presented here may be, unfortunately, of small importance; the author is fully aware of this limitation.

The deluge of survey data gathered over recent years makes it nearly impossible to utilize their full analytical potential. In many cases funded from public sources, the surveying process should be used with care, and sparingly. Moreover, the inflation of surveys is clearly

detrimental to the surveying climate. Therefore, the method of meta-analysis in survey methodology research, while currently gaining in popularity, deserves to be utilized much more.

## Supporting information

**S1 File. PRISMA checklist.** From: [74].  
(DOCX)

**S1 Table. Two-level meta-regression model using NBI as an effect size and three-level meta-regression model using RR6 as an effect size.**  
(CSV)

## Acknowledgments

I would like to thank Piotr Jabkowski and Kamila Szymkowiak for their help in this research.

## Author Contributions

**Conceptualization:** Adam Rybak.

**Data curation:** Adam Rybak.

**Formal analysis:** Adam Rybak.

**Funding acquisition:** Adam Rybak.

**Investigation:** Adam Rybak.

**Methodology:** Adam Rybak.

**Project administration:** Adam Rybak.

**Resources:** Adam Rybak.

**Visualization:** Adam Rybak.

**Writing – original draft:** Adam Rybak.

**Writing – review & editing:** Adam Rybak.

## References

1. de Leeuw ED, Hox JJ, Luiten A. International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data. *Surv Methods Insights Field*. 2018; 1–11. <https://doi.org/10.13094/SMIF-2018-00008>
2. Luiten A, Hox J, de Leeuw ED. Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *J Off Stat*. 2020; 36: 469–487. <https://doi.org/10.2478/jos-2020-0025>
3. Stoop I, Billiet J, Koch A, Fitzgerald R. *Improving Survey Response: Lessons Learned from the European Social Survey*. Chichester: John Wiley & Sons, Ltd; 2010.
4. Jabkowski P. Zróżnicowanie praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie archiwizacji 1537 surveyów zrealizowanych w krajach europejskich w latach 1981–2017. *Przegląd Socjol*. 2019;68: 9–38. <https://doi.org/10.26485/PS/2019/68.1/1>
5. Jabkowski P, Kołczyńska M. Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981–2017. *Methodology*. 2020; 16: 186–207. <https://doi.org/10.5964/meth.2795>
6. Biemer PP. Total survey error: Design, implementation, and evaluation. *Public Opin Q*. 2010; 74: 817–848. <https://doi.org/10.1093/poq/nfq058>
7. Smith TW. Refining the total survey error perspective. *Int J Public Opin Res*. 2011; 23: 464–484. <https://doi.org/10.1093/ijpor/edq052>

8. Groves RM, Lyberg L. Total survey error: Past, present, and future. *Public Opin Q.* 2010; 74: 849–879. <https://doi.org/10.1093/poq/nfq065>
9. Bianchi A, Shlomo N, Schouten B, Da Silva DN, Skinner C. Estimation of response propensities and indicators of representative response using population-level information. *Surv Methodol.* 2019; 45: 231–264.
10. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opin Q.* 2008; 72: 167–189. <https://doi.org/10.1093/poq/nfn011>
11. Groves R. Nonresponse rates and nonresponse bias in household surveys: What Do We Know about the Linkage between Nonresponse Rates and Nonresponse Bias? *Public Opin Q.* 2006; 70: 646–675. <https://doi.org/10.1093/poq/nfl033>
12. Peytchev A, Riley S, Rosen J, Murphy J, Lindblad M. Reduction of nonresponse bias in surveys through case prioritization. *Surv Res Methods.* 2010; 4: 21–29. <https://doi.org/10.18148/srm/2010.v4i1.3037>
13. Beullens K, Loosveldt G. Should high response rates really be a primary objective? *Surv Pract.* 2012; 5: 1–5. <https://doi.org/10.29115/sp-2012-0019>
14. Schouten B, Cobben F, Bethlehem J. Indicators for the representativeness of survey response. *Surv Methodol.* 2009; 35: 101–113.
15. Schouten B, Shlomo N, Skinner C. Indicators for monitoring and improving representativeness of response. *J Off Stat.* 2011; 27: 231–253.
16. Schouten B, Bethlehem J, Beullens K, Kleven Ø, Loosveldt G, Luiten A, et al. Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators. *Int Stat Rev.* 2012; 80: 382–399. <https://doi.org/10.1111/j.1751-5823.2012.00189.x>
17. Koch A, Halbherr V, Stoop IAL, Kappelhof JWS. Assessing ESS sample quality by using external and internal criteria. Mannheim: European Social Survey, GESIS; 2014.
18. Kobilanski FS, Pizzolitto G. Sample Substitutions in the AmericasBarometer 2016/17. Vanderbilt University; 2019.
19. Halbesleben JRB, Whitman MV. Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. *Health Serv Res.* 2013; 48: 913–930. <https://doi.org/10.1111/1475-6773.12002> PMID: 23046097
20. Sodeur W. Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. *ZA-Inf Zentralarchiv Empirische Sozialforschung.* 1997; 58–82.
21. Kohler U. Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Surv Res Methods.* 2007; 1: 55–67. <https://doi.org/10.18148/srm/2007.v1i2.75>
22. Menold N. The influence of sampling method and interviewers on sample realization in the european social survey. *Surv Methodol.* 2014; 40: 105–123.
23. Jabkowski P, Cichocki P. Within-household selection of target-respondents impairs demographic representativeness of probabilistic samples: Evidence from seven rounds of the European social survey. *Surv Res Methods.* 2019; 13: 167–180. <https://doi.org/10.18148/srm/2019.v13i2.7383>
24. Eckman S, Koch A. Interviewer Involvement in Sample Selection Shapes the Relationship between Response Rates and Data Quality. *Public Opin Q.* 2019; 83: 313–337. <https://doi.org/10.1093/poq/nfz012> PMID: 31534276
25. Jabkowski P, Cichocki P, Kołczyńska M. Multi-Project Assessments of Sample Quality in Cross-National Surveys: The Role of Weights in Applying External and Internal Measures of Sample Bias. *J Surv Stat Methodol.* 2021; 00: 1–24. <https://doi.org/10.1093/jssam/smab027>
26. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. *Jama.* 2012; 307: 1805–1806. <https://doi.org/10.1001/jama.2012.3532> PMID: 22550194
27. de Leeuw ED. Choosing the Method of Data Collection. In: de Leeuw ED, Hox JJ, Dillman DA, editors. *International Handbook of Survey Methodology.* New York: L. Erlbaum Associates; 2008. pp. 113–135.
28. de Leeuw ED. *Data Quality in Mail, Telephone and Face to Face surveys.* Amsterdam: TT-Publikaties; 1992.
29. Smyth JD, Olson K, Millar MM. Identifying predictors of survey mode preference. *Soc Sci Res.* 2014; 48: 135–144. <https://doi.org/10.1016/j.ssresearch.2014.06.002> PMID: 25131280
30. Rybak A. Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference. *Pol Sociol Rev.* 2020; 2: 219–28. <https://doi.org/10.26412/psr210.05>
31. Groves RM, Kahn RL. *Surveys by telephone: a national comparison with personal interviews.* New York: Academic Press; 1979.

32. Tourangeau R, Rips LJ, Rasinski KA. The psychology of survey response. Cambridge: Cambridge University Press; 2000.
33. Rybak A. The effects of visual and aural communication on the quality of survey measurement. The proposition of a new theoretical approach to survey mode preference on the basis of the working memory concept. *Przegląd Socjol.* 2019; 68: 207–227. <https://doi.org/10.26485/ps/2019/68.3/8>
34. Villard A, Fitzgerald R. Using mixed modes in survey research: Evidence from six experiments in the ESS. In: Breen MJ, editor. *Values and identities in Europe: Evidence from the European Social Survey*. Abingdon: Routledge; 2017. pp. 273–310. <https://doi.org/10.4324/9781315397146>
35. Fox RJ, Crask MR, Kim J. Mail Survey Response Rate: A Meta-Analysis Of Selected Techniques For Inducing Response. *Public Opin Q.* 1988; 52: 467–491. <https://doi.org/10.1086/269125>
36. Goyder J. Face-to-face interviews and mailed questionnaires: The net difference in response rate. *Public Opin Q.* 1985; 49: 234–252. <https://doi.org/10.1086/268917>
37. Hox JJ, Leeuw ED de. A comparison of nonresponse in mail, telephone, and face-to-face surveys. Applying multilevel modeling to meta-analysis. *Qual Quant.* 1994; 28: 329–344.
38. Chang L, Krosnick JA. Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opin Q.* 2010; 74: 154–167. <https://doi.org/10.1093/poq/nfp090>
39. Sakshaug JW, Yan T, Tourangeau R. Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opin Q.* 2010; 74: 907–933. <https://doi.org/10.1093/poq/nfq057>
40. Daikeler J, Bosnjak M, Manfreda KL. Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *J Surv Stat Methodol.* 2020; 8: 513–539. <https://doi.org/10.1093/jssam/smz008>
41. Shih T-H, Fan X. Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods.* 2008; 20: 249–271. <https://doi.org/10.1177/1525822X08317085>
42. Dillman DA, Smyth JD, Christian LM. *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). Hoboken: John Wiley & Sons Inc; 2014.
43. Couper MP. The future of modes of data collection. *Public Opin Q.* 2011; 75: 889–908. <https://doi.org/10.1093/poq/nfr046>
44. Wolf C, Christmann P, Gummer T, Schnaudt C, Verhoeven S. Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys. *Public Opin Q.* 2021; 85: 623–648. <https://doi.org/10.1093/poq/nfab039>
45. de Leeuw ED, Hox JoopJ, Dillman DA. Mixed-mode Surveys: When and Why. In: de Leeuw ED, Hox JoopJ, Dillman DA, editors. *International Handbook of Survey Methodology*. New York: L. Erlbaum Associates; 2008.
46. de Leeuw ED. Mixed-mode: Past, present, and future. *Surv Res Methods.* 2018; 12: 75–89. <https://doi.org/10.18148/srm/2018.v12i2.7402>
47. de Leeuw ED. To Mix or Not to Mix Data Collection Modes in Surveys. *J Off Stat.* 2005; 21: 233–255.
48. Medway RL, Fulton J. When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opin Q.* 2012; 76: 733–746. <https://doi.org/10.1093/poq/nfs047>
49. Rybak A. The Selection of Survey Modes on the Basis of Respondents' Preferences Analysis. *Ask Res Methods.* 2018; 27: 23–39. <https://doi.org/10.18061/ask.v27i1.0002>
50. Al Baghal T, Kelley J. The Stability of Mode Preferences: Implications for Tailoring in Longitudinal Surveys. *Methods Data Anal.* 2016; 10: 143–166. <https://doi.org/10.12758/mda.2016.012>
51. Olson K, Smyth JD, Wood HM. Does giving people their preferred survey mode actually increase survey participation rates? an experimental examination. *Public Opin Q.* 2012; 76: 611–635. <https://doi.org/10.1093/poq/nfs024>
52. Tourangeau R, Michael Brick J, Lohr S, Li J. Adaptive and responsive survey designs: a review and assessment. *J R Stat Soc Ser A Stat Soc.* 2016. <https://doi.org/10.1111/rssa.12186>
53. Laaksonen S, Heiskanen M. Comparison of three modes for a crime victimization survey. *J Surv Stat Methodol.* 2014; 2: 459–483. <https://doi.org/10.1093/jssam/smu018>
54. Klausch T, Hox J, Schouten B. Selection error in single- and mixed mode surveys of the Dutch general population. *J R Stat Soc Ser A Stat Soc.* 2015; 178: 945–961. <https://doi.org/10.1111/rssa.12102>
55. Fowler FJJ, Gallagher PM, Stringfellow VL, Zaslavsky AM, Thompson JW, Cleary PD. Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members. *Med Care.* 2002; 40: 190–200. <https://doi.org/10.1097/00005650-200203000-00003> PMID: 11880792
56. Dillman DA, Phelps G, Tortora R, Swift K, Kohrell J, Berck J, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Soc Sci Res.* 2009; 38: 1–18. <https://doi.org/10.1016/j.ssresearch.2008.03.007>

57. Luiten A, Schouten B. Tailored fieldwork design to increase representative household survey response: An experiment in the Survey of Consumer Satisfaction. *J R Stat Soc Ser A Stat Soc.* 2013; 176: 169–189. <https://doi.org/10.1111/j.1467-985X.2012.01080.x>
58. Bianchi A, Biffignandi S, Lynn P. Web-face-to-face mixed-mode design in a longitudinal survey: Effects on participation rates, sample composition, and costs. *J Off Stat.* 2017; 33: 385–408. <https://doi.org/10.1515/jos-2017-0019>
59. Cornesse C, Bosnjak M. Is there an association between survey characteristics and representativeness? A meta-analysis. *Surv Res Methods.* 2018; 12: 1–13. <https://doi.org/10.18148/srm/2018.v12i1.7205>
60. Sztabiński PB. Does mixed mode data collection improve the achieved sample? A comparison of the ESS 7 and a mixed mode experiment. In: Sztabiński F, Domański H, Sztabiński PB, editors. *New Uncertainties and Anxieties in Europe: Seven Waves of the European Social Survey.* Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang; 2018. <https://doi.org/10.3726/b13526>
61. Hunter JE, Schmidt FL. *Methods of meta-analysis: Correcting error and bias in research finding.* Thousand Oaks: Sage Publications, Inc.; 2004.
62. Card NA. *Applied Meta-Analysis for Social Science Research.* New York; London: Guilford Press; 2012.
63. GESIS. International Social Survey Programme data and documentation at GESIS archive (GN0001). 2020 [cited 10 May 2021]. Available: [www.gesis.org/issp](http://www.gesis.org/issp)
64. ISSP. ISSP History. 2020 [cited 10 May 2020]. Available: <https://www.gesis.org/issp>
65. Smith TW. The ISSP: History, Organization and Members, Working Principles and Outcomes. In: Haller M, Jowell R, Smith TW, editors. *The International Social Survey Programme, 1984–2009.* London and New York: Routledge; 2009. pp. 2–27.
66. ESS ERIC. ESS Data. 2022 [cited 10 Jan 2021]. Available: <https://www.europeansocialsurvey.org/data/>
67. Jabkowski P. Sondażowe zwierciadło europeizacji Monitorowanie trendów europejskiej opinii publicznej w międzykrajowych sondażach porównawczych jako wyzwanie metodologiczne. Poznań: Wydawnictwo Naukowe UAM; 2021. <https://doi.org/10.14746/amup.9788323239154>
68. Jowell R. How Comparative Is Comparative Research? *Am Behav Sci.* 1998; 42: 168–177. <https://doi.org/10.1177/0002764298042002004>
69. Breen MJ. The Significance of the European Social Survey. In: Breen MJ, editor. *Values and identities in Europe: Evidence from the European Social Survey.* Abingdon: Routledge; 2017. pp. 1–14.
70. Rybak A. Data citations PLOS ONE. OSF; 2022. <https://doi.org/10.17605/OSF.IO/D6PZG>
71. The American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* AAPOR; 2016.
72. Rybak A. Dataset for article PLOS ONE. OSF; 2022. <https://doi.org/10.17605/OSF.IO/VTWN7>
73. Rybak A. SPSS syntaxes for NBI calculation PLOS ONE. OSF; 2022. <https://doi.org/10.17605/OSF.IO/JKE5H>
74. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021; n71. <https://doi.org/10.1136/bmj.n71> PMID: 33782057
75. Jabkowski P. Technical Report, Version 1.0. Comparative analysis of the quality of survey samples in the cross-national studies on the basis of external and internal criteria of representativeness: survey archiving and meta-base of results. National Science Centre: Poland. Grant No. 2017/01/X/HS6/01304; 2018. <https://doi.org/10.13140/RG.2.2.14548.24963>
76. PRISMA. PRISMA. Transparent Reporting of Systematic Reviews And Meta-Analyses. 2015. Available: <http://www.prisma-statement.org/>
77. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis.* Chichester: Wiley; 2009. [https://doi.org/10.1007/978-3-319-14908-0\\_2](https://doi.org/10.1007/978-3-319-14908-0_2)
78. Hartung J, Knapp G, Sinha BK. *Statistical Meta-Analysis with Applications.* Hoboken: Wiley; 2008. <https://doi.org/10.1002/9780470386347>
79. Cumming G. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* New York, London: Routledge; 2012. [https://doi.org/10.1111/j.1751-5823.2012.00187\\_26.x](https://doi.org/10.1111/j.1751-5823.2012.00187_26.x)
80. Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. *J Epidemiol Community Health.* 2013; 67: 974–978. <https://doi.org/10.1136/jech-2013-203104> PMID: 23963506
81. Baker WL, Michael White C, Cappelleri JC, Kluger J, Coleman CI. Understanding heterogeneity in meta-analysis: The role of Meta-regression. *Int J Clin Pract.* 2009; 63: 1426–1434. <https://doi.org/10.1111/j.1742-1241.2009.02168.x> PMID: 19769699

82. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003; 22: 2693–2710. <https://doi.org/10.1002/sim.1482> PMID: 12939780
83. Huizenga HM, Visser I, Dolan CV. Testing overall and moderator effects in random effects meta-regression. *Br J Math Stat Psychol*. 2011; 64: 1–19. <https://doi.org/10.1348/000711010X522687> PMID: 21506942
84. Viechtbauer W, López-López JA, Sánchez-Meca J, Marín-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychol Methods*. 2015; 20: 360–374. <https://doi.org/10.1037/met0000023> PMID: 25110905
85. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med*. 2004; 23: 1663–1682. <https://doi.org/10.1002/sim.1752> PMID: 15160401
86. Gooty J, Banks GC, Loignon AC, Tonidandel S, Williams CE. Meta-Analyses as a Multi-Level Model. *Organ Res Methods*. 2019; 24: 389–411. <https://doi.org/10.1177/1094428119857471>
87. Fernández-Castilla B. The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behav Res Methods*. 2020; 52: 2031–2052. <https://doi.org/10.3758/s13428-020-01373-9> PMID: 32162276
88. Cheung MWL. A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes. *Neuropsychol Rev*. 2019; 29: 387–396. <https://doi.org/10.1007/s11065-019-09415-6> PMID: 31446547
89. Van den Noortgate W, López-López JA, Marín-Martínez F, Sánchez-Meca J. Three-level meta-analysis of dependent effect sizes. *Behav Res Methods*. 2013; 45: 576–594. <https://doi.org/10.3758/s13428-012-0261-6> PMID: 23055166
90. Viechtbauer W. A Comparison of the rma() and the lm(), lme(), and lmer() Functions. 2021 [cited 4 Dec 2021]. Available: [https://www.metafor-project.org/doku.php/tips:rma\\_vs\\_lm\\_lme\\_lmer](https://www.metafor-project.org/doku.php/tips:rma_vs_lm_lme_lmer)
91. Fernández-Castilla B, Maes M, Declercq L, Jamshidi L, Beretvas SN, Onghena P, et al. A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behav Res Methods*. 2019; 51: 1286–1304. <https://doi.org/10.3758/s13428-018-1063-2> PMID: 29873036
92. Shih T-H, Fan X. Response rates and mode preferences in web-mail mixed-mode surveys: a meta-analysis. *Int J Internet Sci*. 2007; 2: 59–82.
93. Manfreda KL, Bosnjak M, Berzelak J, Haas I, Vehovar V, Bosnjak M, et al. Web surveys versus other survey modes: A meta-analysis comparing response rates. *Int J Mark Res*. 2008; 50: 79–104. <https://doi.org/10.1177/147078530805000107>
94. Church AH. Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opin Q*. 1993; 57: 62. <https://doi.org/10.1086/269355>
95. Sturgis P, Williams J, Brunton-Smith I, Moore J. Fieldwork Effort, Response Rate, and the Distribution of Survey Outcomes: A Multilevel Meta-analysis. *Public Opin Q*. 2017; 81: 523–542. <https://doi.org/10.1093/poq/nfw055>
96. Jabkowski P. A Meta-analysis of Within-Household Selection Impact on Survey Outcome Rates, Demographic Representation and Sample Quality in the European Social Survey. *Ask Res Methods*. 2017; 26: 31–60. <https://doi.org/10.18061/1811/81932>
97. Yammarino FJ, Skinner SJ, Childers TL. Understanding mail survey response behavior a meta-analysis. *Public Opin Q*. 1991; 55: 613–639. <https://doi.org/10.1086/269284>
98. Cehovin G, Bosnjak M, Lozar Manfreda K. Meta-Analyses in Survey Methodology: A Systematic Review. *Public Opin Q*. 2018; 82: 641–660. <https://doi.org/10.1093/poq/nfy042>
99. Jabkowski P. The Effect of Using Different Types of Survey Sample on Sample Execution in the European Social Survey. *Pol Sociol Rev*. 2018; 2: 265–278. <https://doi.org/10.26412/psr202.07>
100. Gaziano C. Comparative Analysis of Within-Household Respondent Selection Techniques. *Public Opin Q*. 2005; 69: 124–157. <https://doi.org/10.1093/poq/nfi006>
101. Rybak A. R script for replication PLOS ONE. OSF; 2023. <https://doi.org/10.17605/OSF.IO/QM5RU>
102. Bréchon P. A breakthrough in comparative social research: the ISSP compared with the Eurobarometer, EVS and ESS surveys. In: Haller M, Jowell R, Smith TW, editors. *The International Social Survey Programme 1984–2009 Charting the globe*. London and New York: Routledge; 2009.
103. Smith TW, Fu Y. The Globalization of Surveys. *The SAGE Handbook of Survey Methodology*. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd; 2016. pp. 680–692. <https://doi.org/10.4135/9781473957893.n43>
104. Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. 2010; 1: 112–125. <https://doi.org/10.1002/jrsm.11> PMID: 26061377
105. Viechtbauer W. Model Checking in Meta-Analysis. In: Schmid CH, Stijnen T, White I, editors. *Handbook of Meta-Analysis*. Boca Raton: CRC Press; 2020. pp. 219–254. Available: <https://www.taylorfrancis.com/books/9781498703994>

106. Wang MC, Bushman BJ. Using the Normal Quantile Plot to Explore Meta-Analytic Data Sets. *Psychol Methods*. 1998; 3: 46–54. <https://doi.org/10.1037/1082-989X.3.1.46>
107. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw*. 2010; 36: 1–48. <https://doi.org/10.18637/jss.v036.i03>
108. Rybak A. Software citations PLOS ONE. OSF; 2022. <https://doi.org/10.17605/OSF.IO/RX7Q5>
109. Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. *Public Opin Q*. 2004; 68: 2–31. <https://doi.org/10.1093/poq/nfh002>
110. Zillmann D, Schmitz A, Skopek J, Blossfeld HP. Survey topic and unit nonresponse: Evidence from an online survey on mating. *Qual Quant*. 2014; 48: 2069–2088. <https://doi.org/10.1007/s11135-013-9880-y>
111. Adua L, Sharp JS. Examining survey participation and response quality: The significance of topic salience and incentives. *Surv Methodol*. 2010; 36: 95–109.
112. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London etc.: Sage Publishers; 2012.
113. Lorah J. Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-Scale Assess Educ*. 2018. <https://doi.org/10.1186/s40536-018-0061-2>
114. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods*. 2011; 2: 61–76. <https://doi.org/10.1002/jrsm.35> PMID: 26061600
115. Cheung MWL. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods*. 2014; 19: 211–229. <https://doi.org/10.1037/a0032968> PMID: 23834422
116. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003; 327: 557–560. <https://doi.org/10.1136/bmj.327.7414.557> PMID: 12958120
117. Harrer M, Cuijpers P, A FT, Ebert DD. *Doing Meta-Analysis With {R}: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press; 2021.
118. Viechtbauer W. metafor: Meta-Analysis Package for R, v. 3.4–0. 2022.
119. Hox JJ, de Leeuw ED, Zijlman EAO. Measurement equivalence in mixed mode surveys. *Front Psychol*. 2015;6. <https://doi.org/10.3389/fpsyg.2015.00087> PMID: 25699002
120. West BT, Blom AG. Explaining interviewer effects: A research synthesis. *J Surv Stat Methodol*. 2017; 5: 175–211. <https://doi.org/10.1093/jssam/smw024>
121. Wagner J, Stoop IAL. Comparing Nonresponse and Nonresponse Biases in Multinational, Multiregional, and Multicultural Contexts. In: Johnson TP, Pennell B-E, Stoop IAL, Dorer B, editors. *Advances in Comparative Survey Methods*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2018. pp. 807–833. <https://doi.org/10.1002/9781118884997.ch37>

ADAM RYBAK 

Uniwersytet im. Adama Mickiewicza w Poznaniu

## KOMUNIKACJA OPARTA NA KANAŁE WIZUALNYM I AUDIALNYM A JAKOŚĆ POMIARU SONDAŻOWEGO. W STRONĘ NOWEGO UJĘCIA HIPOTEZY PREFERENCJI TECHNIKI PRZY WYKORZYSTANIU KONCEPCJI PAMIĘCI ROBOCZEJ

### Streszczenie

Celem artykułu jest przedstawienie potencjalnego kierunku rozwoju badań metodologicznych nad konstrukcją sondaży. W centrum rozważań postawiono kwestię projektowanej koncepcji integrującej zaobserwowane konsekwencje preferowania techniki sondażu przez respondentów (z uwzględnieniem wpływu sondażu na poziom realizacji badania i jakość uzyskanych odpowiedzi) przy wykorzystaniu kognitywistycznej koncepcji pamięci roboczej. Wskazano bowiem na istnienie potencjalnej relacji pomiędzy procesem starzenia się respondentów, czy odczuwanym przez nich lękiem, a różnicowaniem w sposobie przetwarzania informacji w modalności wzrokowej i słuchowej, zarysowując tym samym, jakie problemy stoją przed konstruującymi badania łączące różne techniki sondażu, między innymi w starzejących się populacjach zachodnich.

**Słowa kluczowe:** pamięć robocza, preferencja techniki sondażu, *satisficing*, metodologia badań sondażowych

---

Doktorant, Instytut Socjologii; e-mail: [adam.rybak@amu.edu.pl](mailto:adam.rybak@amu.edu.pl);  
<https://orcid.org/0000-0003-3981-1081>

## WSTĘP

Opozycja pomiędzy technikami sondażu, które wykorzystują aktywny udział osoby ankietera (np. CAPI, PAPI<sup>1</sup>), a tymi, które poprzestają na dostarczeniu materiałów tekstowych respondentowi, w większości pokrywa się z podziałem modalności sensorycznej komunikacji na audialną (czyli słuchową) i wizualną [por. np. Krzewińska, Grzeszkiewicz-Radulska 2013: 24–26]. Płynące z tego faktu konsekwencje przedstawione zostaną w artykule jako kontekst dla potencjalnych związków między hipotezą względnie stałych preferencji respondentów względem technik sondażu a psychologiczną koncepcją pamięci roboczej. Propozycje dotyczące potencjalnej obecności tych związków nie pojawiały się dotychczas w literaturze światowej. Już na samym wstępie należy jednak podkreślić, że nawet w sytuacji rzeczywistej obecności relacji pomiędzy wspomnianymi czynnikami psychologicznymi a preferencją techniki nie są one jedynym źródłem różnicowania preferencji.

Dla porządku należy tu wtrącić uwagę dotyczącą wymienionej na początku artykułu opozycji między komunikowaniem ustnym i pisemnym w badaniu sondażowym. Jak twierdzi Franciszek Sztabiński, w przeciwieństwie do sytuacji standaryzowanego wywiadu kwestionariuszowego, kiedy dochodzi do bezpośredniego spotkania, ankieter przekazuje badanemu pytania badacza i zbiera jego odpowiedzi – „dialog [pisemny] [...] jest [...] rozciągnięty w czasie i przestrzeni. Nie zachodzi między osobami znajdującymi się w jednym punkcie czasoprzestrzennym, które mogą się wzajemnie obserwować, oddziaływać na siebie pozawerbalnymi środkami przekazu, lecz między nadawcą i odbiorcą, którzy oddzieleni są od siebie dystansem czasowym, przestrzennym oraz sytuacyjnym” [Sztabiński 1990: 62]. Sam ten fakt niesie za sobą liczne konsekwencje. Jedną z nich może być większa tendencja do formułowania przez respondentów w sytuacji interakcji z ankieterem wymuszonych okolicznościami i presją czasu sądów *ad hoc* – w opozycji do potencjalnie opartych na „społecznym zakotwiczeniu” odpowiedzi generowanych w mniej naglących okolicznościach ankiety pocztowej czy internetowej [Sztabiński 1997: 86].

W pierwszej kolejności przedstawię różne propozycje teoretycznego umocowania samego zjawiska preferowania techniki sondażu, szczególnie na tle istniejących teorii partycypacji w sondażu. W dalszej części opiszę sposoby badania *mode preference* i ich konsekwencje dla analiz. Następnie zarysuję

---

<sup>1</sup> *Computer Assisted Personal Interviewing, Paper and Pencil Interviewing* (ang.) – techniki standaryzowanego wywiadu kwestionariuszowego ze wspomaganie komputerowym oraz bez wspomaganie.

problematykę jakości odpowiedzi sondażowych, odwołując się do koncepcji *satisficingu*. Jej wartość w mojej opinii polega na potencjalnym integrowaniu kwestii błędu braku odpowiedzi (przynajmniej w jego komponencie pochodzącym z braku kooperacji) i błędu pomiaru w sondażu. Wreszcie przedstawię teorię pamięci roboczej i zaobserwowane przykłady zmian jej efektywności w celu zaprezentowania potencjalnego związku pomiędzy nią a zjawiskami *satisficingu* i preferencji techniki sondażu. Etapowe opisywanie kolejnych pojęć umożliwi lepsze ukazanie mojej propozycji w kontekście konstrukcji teorii i przyszłych badań.

### TEORETYCZNE RAMY ANALIZY HIPOTETYCZNEGO ZJAWISKA PREFEROWANIA TECHNIKI SONDAŻU

Autorami, którzy po raz pierwszy *explicite* podjęli temat potencjalnego istnienia wśród badanych względnie trwałych preferencji dotyczących jednej techniki sondażowej kosztem innych, byli Robert Groves i Robert Kahan [1979]. Problemem tym zajęli się, analizując badania porównujące udział w sondażach bezpośrednich i telefonicznych w warunkach amerykańskich. Jednak należy wspomnieć, że jednocześnie to właśnie ci badacze wskazali jedną z podstawowych trudności związanych z tą hipotezą, na którą zwracali uwagę również jej krytycy. Mianowicie dostrzegli, że respondenci mają tendencję do wskazywania jako preferowanej tej techniki, za pośrednictwem której uczestniczyli w badaniu. Jednak zjawisko to, tak jak i inne związane z *mode preference*, nie zostało zadowalająco wyjaśnione, gdyż, choć hipoteza obecna jest w nauce od czterech dekad, wciąż brakuje jej adekwatnej struktury pojęciowej i wyczerpujących analiz empirycznych [Olson, Smyth, Wood 2012: 612]. Już we wspomnianym pierwszym badaniu wskazano (uzyskane od respondentów) uzasadnienia preferowania przez nich konkretnej techniki – w przypadku wywiadu kwestionariuszowego respondenci podkreślali przyjemność płynącą z kontaktu międzyosobowego, w przypadku badań telefonicznych – ich szybki i niewymagający zaangażowania charakter, zaś w przypadku technik pocztowych – brak pośpiechu i możliwość przemyślenia odpowiedzi. Choć hipoteza preferencji budzi pewne zastrzeżenia, to jednak istnieją znaczące argumenty na rzecz jej przyjęcia. Przykładowo cytowane wyżej Kristen Olson, Jolene Smyth i Heather Wood [2012], opierając się na dwóch rundach sondażu *Nebraska Annual Social Indicators*, stwierdziły istnienie istotnego pozytywnego wpływu czynnika otrzymania propozycji udziału w sondażu w zadeklarowanej rok wcześniej technice na prawdopodobieństwo partycypacji. Pierwsza runda z 2008 roku została przeprowadzona techniką wywiadu telefonicznego i za-

wierała pytanie o to, jaką technikę preferowaliby respondenci w przypadku udziału w przyszłych badaniach sondażowych. W drugiej rundzie respondenci zostali przyporządkowani do technik odgórnie – z losowym podziałem na grupy, w których technika była zgodna z wcześniejszą deklaracją lub odmienna. To samo badanie posłużyło także [Smyth, Olson, Kasabian 2014] do stwierdzenia istotnego negatywnego związku między byciem badanym preferowaną techniką a wykorzystywaniem możliwości wynikających z konkretnej konstrukcji kwestionariusza do wypełniania go „na chybcika” (ang. *satisficing* – to trudno przetłumaczalne pojęcie zostanie omówione w dalszej części artykułu).

Poszukując teoretycznych ram podjętej tu analizy, warto dokonać skrótowego przeglądu literatury przedmiotu. Przydatne w pierwszym rzędzie wydaje się sięgnięcie do dorobku metodologów sondażu, którzy stworzyli (niekiedy *implicite*) teorię partycypacji w sondażu. Poniżej przedstawię koncepcje Roberta Grovesa, Dona Dillmana i Rogera Tourangeau (w roli reprezentanta ruchu Cognitive Aspects of Survey Methodology), jako że są one dosyć rozpowszechnione i jednocześnie posiadają (zróżnicowany) potencjał budowania teorii *mode preference*. Następnie uzupełnię to zestawienie o bardziej robocze koncepcje, za to dotyczące już ściśle kwestii preferowania techniki sondażu.

Według Roberta Grovesa i Micka Coupera [1998] cechy i doświadczenia respondenta mogą być jedynie pośrednim źródłem dyspozycji psychicznych respondenta do udziału w sondażu – tak więc proces przewidywania na ich podstawie preferowanej przez daną osobę techniki wydaje się problematyczny, choć zapewne w jakimś stopniu wykonalny. Należy podkreślić, że dla większości respondentów udział w sondażu jest doświadczeniem względnie nowym i/lub rzadkim. Nie mogą oni więc odwołać się do swoich doświadczeń dotyczących badania w ogóle czy w konkretnej jego formie (np. proponowanej techniki). W swoich rozważaniach autorzy odwołali się do teorii skryptów poznawczych Roberta Cialdiniego [1984]. Podkreślono w niej szczególnie istotną dynamiczną rolę ankietera, który, wykorzystując swoją socjologiczną wrażliwość, dostosowuje swoje zachowania, argumenty czy będące w jego dyspozycji nagrody do spodziewanych oczekiwań jednostki. Od razu można zauważyć, że powoduje to trudność w porównywaniu za pomocą tej koncepcji techniki wywiadu kwestionariuszowego z innymi, w których nie ma kontaktu z ankieterem „oko w oko”. Ogranicza to jej stosowalność względem sondaży prowadzonych w różnych technikach. Groves wspomina o możliwościach uprzedniego skierowania na przykład szczególnych zachęt do konkretnych kategorii respondentów, co do których można przypuszczać, że będą mniej zainteresowani partycypacją. Ankieter może też zaproponować techniki badania bardziej odpowiadające preferencjom

respondenta [Groves, Singer, Corning 2000]. Nie są to dla niego jednak kwestie centralne.

D. Dillman, budując swoją koncepcję partycypacji w sondażu [Dillman, Smyth, Melani 2014], odwołuje się do teorii wymiany, którą można streścić w zdaniu, że jednostka wtedy podejmie konkretne działanie, gdy będzie żywiła przekonanie, że będące jego efektem zyski przeważą straty. Należy tu jednak zauważyć, że nie oznacza to traktowania jednostkowej kalkulacji jako w ścisłym sensie racjonalnej – zyski i straty mogące płynąć z udziału w badaniu są niedookreślone, niepewne, oparte na wierze (zaufaniu), często odłożone na przyszłość. Także krótki czas na podjęcie decyzji – choć jego długość może wahać się od błyskawicznej odpowiedzi w badaniu telefonicznym przez szybki namysł podczas odwiedzin ankietera aż do potencjalnie dłuższego zastanowienia w badaniach internetowych czy pocztowych (choć i tu decyzja często będzie podjęta już przy odsiewaniu ulotek reklamowych od ważnych listów ze skrzynki) – nie sprzyja w pełni racjonalnej decyzji. W ujęciu Dillmana proces planowania badania winien polegać na takim dopasowaniu form jego realizacji do poszczególnych kategorii respondentów, by zwiększyć ich zaufanie do badacza/ankietera, zwiększyć spodziewane przez nich zyski i zmniejszyć spodziewane koszty. W tym ujęciu poszukiwanie optymalnego dopasowania pomiędzy technikami sondażu i stawianymi przez nie wymaganiami a cechami, zasobami (np. kompetencjami językowymi, dobrym wzrokiem itd.) i otoczeniem (np. liczbą osób w gospodarstwie domowym – gromadka hałaśliwych dzieci lub cichy współlokator) respondenta, tak by konkretna propozycja udziału w badaniu jawiła mu się jako najmniej kosztowna, najwygodniejsza, zdaje się być szczególnie istotne. Zgodnie z tą intuicją D. Dillman w najnowszej edycji swojego słynnego podręcznika prowadzenia badań sondażowych zaleca (ostrożne) korzystanie z informacji dotyczących rozkładu preferencji techniki w projektowaniu przyszłych badań [Dillman, Smyth, Christian 2014: 424–427]

CASM to zapoczątkowany w późnych latach 70. serią konferencji współorganizowanych przez amerykańską National Research Council, brytyjskie Royal Statistical Society and Social Science Research Council i zachodnioniemieckie Zentrum für Umfragen, Methoden und Analysen ruch na rzecz badania procesów poznawczych zachodzących w trakcie udziału w sondażu [Krzewińska 2006: 57]. Wpisujący się w ten nurt R. Tourangeau uznaje, że najważniejsze elementy różnicujące sondaże prowadzone przy użyciu poszczególnych technik i przez nie wywoływane to: trudność poznawcza (ang. *cognitive burden*), bezosobowość (ang. *impersonality*) i poczucie uprawnomożenia (ang. *legitimacy*) [Tourangeau, Rips, Rasinski 2000]. Ostatni z aspektów nie pokrywa się z analizowaną tutaj

opozycją między modalnością wizualną i słuchową. Do podkreślenia powagi instytucji i wartości badania najlepiej nadają się odpowiednio przygotowani i wyposażeni ankieterzy i oficjalne listy, najmniej – kontakt telefoniczny (poruszając ten problem przed niemal 20 laty, autor sugerował, że używanie nowoczesnych technik komputerowych jest w jakimś stopniu prestiżowe, dziś natomiast istnieje chyba większe prawdopodobieństwo, że komunikaty przesyłane drogą elektroniczną zostaną przez respondenta zaklasyfikowane jako spam). Bezosobowość badania jest natomiast bezpośrednio związana z pojęciem pytań czy tematów drażliwych. Z jednej strony najczęściej za najbardziej bezosobowe w tym ujęciu uznaje się techniki wykorzystujące komunikację za pomocą tekstu (głównie technikę pocztową i internetową), jednak należy uwzględnić trudności, jakie może napotkać respondent podczas zapoznawania się z przesłanymi materiałami i ich wypełniania, na przykład jeśli mieszka na niewielkiej przestrzeni z dużą liczbą innych osób. Dodatkowo może się on spotkać z naciskiem z ich strony, by pokazał im otrzymaną korespondencję (papierową lub elektroniczną). W takiej sytuacji ankieter, podkreślając konieczność zachowania poufnego charakteru badania, może na przykład skłonić inne osoby do opuszczenia mieszkania/pokoju, co stworzy poczucie większej bezosobowości. Drugim aspektem zagadnienia bezosobowości, który należy uwzględnić przy projektowaniu badania, jest fakt, że różne tematy mogą być drażliwe dla różnych grup – na przykład pytanie o aborcję może być bardziej drażliwe dla kobiet (jako że zjawisko to dotyczy ich bezpośrednio) niż dla mężczyzn, osoby starsze mogą mniej chętnie mówić o używkach czy zachowaniach seksualnych niż osoby młodsze. Dlatego też należy podkreślić, że specyficznie definiowana przez Tourangeau bezosobowość badania może być różna dla różnych kategorii respondentów.

Najistotniejszy dla prowadzonych tu rozważań wydaje się temat trudności poznawczych, które mogą wynikać z wymagań konkretnej techniki i poznawczych zdolności respondenta. Tourangeau zwraca szczególną uwagę na biegłość badanych w mowie i piśmie. Według danych z 1993 roku, na które się powoływał, 1/4 Amerykanów ma bardzo ograniczoną sprawność w posługiwaniu się językiem angielskim. Dzisiaj temat ten zdaje się szczególnie istotny ze względu na bardzo duże nasilenie procesów migracyjnych związanych z kryzysami w krajach Trzeciego Świata. Pomocny w rozwiązywaniu problemów z przetworzeniem informacji i sformułowaniu odpowiedzi w danym języku w mowie lub w piśmie może być odpowiednio przygotowany, wspomagający ankieter (możliwe jest na przykład skorzystanie z pomocy ankieterów znających inne języki, obecnych w studiu CATI, którzy nawiązują kontakt z badanymi po stwierdzeniu problemów z komunikacją); z kolei kwestionariusz papierowy lub elektroniczny może

dawać osobom mającym problemy z komunikacją więcej czasu do namysłu i stwarzać mniejsze poczucie presji. W tym przypadku jednak istotną przeszkodę może stanowić z kolei brak kompetencji cyfrowych w przetwarzaniu informacji elektronicznych i odpowiednim wykorzystaniu interfejsu.

Część badaczy tworzy mniej rozbudowane propozycje czynników, które jednak w odróżnieniu od powyższych koncepcji w pełni odnoszą się do problematyki *mode preference*. Według Smyth, Olson i Millar [2014] samotność, depresja, niska samoocena czy obawa o anonimowość ze strony respondenta mogą być predyktorami preferowania technik sondażu niezakładających kontaktu z ankietą – ze względu na możliwość łatwiejszej kontroli nad tym, co jest ujawniane. Kieren Diment i Sam Garrett-Jones [2007] sugerują, że większy entuzjazm wobec korzystania z Internetu w ogóle przejawiany przez mężczyzn może wpływać na to, że częściej wybierają oni internetową technikę sondażu. Verma i inni [2014] podkreślają, że osoby pochodzące z dyskryminowanych mniejszości etnicznych mogą wykazywać mniejszą chęć uczestnictwa w badaniach internetowych. Po bieżne traktowanie problematyki preferowania techniki sondażu przez najbardziej rozbudowane koncepcje teoretyczne i fragmentaryczność tych koncepcji, które dotyczą jej w ścisłym sensie, uzmysławia, że adekwatna teoria zjawiska *mode preference* jest jeszcze przed nami.

### SPOSOBY BADANIA *MODE PREFERENCE*

Powyżej zarysowane zostały (wciąż niewystarczające) podstawy, które jednak w swym aspekcie teoretycznym mogą stanowić punkt wyjścia dla analiz preferowania techniki sondażu przez respondentów. Dla rozjaśnienia tej problematyki konieczne jest przedstawienie również empirycznych możliwości jej badania. Poniżej opiszę obecne w literaturze przedmiotu podejścia do badania preferowanej przez respondentów techniki sondażu, a w dalszej kolejności przywołam przykłady badań, w których je zastosowano. Pierwszą z metod jest zadanie respondentom w trakcie badania prostego pytania o technikę, którą – zgodnie ze swoimi przewidywaniami – będą preferowali w przyszłości, co wymaga przyjęcia założenia o względnej stabilności tej preferencji. Trudno jednak tak naprawdę zakładać dużą stabilność czasową, skoro, jak się niżej przekonamy, istotnym predyktorem preferencji może być na przykład wiek. Jak już wspomniałem wyżej, taki sposób postępowania powoduje problem w postaci wyraźnie większej skłonności do wskazywania jako preferowanej tej techniki, w której się uczestniczy. Może to wynikać z różnych przyczyn, na przykład z efektu selekcji (osoby preferujące w sposób szczególny inne techniki niż ta stosowana w badaniu odmówiły w nim

udziału) albo też z efektu uprzejmości względem ankietera w wywiadzie bezpośrednim lub telefonicznym. Kolejne założenie związane z tym podejściem jest takie, że jednostka jest na tyle świadoma swoich preferencji, że może udzielić kompetentnej odpowiedzi na pytanie o nie. Ten sposób badania *mode preference* może być rozszerzony poprzez dodanie drugiej tury, w której (po odpowiednio długim czasie) respondenci będą losowo przydzielani bądź nieprzydzielani do technik preferowanych, co umożliwi oszacowanie wartości predykcyjnych uzyskanych wcześniej deklaracji. To podejście zostało wykorzystane w przywoływanym we wstępie badaniu opartym na sondażu NASIS.

Inną strategią badania preferowanej przez respondentów techniki sondażu jest umożliwienie im wyboru techniki odpowiedzi – przykładowo poprzez zamieszczenie adresu internetowego w kwestionariuszu pocztowym, podanie go podczas screeningu w CATI czy umożliwienie respondentowi za pośrednictwem ankietera własnoręcznego wypełnienia kwestionariusza. Zmiana sposobu badania z prośby o deklarację na obserwację zachowania eliminuje problem ewentualnej niezgodności między deklaracją co do preferowanej techniki a przyszłym zachowaniem w tym względzie. Jednak może mieć też wady. Po pierwsze, zazwyczaj tylko częściowo znosi problem związany z metodą pierwszą, dotyczący wspomnianego efektu selekcji. Choć możemy sobie wyobrazić sytuację, w której pierwszy kontakt nawiązywany jest jednocześnie w kilku technikach, to w praktyce badawczej zazwyczaj dzieje się inaczej – jest wykorzystywana tylko jedna. A mniej prawdopodobne może być właśnie już samo nawiązanie kontaktu z respondentem i skupienie jego uwagi przez badacza za pośrednictwem kanału, którego respondent nie preferuje. Jeśli niechęć do badań *face-to-face* wynika u kogoś ze strachu przed obcymi, szansa na zaoferowanie mu przez ankietera kwestionariusza pocztowego będzie mniejsza (gdyż na przykład może nie otworzyć mu drzwi). Jeśli brak czasu (lub odległość do urzędu pocztowego czy skrzynki) zniechęca go do udziału w badaniach pocztowych – może nie doczytać w liście, że istnieje alternatywa w postaci badania internetowego. Po drugie, być może bardziej istotne, danie respondentowi wyboru według większości eksperymentów metodologicznych zmniejsza wskaźniki realizacji badania [por. np. Medway, Fulton 2012]. Przyczyny tego zjawiska nie są wyjaśnione. Może ono wynikać na przykład z faktu, że konieczność porównania różnych możliwości powoduje uświadomienie sobie kosztów związanych z każdą z technik.

Wreszcie mniej popularne opcje – możliwa jest metaanaliza cech składu respondentów biorących udział w sondażach prowadzonych różnymi technikami, jednak wymaga ona konstruowania bardzo podobnych badań przeprowadzanych na bardzo podobnych populacjach w celu zapewnienia ich porównywalności.

Z tego powodu jej stosowanie w praktyce może być dość trudne. Możliwe jest także porównanie deklaracji i wyborów – na wzór metody wymienionej na początku tej części artykułu – ale w dwóch bardzo zbliżonych badaniach.

Przykładem zastosowania pierwszego sposobu badania preferencji jest wspomniany wyżej sondaż telefoniczny NASIS 2008. J. Smyth, K. Olson i M. Millar [2014] zbudowały modele wielomianowej regresji logistycznej analizujące wpływ poszczególnych cech respondentów na deklarowane przez nich preferencje. Pozytywny istotny wpływ na wskazywanie modalności wizualnej jako preferowanej względem audialnej miały: dla kwestionariusza pocztowego – płeć żeńska i strach przed kradzieżą, dla ankiety internetowej – wyższe wykształcenie i korzystanie wcześniej z Internetu. Istotnie negatywny wpływ na wybór modalności wizualnej w stosunku do audialnej miały: dla ankiety internetowej – niskie zdolności komputerowe, dla obu technik – wyższy wiek. Wiek jest w tym modelu istotnym czynnikiem w preferowaniu technik audialnych przy kontroli zdolności komputerowych i doświadczenia z Internetem. Wpływa także negatywnie na wskazywanie obu technik wizualnych, należy więc przypuszczać, że zmienna ta nie dotyczy efektu kohortowego związanego z dorastaniem w innych warunkach technologicznych, a pozostaje w relacji do jakichś efektów procesu starzenia.

Przykładem drugiego podejścia jest badanie przeprowadzone w Holandii na podstawie danych z *European Social Survey* 2010 przez Marieke Haan, Yfke Onega i Keesa Arts [2014]. W próbie celowo nadreprezentowane były kategorie respondentów uważanych za trudno dostępnych – to jest osoby w wieku 15–35 lat, mieszkańcy dużych miast, osoby pracujące na pełen etat i emigranci spoza Europy Zachodniej. Z wylosowanymi jednostkami kontaktował się ankietier proponujący im udział w badaniu CAPI lub internetowym. W analizie Chi-kwadrat istotne okazały się różnice między odsetkiem respondentów wybierających technikę wspomaganych komputerowo wywiadów standaryzowanych i tych wybierających kwestionariusz internetowy. Osoby poniżej 35 roku życia częściej wybierały badanie wizualne niż osoby starsze od nich, także pracujący na pełny etat wybierali je częściej niż zatrudnieni w niepełnym wymiarze godzin i niezatrudnieni.

## JAKOŚĆ ODPOWIEDZI

Błąd wynikający ze sposobu, w jaki respondent odpowiada na pytania, to błąd pomiaru, który może być związany z respondentem, ankietierem czy konstrukcją narzędzia badawczego. Sytuuje się w opozycji do błędów pokrycia, próbkowania, braku odpowiedzi (i tych wynikających z niewłaściwego przetwarzania informacji – na przykład pomyłek w kodowaniu odpowiedzi). Błąd pomiaru możemy

rozumieć na dwa sposoby. Statystyczne ujęcie definiuje go jako odchylenie wyników badania od pewnej wartości „prawdziwej”, natomiast psychometryczne – jako „zanieczyszczenie” wyniku pomiaru spowodowane wpływem jakiegoś czynnika. W tym rozumieniu „prawda”, do której chce się zbliżyć badanie, jest niemierzalnym konstruktem teoretycznym [Jabkowski 2015].

Analizę jakości odpowiedzi udzielanych przez respondentów, która ma kluczowy wpływ na jakość pomiaru, można przeprowadzić na trzy sposoby. Po pierwsze, jeśli posiadamy dostęp do danych na temat wartości prawdziwych badanych zmiennych – na przykład dostęp do dokumentów księgowych w pytaniach o zarobki, możemy dokonać walidacji przez porównanie rozkładu zmiennej osiągniętego w badaniu i znanego z dokumentów. Wymaga to jednak kontroli osobnego wpływu braków odpowiedzi na rozkład – jeśli błąd płynący z niepełnej realizacji okaże się błędem systematycznym. Ponadto warto zwrócić uwagę na istotną trudność, która się z tym wiąże, otóż taką praktyczną możliwość kontroli posiadamy niezwykle rzadko – ze względu na brak dostępu do poufnych danych (gdybyśmy go mieli, celowość przeprowadzania sondażu byłaby ograniczona) oraz ze względu na charakter mierzonych zmiennych – które często nie dotyczą empirycznie obserwowalnych zjawisk czy obiektów, a teoretycznych konstruktów [Sztabiński 2011: 85–90].

Łatwiejsza do przeprowadzenia jest kontrola udzielania odpowiedzi na pytania. Często porównuje się liczbę braków odpowiedzi na poszczególne pytania (ang. *item-nonresponse*); respondent może nie udzielić odpowiedzi z różnych przyczyn – z powodu zmęczenia, braku motywacji, niezrozumienia pytania, niezapewnienia mu poczucia prywatności i innych. Porównywać można także ilość treści w odpowiedziach na pytania otwarte. Bardziej złożoną koncepcją jest tzw. *satisficing*<sup>2</sup> [Krosnick 1991] – zjawisko związane z przyjmowaniem przez respondenta najmniej męczącej dla niego strategii udziału w badaniu. Może się to przejawiać na przykład udzielaniem identycznych odpowiedzi w linii pionowej w zbiorczych tabelkach odpowiedzi kwestionariuszowych, zaznaczaniem wszystkich dostępnych opcji, albo tylko pierwszej lub ostatniej, szczególnie częstym odpowiadaniem „nie wiem”, ale również brakiem odpowiedzi na pytanie. Ciekawy, szczególnie w kontekście porównywania technik sondażu wykorzystujących modalność wizualną i audialną, jest efekt pierwszeństwa/świeżości (*primacy/recency*). Zauważono bowiem skłonność do częstszego wskazywania pierwszej

---

<sup>2</sup> Jan Krosnick zapożyczył to pojęcie od ekonomisty Herberta Simona. Oryginał angielski „jest zlepkiem słów *to satisfy*, które oznacza ‘zaspokajać’ oraz *to suffice*, znaczące ‘wystarczać’” i tym samym jest trudny do przetłumaczenia [Meyer 2007: 113].

odpowiedzi w kwestionariuszu, który mamy przed oczami i ostatniej w wysłuchiwanej kafeterii niezależnie od ich treści. Jest to często interpretowane jako jedna ze strategii (z punktu widzenia badacza natomiast jako jeden ze wskaźników) *satisficingu*, która może być związana z innym sposobem przetwarzania w pamięci informacji wzrokowych i słuchowych przy ograniczeniu zasobów pamięciowych.

Trzecią drogą, która ze względu na swoją złożoność i problematyczne założenia przekracza zakres tego artykułu, jest metoda eksperymentalna *Split-Ballot Multitrait–Multimethod Designs* (analizy „wielu cech – wielu metod”), w której porównujemy odpowiedzi na pytania powtarzające się w obrębie jednego kwestionariusza i kilku technik. Wykorzystywane jest do tego również modelowanie równań strukturalnych (*SEM*) [por. np. Vannieuwenhuyze, Revilla 2013].

Przykładem zastosowania walidacji są badania Josepha Sakshauga, Yan Tinga i wspomnianego Rogera Tourangeau [2010]. Badacze mieli dostęp do danych Uniwersytetu Maryland, co umożliwiło im nie tylko wylosowanie reprezentatywnej próby badawczej absolwentów, ale również porównanie udzielanych przez nich odpowiedzi na temat osiągnięć uniwersyteckich i udziału w formach wsparcia uniwersytetu przez absolwentów ze stanem faktycznym. Autorzy podzielili zmienne na społecznie wrażliwe pozytywne, związane z dobrymi wynikami i wsparciem uczelni, wrażliwe negatywne, związane z niskimi wynikami czy niezaliczeniami przedmiotów i neutralne – kontrolne. Operat umożliwił zastosowanie kilku technik badawczych. Porównywano więc odpowiedzi osób losowo przypisanych do CATI, *Interactive Voice Response* (badania prowadzonego przez automat telefoniczny w celu zwiększenia poczucia poufności odpowiedzi) i kwestionariuszy internetowych. Okazało się, że Internet istotnie poprawił jakość pomiaru (prawdziwość odpowiedzi), ale kosztem zwiększenia liczby braków odpowiedzi. Niestety autorzy nie uwzględnili zjawiska preferencji techniki sondażu. Zrobiły to natomiast we wspomnianym już badaniu J. Smyth, K. Olson i A. Kasabian, porównując stopień wykorzystywania oferowanych przez konstrukcję kwestionariusza zachęt do *satisficingu* wśród osób, które zostały przyporządkowane do techniki zgodnie ze zgłoszoną wcześniej preferencją lub niezgodnie z nią. Różnice w budowie kwestionariuszy polegały na stosowaniu na przykład poleceń typu „zaznacz wszystkie odpowiedzi prawdziwe” zamiast przedstawiania listy odpowiedzi „tak/nie”, tabel z odpowiedziami opisanymi tylko w nagłówkach kolumn zamiast takich, w których byłaby opisana każda kratka oraz używaniu małych pól na odpowiedzi otwarte zamiast pól dużych. Każda z pierwszych wersji konstrukcji tych pytań ułatwiała *satisficing* i być może także skłaniała do niego. Drugie wersje konstrukcji utrudniały takie podejście. Wykazano pozytywne ko-

relacje pomiędzy brakiem preferowanej techniki a wykorzystywaniem okazji do *satisficingu*. Jednak trzeba zaznaczyć, że autorki porównywały wyłącznie badania w technikach opartych na kanale wizualnym. Widzimy więc, że dopiero analiza zestawiająca cechy obu opisanych badań, a więc uwzględnienie *mode preference* i obu modalności (wzrokowej lub słuchowej) technik sondażowych, a najlepiej wykorzystująca w zależności od rodzaju zmiennej walidację lub ocenę jakości odpowiedzi, byłaby szczególnie użyteczna dla zgłębianej tutaj problematyki.

Jakie mogą być więc czynniki pogarszające jakość odpowiedzi? Słaba pamięć powoduje błędy w przywoływaniu właściwych odpowiedzi w pytaniach o fakty z przeszłości. Efekt ankierski i pytania drażliwe mogą powodować niechęć respondenta do odpowiadania zgodnie z faktami lub własnymi przekonaniem. Wreszcie najbardziej interesujący, bo będący pojemną koncepcją teoretyczną *satisficing* jest przedstawiany przez Jona Krosnicka [1991] jako zależny z jednej strony od trudności zadania (1), a z drugiej od zdolności (2) i motywacji (3) respondenta. Widzimy, że w takim ujęciu w pewnym zakresie możliwa jest interpretacja dwóch pierwszych elementów jako *cognitive burden* i *cognitive abilities* R. Tourangeau. Trzeci może być związany z kwestią odpowiadania w preferowanej technice – przynajmniej w odniesieniu do koncepcji D. Dillmana, gdyż autor ten podkreśla podejmowanie świadomej decyzji o udziale w badaniu przez respondenta na podstawie ogólnikowego rachunku zysków i strat – a więc preferowana technika wiąże się z bardziej pozytywnym bilansem i jest bardziej motywująca. Idąc za intuicją Tourangeau, proponuję uznać czynnik trudności poznawczej za wpływający nie tylko na jakość odpowiedzi, ale i na decyzję o partycypacji – najskrajniejszą formą zmniejszania wysiłku respondenta przez *satisficing* jest bowiem odmowa kooperacji. Wywołuje to dwa dodatkowe problemy. Pierwszym z nich jest kwestia uświadomienia sobie własnej preferencji. Przywoływane w tym artykule badania sugerują, że preferencja techniki jest przynajmniej częściowo uświadomiona, ale nie jest to do końca pewne. Uświadomienie sobie preferencji przez respondenta rozumiem jako zgodność pomiędzy skłonnością jednostki do wybierania raczej techniki A niż B, C, D... w sytuacji postawienia jednostki przed wyborem techniki a wiedzą jednostki o tym, jaką decyzję podjęłaby, gdyby umożliwiono jej wybór technik A, B, C, D..., za pośrednictwem których może wziąć udział w badaniu. Wiedza ta umożliwia jednostce udzielenie odpowiedzi na pytanie o preferencję.

Sprawa komplikuje się bardziej w odniesieniu do drugiego problemu – formy pierwszego kontaktu z respondentem. Jeśli jest ona tożsama ze sposobem zbierania danych (pod względem kanału komunikacji), to już krótki kontakt może

wystarczyć respondentowi do uświadomienia sobie czekających go trudności (między innymi poznawczych). Jeśli natomiast pierwszy kontakt następuje za pomocą innej techniki niż zbieranie danych (jak może się zdarzyć na przykład w przypadku sondażu używającego równoległego schematu mieszania technik z wyborem pozostawionym respondentowi), efekty związane ze stosunkiem respondenta do obu tych technik mogą się nakładać w nieznany sposób. Mimo tych wątpliwości chciałbym posłużyć się przywołanymi dotychczas koncepcjami w celu zestawienia ich z używanym w obszarze psychologii poznawczej pojęciem pamięci roboczej.

### PAMIĘĆ ROBOCZA A UDZIAŁ W SONDAŻU

Pamięć robocza lub operacyjna (ang. *working memory*) to koncepcja stworzona jako rozwinięcie wcześniejszej teorii pamięci krótkotrwałej. Postulowany system miał być odpowiedzialny nie tylko za przechowywanie danych, jak pamięć krótkotrwała, ale umożliwiał także przetwarzanie i manipulowanie ograniczoną ilością informacji. Najpopularniejszą współcześnie teorią pamięci roboczej jest tzw. Model Baddeleya, choć nie brak i znaczących teorii konkurencyjnych – na przykład pamięci roboczej traktowanej jako aktywna sieć w obrębie pamięci długotrwałej w modelu Cowana. Model Baddeleya – w swojej najprostszej wersji – składa się z 3 powiązanych ze sobą, ale działających odrębnie elementów: centralnego systemu wykonawczego, pętli fonologicznej i szkicownika wzroko-wo-przestrzennego. Model był później wielokrotnie rozwijany, między innymi poprzez dodanie dodatkowego elementu – bufora epizodycznego, czy dokonanie podziału pętli i szkicownika na struktury retencyjne i aktywnego podtrzymywania<sup>3</sup> [Piotrowski i in. 2009]. System centralny pełni w omawianym modelu funkcje koordynacyjne, zarządzając procesami przetwarzania informacji przez różne formy pamięci, ale nie ingerując w nie bezpośrednio. Pętla fonologiczna odpowiadać ma za przechowywanie i przetwarzanie informacji fonologicznych, szkicownik – tych dotyczących właściwości wizualnych i przestrzennych obiektów. System ten jest wykorzystywany do takich czynności, jak uczenie się, argumentowanie, planowanie, tym samym uczestniczy on także w procesach umysłowych zachodzących podczas podejmowania decyzji o uczestnictwie w sondażu i tych w trakcie uczestnictwa w nim.

---

<sup>3</sup> Zagadnienia te w tym miejscu jedynie sygnalizuję, zainteresowanych czytelników odsyłam do literatury kognitywistycznej.

Przyporządkowanie poszczególnych technik według ich modalności (wzrokowej lub słuchowej) do poszczególnych podsystemów nie jest jednak tak proste, jak mogłoby się początkowo zdawać. O ile w technikach sondażu wykorzystujących komunikację werbalną przez respondenta używana jest przede wszystkim pętla fonologiczna (jeśli skupimy się na projektowanej części badania, a usuniemy poza nawias niewerbalne sygnały, które mogą być wysyłane przez ankietera) – do przetwarzania pytań i formułowania odpowiedzi ustnych, o tyle w sondażu wykorzystującym informacje tekstowe jest to bardziej złożone. Takie czynności, jak czytanie [Pham, Hasson 2014], układanie zdań pisemnych [Olive, Kellogg, Piolat 2008] czy proste czynności arytmetyczne [Clearman, Klinger, Szűcs 2017], związane z odpowiedziami dotyczącymi dochodów, wydatków czy częstości zachowań w danym okresie, angażują oba podsystemy – pętlę i szkicownik. Może to prowadzić do hipotezy, że osłabienie funkcjonowania szkicownika spowodowane jakimś czynnikiem, przy stałym poziomie działania pętli, może prowadzić do utrudnienia udziału w sondażu w technikach opartych na komunikacji w kanale wzrokowym. Tym samym zwiększy to preferencję względem technik opartych na komunikacji w kanale słuchowym. Wynika to z potraktowania konieczności przetwarzania informacji w danej modalności jako *cognitive burden*, a poziomu funkcjonowania subsystemu pamięci roboczej jako *cognitive ability* w ujęciu Tourangeau. Zgodnie z modelem Krosnicka skutkiem niedostosowania wartości obu zmiennych jest wzrost skłonności (potencjalnego) respondenta do stosowania strategii *satisficingu*, którego, co oczywiste, najpełniejszą formą jest właśnie odmowa udziału w badaniu. Tłumaczyłoby to więc jednoczesne oddziaływanie preferencji techniki na poziom braków odpowiedzi i jakość odpowiedzi udzielanych w sondażu – teoria pamięci roboczej integruje te pozornie odmienne zjawiska.

Należy się zastanowić, jakie czynniki mogą powodować zróżnicowanie poziomu sprawności subsystemów pamięci roboczej odpowiedzialnych za przetwarzanie informacji wzrokowych i słuchowych. Według niektórych badań w obrębie psychologii poznawczej, poza chorobami, które obejmują niewielką część populacji i nie są (na szczęście) uwzględniane w operatach, takim czynnikiem mógłby być na przykład proces starzenia się. Przeprowadzone przez Fofi Constantinidou i Susan Baker [2002] badanie dotyczyło zapamiętywania 15 krótkich słów przedstawianych badanym za pomocą trzech trybów – odsłuchiwania, pokazywania czarno-białych obrazków – reprezentacji słów, lub obu form na raz. Próba (50 osób) składała się wyłącznie z mężczyzn – połowy w wieku 19–38 i drugiej połowy w wieku 50–77 lat. Uczestnicy zapamiętywali serie słów kilka razy w ramach treningu, by następnie przywoływać je z pamięci. Istotny okazał

się wynik dla interakcji pomiędzy wiekiem a trybem w teście po 30-minutowej przerwie od pokazania serii słów – osoby z grupy w średnim i starszym wieku gorzej przypominały sobie informacje prezentowane w trybie audialnym niż osoby z grupy młodszej. Choć osłabienie działania pętli fonologicznej niesie za sobą skutki dla udziału w sondażu we wszystkich technikach, to jednak w odniesieniu do technik wizualnych można zaryzykować hipotezę, że współudział sprawniejszego szkicownika wzrokowo-przestrzennego w przetwarzaniu informacji miałby wartość wspomagającą. W odniesieniu do powyższych badań wartościowe mogłoby być także uwzględnienie wspomagających materiałów wizualnych dla starszych respondentów w technikach audialnych.

Chciałbym wspomnieć także o innych badaniach dotyczących tej samej tematyki, które przyniosły odmienny wynik. Navnit Kumar i Brajesh Priyadarshi [2013] dali 80 uczestnikom podobne zadanie – zapamiętanie kolejności 7 usłyszanych cyfr lub pokazanych symboli (w postaci kropki umieszczonej na siatce 5 x 5). Od razu należy zaznaczyć, że symbole te wydają się dużo bardziej skomplikowane niż wszystkie inne formy prezentacji informacji w obu badaniach – być może czynnik ten mógł mieć wpływ na wynik. Badani zostali podzieleni na 5 grup w wieku od 40 do ponad 80 lat. W pierwszych trzech grupach (do 70 lat) krzywe spadku sprawności zapamiętywania były niemal równoległe, jednak powyżej 70 lat (czyli w kategorii wiekowej częściowo nieuwzględnionej w poprzednio przywołanym badaniu) zapamiętywanie sygnałów słuchowych utrzymywało się na stałym poziomie, natomiast wartości dla sygnałów wzrokowych spadały nadal.

Gdyby próbować wyciągnąć wniosek z tych wyników, należałoby przyjąć, że osoby „bardzo stare” (jak na przedstawiony powyżej zakres wieku badanych) będą chętniej i lepiej uczestniczyły w sondażach prowadzonych techniką operującą na sygnałach słuchowych, natomiast osoby w wieku średnim i we „wczesnej starości” łatwiej poradzą sobie w badaniach wykorzystujących wizualny kanał komunikacji. Opisane badania eksperymentalne przeprowadzone były na niezbyt dużych próbach, a pierwsze z nich stosuje wyłącznie dychotomiczny podział grup wiekowych, więc wnioskowanie na ich podstawie może wydawać się słabo ugruntowane. Jednak przedstawione wnioski okazują się zgodne z wynikami badań dotyczących preferencji techniki sondażu w Polsce. Adam Rybak [2018] na podstawie analiz reprezentatywnego sondażu *mixed-mode* z wyborem techniki przez respondentów zauważył, że zależność pomiędzy wiekiem a preferowaniem techniki pocztowej (kanał wzrokowy) była krzywoliniowa. Najwyższe prawdopodobieństwo wyboru techniki pocztowej występowało w kategorii osób w wieku 45–64 lat, natomiast powyżej tego wieku gwałtownie spadało (na korzyść

przede wszystkim wywiadu kwestionariuszowego – techniki opartej głównie na kanale audialnym).

Wnioski te mogą być ogólnie słuszne, nie należy jednak bezrefleksyjnie uwzględniać wyłącznie kwestii funkcjonowania pamięci roboczej – różnice w sposobach uczestnictwa (lub nieuczestnictwa) w sondażach mogą w tak zaawansowanym wieku wynikać z przyczyn bardziej prozaicznych, to jest degradacji narządów zmysłów. Zgodnie z badaniami GUS „Stan zdrowia ludności Polski w 2014 r.” przeprowadzonymi w ramach *European Health Interview Survey* wśród badanych w wieku 80 lat i więcej pewne lub duże trudności z widzeniem deklarowało 63,4%. W tej samej grupie wiekowej 50,3% respondentów deklarowało, że ma pewne lub duże trudności z usłyszeniem rozmowy w cichym pomieszczeniu – a więc w takim, w jakim najczęściej odbywa się sondaż. Różnica między obiema wartościami wynosi ponad 13 punktów procentowych. Choć informacje te mogą nieść istotną wartość w projektowaniu badań sondażowych, to należy zadbać, by badając zmienne mogące pozostawać w związku zarówno z pamięcią roboczą, jak i ze zdrowiem narządów zmysłów, kontrolować oba czynniki.

Kolejną rozpoznaną przez badania empiryczne sytuacją, w której może nastąpić nieproporcjonalne obniżenie efektywności jednego z podsystemów pamięci roboczej, jest zjawisko lęku związanego z testem (ang. *test anxiety*). W literaturze dotyczącej metodologii sondażowej niekiedy przywoływane jest twierdzenie, że sytuacja sondażu przez niektóre kategorie respondentów może zostać potraktowana niemal jak egzamin. Można przypuszczać, że najczęściej takie skojarzenia będą towarzyszyć osobom uczącym się, ze względu na ich doświadczenia egzaminacyjne. Określenie skali zjawiska w populacji nie jest łatwe, jednak szacuje się, że silny lęk związany z sytuacją egzaminacyjną może dotyczyć ponad 1/4 studentów [Tahmasbipour, Ahmady 2011: 719]. Nie jest to więc zjawisko marginalne, dodatkowo dotyczy tej kategorii, z którą względnie trudno nawiązać kontakt i kooperację w badaniu sondażowym.

Większość dotychczasowych badań sugeruje, że sytuacje lękowe wywierają większy wpływ na przetwarzanie danych werbalnych, a więc osłabiają funkcjonowanie pętli fonologicznej. Ikeda Maki, Iwanaga Makoto i Seiwa Hidetoshi [1996], przeprowadzając badania na 302 studentach psychologii, ustalili, że wysoki lęk powodował znaczące wydłużenie przetwarzania informacji werbalnych, ale nie obniżył ogólnego wyniku testu, jeśli porównać go z wynikami osób odczuwających niski lęk lub nieodczuwających go wcale. Kwestię komplikuje ustalenie Heather Miller i Jacqueline Bichsel [2004] sugerujące, że specyficzna odmiana lęku związana z wykonywaniem zadań matematycznych silniej osłabia funk-

cjonowanie szkicownika wzrokowo-przestrzennego. Nie pozwala to wyciągnąć jednoznacznych wniosków dotyczących przewidywań preferencji techniki, jednak ostrożny wniosek, że osłabione przetwarzanie informacji werbalnych w sytuacji interpretowanej jako egzaminacyjna może zmniejszać preferencję osób młodych względem udziału w badaniach sondażowych *face-to-face* wydaje się zgodny z dotychczasowymi ustaleniami [Smyth, Olson, Millar 2014].

## PODSUMOWANIE

W powyższym artykule starałem się zarysować ogólną koncepcję systemu zależności pomiędzy konsekwencjami (nie)zgodności techniki badania z preferencją respondenta – jej wpływem na braki odpowiedzi i błąd pomiaru sondażowego. Wykorzystałem w tym celu zjawisko tzw. *satisficingu*. Zaproponowałem jego konceptualne rozbudowanie poprzez odwołanie się do powstałej na gruncie psychologii poznawczej koncepcji pamięci roboczej. Jest to zgodne z formułowanymi przez członków ruchu CASM twierdzeniami, że badanie problematyki metodologii badań sondażowych nie może być owocne w sytuacji ograniczenia się do instytucjonalnie wyróżnionych obszarów nauki. Wymaga współdziałania i wymiany intelektualnej między socjologami, statystykami, demografami i psychologami. Tak jak w obszarze ścisłej tradycji badań metodologii sondażowej wskazane byłoby łączne badanie zjawisk preferencji techniki i różnic modalności przetwarzanych informacji, wraz ze skutkami tych aspektów dla jakości realizowanych badań, tak też wymiana pomiędzy obszarem ilościowych badań socjologicznych a badaniami psychologii poznawczej winna być zintensyfikowana. Potencjalne korzyści są obopólne – psychologowie zyskaliby dostęp do dużych prób reprezentatywnych, na których w ramach większych badań można by przeprowadzać proste testy poznawcze, co dawałoby możliwość rzeczywistej weryfikacji skali występowania postulowanych przez dyscyplinę zjawisk. Badacze sondażowi natomiast skorzystaliby niezmiernie ze względu na pogłębienie kognitywistycznych aspektów badania sondażowego, traktowanych obecnie zazwyczaj niezwykle swobodnie (nie licząc być może kilku flagowych projektów międzykrajowych). Także uwzględnienie w szerszym zakresie (aktualnie niemal nieobecnej) problematyki psychologicznych aspektów konstruowania kwestionariuszy w programach polskich studiów socjologicznych wydaje się wskazane.

Uszczegółowieniem powyższych problemów, na które nacisk położyłem szczególnie w ostatniej części artykułu, jest kwestia starzenia się populacji zachodniej, która wciąż stanowi największe pole badań sondażowych. Choć rozwój analiz opartych na *Big Data* ogranicza w pewnym zakresie zapotrzebowanie na

reprezentatywne badania kwestionariuszowe, to jednak wciąż brak całościowej alternatywy dla tej metody badawczej do formułowania wartościowych wniosków na temat dystrybucji w populacji wielu społecznie istotnych cech. Dlatego istnieje konieczność dostosowywania sposobów konstrukcji i prowadzenia badań do wymagań tej, starszej kategorii potencjalnych respondentów. Jednym z aspektów takich analiz może być kwestia, w jaki sposób procesy starzenia wpływają na funkcjonowanie pamięci roboczej, co przekłada się na sposoby przetwarzania przekazywanych w sondażu informacji w różnych modalnościach. Oczywiście, jak zostało zaznaczone na początku, nie sugeruję, iż całość zróżnicowania preferencji techniki pomiędzy respondentami starszymi a resztą, jak i w obrębie tej kategorii, wynika ze zmian zachodzących w pamięci roboczej. Nie chciałbym także ograniczać możliwych zastosowań koncepcyjnych tego konstruktu teoretycznego wyłącznie do wspomnianych procesów starzenia i sytuacji lękowych. Wartościowe byłoby także poszukiwanie innych czynników mogących mieć wpływ na różnicowanie funkcjonowania elementów pamięci roboczej i przekładanie się tego na przetwarzanie informacji w różnych modalnościach. Ze względu na ich obecność, lub związek z informacjami zawartymi w obecnie istniejących lub budowanych w przyszłości operatach, przyczyniłoby się to do ulepszenia konstrukcji przeprowadzanych badań ilościowych.

Badania, które pogłębiłyby rozumienie omawianej w artykule problematyki i umożliwiły integrowanie poruszonych w nim zagadnień, mogą koncentrować się przede wszystkim w trzech obszarach. Po pierwsze, potrzebne są liczniejsze eksperymenty kognitywistyczne dotyczące potencjalnych czynników powodujących różnicowanie jakości funkcjonowania podsystemów pamięci roboczej – ze szczególnym uwzględnieniem przywołanego w tekście czynnika wieku. Dotychczasowe badania wykonane były na niewielkich, lokalnych, niezbyt zróżnicowanych próbach, co utrudnia wyciąganie solidnych szczegółowych wniosków. Dodatkowo należy pamiętać, że nauki o poznaniu są świeżą i dynamiczną dyscypliną, której koncepcje teoretyczne mogą w szybkim czasie ulec znaczącym modyfikacjom. Badacze chcący uwzględnić pojęcie pamięci roboczej w innych dyscyplinach muszą śledzić zmiany dotyczące tego pojęcia.

Po drugie, wydaje się, że wywiady kognitywne, będące specyficzną metodą sondażowych badań pilotażowych, mogą umożliwiać badanie związków pomiędzy sposobami przetwarzania przez respondentów informacji przekazywanych różnymi kanałami a sposobami uczestnictwa w badaniu i odczuciami w jego trakcie u respondentów. Jednocześnie mogą badać tę problematykę w bardziej „naturalnych” z punktu widzenia metodologów sondażowych warunkach niż abstrakcyjne testy kognitywistyczne. Potencjalnie wartościowe byłoby łączenie

eksperymentalnych schematów przyporządkowujących pilotażowych respondentów do różnych technik sondażu z nowymi, pomysłowymi sposobami sondowania przykładowo przeżywanych przez respondentów trudności poznawczych.

Wreszcie, po trzecie, badania zróżnicowania preferencji techniki sondażu w populacjach różnych krajów są wciąż nieliczne. Utrudnia to przykładowo odpowiedź na pytanie, czy wzory zależności pomiędzy tymi preferencjami a podstawowymi zmiennymi społeczno-demograficznymi są stałe, czy zmieniają się pod wpływem różnych warunków kulturowych, instytucjonalnych czy infrastrukturalnych. Stałość tych wzorów bez względu na inne warunki mogłaby być częściową przesłanką sugerującą wagę bardziej uniwersalnych procesów psychologicznych w determinowaniu preferencji.

## BIBLIOGRAFIA

- Cialdini Robert.** 1984. *Influence: The new psychology of modern persuasion*. New York: Quill.
- Clearman Jack, Vojtěch Klinger, Dénes Szűcs.** 2017. "Visuospatial and verbal memory in mental arithmetic". *Quarterly Journal of Experimental Psychology* 70(9): 1837–1855.
- Constantinidou Fofi, Susan Baker.** 2002. "Stimulus modality and verbal learning performance in normal aging". *Brain and Language* 82(3): 296–311.
- Dillman Don A., Jolene D. Smyth, Leah Melani Christian.** 2014. *Internet, phone, mail, and mixed mode surveys: The tailored design method (4th ed.)*. Hoboken: John Wiley & Sons Inc.
- Diment Kieren, Sam Garrett-Jones.** 2007. "How demographic characteristics affect mode preference in a postal/web mixed-mode survey of Australian researchers". *Social Science Computer Review* 25(3): 410–417.
- Groves Robert M., Mick P. Couper.** 1998. *Nonresponse in household interview surveys*. New York: John Wiley & Sons.
- Groves Robert M., Robert L. Kahn.** 1979. *Surveys by telephone : a national comparison with personal interviews*. New York: Academic Press.
- Groves Robert M., Eleanor Singer, Amy Corning.** 2000. "Leverage-saliency theory of survey participation". *Public Opinion Quarterly* 64(3): 299–308.
- Haan Marieke, Yfke P. Ongena, Kees Arts.** 2014. "Reaching hard-to-survey populations: Mode choice and mode preference". *Journal of Official Statistics* 30(2): 355–379.
- Ikeda Maki, Makoto Iwanaga, Hidetoshi Seiwa.** 1996. "Test anxiety and working memory system". *Perceptual and Motor Skills* 82(3): 1223–1231.
- Jabkowski Piotr.** 2015. *Reprezentatywność badań reprezentatywnych: analiza wybranych problemów metodologicznych oraz praktycznych w paradygmacie całkowitego błędu pomiaru*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.
- Krosnick Jon.** 1991. "Response strategies for coping with the cognitive demands of attitude measure in surveys". *Applied Cognitive Psychology* 5: 213–236.
- Krzewińska Aneta, Katarzyna Grzeszkiewicz-Radulska.** 2013. „Klasyfikacja sondażowych technik otrzymywania materiałów”. *Przegląd Socjologiczny* 62(1): 9–32.

- Kumar Navnit, Brajesh Priyadarshi.** 2013. "Differential effect of aging on verbal and visuo-spatial working memory". *Aging and Disease* 4(4): 170–177.
- Medway Rebecca L., Jenna Fulton.** 2012. "When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates". *Public Opinion Quarterly* 76(4): 733–746.
- Miller Heather, Jacqueline Bichsel.** 2004. "Anxiety, working memory, gender, and math performance". *Personality and Individual Differences* 37(3): 591–606.
- Meyer Maciej.** 2007. „Herbert Simon i jego idea ograniczonej racjonalności”. *Decyzje* 7: 111–115.
- Olive Thierry, Ronald Kellogg, Annie Piolat.** 2008. "Verbal, visual, and spatial working memory demands during text composition". *Applied Psycholinguistics* 29(4): 669–687.
- Olson Kristen, Jolene D. Smyth, Heather M. Wood.** 2012. "Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination". *Public Opinion Quarterly* 76(4): 611–635.
- Pham Andy V., Ramzi M. Hasson.** 2014. "Verbal and visuospatial working memory as predictors of children's reading ability". *Archives of Clinical Neuropsychology* 29(5): 467–477.
- Piotrowski Krzysztof, Zbigniew Stettner, Jarosław Orzechowski, Robert Balas.** 2009. Jak działa pamięć robocza? W: *Pamięć robocza*. J. Orzechowski i in. (red.), 25–46. Warszawa: Wydawnictwo Szkoły Wyższej Psychologii Społecznej Academica.
- Rybak Adam.** 2018. „The selection of survey modes on the basis of respondents' preferences analysis”. *Ask: Research and Methods* 27(1): 23–39.
- Sakshaug Joseph W., Ting Yan, Roger Tourangeau.** 2010. "Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items". *Public Opinion Quarterly* 74(5): 907–933.
- Smyth Jolene D., Kristen Olson, Alian Kasabian.** 2014. „The effect of answering in a preferred versus a non-preferred survey mode on measurement”. *Survey Research Methods* 8(3): 137–152.
- Smyth Jolene D., Kristen Olson, Morgan M. Millar.** 2014. „Identifying predictors of survey mode preference”. *Social Science Research* 48: 135–144.
- Sztabiński Franciszek.** 1990. Proces komunikowania się w ankiecie pocztowej. Prezentacja schematu. W: *Analizy i próby technik badawczych w socjologii*. Z. Gostkowski (red.), T. VIII, 59–97. Wrocław–Warszawa–Kraków: Ossolineum.
- Sztabiński Franciszek.** 1997. *Ankieta pocztowa i wywiad kwestionariuszowy*. Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński Franciszek.** 2011. *Ocena jakości danych w badaniach surveyowych*. Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński Paweł.** 1995. „Efekt ankierski: czy zmienia się tylko respondent”. *Ask. Research & Methods* 1(1): 81–99.
- Tahmasbipour Najaf, K. Ahmady.** 2011. "The investigation of prevalence rate and diagnosis of exam anxiety risk factors among Shahid Rajaei University students (SRTTU)". *Procedia – Social and Behavioral Sciences* 30: 717–721.
- Tourangeau Roger, Lance J. Rips, Kenneth A. Rasinski.** 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.
- Vannieuwenhuyze Jorre T.A., Melanie Revilla.** 2013. "Evaluating relative mode effects on data quality in mixed-mode surveys". *Survey Research Methods* 7(3): 157–168.

Verma Santosh K., Theodore K. Courtney, David A. Lombardi, Wen Ruey Chang, Yueng Hsiang Huang, Melanye J. Brennan, Melissa J. Perry. 2014. "Internet and telephonic IVR mixed-mode survey for longitudinal studies: Choice, retention, and data equivalency". *Annals of Epidemiology* 24(1): 72–74.

*Adam Rybak*

**THE EFFECTS OF VISUAL AND AURAL COMMUNICATION ON THE QUALITY  
OF SURVEY MEASUREMENT. THE PROPOSITION OF A NEW THEORETICAL  
APPROACH TO SURVEY MODE PREFERENCE ON THE BASIS  
OF THE WORKING MEMORY CONCEPT**

Abstract

The aim of the article is to present a potential direction for the development of survey design research. It tries to describe the possibility of building a concept that integrates the observed consequences of survey mode preference – its impact on the response rate and the quality of responses – using the cognitive concept of working memory. It indicates the existence of a potential relationship between the aging process, or anxiety, and the diversification of information processing efficiency in the visual and aural modalities, and thus outlines the problems facing researchers planning mixed-mode surveys, especially in aging western populations.

**Keywords:** working memory, survey mode preference, satisficing, survey research methodology

ADAM RYBAK  
Adam Mickiewicz University

## Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference

*Abstract:* Using previously established knowledge about survey mode preferences distribution in a population can be one of many ways of improving representativeness and quality of data gathered by survey research. Apart from mode preference existence and stability, the main problem concerns the question: which of the sources of information about mode preference could be treated as trustworthy. Because real observed choices are typically treated as better predictors of future choices than declaration only, this paper tries to answer the question: 'Are the declarations a good predictor of preference in comparison to real choices?' For this purpose, it uses combined data from 1) the 2015 Polish mixed-mode ESS experiment and 2) data from ESS8 in Poland. The multinomial logistic regression survey preference model includes socio-demographic variables accompanied by declaration/choice control variable. The results suggest significant differences between choices/declarations. Findings could be used to refine the contact strategies used in surveys.

*Keywords:* European Social Survey, survey mode preference, mixed-mode survey design, adaptive survey design.

### Introduction

The analysis of survey mode preference is one of many approaches employed to deal with the continuously rising level of nonresponse error (and additionally with satisficing—Smyth, Olson, and Kasabian 2014). Although there are some definitional discrepancies—especially concerning the difference between observed and declared preference (Al Baghal and Kelley 2016: 145)—it is usually described as a relatively stable propensity toward participation in the survey research in one mode (at the expense of others). It is worth mentioning that the time stability of preferences is sometimes questioned (Al Baghal and Kelley 2016: 162; Wardle and Robinson 2007). Despite the evidence for the usability of the surveying approach based on mode preference predictions being still sparse, it is mentioned in more recent handbooks (Dillman, Smyth, and Christian 2014).

The two main strategies of gathering data about survey mode preference distribution in the population are asking respondents about their preference in the single-mode survey or giving them the choice of mode in the mixed-mode survey. Ostensibly, observing choice could look like the best way of predicting future behavior, but this strategy comes with a cost of reducing the response rate (Medway and Fulton 2012), thus counteracting the primary goal. On the other hand, asking about preferences in one mode leads to a significantly higher number of respondents selecting the mode of the question as preferred

(Groves and Kahn 1979; Tarnai and Paxson 2004), so this is not the perfect solution either.

Gathering data about the preference of both behavioral and declarative origin could lead to the creation of classification models, which can be used for allocating a sampled person for their probable preferred mode without asking. The usefulness of knowledge about mode preference distribution also manifests in the possibility of preliminary selection effect assessment. “Mixed-mode surveys are only advantageous over single-mode surveys if selection effects occur” (Vannieuwenhuyze 2013). The adequately designed mixed-mode survey can lead to a significant decrease in fieldwork costs. Centre of Sociological Research in the Institute of Philosophy and Sociology of Polish Academy of Sciences (Polish ESS Fieldwork Agency) declared that representative survey in mail mode would cost 40% of the analogous survey in face-to-face mode (Villard and Fitzgerald 2017). Nevertheless, the two major examples of mixed-mode survey experiments in central-eastern Europe led to the final unit-cost being very similar (or even higher) than in the corresponding single-mode face-to-face surveys (Ainsaar et al. 2013; Sztabiński 2018). However, both mixed-mode studies had a pioneering character in their countries in contrast to the refined methodology of main (f2f) surveys—in the next rounds achieved cost could be lower.

Rybak (2018) used data about mode choices in the Polish survey experiment to build a model assessing the basic socio-demographic indicators of preference. The main aim of that study was to determine who can be pushed to mail mode in Poland (this limitation comes from the assumption that the preferences are dependent on the local, cultural and infrastructural conditions). This study combines the Polish behavioral data and declaration data about survey mode preferences with the main aim of estimating how much both sources are similar in assessing the impact of preference predictors. Because real choices are assumed to be the better reflection of preference, this question can be transformed into: ‘Are the declarations a good predictor of preference?’

## Hypotheses

Apart from the main question, however, comparing the data from both studies in more specific areas could be useful. The aforementioned first study found support for several hypotheses based on survey methodology literature (Diment and Garrett-Jones 2007; Groves and Couper 1998: 113, 147, 176; Haan, Ongena, and Aarts 2014; Millar, O’Neill, and Dillman 2009; Smyth, Olson, and Millar 2014). They can be used as hypotheses for this research:

*H1:* There will be a higher preference toward the mail mode among the female population.

*H2:* There will be a higher preference toward the Web mode among the male population.

It is worth mentioning that from the first study (2015) to the second study (2017), internet usage in Poland changed from 68% to 76% of the population (International Telecommunication Union 2019).

*H3:* There will be a higher preference toward an interviewer mode survey among residents of villages, and/or small cities, than the bigger ones.

*H4:* There will be a higher preference toward a postal mode survey among residents of bigger cities, than of small ones and/or villages.

Rybak's study also came to two findings contradictory to the previous research, which can also be used as hypotheses:

*H5:* There will be a curvilinear relationship between age and the mail mode preference with middle-aged people preferring it the most.

*H6:* There will be no significant relationship between age and the Web mode preference.

The last hypothesis is based on the central question of this study stated above. As the declarations about preferences are used as the indicator of preference (and future choices) (e.g. [Olson, Smyth, and Wood 2012](#)), so in two samples representative of the Polish population with just a 1.5-year gap, there should be no significant difference in findings.

*H7:* There will be no significant difference between behavioral and declarative data concerning survey preferences.

The use of basic socio-demographic predictors only is determined by the idea of creating the classification model that can be used to allocate respondents to mode before reaching them. This approach is called Adaptive Survey Design (e.g. [Tourangeau et al. 2016](#)). Furthermore, as the sampling frame most usually utilized in Poland (which is the PESEL database) contains information about gender, age, and place of residence, these variables are used in this study.

## Data and Methods

Data for this study comes from two sources. The first one is the Mixed-mode experiment conducted parallelly to European Social Survey Round 7 in Poland by Paweł & Franciszek Sztabiński. Second is the national module attached to European Social Survey Round 8 in Poland. Stated experiment (**S1**) took place between 7<sup>th</sup> April and 30<sup>th</sup> September 2015 ([Sztabiński 2018](#); [Sztabiński and Sztabiński 2016](#)) and used an identical sampling scheme as main ESS7 study ([European Social Survey 2018a](#)). ESS8 (**S2**) in Poland took place between 7<sup>th</sup> November 2016 and 22<sup>nd</sup> February 2017 ([European Social Survey 2018b](#)). Both studies resulted in the creation of datasets based on samples representative for the Polish population.

**S1** mixed-mode design was neither strictly sequential nor parallel. The respondent was first contacted by mail. The envelope contained the paper self-administered questionnaire, return envelope, examples of finding from previous ESS research, gift (a notepad with a magnet) and cover letter. The letter stated that the sampled person could answer by mail mode or access the Web questionnaire on the dedicated website. When no answer was given, reminders were sent, and if there was still no response, interviewers visited selected addresses. A sampled person then had the option to participate in a face-to-face interview with interviewers from ESS7 in Poland. If there was no contact or cooperation, a third letter was sent, this time containing only the paper questionnaire with the return envelope.

**S2** was the single-mode face-to-face interview but contained an additional question about the preferred mode of participation: F2F, CATI, mail, or Web (also 'no preference'; 'do not want to participate at all'). As CATI preference was absent in the first study, these responses were excluded from further analysis. Because that could lead to the breach of the assumption of irrelevant alternatives, the test of Hausman-McFadden ([Long and Freese 2014: 407–9](#)) was conducted. Results proved to be statistically insignificant.

In the first study, 800 people were sampled, and 419 questionnaires were completed—the response rate was 55.9%. 164 (39.1%) people answered via mail mode, 24 (5.7%) by Web mode, and 231 (55.2%) by the mode with the interviewer present. In the second study, the number of sampled individuals was 2675, and 1694 questionnaires were completed—the response rate was 69.6%. 1213 (71.6%) respondents declared face-to-face mode preference, 96 (5.7%) mail mode preference and 190 (11.2%) Web mode preference. Frequencies for both studies are presented in tables 1 and 2.

Table 1

**Frequencies of mode choices in ESS7 mixed-mode experiment in Poland**

| Study 1 ESS7PL mixed-mode |            |              |
|---------------------------|------------|--------------|
| Choice                    | Frequency  | Percentage   |
| Face2face                 | 231        | 55.2         |
| Mail                      | 164        | 39.1         |
| Web                       | 24         | 5.7          |
| <b>Total</b>              | <b>419</b> | <b>100.0</b> |

Table 2

**Frequencies of mode preference declarations in ESS8 in Poland**

| Study 2 ESS8PL |             |              |
|----------------|-------------|--------------|
| Declaration    | Frequency   | Percentage   |
| Face2face      | 1213        | 71.6         |
| Mail           | 96          | 5.7          |
| Web            | 190         | 11.2         |
| Phone          | 27          | 1.6          |
| Not interested | 74          | 4.4          |
| No preference  | 81          | 4.8          |
| Refusal        | 2           | 0.1          |
| No answer      | 11          | 0.7          |
| <b>Total</b>   | <b>1694</b> | <b>100.0</b> |

The combined dataset, which included information about age, gender, size of settlement of residence, weights, and preference of each respondent from both studies, was created. Preferences were coded dependent on the source study as participation in one mode in the mixed-mode experiment (the behavioral indicator of preference) or as a declaration of preferring one of the possible modes of contact in the single-mode survey (the declarative indicator of preference).

The multinomial logistic regression model was computed in STATA 15, containing age, gender, size of settlement, and source dataset as independent variables and preference (face-to-face, mail, or Web) as the dependent variable. Face-to-face was treated as the base outcome. Age and urbanicity were treated as categorical variables because of a supposed non-linear relationship. Design weights from both surveys were used in the model. In the

tested models, which included interactions between variables, these interactions turned out to be not significant.

## Results

The odds of mail mode preference in comparison to face-to-face were significantly higher for females, individuals 15–19 and 25–64 years old (as opposed to 75 and older) living in cities having more than 10k inhabitants (as opposed to the residents of villages) and for the observed, behavioral data. Therefore, there is clear support for *H1*, *H3* and *H4*, but no support for the *H7*. The odds for people 45–54 are the highest from each category and are significantly higher than the value of base-level (75+) so there is also some support for *H5*.

The odds of Web mode preference in comparison to face-to-face were higher for people living in cities having 10–19k inhabitants and cities with 50k inhabitants or more as opposed to residents of villages. They were also significantly higher for respondents 15–54 years old in comparison to people 75 and older. The ratio was also lower for behavioral data, but the difference was at the limit of statistical significance ( $p < 0.051$ ). Therefore, there is clear support for the *H3*, unclear for *H7*, but no support for *H2* and *H6*.

McFadden's pseudo- $R^2$  of the model was 0.1824 and the regression was significant as a whole. Model excluding type-of-data variable had an  $R^2$  value of 0.0938. The model is presented in [table 3](#).

Table 3

Multinomial logistic regression: Mail and Web survey mode preference relative risk ratios

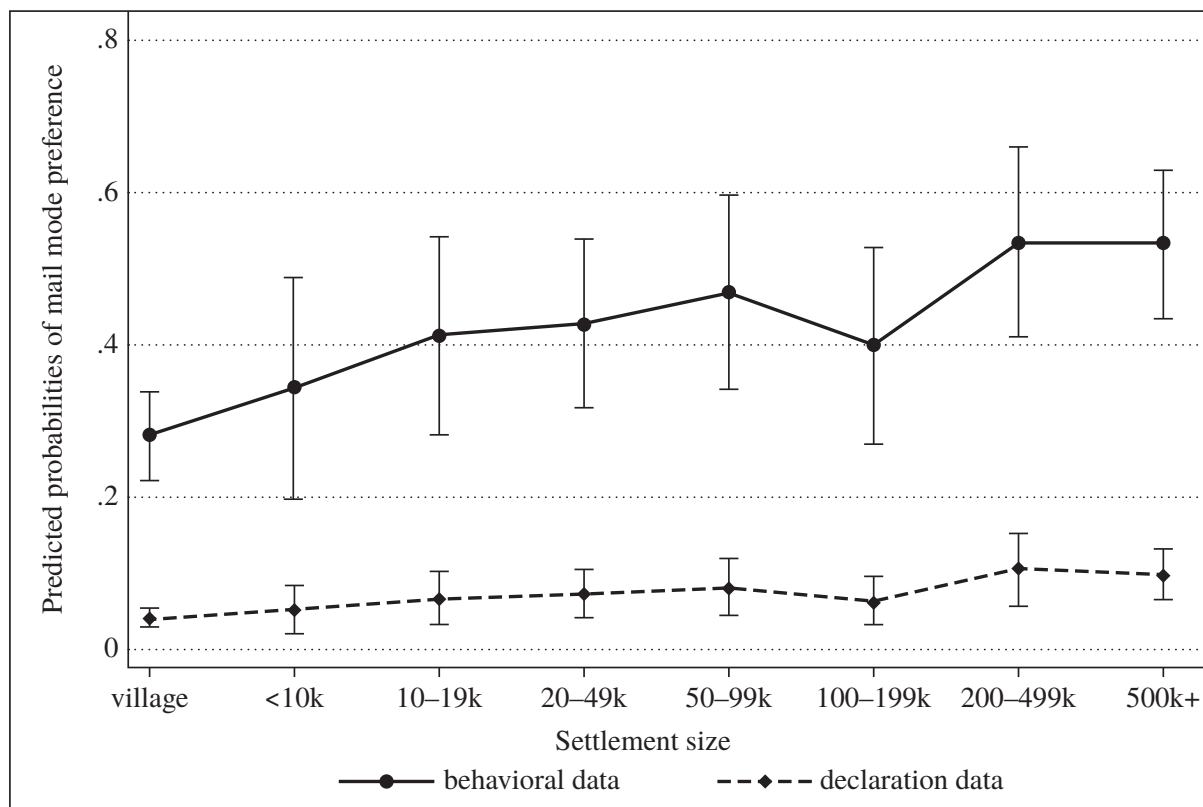
| Model (base=F2F)                       | Mail RRR | Web RRR |
|----------------------------------------|----------|---------|
| <b>Female</b>                          | 2.149**  | 1.196   |
| <b>Age (base 75+)</b>                  |          |         |
| 15–19                                  | 3.012*   | 21.56** |
| 20–24                                  | 2.429    | 27.99** |
| 25–34                                  | 2.669*   | 14.92** |
| 35–44                                  | 2.907*   | 12.12** |
| 45–54                                  | 3.758**  | 6.944*  |
| 55–64                                  | 2.903**  | 4.095   |
| 65–74                                  | 2.198    | 1.374   |
| <b>Urbanicity (base village)</b>       |          |         |
| city <10k                              | 1.330    | 0.969   |
| city 10–19k                            | 1.859*   | 2.069*  |
| city 20–49k                            | 1.950**  | 1.479   |
| city 50–99k                            | 2.456**  | 3.517** |
| city 100–199k                          | 1.806*   | 2.744** |
| city 200–499k                          | 3.186**  | 3.363** |
| city 500k+                             | 3.547**  | 7.164** |
| <b>Type of data (base declarative)</b> |          |         |
| behavioral data                        | 9.302**  | 0.620'  |
| <b>Constant</b>                        | 0.011**  | 0.008** |

\*\*  $p < 0.01$ , \*  $p < 0.05$ , '  $p < 0.051$

Predicted probabilities of mail mode preference for age and urbanicity categories were computed. The charts below show the probabilities with 95% confidence intervals for the variable “dataset” held at values “behavioral data” and “declaration data.” There are significant differences between the age and urbanicity levels, but the differences between values for different datasets are much more apparent.

Chart 1

**Predicted probabilities—mail mode propensities for urbanicity categories.**  
**‘Dataset’ variable held at ‘behavioral’ and ‘declaration’ level**

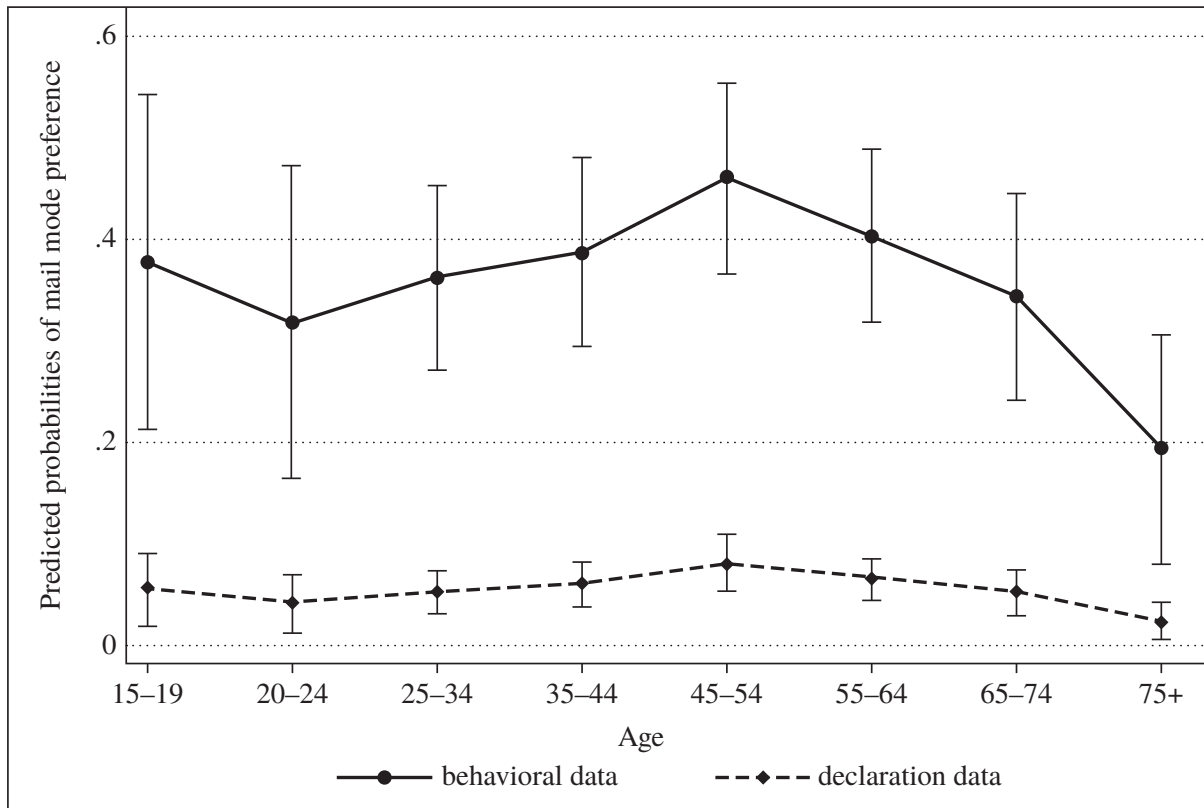


### Discussion and Conclusion

The primary result of this study is that according to the model declarations do differ from actions (choices), especially concerning the mail mode. There are some relationships shapes maintained between results from both sources of data, and between the current model and previous model by Rybak (2018). Predicted probabilities of mail mode preference as collated with age categories still have the curvilinear shape, which can be seen in the odds ratio values or the predicted probabilities charts. The people aged 45–54 still have the highest mail mode propensities. Nevertheless, it cannot be omitted, that differences between declarations and choices are important. The odds of mail mode preference were nine times higher in the mixed-mode experiment than in the ESS8 preference question. This leads to the answer to the question asked in the introduction—can we really treat declarations as a good predictor of the preference? Probably not. They could be used in the absence of better auxiliary data, but they are imperfect.

Chart 2

**Predicted probabilities—mail mode propensities for age categories.**  
**'Dataset' variable held at 'behavioral' and 'declaration' level**



The study has many limitations, especially concerning the mixed-mode experiment part. The design of it was not strictly parallel, which can suggest that mail mode preference has been artificially inflated in comparison to face-to-face mode (because of the fact that interviewers contacted only nonrespondents in mail/Web modes). On the other hand, the declarations in favor of the mode with the interviewer present may be raised by the mode of the question itself. Another difficulty comes from the probable instability of mode preference. All of these problems were addressed in the previous sections of the article.

However, what can be seen in this study is that there are essential differences between mail and Web preferences. Mail preference was observed as much more frequent in real choices, as Web mode was much more frequent in declarations. This difference cannot be attributed to the defect in design, because Web and mail modes were offered simultaneously in the experiment. Change in Internet usage in Poland in the 1.5-year period between the studies also was not that big. One of the possible explanations of this fact could be based on treating the mode preference question as a sensitive question.<sup>1</sup> People asked by the interviewer can try to appear as more modern and tech-savvy by declaring Web mode preference. In the real situation, however, the usage of mail mode can be much more convenient to them, because of lack of digital literacy, the necessity of mode switch (when the sample is contacted by mail, because of the address frame) or other reasons.

The cost of face-to-face survey research in Poland is still relatively low compared to the cost in western countries. Because of this, mixed-mode surveys are still seen as one

<sup>1</sup> I would like to thank dr. Edith de Leeuw for suggesting me this interpretation.

of the possible options, while in the west they start to become a necessity. We can expect that the processes of the response rate reduction and fieldwork cost increase will accelerate in Poland in the future. Because of that, the preparation for the upcoming situation, also by refining the mixed-mode designs best fitted for the local environment (especially in the adaptive survey design framework) could be beneficial.

### Acknowledgments

I would like to thank Paweł Sztabiński for the data on which this study is based and Piotr Jabkowski for supervision over my research.

### References

- Ainsaar, M., Lilleoja, L., Lumiste, K., and Roots, A. 2013. *ESS Mixed Mode Experiment Results in Estonia (CAWI and CAPI Mode Sequential Design)*. Tartu: University of Tartu, Institute of Sociology and Social Policy.
- Al Baghal, T., and Kelley, J. 2016. The Stability of Mode Preferences: Implications for Tailoring in Longitudinal Surveys, *Methods, Data, Analyses* 10 (102): 143–66.
- Dillman, D.A., Smyth, J.D., and Christian, L.M. 2014. *Internet, Phone, Mail, and Mixed Mode Surveys: The Tailored Design Method (4<sup>th</sup> Ed.)*. Hoboken: John Wiley & Sons Inc.
- Diment, K., and Garrett-Jones, S. 2007. How Demographic Characteristics Affect Mode Preference in a Postal/Web Mixed-Mode Survey of Australian Researchers, *Social Science Computer Review* 25 (3): 410–17.
- European Social Survey. 2018a. *ESS-7 2014 Documentation Report*. Edition 3.2. Bergen, European Social Survey Data Archive, NSD—Norwegian Centre for Research Data for ESS ERIC.
- European Social Survey. 2018b. *ESS-8 2016 Documentation Report*. Edition 2.1. Bergen, European Social Survey Data Archive, NSD—Norwegian Centre for Research Data for ESS ERIC.
- Groves, R.M., and Couper, M.P. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Groves, R.M., and Kahn, R.L.. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Haan, M., Ongena, Y.P., and Arts, K. 2014. Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference, *Journal of Official Statistics* 30 (2): 355–79.
- International Telecommunication Union. 2019. World Telecommunication/ICT Development Report and Database, Retrieved (<https://data.worldbank.org/indicator/IT.NET.USER.ZS>).
- Long, J.S., and Freese, J. 2014. *Regression Models for Nominal and Ordinal Outcomes*. College Station, Texas: Stata Press.
- Medway, R.L., and Fulton, J. 2012. When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates, *Public Opinion Quarterly* 76 (4): 733–46.
- Millar, M.M., O’Neill, A.C., and Dillman, D.A. 2009. *Are Mode Preferences Real?* Washington: Social and Economic Sciences research Center.
- Olson, K., Smyth, J.D., and Wood, H.M. 2012. Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination, *Public Opinion Quarterly* 76 (4): 611–35.
- Rybak, A. 2018. The Selection of Survey Modes on the Basis of Respondents’ Preferences Analysis, *Ask: Research and Methods* 27 (1): 23–39.
- Smyth, J.D., Olson, K., and Kasabian, A. 2014. The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement, *Survey Research Methods* 8 (3): 137–52.
- Smyth, J.D., Olson, K., and Millar, M.M. 2014. Identifying Predictors of Survey Mode Preference, *Social Science Research* 48: 135–44.
- Sztabiński, P.B. 2018. Does Mixed Mode Data Collection Improve the Achieved Sample? A Comparison of the ESS 7 and a Mixed Mode Experiment, in: F. Sztabiński, H. Domański, and P.B. Sztabiński (eds.), *New Uncertainties and Anxieties in Europe: Seven Waves of the European Social Survey*. Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang.

- Sztabiński, P.B., and Sztabiński, F. 2016. Does Mixed Mode Data Collection Improve the Achieved Sample? A Comparison of the ESS PAPI Survey and a Mixed Mode Experiment, Retrieved 22 September 2017 ([www.ifispan.pl/wp-content/uploads/2016/03/PBS\\_FS\\_Sztabinski\\_ESS7\\_dissconf.pdf](http://www.ifispan.pl/wp-content/uploads/2016/03/PBS_FS_Sztabinski_ESS7_dissconf.pdf)).
- Tarnai, J., and Paxson, M.C. 2004. Survey Mode Preferences of Business Respondents, *59<sup>th</sup> Annual Conference, American Association for Public Opinion Research*.
- Tourangeau, R.J., Brick, M., Lohr, S., and Li, J. 2016. Adaptive and Responsive Survey Designs: A Review and Assessment, *Journal of the Royal Statistical Society. Series A: Statistics in Society*.
- Vannieuwenhuyze, J.T. a. 2013. On the Relative Advantage of Mixed-Mode versus Single-Mode Surveys, *Survey Research Methods* 8 (1): 31–42.
- Villard, A., and Fitzgerald, R. 2017. Using Mixed Modes in Survey Research: Evidence from Six Experiments in the ESS, in: M.J. Breen (ed.), *Values and identities in Europe: Evidence from the European Social Survey*. Abingdon: Routledge, pp. 273–310.
- Wardle, H., and Robinson, C. 2007. Choosing Web Surveys: Mode Choices among Youth Cohort Study Respondents, in: M. Trotman, T. Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, K. Morris, D. Birks, A. Johnson, R. Baker, M. Rigg, S. Taylor, and A. Westlake (eds.), *The Challenges of a Changing World. Proceedings of the Fifth ASC International Conference*. Southampton: Association for Survey Computing, pp. 193–202.

*Biographical Note:* Adam Rybak is a Ph.D. student at the Faculty of Sociology, Adam Mickiewicz University in Poznań, Poland. His current research focuses on survey methodology, with using meta analysis of comparative survey projects, in particular.

ORCID iD: <https://orcid.org/0000-0003-3981-1081>

E-mail: [adam.rybak@amu.edu.pl](mailto:adam.rybak@amu.edu.pl)

## Aneks

### **1. Charakterystyki sondaży z bazy cd.**

#### **1.1. Wiek włączający do próby**

Kryteriów doboru do próby nie spełniały badania kierowane wyłącznie do określonych grup wiekowych, natomiast należy uwzględnić, że zarówno maksymalny, jak i szczególnie minimalny wiek włączający do populacji wahały się. Jest to istotne w przypadku porównywania rozkładu wieku pomiędzy zrealizowaną w badaniu próbą, a populacją krajową. Najczęściej wiek minimalny przypadał na 15 (7586) i 18 (2211) rok życia. Duża część operatów imiennych opiera się na spisie wyborców w danym kraju, dlatego też wiek lat 18 był częstszy niż 15 w przypadku badań w Oceanii i Ameryce Północnej. Część programów, jak np. ESS i EB, narzuca przedział badanej populacji we wszystkich krajach (w obu od 15 lat), a część, jak np. CSES, ISSP i WVS pozostawia w tym względzie wolną rękę. W tych dwóch ostatnich programach wiek minimalny przyjmował wszystkie możliwe wartości z przedziału 14-21. Jeśli chodzi o wiek maksymalny, to jego stosowanie było sporadyczne. Nie jest także wykluczone, że jego raportowanie wynikało z błędu krajowej organizacji badawczej, która podała wiek najstarszego respondenta jako limit wieku.

#### **1.2. Dobór wewnątrzspołowy**

Brak potrzeby losowania wewnątrzspołowego udokumentowany był w 621 przypadkach. W sytuacji użycia operatów innych niż imienny (lub przypadku operatu telefonów komórkowych traktowanego jak imienny) konieczne jest przeprowadzenie losowania respondenta wśród członków gospodarstwa domowego. Najczęściej stosowaną metodą (980 sondaży) był dobór wg daty urodzin (Salmon i Nichols 1983) – respondentem zostaje osoba, której urodziny przypadły jako ostatnie, lub będą pierwszymi po momencie wywiadu – konieczny jest wybór jednego z tych dwóch rozwiązań. Metoda ta z jednej strony może skutkować zwiększeniem odsetka kooperacji, gdyż przez brak konieczności sporządzenia spisu domowników jawi się jako mniej inwazyjna, z drugiej strony stanowi pewne zagrożenie dla reprezentatywności próby. (Jabkowski 2015:144). Metoda siatki Kisha (Kish 1949) (557 sondaży) odwrotnie, losowanie następuje po sporządzeniu spisu wszystkich domowników, co przynosi też odwrotne niż metoda urodzinowa skutki.

Inne metody doboru losowego – niekiedy mowa o przypadkach słabiej opisanych, stosowano np. opis: „losowanie wśród domowników przeprowadziła aplikacja komputerowa” – wystąpiło 202 razy. Oryginalną metodę doboru wewnątrzspołowego, również kodowaną w kategorii „inne losowe” zastosowano w AB. Instrukcja realizacji terenowej nakazuje w większości krajów (wyjątek często dotyczy krajów muzułmańskich, ze względu zapewne na ewentualne skojarzenie z hazardem) użycie kart do losowania. Osoba, z którą nawiązano kontakt podaje imiona domowników, z których każdy zapisywany jest na osobnej karcie, a następnie respondent losowany jest z potasowanej talii. Według autorów instrukcji metoda ta pozwala łatwiej wytłumaczyć, dlaczego ankietę chce rozmawiać na osobności z konkretną osobą, szczególnie jeśli dotyczy to „kobiety, albo osoby młodej” (Afrobarometer 2011). Należy zaznaczyć, że AB uzupełnia to stosowaniem kwot płciowych i substytucją gospodarstw domowych. Możliwy jest także dobór nielosowy (549), który może polegać na stosowaniu wyłącznie kwot demograficznych, albo podobnych metod, jak np. metoda „najmłodszego

mężczyzny, najstarszej kobiety” będąca odmianą metody Trolldahla-Cartera (Gaziano 2005), albo stosowany w niemal całym LAPOP „frequency matching” (Castorena 2021). Zdecydowanie najczęściej metoda doboru wewnątrzspołowego w ogóle nie była opisana – dotyczy to aż 7584 przypadków.

### **1.3. Stosowanie doboru kwotowego i substytucji nierespondentów**

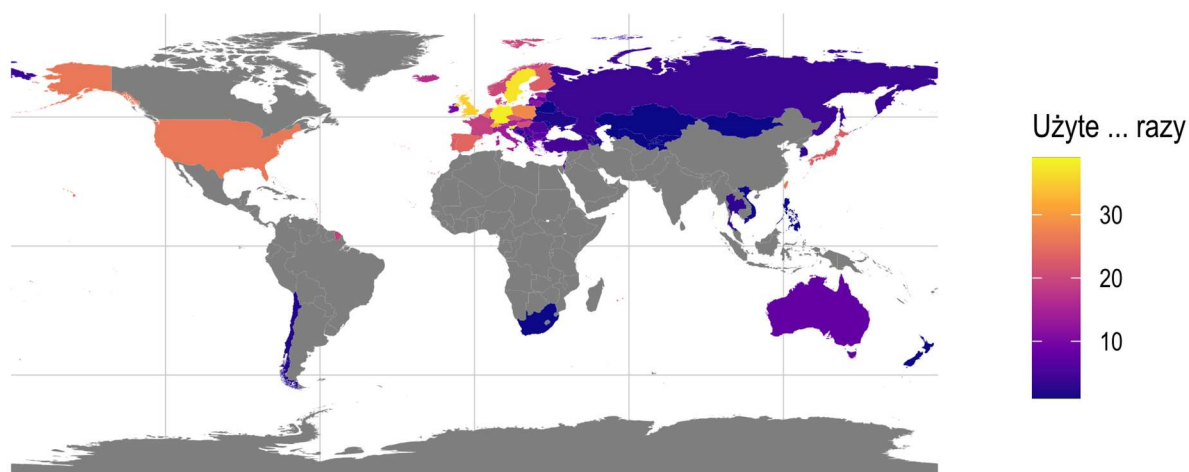
Dobór kwotowy może wystąpić na różnych etapach losowania. Jeśli w ogóle rezygnowano z losowania respondentów, to operat kodowany był jako nielosowy (323 sondaże). Jeśli kwoty (lub któraś z ich wspomnianych odmian) występowały jako jedyna metoda doboru respondenta wśród członków gospodarstwa domowego, to dobór wewnątrzspołowy kodowany był jako nielosowy (549). Kwoty mogły wystąpić także jako uzupełnienie któregoś z etapów losowania, jak w przypadku AB. Użycie kwot z definicji zakłada substytucję respondentów odmawiających lub tych, z którymi nie można nawiązać kontaktu, jednak również ona może wystąpić jako uzupełnienie losowych metod doboru próby.

Zarówno kwoty, jak i substytucja respondentów sprawiają, że trudne jest zaklasyfikowanie badania jako reprezentatywnego. Kwoty lub substytucje zastosowano w 7582 przypadkach. Dokumentacja pozwala stwierdzić, że wolne od tych metod było 1280 sondaży. Najwięcej sondaży, które miały nie stosować substytucji było w ISSP – 641. Kolejne to ESS (233), EQLS (125), CSES (84) i WVS (56). W przypadku LITS brak substytucji i kwot występuje w 35 przypadkach, jednak specyficzna i słabo opisana metoda doboru wewnątrzspołowego nie pozwala określić ich, według dokumentacji, jako badań w pełni losowych. Wśród badań niestosujących substytucji i kwot najwięcej jest sondaży z Europy (907) i Azji (191), jednak biorąc pod uwagę proporcje, w obrębie kontynentów największy odsetek badań, które ich nie stosowały występuje w Oceanii (50%) i Ameryce Północnej (40%) w porównaniu do 13% w Europie i 12% w Azji.

### **1.4. Kwestie relacji z respondentem**

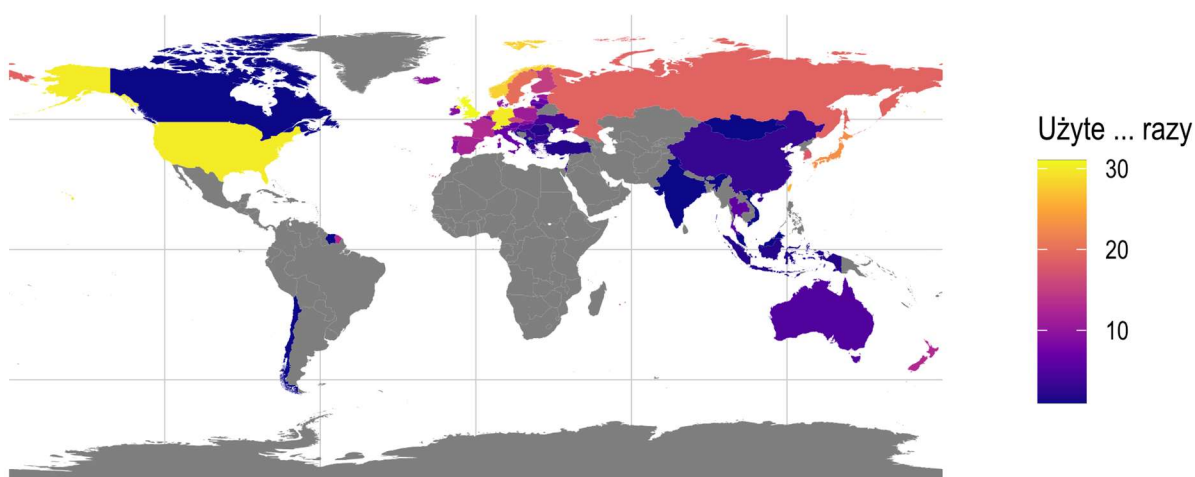
W celu zwiększenia odsetka zrealizowanych wywiadów w sondażach stosuje się różne metody zachęcania respondentów do wywiadu. Jedną z nich jest wysłanie (lub pozostawienie przez ankietera) listu zapowiedniego, który informuje członków gospodarstwa, albo w przypadku operatu imiennego konkretną osobę, o przyszłej wizycie ankietera. Często wyjaśnia on także cele badania, sens stosowania losowego doboru, może także zawierać niektóre wyniki poprzednich rund programu sondażowego. Listy zapowiednie zastosowano w 743 sondażach. W 694 przypadkach zakodowano ich brak – mowa o sytuacji, gdy w dokumentacji występowała odpowiednia rubryka lub kwestia ta (rzadko) była omówiona w instrukcji badania.

Listy zapowiednie najczęściej stosowano w programie ISSP (254), ESS (197), EQLS (125) CSES (60) i LITS (35). Najczęściej stosowano je w Europie (611) – szczególnie w Słowenii (39), Niemczech (38), Szwecji (37), Wielkiej Brytanii (35) i Szwajcarii (32). W drugiej kolejności najpopularniejsze były one w Azji (94), przede wszystkim na Tajwanie (26) i w Japonii (23). Warto także odnotować, że 26 razy zastosowano je w Stanach Zjednoczonych. Częstość zastosowania listów zapowiednich w poszczególnych krajach przedstawia wykres 8.



Wykres 8. Listy zapowiednie wg krajów – liczba sondaży, w których je stosowano

Inną metodą mającą zwiększyć odsetek realizacji są zachęty materialne. Może to być zarówno gotówka – w 4 rundzie ESS w Szwajcarii część próby dostawała nawet 60 franków, jak i drobny upominek – breloczek do kluczy w ESS7 w Polsce, czy składana parasolka w 3 rundzie CSES w Korei. Zachęty materialne najczęściej stosowano w ISSP (222 razy), ESS (159), CSES (51) i EVS (41). Najczęściej stosowano je w Europie (373 razy) – szczególnie w Wielkiej Brytanii i Szwajcarii (oba kraje po 31), oraz w Niemczech (30). Druga w kolejności była Azja (89) – 26 razy na Tajwanie, 23 razy w Japonii i 18 razy w Korei. W Stanach Zjednoczonych zastosowano je 30 razy. Częstość użycia zachęt materialnych dla respondentów przedstawia wykres 9.



Wykres 9. Zachęty materialne wg krajów - liczba sondaży, w których je stosowano

Użycie zarówno zachęt materialnych, jak i listów zapowiednich notowane jest w dokumentacji od 1996 roku. Najwięcej razy listów zapowiednich użyto w roku 2012 (70), a zachęt materialnych w 2016 (52).

Wpływ na odsetek odpowiedzi ma także liczba prób kontaktu, jaką podejmują ankieterzy, zanim zrezygnują z przeprowadzenia wywiadu z konkretnym respondentem. Niekiedy dokumentacja zawiera informacje na temat godzin lub dni tygodnia, w których miały

być podejmowane kontakty, jednak w bazie zakodowana jest sama minimalna liczba kontaktów. Największe wartości notujemy w Szwajcarii, 100 kontaktów (2 runda CSES) w przypadku sondażu *mixed-mode* łączącego wywiad telefoniczny z ankietą pocztową, oraz 80 (5 runda WVS) w przypadku sondażu stosującego wywiad kwestionariuszowy – jednak i tu możemy podejrzewać, że osoba wypełniająca formularz metodologiczny mogła mieć na myśli kontakt telefoniczny, badania CATI są w Szwajcarii niezwykle popularne (Stoop i in. 2010). Także w przypadku 75 prób kontaktu w Norwegii (CSES3) i 20 prób kontaktu w trzech sondażach w Szwecji (7, 8 i 9 runda ESS), choć sondaż stosował technikę wywiadu kwestionariuszowego, to komentarz w dokumentacji uściśla, że chodziło o kontakty telefoniczne. Wreszcie w EES w 2009 roku w 17 sondażach CATI stosowano minimum 15 kontaktów.

Zdecydowanie najczęściej stosowanym standardem były 3 próby kontaktu – zastosowano je w 2853 sondażach. Drugie w kolejności były 4 próby kontaktu – zastosowane w 594 przypadkach. Jedynie 148 badań, których dokumentacja poruszała kwestię liczby prób kontaktu określiło, że nie było ustalonego minimum. W badaniach stosujących ankietę pocztową zliczano ilość przesłanych kwestionariuszy – jeśli było to możliwe. Tutaj najczęstszą wartością były 2 kwestionariusze (38 sondaży), a drugie w kolejności 3 kwestionariusze (30). Jeśli porównać poszczególne programy, to częstość większą lub równą 20 kontaktom najczęściej odnaleźć można w dokumentacji CSES – w 7 sondażach przeprowadzonych różnymi technikami.

### **1.5. Praca ankierów**

Aspektami pracy ankierów i używanych przez nich narzędzi uwzględnionymi w bazie są sposób wynagradzania, liczba ankierów uczestniczących w badaniu, oraz czas potrzebny na przeprowadzenie przez nich badania w oparciu o kwestionariusz. Odnośnie pierwszej zmiennej – pytanie dotyczy tego, czy ankierzy pracują za stałą stawkę, czy też są opłacani od wykonanej ankiety. Takie zróżnicowanie może mieć dwojaki skutek, osoby dostające wynagrodzenie za wypełnione ankiety będą z jednej strony bardziej zmotywowane, żeby nawiązać kontakt z respondentem i przekonać go do udzielenia zgody na udział w badaniu. Jednocześnie mogą być przez to motywowane do stosowania rozmaitych nadużyć, od naruszania procedury doboru wewnątrzzespołowego, po wypełnianie kwestionariuszy „z sufitu”. Temat kontroli nadużyć porusza następny podrozdział.

Należy dla porządku zaznaczyć, że niekiedy opisany sposób wynagradzania łatwo wpasowywał się w jedną z dwóch kategorii, kiedy indziej był na tyle skomplikowany, że potrzebna była arbitralna decyzja – do której kategorii mu bliżej. Jedynie pięć programów badawczych zawiera tę informację w dokumentacji: ESS, EVS, ISSP, ISSP-SOR i WVS. Dla wszystkich programów liczba sondaży, przy których ankierów opłacano od kwestionariusza była istotnie większa od sytuacji przeciwnej. Waha się to od około dwukrotności w przypadku ISSP-SOR, przez ponad trzykrotność w przypadku WVS, ponad czterokrotność w ISSP i ESS, po sześciokrotność w przypadku EVS. W podziale na kontynenty otrzymujemy niemal siedmiokrotność w Azji i sześciokrotność w Europie, ponad dwukrotność w Afryce i Ameryce Łacińskiej. Jedynie w Oceanii trzy na czternaście badań było opłacanych „od kwestionariusza”, a w Ameryce Północnej – żadne z 23.

Ilość ankietników przeprowadzających sondaż nie wydaje się przekładać bezpośrednio na jakość pomiaru. W szczególnym przypadku, gdy w trakcie badań terenowych pojawią się jakieś wydarzenia istotne ze względu na badane zmienne – szybsze zakończenie badań w wyniku większej siły roboczej może zmniejszyć szkody wynikające z takiego zdarzenia. Większa ilość ankietników pozwala także na większą elastyczność w podziale pracy, między innymi wyznaczenie ankietników doświadczonych, wysyłanych do jednostek trudnych do przekonania do udziału w badaniu (Tourangeau i in. 2016). Nie posiadam żadnych informacji o ilości ankietników w sondażach stosujących technikę *self-completion*, lub jej połączenie z wywiadem kwestionariuszowym. Dla ankiet pocztowych i internetowych oczywiście także nie zliczam ich ilości.

Jeśli chodzi o wywiad kwestionariuszowy, to najmniejsza ilość ankietników występowała w badaniach w bardzo małych krajach, i.e., Luksemburgu i Malcie (EQLS odpowiednio 1 i 3) – 20 osób. Największa zaś w ESS1 w Czechach – 409. Średnia wartość dla wszystkich wywiadów kwestionariuszowych to 115 ankietników na badanie. W przypadku badań CATI średnia wartość to 72 ankietników telefonicznych. Waha się ona od 15 w EES99 w Finlandii, do 400 w CSES1 w Szwajcarii. W przypadku obu rodzajów badań *mixed-mode* mamy jedynie po 3 przykłady dla każdego. W badaniu CSES3 w Norwegii, łączącym wywiad kwestionariuszowy i telefoniczny brało udział 84 ankietników. W badaniu CSES4 w Wielkiej Brytanii, łączącym wywiad kwestionariuszowy, *self-completion* i ankietę internetową ankietników było 202. Natomiast w fali 5 tego samego badania w tym samym kraju było ich 208, a do pozostałych technik w schemacie doszła jeszcze ankietę pocztową. Wśród schematów mieszanych bez wywiadu kwestionariuszowego wyróżnia się badanie ABS5 z Australii, które połączyło techniki ankietę pocztową i internetową - konkretnie był to model sekwencyjny, z *push-to-web*, a w razie jego niepowodzenia przesłaniem kwestionariusza pocztowego, połączony z ponagleniami telefonicznymi. W jego przypadku liczba 14 ankietników dotyczy w istocie osób dokonujących ponagleń telefonicznych. W badaniu CSES3 w Szwajcarii, łączącym wywiad telefoniczny z ankietą pocztową ankietników było już 280. Największą liczbę ankietników (telefonicznych) w bazie wykorzystało badanie w ramach CSES2 w Szwajcarii, łączące wywiad telefoniczny z ankietą pocztową – było ich aż 800.

Jeśli chodzi o długość trwania wywiadów, to zdecydowanym rekordzistą jest tu sondaż będący sondażem suplementarnym (ale niezależnie wykonany), którego potrzeba miała wynikać ze złego sformułowania pytań w badaniu EES14. Był on przeprowadzony techniką wywiadu kwestionariuszowego w Słowenii, Rumunii, Portugalii, Irlandii i Wielkiej Brytanii i trwał średnio 8 minut. Najdłużej trwały wywiady w trzeciej fali WVS w Rosji, które mogły dochodzić aż do 325 minut – w bazie przyjęta została uśredniona wartość 182. Średnio długość wywiadów (z tych, o których posiadam informacje) wynosiła ok 51 minut, najwięcej w przypadku WVS – średnio około 83 minuty, najmniej EES – 25 i ARAB – 29.

## 1.6. Kontrola wywiadów

Zainteresowanie może budzić kontrola tego, czy uwzględnione w wynikach sondażu wywiady zostały w rzeczywistości przeprowadzone. Popularną metodą kontroli pracy ankietników jest tak zwany *back checking*, który polega na ponowieniu kontaktu z losową podpróbą respondentów i wykorzystaniu niewielkiego kwestionariusza, którego celem jest ustalenie, czy właściwy respondent rzeczywiście udzielił danych odpowiedzi na pytania

sformułowane w badaniu pierwotnym. Sens takiej kontroli dotyczy przede wszystkim wywiadów, gdyż w ich przypadku możemy mieć do czynienia z ankierską nierzetelnością. W przypadku ankiety pocztowej i internetowej kontrola może jednak dotyczyć tego, czy odpowiedzi udzielił właściwy członek gospodarstwa domowego, jednak należy wziąć pod uwagę, że przykładowo dokumentacja ISSP zawiera informację o *back checkingu* wyłącznie w przypadku sondaży wykorzystujących wywiad przynajmniej jako jedną z technik. Jeśli informacja o tej procedurze była uwzględniona w dokumentacji, to w zdecydowanej większości sondaży miała być ona przeprowadzona – jej brak raportowano w przypadku 282 badań, a obecność w 1676. Badaniami, które szczegółowo raportowały użycie, bądź nie, tego rodzaju kontroli, a jednocześnie wskazywały istotny odsetek badań, w których nie została ona przeprowadzona były CSES i ISSP. W przypadku pierwszego kontrolę przeprowadzono w 85 sondażach, a w 34 nie, w przypadku drugiego była ona zastosowana w 555 sondażach, a w 219 nie. Zróżnicowanie na wystąpienie kontroli bądź nie możemy odnaleźć jeszcze tylko w ESS, EVS i WVS, jednak tutaj brak kontroli raportowano dla bardzo niewielkiego odsetka badań (odpowiednio: 6 nie, 236 tak; 8 nie, 96 tak; 15 nie, 163 tak). Niezróżnicowaną informację, że *back checking* przeprowadzono w większości uwzględnionych w bazie wywiadów posiadamy dla AB, ABS, ARAB, CRRC, EQLS i ISSP-SOR.

### 1.7. Współczynniki realizacji próby

Odsetek realizacji raportowany był na dwa sposoby. Z jednej strony spotkać można było samą wartość *response rate*, często o niezdefiniowanym rodzaju. Z drugiej strony badania raportowały liczebności poszczególnych kategorii przeprowadzonych w całości lub częściowo wywiadów, braków kontaktu, odmów, innych przyczyn braku wywiadu, a także – najrzadziej – przypadków o niepewnej klasyfikacji (The American Association for Public Opinion Research 2023). Pozwalało to na samodzielne obliczenie wybranego rodzaju wskaźnika *response rate*. Tylko część badań podawała w dokumentacji jednocześnie obie wartości. Dodatkowo zdecydowana większość badań raportowała wyłącznie liczbę przeprowadzonych wywiadów.

Jedyne programy badawcze, w których dla żadnego sondażu liczba przeprowadzonych wywiadów nie była podana, to ASES i EUNB. Jednocześnie w przypadku AB i ABS brak danych na ten temat występował w odniesieniu do większości sondaży w programie. Ogółem nie posiadamy tej informacji jedynie o 996 badaniach. Część szczególnie niskich wartości dla tej zmiennej wynika z pewnych niespójności dokumentacji, gdy fragmenty badania zostały przeprowadzone oddzielnie. Nie ma wątpliwości, że najmniej wywiadów w całej próbie przeprowadzono w ramach LAPOP z roku 1978 w Kostaryce – 201. Niskie wartości można również odnotować w przypadku wczesnych badań EB w Luksemburgu – przykładowo w EB5 z 1976 roku przeprowadzono tam 268 wywiadów. Szczególnie wysokie liczby wynikają natomiast niekiedy ze zsumowania badań przeprowadzonych w kilku fazach. Ponadto duża liczba respondentów jest osiągalna w przypadku sondaży internetowych albo telefonicznych. Mimo to największą wartość omawiana zmienna osiąga w przypadku EB44.2bis z 1996 roku, przeprowadzonego w Niemczech, który w dokumentacji określony jest jako „*Mega-Survey*”, mający w zamierzeniu umożliwić konstrukcję precyzyjniejszych wskaźników dla określonych subpopulacji. Sondaż ten korzystał z techniki wywiadu kwestionariuszowego, natomiast liczba zrealizowanych w jego trakcie wywiadów to 9934.

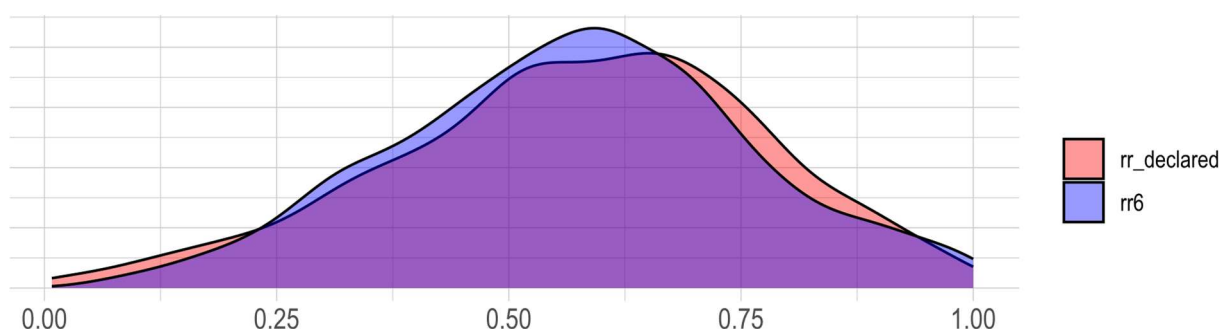
Programy, o których nie posiadamy informacji dotyczących zarówno współczynnika realizacji, jak i liczebności poszczególnych kategorii respondentów i osób nieuczestniczących w badaniu, to ASES, CCEB, CEEB, CONS, EB, ECS, EUNB, GALLUP, NEB/NDB. Wartości *response rate* dla wszystkich badań podają jedynie CCS i EASS. Dla większości badań możemy je odnaleźć w odniesieniu do programów: AB, CNEP, CRRC, CSES, EQLS, ESS, EVS, INTUNE, LITS, MEDW i NBB. Ogółem 1479 badań raportuje jakiś współczynnik realizacji. Sposób raportowania waha się pomiędzy podawaniem precyzyjnie zdefiniowanych wartości *contact*, *cooperation*, *refusal* i *response rate*, jak w przypadku AB, przez stosowanie częściowo opisanych (ale nieodpowiadających w pełni kategoriom AAPOR) wskaźników, jak na przykład w przypadku ESS i części ISSP oraz rundy 21 LAPOP, po prezentowanie całkowicie nieopisanych wartości w wielu innych programach.

Średnia wartość deklarowanego wskaźnika realizacji wynosi 0,57. Zarówno w tym przypadku, jak i w poniższych zestawieniach uwzględnione zostały wszystkie sondaże, dla których wartość współczynnika była podana lub możliwe było jej obliczenie – a więc również te, w których stosowano substytucję nierespondentów lub dobór kwotowy na dowolnym etapie losowania. To, czy sondaże takie będą miały zawyżony czy zaniżony współczynnik odpowiedzi w stosunku do badań stosujących rygorystyczne kryteria losowości, zależy od przyjętego sposobu raportowania braków odpowiedzi. Zazwyczaj jednak nie jest on sprecyzowany w dokumentacji. Teoretycznie sondaże stosujące dobór kwotowy powinny mieć *response rate* zawyżony w stosunku do reszty, gdyż kwoty demograficzne są mniej rygorystyczne we wskazywaniu osoby wyznaczonej do odpowiedzi niż każda inna metoda losowa. Natomiast praktyczny skutek stosowania substytucji może być różny, w zależności od tego, czy osoby wskazane w wyniku substytucji zostaną respondentami, czy też zwiększą liczbę braków odpowiedzi.

Najmniejsza wartość deklarowanego wskaźnika realizacji została zanotowana dla sondażu przeprowadzonego techniką wywiadu telefonicznego w Meksyku (LAPOP21) i wyniosła zaledwie 0,008. Z powodu pandemii COVID-19 w programie LAPOP zastosowano wtedy CATI po raz pierwszy w jego historii. Wartość 1 (czyli 100%), a więc nietypowa sytuacja, kiedy wszystkie wylosowane osoby wzięły udział w badaniu, nastąpiła, według dokumentacji, w 5. rundzie AB w Sierra Leone i w CNEB04 w Urugwaju. Oba sondaże zostały przeprowadzone techniką wywiadu standaryzowanego. Zbliżoną wartość osiągnęły tu 2 sondaże w Tajlandii (3. i 4. runda CSES). Warto o nich wspomnieć jednak w kontekście, poruszonego szerzej w Publikacji 1, problemu jakości dokumentacji. W obydwu przypadkach z podanych kategorii odmów i kompletnych lub częściowych wywiadów wynika, że *response rate* powinien wynosić dokładnie 1. Osoba przygotowująca raport ewidentnie oblicza wskaźnik realizacji w oparciu o stosunek pomiędzy wszystkimi wylosowanymi adresami a tymi, które spełniały warunki uwzględnienia w próbie i otrzymuje 0,993 (CSES3) oraz 0,984 (CSES4). Taki absurd zakodowano wyłącznie dlatego, że zmienna ta dotyczy dowolnego podanego *response rate*.

Dużo więcej informacji możemy wydobyć z sytuacji, kiedy dane o liczebnościach wszystkich kategorii zostały podane zgodnie z normami AAPOR lub też przelicznik wartości autorskich na wartości AAPOR można było uzyskać w wyniku personalnego kontaktu z przedstawicielem organizacji (ESS). Dla sondaży, które podały te liczebności, obliczony został współczynnik RR6. Dla 1650 sondaży umożliwiających to obliczenie średnia wartość

wskaźnika wynosiła 0,57, była więc niemal tożsama ze średnią wartością współczynnika deklarowanego w raporcie. Korelacja Pearsona pomiędzy tymi dwiema wartościami wynosiła 0,89 ( $p < 0.000$ ), ich rozkład przedstawia wykres 10.



Wykres 10. Porównanie rozkładu *response rate* deklarowanego w dokumentacji i obliczonego samodzielnie wg wzoru RR6 (skala osi y w przypadku użytej wizualizacji gęstości prawdopodobieństwa nie ma interpretacji)

Wśród programów, dla których obliczenie tego wskaźnika było możliwe, najniższą wartość przyjmuje on dla CRRC i EES. W przypadku CRRC mowa jedynie o dwóch sondażach w 11. fali programu, które były przeprowadzone techniką wywiadu kwestionariuszowego. Autorzy raportu tłumaczą, że bardzo niski *response rate* w przypadku Armenii wynikał z częściowego pokrycia się okresu badań z obostrzeniami wynikającymi z pandemii COVID-19. Dla Gruzji, w której wskaźnik *response rate* był jeszcze niższy, nie oferują takiego wyjaśnienia. Jeżeli chodzi o badanie EES, które uzyskało najniższy uśredniony *response rate* w bazie, obliczenie wskaźnika realizacji było możliwe wyłącznie dla dwóch fal programu (09 i 99), realizowanych głównie techniką wywiadu telefonicznego, a – jak wspomniano powyżej – przy jej stosowaniu należy zawsze spodziewać się bardzo wysokiego wskaźnika odmów. Programem, dla którego wszystkich fal podano informacje umożliwiające obliczenie wskaźników i który w tej kategorii ma najniższy *response rate*, jest EQLS. W ramach tego badania wywiady kwestionariuszowe były jedyną stosowaną techniką i według dokumentacji nie stosowano substytucji ani kwot. Może więc niski średni wskaźnik wynika w tym przypadku z rygorystycznie stosowanego doboru losowego, przy niedostatecznych nakładach na konwersję jednostek odmawiających – gdyż współczynnik kontaktów znajduje się w normie, a współczynnik odmów jest dla tego programu bardzo wysoki. Tabela 12 pokazuje średnie wartości RR6 oraz dodatkowo współczynników kooperacji, odmów i kontaktów w podziale na programy.

Tabela 12. Średnie wartości wskaźników realizacji sondaży, dla których dało się je samodzielnie obliczyć, oraz liczba takich sondaży w podziale na programy sondażowe

| Program  | RR6  | COOP2 | REF3 | CON3 | Liczba |
|----------|------|-------|------|------|--------|
| ABS      | 0.50 | 0.68  | 0.22 | 0.73 | 2      |
| ASIAB    | 0.54 | 0.63  | 0.29 | 0.82 | 2      |
| CCS      | 0.71 | 0.77  | 0.12 | 0.91 | 4      |
| CNEP     | 0.75 | 0.80  | 0.09 | 0.92 | 2      |
| CRRC     | 0.38 | 0.48  | 0.40 | 0.79 | 2      |
| CSES     | 0.53 | 0.60  | 0.29 | 0.88 | 153    |
| EASS     | 0.56 | 0.66  | 0.23 | 0.84 | 12     |
| EES      | 0.35 | 0.39  | 0.51 | 0.91 | 40     |
| EQLS     | 0.49 | 0.57  | 0.34 | 0.87 | 124    |
| ESS      | 0.60 | 0.65  | 0.26 | 0.93 | 233    |
| EVS      | 0.60 | 0.68  | 0.23 | 0.88 | 102    |
| ISSP     | 0.57 | 0.71  | 0.19 | 0.81 | 796    |
| ISSP-SOR | 0.79 | 0.83  | 0.10 | 0.94 | 12     |
| LAPOP    | 0.64 | 0.87  | 0.10 | 0.74 | 1      |
| LITS     | 0.67 | 0.76  | 0.18 | 0.87 | 29     |
| NBB      | 0.59 | 0.68  | 0.14 | 0.87 | 6      |
| WVS      | 0.64 | 0.74  | 0.19 | 0.86 | 123    |

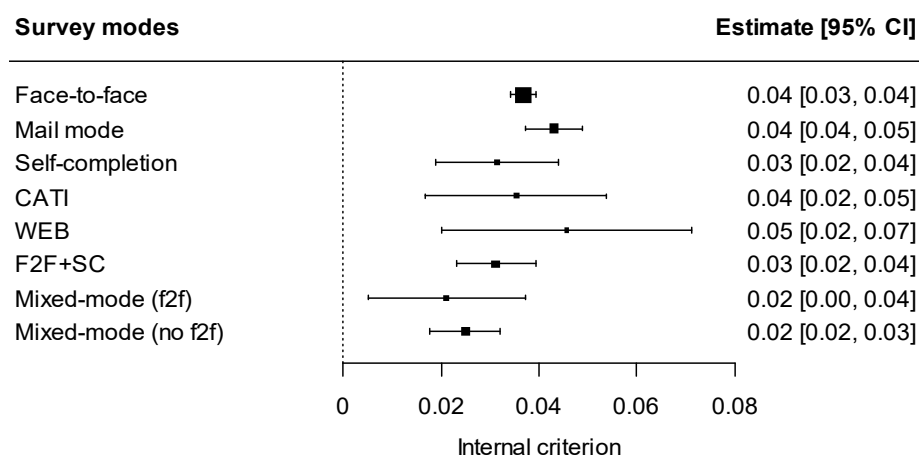
Najwyższy średni wskaźnik odpowiedzi osiąga ISSP-SOR, który był niezależnym od głównego programu powtórzeniem kwestionariusza ISSP z 2018 roku, dotyczącego głównie tematyki religijnej. Został on przeprowadzony techniką wywiadu kwestionariuszowego wyłącznie w Azji i Afryce, przy czym większość sondaży stosowała substytucje nierespondentów. Tak wysoki wynik może być spowodowany połączeniem luźnego podejścia do doboru losowego z prowadzeniem sondaży w krajach, które nie doświadczyły jeszcze tak wyczuwalnej na Zachodzie inflacji badań (ang. *oversurveying*). Drugi w kolejności jest CNEP, jednak tutaj znów mowa tylko o dwóch badaniach, z których jedno osiąga *response rate* 0,5, a drugie 1. Badanie drugie, przeprowadzone w 2014 w RPA, stosowało kwoty i substytucje. Możemy w tym raporcie dostrzec przykład jasnej sprzeczności na dwóch sąsiadujących stronach. Najpierw autor w oryginalny sposób oblicza *response rate* poprzez wskazywanie, ile kolejnych substytucji musiało zostać użytych do osiągnięcia każdego wywiadu – otrzymując wynik 0.34. Jednak poniżej, w odniesieniu do kategorii według podziału AAPOR, w rubrykach *odmowy*, *braki kontaktu* i *inne braki odpowiedzi* figurują wpisane zera – zatem obliczony samodzielnie wskaźnik realizacji musi wynieść 1.

Trzecim w kolejności programem z najwyższym wskaźnikiem odpowiedzi jest CCS z 1959 roku. Tylko jedno z pięciu badań stosowało tutaj substytucje, wszystkie korzystały z wywiadów kwestionariuszowych. Tak wysoki wskaźnik realizacji może wynikać z odległego czasu badania. Dawniej sondaż był wydarzeniem rzadkim, ciekawostką traktowaną z entuzjazmem (CCS był w końcu pierwszym w historii programem międzykrajowym).

## 2. Metaanalizy cd.

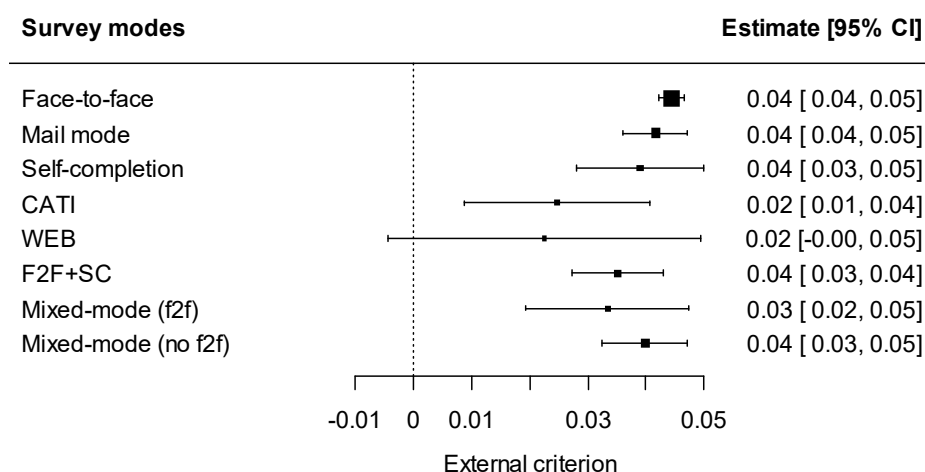
### 2.1. Metaanalizy w podgrupach cd.

W porównaniu sondaży stosujących różne techniki przy użyciu wewnętrznego wskaźnika reprezentatywności (wykres 11) najlepiej wypadają sondaże mixed-mode (bez f2f), które jak było wspomniane powyżej najczęściej wykonywane były w krajach nordyckich, które często korzystały z ciekawych operatów i wykorzystywały większość możliwych środków dotarcia do respondenta. Wartość górnej granicy przedziału ufności dla takich badań to 0,3. Jednoznacznie zły wynik uzyskały badania ankietą pocztową – jednak czytelnik przekonał się już, że jest to obraz niejednoznaczny i prawidłowa ocena wymaga co najmniej dodania do modelu operatu i interakcji pomiędzy operatem a techniką. Przypominam o problematycznych badaniach używających techniki pocztowej we Francji, gdzie stosowano operat adresowy, co zmuszało odbiorcę listu do samodzielnego przeprowadzenia prawidłowego doboru wewnątrzspołowego ostatecznego respondenta. Wszystkie techniki, które zakładały udział ankietera (F2F, S-C, F2F+S-C, *mixed-mode* (z f2f)) również uzyskały akceptowalną wartość górnego przedziału wskaźnika – 0,4.



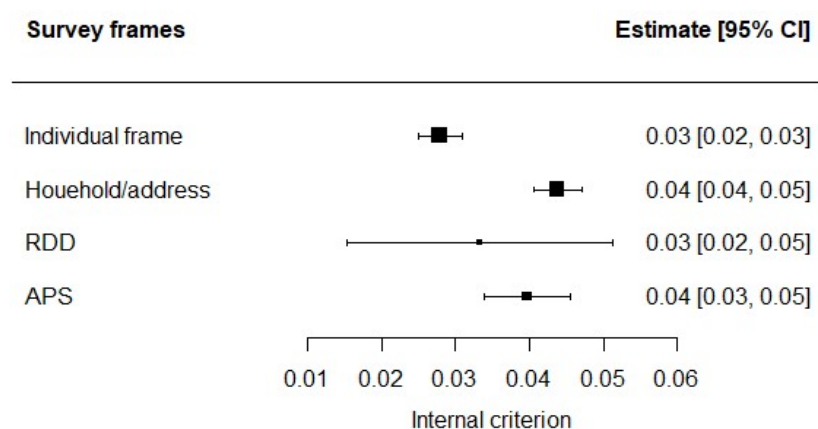
Wykres 11. Wewnętrzne kryt. reprezentatywności wg techniki sondażu - wyniki metaanalizy w podgrupach

Jeśli chodzi o ocenę za pomocą wskaźnika zewnętrznego (wykres 12), to paradoksalnie najlepiej wypadają badania CATI – co może wynikać z nieraportowania użycia kwot demograficznych w dokumentacji i tym samym błędnym zaliczeniu ich do badań w pełni losowych. Odnośnie reszty technik, to ich wyniki nakładają się tak bardzo, że ciężko mówić tu o wyraźnym rozróżnieniu jakości pomiaru.

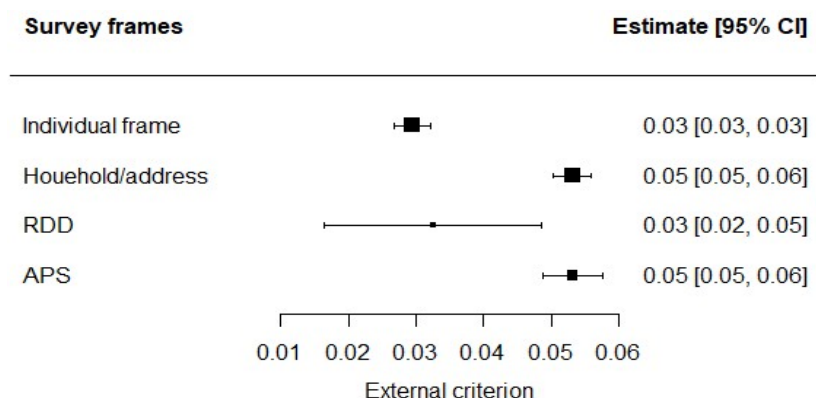


Wykres 12. Zewnętrzne kryt. reprezentatywności wg techniki sondażu - wyniki metaanalizy w podgrupach

W przypadku porównania operatów za pomocą wewnętrznego i zewnętrznego wskaźnika reprezentatywności (wykresy 13 i 14), to wyjątkowo uderzającym wnioskiem jest drastyczna przewaga operatu imiennego nad resztą, oraz brak zauważalnej różnicy pomiędzy próbą adresową a losowaniem przestrzennym i to w ocenie za pomocą obydwu kryteriów. Ugruntowuje to wniosek z Publikacji 2, że jeśli chodzi o sondażową infrastrukturę badawczą, to priorytetem w działaniach środowisk naukowych, szczególnie na styku ze środowiskami politycznymi, powinno być dążenie do tworzenia dobrej jakości operatów imiennych (Rybak 2023).



Wykres 13. Wewnętrzne kryt. reprezentatywności wg operatu - wyniki metaanalizy w podgrupach



Wykres 14. Zewnętrzne kryt. reprezentatywności wg operatu - wyniki metaanalizy w podgrupach

## 2.2. Wpływ fieldwork efforts na reprezentatywność – metaanaliza

Przedstawiona w tabeli 13 meta-regresja analizuje wpływ listów zapowiednich, zachęt materialnych i minimalnej liczby kontaktów z respondentem na reprezentatywność. Ograniczona jest do sondaży przeprowadzonych techniką wywiadu kwestionariuszowego, kwestionariusza dostarczonego przez ankietera i mixed-mode łączącego te techniki, oraz, jak wspomniano na początku, stosujących dobór losowy. Używając wewnętrznego kryterium reprezentatywności jako miary wielkości efektu osiągnęła niewielki pseudo- $R^2$  3,34%. Test moderatorów okazał się istotny statystycznie ( $F(df1 = 3, df2 = 871) = 7,9286; p < 0.0001$ ). Wpływ stosowania zachęt materialnych i listów zapowiednich w niewielkim stopniu, ale istotnie statystycznie zmniejszał wartość błędu. Jest to o tyle interesujące, że analizy I. Stoop (Stoop i in. 2010:102–13) odnoszące się do współczynnika realizacji znajdowały taki wpływ wyłącznie przy porównaniu stanu sprzed wprowadzeniem i po wprowadzeniu interwencji. Z drugiej strony, jak było wspomniane, większość dotychczasowych badań sugeruje, że nie ma wyraźnego związku między współczynnikiem realizacji, a błędem braku odpowiedzi. Możliwe więc, że wpływ *fieldwork efforts* na błąd braku odpowiedzi jest łatwiejszy od uchwycenia, niż na *response rate*.

Tabela 13. Wewnętrzne kryt. reprezentatywności sondażu - metaregresja  $N=875$

| ES = Int. Crit.   | estimate | se     | t       | p.value | conf.low | conf.high |     |
|-------------------|----------|--------|---------|---------|----------|-----------|-----|
| Intercept         | 0.0432   | 0.0021 | 20.8198 | 0.0000  | 0.0391   | 0.0473    | *** |
| Incentives1       | -0.0066  | 0.0025 | -2.6658 | 0.0078  | -0.0114  | -0.0017   | **  |
| Adv_letters1      | -0.0069  | 0.0025 | -2.7173 | 0.0067  | -0.0118  | -0.0019   | **  |
| Min_n_of_contacts | -0.0002  | 0.0002 | -0.7746 | 0.4388  | -0.0005  | 0.0002    |     |

W tabeli 14 znajduje się meta-regresja przeprowadzona na zbiorze o identycznych kryteriach włączania, jak poprzednia (różnice liczebności wynikają z braków danych), ale używająca kryterium zewnętrznego jako miary wielkości efektu. Pseudo- $R^2$  to także skromne 5,33%, a test moderatorów również jest istotny statystycznie ( $F(df1 = 3, df2 = 854) = 11.8908; p < 0.0001$ ). Wyniki są bardzo zbliżone do poprzedniego modelu i tym samym potwierdzają wnioski z niego płynące. Obie miary, oparte na niezależnej idei konstrukcji wskaźnika błędu, wskazały takie same zależności.

Tabela 14. Zewnętrzne kryt. reprezentatywności sondażu - metaregresja N=858

| ES = Ext. Crit.   | estimate | se     | t       | p.value | conf.low | conf.high |     |
|-------------------|----------|--------|---------|---------|----------|-----------|-----|
| Intercept         | 0.0510   | 0.0018 | 27.9765 | 0.0000  | 0.0475   | 0.0546    | *** |
| Incentives1       | -0.0051  | 0.0023 | -2.2351 | 0.0257  | -0.0097  | -0.0006   | *   |
| Adv_letters1      | -0.0093  | 0.0023 | -4.0101 | 0.0001  | -0.0139  | -0.0048   | *** |
| Min_n_of_contacts | -0.0002  | 0.0002 | -0.9297 | 0.3528  | -0.0006  | 0.0002    |     |

### 3. Zmienne z inwentaryzacji i ich rozkłady

Tabela 15. Opis zinwentaryzowanych zmiennych zawartych w bazie

| Nazwa                   | Opis                                                                                                                                                              | Wartości                                                                                                             | Liczebności albo rozkład                                                                                 |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| id                      | numer obserwacji                                                                                                                                                  | 1-10493                                                                                                              |                                                                                                          |
| survey                  | kod obserwacji:<br>'program"fala' 'kraj'                                                                                                                          | np. WVS7_NZ                                                                                                          |                                                                                                          |
| program                 | nazwa programu<br>sondażowego                                                                                                                                     | 30 skrótów (patrz Publikacja 1)                                                                                      |                                                                                                          |
| wave                    | nazwa fali programu                                                                                                                                               | często rok - nie zawsze tożsamy z rokiem badania!                                                                    |                                                                                                          |
| country                 | kraj wg Normy ISO<br>3166-1 alfa-2                                                                                                                                | 179 krajów (oraz dwa kody zawierające _P, oznaczające<br>badanie części kraju)                                       |                                                                                                          |
| year                    | rok rozpoczęcia<br>fieldworku                                                                                                                                     | 1959-2022                                                                                                            |                                                                                                          |
| nonrandom               | czy użyte substytucje<br>nierespondentów, albo<br>kwoty, albo inne<br>metody<br>nieprobabilistyczne na<br>dowolnym etapie?<br>(jeśli brak info., to<br>nielosowe) | 0-losowe<br>1-nielosowe                                                                                              | Losowe: 1134<br>Nielosowe: 9359                                                                          |
| substitution <br>quotas | stosowanie substytucji<br>nierespondentów, lub<br>kwot na dowolnym<br>etapie losowania                                                                            | "-9999"-brak<br>informacji<br>0-niezastosowane<br>1-zastosowane                                                      | Brak info.: 1631<br>Niezasosowane: 1280<br>Zastosowane: 7582                                             |
| hh_selection            | metoda doboru<br>wewnątrzspołowego                                                                                                                                | "-9999"-brak<br>informacji<br>0-nie dotyczy<br>1-siatka kisha<br>2-inna losowa<br>3-metoda urodzinowa<br>4-nielosowa | Brak info.: 7584<br>Nie dotyczy: 621<br>Kish: 557<br>Inna losowa: 202<br>Urodziny: 980<br>Nielosowa: 549 |
| lower_age_<br>limit     | minimalny wiek<br>potencjalnego<br>badanego                                                                                                                       | "-9999"-brak<br>informacji                                                                                           | Brak info.: 432<br>min: 14 max: 21 średnia: 15,72                                                        |

|                 |                                                                                                                   |                                                                                                            |                                                                                             |
|-----------------|-------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| upper_age_limit | maksymalny wiek potencjalnego badanego                                                                            | "-9999"-brak informacji<br>100-brak limitu                                                                 | Brak info.: 450<br>min: 50 max: 100 średnia: 99,39                                          |
| frame           | typ operatu                                                                                                       | "-9999"-brak informacji<br>1-imienny<br>2- adresowy/gospodarst w domowych<br>3-RDD<br>4-APS<br>5-nielosowy | Brak Info.: 630<br>Imienny: 763<br>Adresowy: 746<br>RDD: 788<br>APS: 7243<br>Nielosowy: 323 |
| sampling_stages | czy więcej niż jeden etap losowania?                                                                              | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 513<br>Nie: 275<br>Tak: 9705                                                    |
| f2f             | czy wywiad kwestionariuszowy face-to-face?                                                                        | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 1180<br>Tak: 9176                                                   |
| cati            | czy wywiad telefoniczny?                                                                                          | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 9479<br>Tak: 877                                                    |
| web             | czy ankieta internetowa?                                                                                          | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 10197<br>Tak: 159                                                   |
| mail            | czy ankieta pocztowa?                                                                                             | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 10097<br>Tak: 259                                                   |
| self-completion | czy technika, w której respondent samodzielnie wypełnia kwestionariusz dostarczony mu i odebrany przez ankietera? | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 10168<br>Tak: 188                                                   |
| mixed-mode      | jeśli co najmniej 2 z powyższych użyte                                                                            | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                  | Brak Info.: 137<br>Nie: 10083<br>Tak: 273                                                   |

|                    |                                               |                                                                                                                                                                              |                                                                                                                                                                             |
|--------------------|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mode               | technika w postaci zmiennej kategorialnej     | "-9999"-brak informacji<br>1-f2f<br>2-mail<br>3-self-c.<br>4-cati<br>5-web<br>6-f2f+s-c<br>7-mixed-mode (z f2f), poza specyficzną sytuacją powyżej<br>8-mixed-mode (bez f2f) | Brak info.: 137<br>Face-to-face: 9003<br>Mail mode: 155<br>Self-completion: 62<br>CATI: 792<br>WEB: 71<br>F2F+S.C.: 111<br>Mixed-mode (f2f): 62<br>Mixed-mode (no f2f): 100 |
| incentives         | czy zachęty materialne dla respondentów?      | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                                                                                    | Brak info.: 9165<br>Nie: 816<br>Tak: 512                                                                                                                                    |
| adv_letters        | czy listy zapowiednie?                        | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                                                                                    | Brak info.: 9056<br>Nie: 694<br>Tak: 743                                                                                                                                    |
| back_check         | czy back checki?                              | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                                                                                    | Brak info.: 8535<br>Nie: 282<br>Tak: 1676                                                                                                                                   |
| min_n_of_contacts  | minimalna liczba prób kontaktu z respondentem | "-9999"-brak informacji<br>0-brak minimalnej liczby                                                                                                                          | Brak info.: 6362<br>Brak minimum: 148<br>min: 2 max: 100 średnia: 3,48                                                                                                      |
| paidby             | czy ankieter opłacany od kwestionariusza?     | "-9999"-brak informacji<br>0-nie<br>1-tak                                                                                                                                    | Brak info.: 9410<br>Nie: 205<br>Tak: 875                                                                                                                                    |
| no_of_interviewers | liczba ankieterów                             | "-9999"-brak informacji<br>0-nie dotyczy<br>...A-liczba w całej organizacji                                                                                                  | Brak info.: 9886<br>Nie dotyczy: 54<br>Podana liczba w całej sieci ankieter.: 6<br>min: 14 max: 800 średnia: 114,93                                                         |
| avr_time_of_int    | średni czas wywiadu (w min.)                  | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 7209<br>min: 8 max: 182 średnia: 51,15                                                                                                                          |
| deff_p             | wartość współczynnika DEFFp                   | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 10346<br>min: 1 max: 2,18 średnia: 1,17                                                                                                                         |
| rr_declared        | dowolnie obliczany deklarowany response rate  | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 9014<br>min: 0,008 max: 1 średnia: 0,57                                                                                                                         |
| n_of_interviews    | liczba wywiadów (w tym „partial”)             | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 997<br>min: 200 max: 9934 średnia: 1173                                                                                                                         |
| n_of_refusals      | liczba odmów                                  | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 8736<br>min: 0 max: 54654 średnia: 1004                                                                                                                         |
| n_of_non_contact   | liczba braków kontaktu                        | "-9999"-brak informacji                                                                                                                                                      | Brak info.: 8822<br>min: 0 max: 22313 średnia: 575                                                                                                                          |

|                       |                                                                                   |                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-----------------------|-----------------------------------------------------------------------------------|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n_other_eligibile     | liczba braków odpowiedzi z kategorii inne                                         | "-9999"-brak informacji    | Brak info: 8835<br>min: 0 max: 9677 średnia: 184,6                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| n_unknown_eligibility | liczba przypadków, o których nie można stwierdzić, czy należały do próby          | "-9999"-brak informacji    | Brak info.: 8908<br>min: 0 max: 32205 średnia: 124,8                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| rr6                   | współczynnik odpowiedzi RR6 wg AAPOR                                              | bład-niemожność obliczenia | Niemожność: 8843<br>min: 0,035 max: 1 średnia: 0,57                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| rr6_se                | błąd standardowy RR6                                                              | bład-niemожność obliczenia | Niemожność: 8843<br>min: 0 max: 0,017 średnia: 0,009                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| coop2                 | współczynnik COOP2 wg AAPOR                                                       | bład-niemожność obliczenia | Niemожność: 8850<br>min: 0,079 max: 1 średnia: 0,673                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| ref3                  | współczynnik REF3 wg AAPOR                                                        | bład-niemожność obliczenia | Niemожność: 8850<br>min: 0 max: 0,896 średnia: 0,229                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| con3                  | współczynnik CON3 wg AAPOR                                                        | bład-niemожność obliczenia | Niemожność: 8850<br>min: 0,154 max: 1 średnia: 0,853                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| internal              | wewnętrzne kryterium reprezentatywności (liczone jak Rybak 2023)                  | bład-niemожność obliczenia | Niemожność: 7687<br>min: 0 max: 0,492 średnia: 0,044                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| internal_se           | błąd standardowy (liczony jak Rybak 2023)                                         | bład-niemожność obliczenia | Niemожność: 7687<br>min: 0,012 max: 0,224 średnia: 0,033                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| external              | różnica w proporcji kobiet w wieku 25-64 pomiędzy danymi UN dla kraju/roku i bazą | bład-niemожność obliczenia | Niemожność: 6485<br>min < 0,001 max: 0,356 średnia: 0,033                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| external_se           | błąd standardowy                                                                  | bład-niemожność obliczenia | Niemожność: 6485<br>min: 0,007 max: 0,038 średnia: 0,017                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| region                | region świata wg UN                                                               | 19 regionów                | Australia i Nowa Zelandia: 111<br>Karaiby: 90<br>Ameryka Środkowa: 252<br>Azja Środkowa: 101<br>Afryka Wschodnia: 261<br>Azja Wschodnia: 294<br>Europa Wschodnia: 1440<br>Afryka Środkowa: 80<br>Afryka Północna: 173<br>Ameryka Północna: 134<br>Europa Północna: 2091<br>Azja Południowo-Wschodnia: 252<br>Ameryka Południowa: 377<br>Afryka Południowa: 106<br>Azja Południowa: 168<br>Europa Południowa: 1827<br>Afryka Zachodnia: 268<br>Azja Zachodnia: 723<br>Europa Zachodnia: 1745 |

|         |                                                                                                  |                                                             |                                                                                                                                                                       |
|---------|--------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| cont    | kontynent wg UN                                                                                  | 6 kontynentów                                               | Afryka: 888<br>Azja: 1538<br>Europa: 7103<br>Ameryka Łacińska i Karaiby: 719<br>Ameryka Północna: 134<br>Oceania: 111                                                 |
| income  | grupa dochodu kraju wg WB '21                                                                    | wysoki, średni i niski dochód                               | Brak danych: 16<br>Wysoki dochód: 7204<br>Średni dochód: 2803<br>Niski dochód: 470                                                                                    |
| gdpperc | PKB per capita w USD dla kraju wg roku badania (za WB '21)                                       | blad-brak danych                                            | Brak danych: 967<br>min: 163,6 max: 118823,6<br>średnia: 23366,1                                                                                                      |
| panel   | zmienna pomocnicza dla niektórych programów, gdzie niektóre sondaże były niekiedy częścią panelu | 0-nie<br>1-tak<br>2-tak, ale pierwsza runda, więc jakby nie | Wartości tylko dla programów, w których występowała istotna liczba sondaży o losowym charakterze (przede wszystkim CSES).<br><br>panel: 78<br>panel-pierwsza runda: 5 |
| *_info  | 16 zmiennych dotyczących kompletności dokumentacji opisanych w Publikacji 1                      |                                                             |                                                                                                                                                                       |

### **Oświadczenie autora o wkładzie w przygotowanie publikacji naukowych**

Oświadczam, że mój wkład w powstanie publikacji zawartych w prezentowanym cyklu wynosi 100% i obejmuje:

stworzenie koncepcji badań, opracowanie metodologii badań, pozyskanie funduszy, zebranie danych zastanych, zarządzanie danymi, przeprowadzenie badań, wykonanie analiz statystycznych, wykonanie wizualizacji, napisanie maszynopisu pracy, odpowiedzi na recenzje pracy, wprowadzenie poprawek recenzyjnych i redakcyjnych.

Adam Rybak