

Reprezentatywność i trafność prognostyczna danych sondażowych CBOS w kontekście wyborów prezydenckich w 2025 roku

Opracowali
Ireneusz Sadowski i Dominik Batorski
(członkowie Rady CBOS)



Znak jakości przyznany CBOS przez
Organizację Firm Badania Opinii i Rynku



Fundacja Centrum Badania Opinii Społecznej
ul. Świętojerska 5/7, 00-236 Warszawa
e-mail: sekretariat@cbos.pl; info@cbos.pl
<http://www.cbos.pl>
(48 22) 629 35 69

Podstawowym celem Centrum Badania Opinii Społecznej (CBOS) jest prowadzenie badań opinii na użytek publiczny. Realizacja tego celu nierozdzielnie wiąże się z obowiązkiem zapewnienia najwyższej możliwej wiarygodności dostarczanych informacji, które dla obywateli stanowią istotne źródło wiedzy o całym społeczeństwie. Co więcej, do zadań Centrum należy także „inspirowanie, organizowanie i prowadzenie prac dotyczących badania opinii społecznej”. Niniejszy raport jest wyrazem realizacji obu tych zobowiązań. Stałe doskonalenie warsztatu badawczego jest wymuszone fundamentalnymi zmianami w sposobie realizacji sondaży, jakie nastąpiły w ostatnich kilkunastu latach, tak w Polsce, jak i na świecie. Istotną barierą jest tu jednak to, że ogłaszane publicznie wyniki badań opinii stosunkowo rzadko mogą być poddane bezpośredniej weryfikacji. Wyjątkową okazję w tym zakresie stwarzają wybory powszechne, pozwalające ocenić trafność badań przedwyborczych poprzez konfrontację z oficjalnymi rezultatami. Prognostyczne wykorzystanie rutynowo zbieranych danych daje sposobność do oceny wiarygodności stosowanej metodologii oraz rozpoznania tych jej elementów, które wymagają dalszej poprawy. Z tego założenia wyszła powołana w lipcu 2024 roku komisja Rady CBOS ds. audytu metodologicznego, która postanowiła, że wynik analiz prognostycznych zostanie opublikowany *ex ante*¹, a sam proces i jego metodologiczne aspekty omówione w niniejszym raporcie.

Na wstępie wypada podkreślić, że CBOS pełni również istotną rolę w odniesieniu do całego polskiego sektora badań opinii. Odpowiedzią na wspomniany wcześniej kryzys tradycyjnych metod realizacji badań sondażowych były fundamentalne zmiany technik ich prowadzenia przez agencje badawcze, zaś rosnącej złożoności procedur nie towarzyszyła proporcjonalnie większa przejrzystość i dostępność informacji o sposobie uzyskiwania publikowanych wyników. Metody doboru respondentów czy ważenia uzyskanych odpowiedzi traktowane są często jak sekrety przedsiębiorstwa. W tym kontekście CBOS, ze względu na swoją misję publiczną, powinien stanowić niezależny punkt odniesienia, nie tylko w warstwie wyników, ale również informacji o sposobie ich uzyskania. Potrzeba ta jest istotna dziś o tyle, że na przestrzeni ostatnich kilkunastu lat obserwujemy „efekt peletonu”, polegający na zmniejszającym się zróżnicowaniu rozkładów wynikowych uzyskiwanych przez różne pracownie badawcze². Może to być zarówno symptom konwergencji stosowanych metod (CATI lub CAWI w połączeniu ze zbieżnymi wagami analitycznymi), jak również zjawiska *herdingu*³, czyli uwzględniania wyników innych badań we własnych procedurach badawczych lub analitycznych (co redukuje ryzyko z punktu widzenia pojedynczej pracowni, ale zwiększa ryzyko porażki prognostycznej całego sektora badawczego). Z tego punktu widzenia istotne pozostaje, aby CBOS korzystał wyłącznie z własnych danych oraz dostarczał szczegółowej informacji na temat metodologii

¹ CBOS Flash: „Głosowanie w pierwszej turze wyborów prezydenckich (18 maja 2025) – prognoza wyników” ukazał się 15 maja 2025 roku (www.cbos.pl/PL/publikacje/flash/pliki/2025/fl_030_2025.pdf), zaś CBOS Flash: „Głosowanie w drugiej turze wyborów prezydenckich” został opublikowany 29 maja 2025 roku (www.cbos.pl/PL/publikacje/flash/pliki/2025/fl_033_2025.pdf).

² W latach 2011–2023 odnotowany został około 40-procentowy spadek wzajemnego odchylenia wyników.

³ Zob. definicja American Association for Public Opinion Research: www.aapor.org/wp-content/uploads/2022/12/Herding-508.pdf.

przeprowadzonych badań. Z drugiej strony badania prowadzone przez Centrum mogą być obarczone tzw. *house effect*, to jest specyficznym efektem dotyczącym postrzegania go jako podmiotu z sektora państwowego (Centrum działa na podstawie osobnej ustawy jako fundacja i posiada publiczne dofinansowanie). Test prognozy wyborczej pozwala na sformułowanie wstępnych wniosków w tej materii.

REPREZENTATYWNOŚĆ DWÓCH TYPÓW SONDAŻY

Wyjątkowość warsztatu CBOS wiąże się obecnie z tym, że prowadzi zarówno rozpowszechnione dziś w sektorze badawczym sondaże telefoniczne i internetowe, jak również badania bardziej tradycyjne, oparte na losowym doborze respondentów z operatu PESEL i osobistym kontakcie ankietera z respondentem (te realizowane są aktualnie w Polsce względnie rzadko). Te dwa typy badań prowadzone są niezależnie, zaś obserwowane między ich wynikami zbieżności i rozbieżności stanowią cenną informację diagnostyczną. Dzieje się tak dlatego, że reprezentatywność uzyskuje się w procedurach o znacząco odmiennej logice.

Kluczem do zrozumienia wartości porównawczej badań CBOS jest fakt, że opierają się one na dwóch odmiennych filozofiach zapewniania reprezentatywności. Pierwsza, określana mianem reprezentatywności proceduralnej, zakłada, że o wartości próby decyduje losowy mechanizm jej doboru – to on jest gwarantem braku systematycznych obciążeń. Na tej logice opiera się dobór z operatu PESEL. Druga koncepcja, reprezentatywność strukturalna, koncentruje się na rezultacie zgodności składu próby ze składem populacji pod względem znanych cech. W tym ujęciu badacz konstruuje swoiste „rusztowanie” dla próby, stosując dobór kwotowy i wagi analityczne, aby skompensować odchylenia powstałe na etapie rekrutacji respondentów. Ten model jest podstawą badań telefoniczno-internetowych. Naszym celem było zatem zderzenie wyników pochodzących z tych dwóch, znacząco różniących się, podejść metodologicznych.

Badanie na próbie losowej PESEL jest najdłużej stosowanym przez CBOS sposobem doboru respondentów. Nie jest to prosta próba losowa, ponieważ dobierana jest w sposób dwustopniowy – warstwowanie zapewnia odpowiednią reprezentację geograficzno-wiekową. Wylosowani respondenci otrzymują list zapowiedni, który pozwala im zdecydować, czy ma ich odwiedzić ankieter, czy ma on do nich zadzwonić, czy też wolą samodzielnie wypełnić ankietę przez internet⁴. Proporcje wyboru tych możliwości udzielenia odpowiedzi kształtują się od dłuższego czasu w granicach 65%/20%/15%. W tym rodzaju badań stosunkowo wielu respondentów nie udziela odpowiedzi na pytania o preferencje polityczne (zwykle jest to aż kilkanaście procent), co można hipotetycznie łączyć ze wzmiankowanym *house effect*.

⁴ Techniki realizacji wywiadu określa się angielskimi skrótami CAPI (*computer-assisted personal interview*), CATI (*computer-assisted telephone interview*), CAWI (*computer-assisted web interview*).

W badaniach telefoniczno-internetowych stosowana jest fundamentalnie inna logika, w której reprezentatywność jest aktywnie konstruowana, a nie dana „na wejściu” jak w próbie z operatu PESEL. Jest to dwupoziomowy mechanizm kontroli próby. Poziom pierwszy, realizowany w trakcie badania, polega na zastosowaniu doboru kwotowego. Choć pierwotny kontakt z respondentem w części telefonicznej następuje losowo (*Random Digit Dialing*), to o włączeniu go do próby decyduje spełnienie określonych kryteriów społeczno-demograficznych w ramach zdefiniowanych kwot. Poziom drugi, stosowany po zakończeniu etapu zbierania danych, to ważenie analityczne, które ostatecznie kalibruje strukturę próby do znanych parametrów populacji. W obu tych mechanizmach – kwotowaniu i ważeniu – kontrolowany jest rozbudowany zestaw cech, obejmujący m.in. wiek, miejsce zamieszkania, wykształcenie, status zawodowy, region, a także deklarowany udział w poprzednich wyborach. Wypada podkreślić, że jest to rozpowszechniona forma zapewniania reprezentatywności próbom telefonicznym, które stanowią współcześnie „produkt wysokoprzetworzony”. Warto jednocześnie nadmienić, że stosowanie bardziej forsownego ważenia silniej wpływa na zmniejszenie efektywnej wielkości próby (a więc powiększenie błędu standardowego). CBOS rozpoczął systematyczną realizację tego typu sondaży pod koniec 2023 roku. Także ta próba realizowana jest w modelu *mixed-mode*, gdzie komponent telefoniczny (ok. 90%) uzupełniany jest przez wywiady z panelu internetowego CAWI (ok. 10%).

W tego typu badaniach mniej jest odpowiedzi „trudno powiedzieć”, po części ze względu na to, że na udział w badaniu godzą się przeważnie osoby przeciętnie silniej przekonane do wyrażenia swojej opinii. Jest to jedna z istotnych różnic pomiędzy dwoma omawianymi typami badań. Inna kluczowa różnica dotyczy stopnia pokrycia populacji doбором respondentów do próby – zarówno obecność w rejestrze PESEL, jak i możliwość dotarcia do obywateli za pomocą połączenia telefonicznego jest dziś dalece niekompletna. Jednocześnie błędy pokrycia są w każdym przypadku odmienne (dotyczą nieco innych kategorii społeczno-demograficznych). Informacje na temat sondaży zrealizowanych przed pierwszą turą wyborów prezydenckich przedstawia tabela 1. Zaprezentowana jest w niej również przykładowa korekta rozkładu wieku i płci z zastosowaniem wag analitycznych (jak widać w próbie PESEL jest ona mniejsza).

Należy nadmienić, że reprezentatywność ze względu na jedną cechę nie oznacza koniecznie reprezentatywności próby ze względu na inne cechy. Skompensowanie rozkładu cech społeczno-demograficznych nie oznacza więc automatycznie, że rozkład opinii i preferencji w próbie stanie się jednakowy jak w populacji. Skuteczność takiej korekty jest tym większa, im silniej zmienne użyte do ważenia są skorelowane ze zmienną wynikową – w tym wypadku z rzeczywistymi preferencjami politycznymi. Jeśli więc na przykład osoby pomijane w doborze próby (błędy pokrycia) lub odmawiające udziału (autoselekcja) mają specyficzne poglądy, to ważenie według samej demografii pomoże tylko w takim stopniu, w jakim demografia te poglądy wyjaśnia. Dlatego kluczowe pozostaje określenie zakresu błędów pokrycia i charakterystyki autoselekcji respondentów, aby móc dobrać do modelu korekcyjnego jak najtrafniejsze zmienne.

TABELA 1. Charakterystyka sondaży CBOS zrealizowanych w pierwszej połowie maja 2025 roku

	Badanie na próbie losowej PESEL	Badanie CATI na próbie kwotowo-losowej RDD
Okres realizacji	5–14 maja 2025	12–13 maja 2025
Wielkość próby	1080	1001
Udział technik realizacji wywiadu	CAPI – 64,4% CATI – 23,7% CAWI – 11,9%	CATI – 91,2% CAWI – 8,8%
Udział kobiet w wieku 18–44 (dane surowe)	9,17%	7,50%
Udział kobiet w wieku 18–44 (dane ważone ^a)	10,95%	10,39%
Liczba respondentów, którzy odmówili odpowiedzi na pytanie o preferencje wyborcze	191	70

^a Wyłącznie podstawowe wagi społeczno-demograficzne CBOS

MODEL AUTOSELEKCJI ORAZ IMPUTACJI DANYCH

Wartość informacyjna surowych rozkładów z próby jest często niewielka ze względu na liczne błędy systematyczne powstające w toku badania. Perspektywa teoretyczna błędu całkowitego badania (*Total Survey Error*) pozwala zidentyfikować ich główne źródła. W naszym przypadku kluczowe znaczenie mają błąd pokrycia (związany z niekompletnością operatu PESEL czy bazy numerów RDD) oraz błąd braku odpowiedzi, zarówno na poziomie całego wywiadu, jak i poszczególnych pytań. Ponieważ grupy osób, których błędy te dotyczą, niemal nigdy nie są losową podpróbą populacji, ich pominięcie prowadzi do obciążonych wyników. Aby przeciwdziałać tym obciążeniom, konieczne jest modelowanie mechanizmów leżących u ich podstaw. Nasze prace prognostyczne poprzedziła więc analiza wzorców odpowiedzi oraz praca teoretyczna nad modelem, który roboczo nazwaliśmy „kaskadowym modelem autoselekcji”. Zakłada on, że na poszczególnych etapach realizacji badania – od znalezienia się w operacie, przez decyzję o udziale, aż po odpowiedź na pytanie o preferencje – następuje nielosowa selekcja, której mechanizm staraliśmy się zrozumieć i skompensować.

W przypadku badań PESEL pierwszym progiem jest meldunek zgodny z faktycznym miejscem zamieszkania, drugim elementem jest dostępność pod tym adresem w okresie realizacji badania, kolejnym – ograniczenia czasowe (możliwość rozmowy z ankierem), a następnie osobista motywacja i potrzeba zachowania anonimowości. Realizacja badania w trybie *mixed-mode* sprawia, że także sam wybór trybu udzielania odpowiedzi (CAPI/CATI/CAWI) zależy od niektórych cech społeczno-demograficznych (np. CATI oraz CAWI chętniej wybierają młodszy badani, czyli względnie mobilna kategoria społeczna). Wybiórczość procesów doboru i autoselekcji nie podlega jednak warunkowaniu jedynie ze względu na cechy demograficzne, ale również ze względu na cechy światopoglądowe. Podobnie rzecz ma się w przypadku respondentów, którzy odmawiają odpowiedzi

na niektóre pytania, np. o preferencje polityczne (zob. tabela 1). Zatem na poziomie pojedynczego pytania można mówić o ostatnim progu autoselekcyjnym.

Aby skompensować wspomniane efekty, zespół przygotowujący wyniki sondaży wykorzystał dwie względnie standardowe procedury statystyczne: wspomnianego już wcześniej **ważenia** (stosowanego rutynowo w badaniach CBOS), a także **imputacji**. Ta druga polega na modelowaniu prawdopodobnych odpowiedzi respondentów, którzy ich nie udzielili (braki danych). Opiera się ona na wielonomialnym modelu logistycznym, którego podstawą były zarówno cechy demograficzne (wiek, wielkość miejscowości zamieszkania, region), społeczne (uczestnictwo w nabożeństwach), jak i odpowiedzi na relewantne politycznie pytania, takie jak ocena sytuacji w kraju czy określenie swoich poglądów na skali lewica–prawica. W modelu takim szacowane są, na podstawie pozostałych obserwacji w próbie lub wielu próbach, prawdopodobieństwa udzielenia przez respondenta określonej odpowiedzi. Daje to możliwość probabilistycznego szacowania łącznego rozkładu interesującej nas zmiennej, a dzieje się to bez jednoznacznego przypisywania określonej osobie konkretnych poglądów. Procedura imputacyjna jest ważna o tyle, że alternatywą dla niej jest pominięcie braków danych lub ich proporcjonalne rozdzielenie – co w praktyce daje jednakowy efekt (oznacza nieme przyjęcie założenia, że braki odpowiedzi mają charakter losowy). Analiza cech skorelowanych z odmową odpowiedzi i odpowiedzią „trudno powiedzieć” pokazała natomiast, że ich pominięcie nie prowadzi do trafnego określenia rozkładu preferencji w całym społeczeństwie. W szczególności eksperymenty na starszych danych pokazały, że w myśl przyjętego modelu kaskadowego na ostatnich etapach autoselekcji (udział w badaniu, odpowiedzi na wybrane pytania) słabiej reprezentowani są zwolennicy partii opozycyjnych. Nie jest to fenomen nowy czy nieznan – w literaturze przedmiotu opisane są efekty wpływu kontekstu medialnego. Często przywoływanym przykładem jest tzw. efekt Bradleya, czyli czysto deklaratywne poparcie kandydata silniej akceptowanego w przestrzeni dyskursu medialnego, jednak w naszym modelu teoretycznym przyjęliśmy nie tyle deklaratywne przeniesienie poparcia, co reakcje unikowe (odmowa udziału, odpowiedź „nie wiem / trudno powiedzieć”, zupełny brak odpowiedzi).

Pominięcie dwóch wspomnianych procedur i publikacja wyników surowych oznaczałyby prawdopodobnie (było to przekonanie podzielane w zespole przygotowującym wynik publikacji) przekazanie opinii publicznej informacji o niższej wartości predykcyjnej – bowiem obciążonej opisanymi wyżej efektami. Wypada podkreślić, że rozkłady wynikowe nie były efektem innych, szczególnie „ręcznych”, interwencji w dane.

TRAFNOŚĆ PROGNOSTYCZNA WYNIKÓW

W tabeli 2 zaprezentowany został rozkład preferencji respondentów – osobno dla danych w sondażu na próbie z operatu PESEL i osobno w sondażu na próbie kwotowo–losowej RDD (CATI+CAWI). Pierwsze dwie kolumny prezentują rozkłady „surowe”, czyli odsetki uzyskanych odpowiedzi po oddzieleniu braków danych (odmów odpowiedzi i wskazań „trudno powiedzieć”). Kolejne dwie to

wynik procedur, które kompensują procesy autoselekcyjne (odmiennych w przypadku próby PESEL i próby telefonicznej). Kolumna 5 zawiera wynik na połączonych próbach i procedurach (to on został opublikowany w formie komunikatu przed pierwszą turą wyborów).

Należy podkreślić, że w przypadku badania na próbie PESEL wynik tylko w nieznacznym stopniu poddany został korekcie za pomocą wag społeczno-demograficznych (opartych na zaledwie trzech zmiennych). Większe znaczenie miała tu jednak imputacja, ponieważ relatywnie wyższa była liczba braków odpowiedzi. W sondażu telefoniczno-internetowym wagi opierały się na znacznie bardziej szczegółowej strukturze cech – dotyczyło to ośmiu zmiennych, w tym deklaracji o udziale i sposobie głosowania podczas wyborów parlamentarnych w roku 2023. Powody, dla jakich owe procedury są stosowane, dość dobitnie prezentują wyniki w tabeli. Surowe dane – z uwagi na procesy (auto)selekcyjne – dalece odbiegają od wyników wyborów, jakich można byłoby realnie się spodziewać. Pominięcie braków odpowiedzi prowadziłoby do dalszego pogłębienia różnicy pomiędzy czołowymi kandydatami.

Dalece inną sytuację przedstawiają kolejne kolumny (3–5). W przypadku skompensowania różnic względem znanych rozkładów w populacji i w samej próbie (czyli przy użyciu odpowiednio: ważenia i imputacji) poparcie dla Rafała Trzaskowskiego wynosiło prawie 35% w przypadku estymacji na danych PESEL i 31% w przypadku danych telefonicznych. Jeśli chodzi o poparcie dla Karola Nawrockiego, Sławomira Mentzena i Grzegorza Brauna, wyniki były nieco bardziej zbliżone, rozbiegały się zaś ponownie w przypadku poparcia Adriana Zandberga, Magdaleny Biejat i Joanny Senyszyn. Widać zatem, że dwa zestawiane źródła danych różniły się w szczególności pod względem balansu między notowaniami lidera rankingu oraz kandydatów partii lewicowych. Wrażliwość tych metod na poparcie w ramach szerszych bloków światopoglądowych wydaje się zatem nieco inna.

Jako dane o najwyższej wiarygodności potraktowaliśmy połączone zbiory (kolumna 5) – choćby ze względu na wyższe liczebności oferowały niższy komponent błędu losowego. Warto jednocześnie zwrócić uwagę, że zastosowanie kompletu wag (połączenie społ.-dem. i wyborczych) oraz wspólnej imputacji nie zawsze dawało w rezultacie prostą wypadkową dwóch wyników z kolumn 3 i 4. W szczególności zastosowanie wag wyborczych do danych PESEL wyraźnie zmniejszało szacunek dotyczący poparcia Sławomira Mentzena i Krzysztofa Stanowskiego. Trzeba mieć jednak na względzie, że retrospektywne deklaracje dotyczące głosowania w poprzednich wyborach parlamentarnych zwykle nie są w pełni miarodajne (i pozwalają na tym mniej pewną ekstrapolację, im więcej czasu upłynęło od owych wyborów)⁵. Należy zakładać, że w kolejnych latach pytanie o głosowanie w 2023 roku może być mniej pewnym „rusztowaniem” do zapewniania reprezentatywności wyników.

⁵ Przykładowo, według ekstrapolacji Adama Gendźwiłła, liczba osób, które 18 maja 2025 deklarowały w badaniu *exit poll*, iż w wyborach parlamentarnych w 2023 roku głosowały na Konfederację, była znacznie wyższa niż realny wynik tej partii w owych wyborach.

TABELA 2. Rozkład poparcia kandydatów prezydenckich w próbie PESEL, próbie telefonicznej oraz w obu łącznie (tylko osoby deklarujące udział w wyborach)

	Badanie na próbie PESEL (dane surowe)	Badanie na próbie telefonicznej (dane surowe)	Badanie na próbie PESEL (wagi społ.-dem. + imputacja)	Badanie na próbie telefonicznej (wagi społ.-dem.-wyborcze)	Obie próby razem ^a (wagi + imputacja)	Wynik wyborów	Różnica kolumn 5 i 6
	1	2	3	4	5	6	7
R. Trzaskowski	29,5	38,7	34,8	30,9	32,7	31,4	1,3
K. Nawrocki	23,7	20,9	27,6	26,3	27,5	29,5	-2
S. Mentzen	11,1	13,8	16,4	15,5	13,7	14,8	-1,1
S. Hołownia	4,1	5,8	5,4	5,8	6,2	5,0	1,2
G. Braun	3,3	4,8	4,4	4,0	4,4	6,3	-1,9
A. Zandberg	2,4	6,4	3,1	7,0	5,7	4,9	0,8
M. Biejat	2,7	5,1	3,8	6,0	4,9	4,2	0,7
M. Jakubiak	1,1	1,4	1,4	1,3	1,5	0,8	0,7
K. Stanowski	0,8	1,0	1,4	1,1	0,9	1,2	-0,3
J. Senyszyn	0,4	1,4	0,6	1,6	1,6	1,1	0,5
A. Bartoszewicz	0,4	0,2	0,6	0,2	0,6	0,5	0,1
M. Maciak	0,2	0,2	0,4	0,2	0,2	0,2	<0,1
M. Woch	0,1	0,1	0,1	0,1	0,1	0,1	<0,1
Brak odpowiedzi (odmowa lub brak zdania)	20,2	7,4	-	-	-	-	-
Suma	100	100	100	100	100	100	10,6 ^b

^a Wynik opublikowany w charakterze prognostycznym 15 maja 2025 roku

^b Suma wartości bezwzględnych poszczególnych odchyleń

W podsumowaniu wypada wskazać, że oszacowanie na podstawie połączonych zbiorów dało najbardziej zbliżony rezultat do wyników ogłoszonych przez PKW (łączne odchylenie to 10,6 punktu procentowego na 13 pozycjach), a największy błąd punktowej predykcji wyniósł 2 punkty procentowe. Błędy w przypadku każdego ze zbiorów wziętego osobno były wyższe – wyniosły 13 pp w przypadku próby PESEL (max. 3,4 pp) oraz 12,8 pp w przypadku sondażu telefonicznego (max. 3,2 pp). Jako rodzaj ciekawostki należy wspomnieć, że wynik najbardziej zbliżony do rezultatu wyborów oferował obliczony *ex post* model imputacyjny wyłącznie na danych PESEL, ale z zastosowaniem wag wyborczych (błąd łączny 8,5 pp). Szacowany na podstawie danych sondażowych poziom frekwencji wynosił 66,9% i różnił się tylko nieznacznie od faktycznego (67,3%).

Przed drugą turą wyborów prezydenckich CBOS zrealizował jedynie sondaż telefoniczny, podstawa do estymacji była zatem znacznie skromniejsza (a tym samym wyraźnie szerszy był przedział ufności

tego oszacowania: ponad ± 4 punkty procentowe). Zastosowanie analogicznej metodologii jak w pierwszej turze pozwoliło na predykcję frekwencji na poziomie 71,3%, poparcie dla Rafała Trzaskowskiego na poziomie 50,3%, zaś dla Karola Nawrockiego 49,7%. Według danych PKW wartości te wyniosły w całej populacji odpowiednio: 71,6%, 49,1% i 50,9%. A zatem odchylenie wyniku wynosiło 0,3 pp w przypadku frekwencji i 1,2 pp w przypadku poparcia każdego z kandydatów (mieściło się oczywiście w przedziale ufności). Jak zaznaczyliśmy w komunikacie przedwyborczym: „pewność wskazania zwycięzcy drugiej tury wyborów prezydenckich jest stosunkowo niska” (wysoki błąd standardowy był pochodną zarówno mniejszej próby, jak i silniejszej autoselekcji – a większa dyspersja wag stanowi o zwiększeniu błędu losowego).

WSTĘPNE WNIOSKI

Mądrzejsi o znajomość (ówczesnej) przyszłości możemy dziś skonstatować, że wiele spośród przyjętych założeń na temat procesów autoselekcyjnych potwierdziło się – w szczególności przypuszczenie, że przy poprawnym sformułowaniu modelu generowania danych (*data generating process*) można, a nawet należy stosować korektę danych surowych (ale także tych ważonych jedynie na podstawie wybranych cech demograficznych). Przy poziomie realizacji próby (tzw. *response rate* RR1), która prawie zawsze jest niższa niż 40%, procesy selekcyjne niemal zawsze nie mają prostego charakteru losowego. Oznacza to, że wypaczenia (błędy systematyczne) należy dodatkowo kompensować. Trafność prognostyczna uzyskanych szacunków należy ocenić jako akceptowalną z uwagi na fakt, iż wynik mieścił się w granicach błędu losowego – co stanowi powszechnie stosowane kryterium oceny wiarygodności. Wyniki pokazały jednocześnie, że oszacowanie punktowe zawsze należy traktować w kategoriach probabilistycznych i nie należy spodziewać się, że dla każdego z nich da się uzyskać dokładność zarezerwowaną dla badań *exit poll* (choć, po prawdzie, błąd w przypadku drugiej tury był jednakowy w sondażu CBOS i badaniu *exit poll* IPSOS, a tylko nieznacznie większy niż w *exit poll* OGB).

Przeprowadzony przez nas quasi-eksperyment pokazał, że żadna z technik badawczych i sposobów zapewnienia reprezentatywności nie oferowała jednoznacznie bardziej trafnych wyników. Zarówno sondaż oparty na doborze z operatu PESEL, jak i sondaż telefoniczny (*random digit dialing*) posiadały swoje specyficzne obciążenia. Dalszego ostrożnego badania wymagałaby weryfikacja hipotezy, że najlepsze własności prognostyczne daje zastosowanie dodatkowych wag wyborczych na danych z sondażu na próbie PESEL. W szczególności nie jest przesądzone, czy dane zbierane cyklicznie przez CBOS w ramach trackingu postaw i opinii (seria badań: „Aktualne problemy i wydarzenia”) powinny zostać objęte nowymi procedurami ważenia. Ta kwestia wymaga dalszego, starannego zbadania. Podobnie jak hipoteza dotycząca *house effect* – wprawdzie predykcja z danych Centrum systematycznie, choć nieznacznie, odchyłała się w kierunku kandydatów partii rządzących, to podobne odchylenie zanotowało większość spośród prywatnych agencji badawczych (owo odchylenie było zresztą przeciętnie wyższe niż w przypadku badań CBOS, więc trudno orzec, czy „marka” miała wpływ na reakcje respondentów).

REKOMENDACJE

Przedstawiając opinii publicznej wyniki badań, należy zadbać o ich możliwie wysoką wartość informacyjną. Wielu ekspertów zajmujących się badaniami sondażowymi zdaje sobie z pewnością sprawę, że wyniki zawierające duży odsetek odmów odpowiedzi, w tym odpowiedzi „nie wiem” czy „trudno powiedzieć”, należy traktować bardzo ostrożnie. Procedury ważenia weszły do repertuaru metodologii badań sondażowych już wiele lat temu i są powszechnie stosowane, więc wielu badaczy zdaje sobie również sprawę, że surowe dane sondażowe potrafią dalece odbiegać od standardu reprezentatywności (i wymagają odpowiednich korekt). Wydaje się jednak, że wielu odbiorców oczekuje przede wszystkim łatwej w odbiorze, a jednocześnie możliwie pełnej informacji o wyniku badania.

W przypadku zamieszczania odsetka niezdecydowanych (lub braków odpowiedzi) dobrą praktyką byłoby stosowanie komentarza lub przypisu z informacją, że odsetek ten nie przedstawia ani grupy jednorodnej, ani takiej, która ma kompletnie nieokreślone preferencje. Często w odsetku tym mieści się, przykładowo, więcej osób niechętnych rządowi niż mu sprzyjających (pokazują to dane z wielu lat, a nie tylko te z maja). Druga możliwość polega na zastosowaniu imputacji, czyli probabilistycznego modelowania opinii lub preferencji osoby, która nie udziela konkretnej odpowiedzi na dane pytanie. Ma to sens zwłaszcza w sytuacjach, kiedy wyniki sondażu czytane są przez opinię publiczną w kontekście przedwyborczym – jako rodzaj prognozy. Wówczas informacja oparta na całości posiadanych danych szczegółowych ma dla odbiorcy wartość większą niż same zagregowane liczebności odpowiedzi.

Odpowiedzialne przedstawianie rozkładu opinii w społeczeństwie wymaga również śledzenia go w czasie. Jest to niezmiernie istotny walor danych posiadanych przez CBOS (niektóre serie sięgają lat osiemdziesiątych). W przypadku badań przedwyborczych istotne jest dokładniejsze śledzenie zmian poparcia w czasie. Pojedyncze badania przed każdą z tur ograniczało możliwość analizy trendów, które w okresie kampanii potrafią odgrywać relatywnie duże znaczenie.

Jednocześnie wskazane wydaje się przeprowadzenie dalszej kalibracji sposobu kwotowania stosowanego w badaniach telefonicznych. W przypadku badań na próbie PESEL istotne jest z kolei przeprowadzenie analiz pod kątem niekompletności operatu. Różnice pomiędzy populacją obywateli-rezydentów a wspomnianym rejestrem stały się dziś bardzo znaczące – dotyczą zarówno braku meldunku, nieaktualnego adresu, jak i czasowych oraz stałych migracji zagranicznych.

W końcu – w dłuższej perspektywie czasowej – potrzebna jest dalsza szczegółowa diagnoza reprezentatywności i trafności prognostycznej prowadzonych sondaży. W tym celu należy przeprowadzać dodatkowe analizy i szereg eksperymentów metodologicznych, które powinny objąć nie tylko kwestie wymienione wcześniej, ale również aspekty takie jak konstrukcja kwestionariusza, formuły pytań ankietowych, działanie siatki ankierskiej, sposoby dotarcia do respondentów oraz ich ewentualnej gratyfikacji, a także nieprobabilistyczne schematy badawcze⁶.

⁶ W tym zbadanie w praktyce nowych technik, stosowanych niekiedy z powodzeniem przez ośrodki badawcze, jak losowa rekrutacja cyfrowa (*random digital recruitment*), czyli „celowany” dobór użytkowników internetu.



Powyższa publikacja jest głosem w dyskusji w ramach nowego projektu Centrum – CBOSLab. Refleksja metodologiczna jest niewątpliwie jednym z najważniejszych zadań CBOS ze względu na zapewnienie możliwie największej wiarygodności prezentowanych badań opinii publicznej. Oprócz projektu dotyczącego metodologii badań w ramach CBOSLab we współpracy z naukowcami odbyły się dwie konferencje, na których dyskutowano wyniki badań CBOS:

– *Co powinniśmy wiedzieć o obcokrajowcach w Polsce?*

– *Młodzi Polacy o sobie.*

Badania te spotkały się z dużym zainteresowaniem badaczy i mediów. Kolejne projekty z tej serii są w przygotowaniu.