

## MIĘDZYNARODOWY ROK STATYSTYKI 2013 KONFERENCJA NAUKOWA *STATYSTYKA — WIEDZA — ROZWÓJ*

**Włodzimierz OKRASA**

### Eksperymenty sondażowe w kontekście badań ewaluacyjnych

---

W tradycyjnym rozróżnieniu pomiędzy danymi obserwacyjnymi i danymi eksperymentalnymi zawiera się nie tylko informacja o sposobie, w jaki one powstały (w praktyce dane statystyczne utożsamia się najczęściej z danymi reprezentacyjnymi, jak to ma miejsce w statystyce publicznej). Zawiera się w nim zarazem wskazanie na odrębne procedury analizowania związków pomiędzy zmiennymi, tzn. badanie zależności statystycznych w populacji w pierwszym znaczeniu oraz ustalanie relacji przyczynowo-skutkowych w drugim przypadku.

O ile analizy prowadzone oddzielnie nie nastęrczają problemu, to pojawia się on wraz z dążeniem do interpretowania wyników badania sondażowego dotyczących zależności statystycznych w sposób wykraczający poza samo ich stwierdzenie. W szczególności poprzez odwołanie się do zależności przyczynowo-skutkowych, a to z racji dodatkowych warunków, jakie musiałyby być spełnione: manipulowanie zmienną niezależną (bodźcem), której zmiana musi poprzedzać zmienną wynikową (zjawisko nią opatrywane); losowe przyporządkowanie badanych do grupy poddanej działaniu bodźca (programu/interwencji); wyeliminowanie zależności pozornych pomiędzy zmienną przyczynową i wynikową (kontrolowanie ze względu na trzecią zmienną).

W celu generalizacji interpretowanych przyczynowo zależności niezbędne jest odwołanie się do takiej formuły postępowania badawczego, która spełnia łącznie wymienione warunki. Eksperyment sondażowy dostarcza takiej hybrydowej formuły badań o właściwościach wykraczających z jednej strony ponad cechy jej składników bazowych (Nock, Guterbock, 2010; Mutz, 2011), ale z drugiej strony ma ona też własne ograniczenia związane m.in. z nie do końca rozwiązywalnym problemem obciążenia selekcją (Barabas, Jerit, 2010; Gaines, Kuklinski, 2011) czy zjawiskiem przenoszenia efektów w czasie i przestrzeni poprzez dyfuzję i *spillover* (Transue i in., 2009).

Tezą artykułu jest uznanie, że eksperyment sondażowy (będący czymś jakościowo odmiennym od powszechnie stosowanego w badaniach metodologicznych sondażu eksperymentalnego — Okrasa, Scott, 1998) stanowi proceduralny odpowiednik — na poziomie generowania danych — paradygmatycznej strategii rozumowania przyczynowego odnoszonej do danych obserwacyjnych, traktowanych na podobieństwo danych eksperymentalnych dzięki odpowiedniemu uzupełnieniu pierwszych przez drugie. Dokonywane w ten sposób projekcje na populację (poprzez sondaż) rezultatów eksperymentu — w praktyce najczęściej zbioru różniących się od siebie wersjami wielu eksperymentów — oparte są na założeniu, że to, co wydarzy się w eksperymencie sondażowym odzwierciedla (i może być uogólnione) to, co wydarzy się w realnym świecie. Istotą tego założenia jest zatem trafność zewnętrzna (Mutz, 2011; Gaines i in., 2006). Współwystępowanie elementów konstytutywnych eksperymentu sondażowego jako metody jakościowo odmiennej od elementów składowych — sondażu i eksperymentu, mającej zarazem własności obydwóch, przedstawiono na schemacie (Nock, Guterbock, 2010).

Eksperyment sondażowy uchodzi generalnie w większym stopniu za eksperyment niż sondaż (Mutz, 2011), stanowi metodologicznie alternatywną strategię empiryczną stosowaną w fazie zbierania danych, w odróżnieniu od analitycznych, stosowanych przy przetwarzaniu danych, tj. ujmowania przyczynowości w badaniach w obrębie dyscyplin nieeksperymentalnych z dominującymi danymi obserwacyjnymi. Z interesującego tu punktu widzenia wyróżnić można w literaturze przedmiotu dwa typy strategii rozumowania przyczynowego:

- statystyczna, odwołująca się do pojęcia wyniku potencjalnego w ramach kontrfaktycznego paradygmatu Neymana-Rubina (Rubin, 1974; Holland, 1986; Rosenbaum 2002; Okrasa, 2012); charakteryzująca się brakiem specyfikacji modelu wpływu (zależności) w definicjach merytorycznych oraz bezwarunkowym założeniem dotyczącym losowości przyporządkowań jednostek badanych do źródeł zmienności (bodźca/interwencji) oraz do możliwych efektów (w tym stanów kontrfaktycznych) — wyrażająca podejście ograniczające się do „modelowania efektów przyczyn, a nie przyczyn efektów” (Holland, 1986);

## SCHEMAT

- ekonometryczna, zorientowana na artykulację mechanizmu przyczynowego w definicjach merytorycznych, w tym manipulacji i wyboru stanu kontrfaktycznego oraz generowania efektu w ramach ekonometrycznego modelu przyczynowego rozwijanego w kontekście ewaluacji wpływu i uwzględniającego błąd (samo)selekcji w ocenie skutków interwencji (Heckman, Smith, 1995; Heckman, Vytlačil, 2007).

W jednym i drugim typie podejścia ważne jest dążenie do możliwie dokładnego oszacowania efektu działania bodźca (interwencji, programu) przy uwzględnieniu możliwego obciążenia niedającego się do końca wyeliminować błędu selekcji (z racji dodatkowego pozytywnego lub negatywnego nastawienia uczestników, nawet po ich w pełni losowym przyporządkowaniu do odpowiednich grup poddanych i niepoddanych oddziaływaniu). W tym celu w obydwu ujęciach krytycznie ważna jest konstrukcja tzw. warunku kontrfaktycznego, odpowiadającego pojęciu wyniku potencjalnego w sensie Neymana albo też wyniku kontrfaktycznego w sensie Neymana-Rubina. Jego znaczenie jest ważne (powinno być w odpowiedni sposób symulowane i uwzględnione) w oszacowaniu łącznym wpływu w postaci tzw. przeciętnego efektu oddziaływania.

Przyjmując oznaczenia  $Y_i^t$  jako wynik zmiennej losowej dla jednostki  $i$  przy warunku, że została poddana działaniu  $t$  oraz  $Y_i^c$ , jeśli nie została poddana  $c$ , indywidualny efekt oddziaływania dany jest różnicą:

$$\delta_i = Y_i^t - Y_i^c \quad (1)$$

przy czym obserwowany wynik zmiennej:

$$Y = Y^t \text{ jeśli } D=1 \quad (2a)$$

$$Y = Y^c \text{ jeśli } D=0 \quad (2b)$$

gdzie  $D$  wskazuje przynależność jednostki do danej grupy eksperymentalnej lub kontrolnej.

Formalnie możliwe jest uwzględnienie, obok wyników obserwowalnych, także wyników nieobserwowalnych (kontrfaktycznych):

- grupa eksperymentalna/poddana  $t$  ( $D=1$ ) &  $Y=Y^t \rightarrow$  obserwowane jako  $Y$ ,
- grupa eksperymentalna/poddana  $t$  ( $D=1$ ) &  $Y=Y^c \rightarrow$  warunek kontrfaktyczny,
- grupa kontrolna/niepoddana  $t$  ( $D=0$ ) &  $Y=Y^t \rightarrow$  warunek kontrfaktyczny,
- grupa kontrolna/niepoddana  $t$  ( $D=0$ ) &  $Y=Y^c \rightarrow$  obserwowane jako  $Y$ .

Zagregowany efekt działania (interwencji) w populacji/ATE:

$$E(\delta) = E(Y^t - Y^c) = E(Y^t) - E(Y^c)$$

Wyniki w ramach warunku kontrfaktycznego są możliwe do konstruowania i uwzględniania w tych formułach oszacowywania efektów oddziaływania (interwencji) dzięki rozbudowanym statystycznym procedurom powiązań.

W ślad za wspomnianymi pracami Rubina i innych badaczy, które zainicjowały takie techniki (na czele z *Propensity Score Matching — PSM*) (Rosenbaum, Rubin, 1983), bardziej problemowo zorientowane techniki powiązań proponowane są w literaturze ekonometrycznej poświęconej ewaluacji programów (Heckman i in., 1998). Z kolei w badaniach nad postawami, opiniami, preferencjami i wyborami, w których mimo losowego przyporządkowania do grup istnieje ryzyko obciążenia wyników z racji niejednorodności wewnątrz grup w sensie większego/mniejszego wystawienia (podatności czy wrażliwości) badanych na oddziaływanie, takie jak ogłoszenia i reklamy, akcje propagandowe bądź uświadamiające itp., realizowany w ramach sondażu eksperyment stanowić może dobry sposób na przybliżenie do realnego świata sytuacji hipotetycznej, tzn. na konstruowanie warunku kontrfaktycznego. Niejednokrotnie występuje on wówczas obok eksperymentu naturalnego (Barabas, Jerit, 2010) bądź paralelnie do eksperymentu opartego na wyznaczaniu losowym, w wersji przyzwalającej na autoselekcję. Pozwala to bowiem na oszacowanie wielkości obciążenia wynikającego z pozabodźcowych czynników, jak nastawienie pro lub anty do przedmiotu postawy, w obliczeniach przeciętnych efektów wpływu (ATE) dla podzielonych w ten sposób podgrup (Gaines, Kuklinski, 2011).

Dotychczasowe rozważania można skonstruować stwierdzeniem, że eksperyment sondażowy jest odpowiedzią na wyzwania wobec coraz częściej podejmowanych prób wnioskowania przyczynowego z danych sondażowych (obserwacyjnych) poprzez ich wzbogacenie o brakujący element, jakim jest pomiar eksperymentalny. Specyfikując, są to zatem wyzwania obciążenia selekcją, korelacji pozornej, korelacji błędów pomiarowych, obciążenia danych w czasie oraz braku „prawdziwych” kontrfaktów, a także występowania wzajemnych współzależności przyczynowych.

### *EKSPERYMENT SONDAŻOWY W PRAKTYCE BADAWCZEJ — TYPOLOGIA ZASTOSOWAN*

Omówienie eksperymentu sondażowego (ES) jako postępowania badawczego warto poprzedzić opisami dwóch typów przedsięwzięć. Jedno jest przykładem konkretnego zastosowania tej metody (przykład badania). Drugie przedsięwzięcie informuje o kontynuowanym od 2002 r. projekcie infrastrukturalnym (finansowanym przez Narodową Fundację Nauki Stanów Zjednoczonych) ułatwiającym realizację tego typu badań przy niewielkim nakładzie środków finansowych i operacyjnych (projekt TESS).

***Przykład badania*** — identyfikacja wyznaczników postaw wobec reformy emerytalnej w Niemczech za pomocą eksperymentu osadzonego w ogólnokrajowym sondażu gospodarstw domowych (pn. SAVE, modelowanym m.in. na U.S. Health and Retirement Study (HRS); Scheubel i in., 2013).

Odwołano się do eksperymentu kontrolowanego w ramach panelowego sondażu (realizowanego na danych od 2003 r., wykorzystując rundę badań w latach 2006 i częściowo 2007). Starano się ustalić, jaka jest rzetelność (albo równoważenie, jakie jest obciążenie) informacji otrzymywanych z różnych sondażowych badań opinii dotyczących zmiany wieku uprawniającego do przejścia na emeryturę. Konkretnie, w jaki sposób opinie będące efektem czynników powszechnie rozpoznanych — takich jak zdrowie, ryzyko niezatrudnienia oraz charakterystyka społeczno-demograficzna — są ewentualnie obciążone odpowiedziami strategicznymi, odzwierciedlającymi możliwe uprzednie, negatywne nastawienie niektórych respondentów wobec reformy. W tym celu, dysponując szczegółowymi danymi o tych czynnikach zbieranymi w ramach regularnego badania sondażowego, część respondentów — wybrana losowo do grupy eksperymentalnej — otrzymywała dodatkowo pytanie o oczekiwania dotyczące zdolności wykonywania pracy. Zakładając, że jest to jedyny czynnik, który dla ustalonych wartości (kontrolowanych) zmiennych rozpoznanych jako ważne wynika ze zwiększonego formalnie wieku emerytalnego — z 65 do 67 lat — zastosowano trzy wersje pytania zadawanego respondentom w wieku od 18 do 60 lat, podzielonym na trzy grupy. Respondentów grupy pierwszej pytano o ocenę (na skali 10-punktowej) zdolności kontynuowania pracy po osiągnięciu 63 lat; drugiej — po osiągnięciu 65 lat; trzeciej — po osiągnięciu 67 lat.

Zakładając, że wystąpi tendencja do monotonicznego spadku zdolności do pracy w grupach odpowiadających rosnącemu wiekowi emerytalnemu, to odchylenia od niej traktowano jako zaburzenia spowodowane strategicznym obciążeniem odpowiedzi. Nadto okazały się one silnie zróżnicowane pomiędzy respondentami z Niemiec Wschodnich (d. NRD) i Zachodnich (d. RFN). Wyniki uogólniono na całą populację osób w wieku od 18 do 59 lat (po wykluczeniu samozatrudnionych i emerytowanych wcześniej). Efektem badania była reinterpretacja wyników wielu innych badań sondażowych opinii nt. reformy w Niemczech.

\*  
\*      \*

Na systematycznie rosnące zastosowania ES w badaniach społecznych wpływa, obok wspomnianego dążenia do testowania hipotez przyczynowych, technologiczne zaawansowanie instrumentarium badań sondażowych — w szczególności rozwój techniki internetowej na czele z CATI (*Computer Assisted Telephone Interviewing*), ale też CAWI (*Computer Assisted Web Interviewing*) i CASI (*Computer Assisted Self-Interviewing*). Pozwalają one nie tylko na automatyczne wprowadzanie danych (bez konieczności posługiwania się wieloma wersjami papierowych kwestionariuszy), ale też znacząco poszerzają zakres badań reprezentacyjnych (mimo jeszcze niedopracowanej teorii próbkowania populacji internetowych), dodając do nich wirtualne laboratorium do eksperymentalnego testowania hipotez nt. rozmaitego typu zjawisk społecznych i behawioralnych. Takie potencje ES po raz pierwszy pokazały na szerszą skalę badania eksperymentalne podejmowane w ramach infrastrukturalnego projektu pn. TESS (*Time-Sharing Experiments for the Social Sciences*) — dzielenie się czasem w naukach społecznych (Mutz, 2011; Freese, 2010).

**Projekt TESS.** Populacyjne ES są oparte na sondażu (ogólnokrajowym) typu *online* (Mutz, 2011; Druckman i in., 2006) — służą zasadniczo trzem, częstoć realizowanym w tym samym badaniu, typom usprawnień stawianych badaniom naukowym jako ich cel (King i in., 1994), tj. usprawnieniom dotyczącym: pytań badawczych, teorii (poprzez testowanie), jakości danych (z akcentem na pomiar). Ułatwiający takie badania międzynarodowy internetowy system infrastrukturalny TESS przyczynia się zarazem do popularyzacji analiz przyczynowych w kontekście wielodyscyplinarnym. Są one realizowane przy niskim budżecie — małe grupy badaczy przeprowadzają telefoniczny wywiad, zaprojektowany jako eksperyment, z wykorzystaniem elementów wspólnych — moduł podstawowy zawierający pytania socjodemograficzne, uzupełniony o oryginalne, własne moduły tematyczne (co można uznać za formę integracji badań przy zbieraniu danych).

Projekty takie realizowane są od 2000 r. na skalę globalną. Do 2007 r. przeprowadzono ponad 200 badań eksperymentalnych z wykorzystaniem tego sys-

temu (platformy), demonstrując zarazem coraz wyraźniejsze różnice pomiędzy populacyjnymi eksperymentami sondażowymi a klasycznym wyznaczaniem eksperymentów (typu laboratoryjnego). Aktualnie bez dodatkowych opłat można realizować (pod warunkiem zaakceptowania złożonego na konkurs projektu) badania zawierające od 5 pytań własnych (na próbie do 5000 respondentów) do 40 pytań (na próbie do 300 respondentów) (Freese, 2010).

Powstają rozmaite nowe platformy do badań eksperymentalnych na populacjach osiągalnych przez Internet lub telefon (typowo, technikami mieszanymi, jak CAWI i CATI), np. sprawdzane są komercyjne systemy internetowe typu MTurk oferowane przez firmę Amazon (Berinsky i in., 2012).

\*  
\*      \*

Badania metodą ES rozwinęły się wraz z rozwojem Internetu i wspomnianej, wykorzystującej go, techniki zbierania danych, w wielu dyscyplinach równolegle. Przykładowo:

- nauki polityczne — z inicjatywy Paula Snidermana na początku lat 90. ub. stulecia (Sniderman i in., 2004) zinstytucjonalizowane w postaci specjalizującego się w badaniach metodą ES ośrodka badań sondażowych w Kalifornijskim Uniwersytecie Berkeley. Adekwatność tej metody w odniesieniu do wielu problemów zaowocowała jej rosnącą popularnością, obejmując próby objaśnienia takich kwestii, jak: stosunek do mediów (np. dlaczego ludzie nie lubią mediów — Ladd, 2010); problematyka terroryzmu i źródeł poparcia (Lyall i in., 2013) bądź jego zwalczania (Druckman, Leeper, 2010); badania nad motywacjami postaw wobec imigrantów (np. decyduje bardziej poczucie tożsamości narodowej niż względy ekonomiczne, jak aktualnie w Holandii) czy nad zachowaniami wyborczymi (Cassino, Erisen, 2008);
- nauki ekonomiczne i demograficzne — badania porównawcze dotyczące: wyznaczników zachowań na rynku pracy, satysfakcji z pracy oraz dochodów i subiektywnego dobrostanu (Clark i in., 2007); nierówności dochodów i stosunku do polityki redystrybucji (Kuziemko i in., 2013); zadowolenia z pracy i postawy wobec wprowadzanych reform systemu emerytalnego (*Guidelines...*, 2013), w tym skłonności do dłuższej pracy w krajach Unii Europejskiej (Scheubel i in., 2013); badań nad stosunkiem do imigrantów ze zróżnicowaniem poziomu (wysoki/niski) i profili kwalifikacji zawodowych (np. przydatny/nieprzydatny na lokalnym rynku pracy — Hainmueller, Hiscox, 2010);
- nauki społeczne — z przewagą badań o charakterze interdyscyplinarnym i transdyscyplinarnym. Między innymi są to badania o charakterze metodologicznym, na czele ze statystycznymi i socjologicznymi badaniami nad sondażami i sposobami ich realizacji, związanymi głównie z narzędziami (Hall i in., 2012) i ich implementacją, np. uwzględniającą niejednorodność badanej próby z punktu widzenia treści i procedur zbierania danych, jak pro-

jektowanie kwestionariuszy dostosowanych do różnych kategorii respondentów (np. dzielenie zestawów pytań filtrując je według wstępnego zróżnicowania postaw do przedmiotu badań — Sterngold i in., 1994) czy zbieranie danych w odmienny sposób od poszczególnych grup respondentów różniących się skłonnością do udziału w badaniu w zależności od stosowanych procedur kontaktu i zbierania danych w celu zwiększenia stopy odpowiedzi (Luiten, Schouten, 2013) bądź zmniejszenia braków odpowiedzi (Groves i in., 2006). Dotyczą też badań porównujących motywacje uczestniczenia, takich jak zachęty materialne dla respondentów czy zainteresowanie i normy współpracy (Kropf, Blair, 2005) lub sposoby prezentacji przedmiotu badania w pytaniach (Tourangeau i in., 2009), jak też ewaluacyjnych z koncentracją na identyfikacji wyznaczników ocen wobec ewaluowanych programów (Cruces i in., 2013);

- nauki behawioralne — porównania interpersonalne w perspektywie międzykulturowej (pojęć, wartości, preferencji itd.), badania nad ryzykiem i postrzeganiem zagrożeń oraz wyborem zachowań prewencyjnych (Guterbock, 2010); analiza skłonności do uczestniczenia w badaniach w terminach zmiennych behawioralnych (jako ujęcie alternatywne wobec ujęć typowo skoncentrowanych na korelatach braku odpowiedzi (Esterling i in., 2008) czy badania nad efektem kontekstowym sondażu (korelacyjnym albo kierunkowym), gdy odpowiedzi na ogólniejsze i bardziej szczegółowe pytania (jak np. o wolność słowa i o swobodę głoszenia haseł ekstremistycznych) warunkują się w określony sposób (Tourangeau i in., 2000).

Te przykłady zastosowań ES najwyraźniej dzielą się na związane z badaniami metodologicznymi (dominującymi, jak dotąd, w literaturze), z jednej strony oraz na tematyczne, służące rozwiązaniu problemu — typu testowanie teorii lub na jej konstruowaniu (teoretyzowaniu, w szczególności w badaniach ewaluacyjnych, w ramach budowania modelu logicznego programu) bądź jej aplikacji, z drugiej strony (zestawienie 1). Pokazana w zestawieniu kategoryzacja nie oznacza jednak, że metodologiczne lub substantywne zastosowania eksperymentów sondażowych wykluczają się wzajemnie. Raczej odwrotnie, większość przypadków zastosowania ES może być klasyfikowana jako leżące gdzieś na *continuum*, od skoncentrowanych na głównie metodologicznych zagadnieniach do skoncentrowanych na problemach merytorycznych (Schlueter, Schmidt, 2010).

**ZESTAWIENIE (1) TYPOLOGII ZASTOSOWAŃ EKSPERYMENTÓW SONDAŻOWYCH**

| Cele badań     | Badania podstawowe                                                                                                                                                                                                                        | Badania stosowane i ewaluacyjne                                                                                                                                                                                                                                         |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Metodologiczne | testowanie mocy wyjaśniającej teorii („testy krzyżowe” teorii), np. teorię identyfikacji etnicznej (kulturowej itd.) czy teorie interesu ekonomicznego w objaśnianiu postaw wobec imigrantów; badania nad metodologią badań ewaluacyjnych | alternatywne podejścia dotyczące: <ul style="list-style-type: none"> <li>• zwiększenia stopy odpowiedzi</li> <li>• zmniejszenia obciążenia respondentów</li> <li>• usprawnienia zbierania danych — porównania techniki</li> <li>• polepszenia jakości danych</li> </ul> |

**ZESTAWIENIE (1) TYPOLOGII ZASTOSOWAŃ EKSPERYMENTÓW SONDAŻOWYCH (dok.)**

| Cele badań | Badania podstawowe                                                                                                                                              | Badania stosowane i ewaluacyjne                                                                  |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Tematyczne | testowanie zasięgu teorii, np. teoria wzrostu/spadku tendencji do uprzedzeń etnicznych; testowanie teorii przyczynowo-skutkowych w badaniach ewaluacyjnych itp. | testowanie wyniku/porównania — porównawcza ewaluacja programów/wyznaczenie złożone (czynnikiowe) |

Źródło: na podstawie Nock, Guterbock, 2010.

***Zalety stosowania eksperymentu sondażowego — porównanie z eksperymentem klasycznym w badaniach społecznych***

Eksperyment klasyczny (laboratoryjny) definiowany jest jako badanie (doświadczenie) spełniające następujące warunki: wprowadzenie oddziaływania (interwencji, bodźca) podlegającego manipulacji przez badacza; losowe przyporządkowanie badanych do grupy poddanej tzw. bodźcowaniu (eksperymentalnej) i do grupy kontrolnej; dokonanie pomiaru zmiennej zależnej (zmiennej wynikowej) przed i po zastosowaniu bodźca wraz z określeniem różnic wyników w obydwu grupach — eksperymentalnej i kontrolnej. Porównanie tych wyników traktowane jest jako efekt doświadczenia (bodźca); zakłada się brak zakłóceń spowodowanych wpływem czynników innych niż działanie eksperymentalne (bodziec), na zmienną wynikową w konsekwencji czego przyjmuje się, że eksperyment spełnia warunek trafności wewnętrznej.

Mimo formalnie podobnej struktury operacyjnej, ES odróżnia od klasycznego eksperymentu m.in.:

- wprowadzenie systematycznego źródła zmienności (bodźca) pomiędzy badanymi jako elementu sondażu. Jest nim zazwyczaj informacja (pytanie, wiadomość, instrukcja itp.) różna co do treści lub co do formy, lub w obu aspektach, a także co do sposobu jej dostarczenia — najczęściej drogą kontaktu mieszanego (internetowo-, komputerowo-, mailowo-, telefonicznego) dla grup różniących się swoją rolą w badaniu;
- respondenci przyporządkowani są do grupy poddanej działaniu i niepoddanej działaniu bodźca (programu) w sposób losowy (poza przypadkami celowego obciążenia, tzn. pożądaną nadreprezentacji/podreprezentacji określonej kategorii badanych); badani są członkami próbki wylosowanej z populacji (zamiast grupy podobnych do siebie osób) w sposób losowy, używając techniki próbkowania na odpowiednio dużą skalę (idealnie — ogólnokrajową);
- zazwyczaj brak pomiaru przed bodźcowaniem i w konsekwencji brak możliwości dokonywania porównań „podłużnych”, czyli grup eksperymentalnej i kontrolnej, zaś ocena efektu doświadczenia sprowadza się do porównań równoległych (wyniku pomiaru *ex post* w grupach poddanych i niepoddanych działaniu bodźca); w konsekwencji uznaje się, że wyniki eksperymentu w ramach sondażu można prawomocnie uogólniać na populację (spełniają bowiem warunek trafności zewnętrznej);

- zakłada się możliwość wnioskowania przyczynowego spełniającego 3 podstawowe wymagania dotyczące relacji przyczynowej: obydwie zmienne współwystępują w czasie lub w przestrzeni; zmienna przyczynowa musi poprzedzać zmienną wynikową/efekt w czasie; relacja pomiędzy przyczyną i skutkiem nie może być objaśniona działaniem jakiejś trzeciej zmiennej (jak w korelacji pozornej).

Dzięki swym zaletom ES wykazują przewagę zarówno nad coraz powszechniej stosowanymi sondażami internetowymi, z jednej strony oraz nad eksperymentami klasycznymi typu *quasi-laboratoryjnego* (jak RCT — *Random Controlled Trials*) w naukach społecznych, z drugiej strony. Będąc bliżej realnego świata zapewniają większą trafność zewnętrzną, pozwalając zarazem na koncentrację badania na określonej kategorii osób w populacji, szczególnie w badaniach panelowych z użyciem techniki internetowej. Z kolei zwiększenie znaczenia badań typu eksperymentalnego prowadzi do zaawansowania metodologicznego nauk społecznych, m.in. dzięki możliwościom porównawczej ewaluacji alternatywnych hipotez przyczynowych (np. czy wzrost zaufania społecznego prowadzi do wzrostu zakupów poprzez Internet czy odwrotnie — badania potwierdzają zależność dwukierunkową).

## ZASTOSOWANIA ES W BADANIACH METODOLOGICZNYCH I TEMATYCZNYCH

### Rodzaje danych i analiz

Przykładowe omówienie zastosowania ES warto poprzedzić uwagą o (pokazanej na schemacie) mieszanej kategorii danych generowanych za pomocą tej metody, jakimi są eksperymentalne dane sondażowe oraz obserwacyjne dane eksperymentalne (w zestawieniu (2) cudzysłów odróżnia metajęzykowy poziom, na jakim mówi się o obydwu typach danych). Skrzyżowanie tej typologii z podziałem możliwej taktyki analitycznej na analizę sondażową, skoncentrowaną na eksploracji zależności statystycznych oraz analizę eksperymentalną, nastawioną na weryfikację zależności przyczynowych, przedstawia zestawienie (2).

**ZESTAWIENIE (2) RODZAJÓW DANYCH I ANALIZ**

| Rodzaj danych                                | Analiza sondażowa<br>(zależności statystyczne) | Analiza eksperymentalna<br>(zależności przyczynowe) |
|----------------------------------------------|------------------------------------------------|-----------------------------------------------------|
| Obserwacyjne .....                           | √                                              | •                                                   |
| Eksperymentalne .....                        | •                                              | √                                                   |
| „Eksperymentalne” — <i>dane obserwacyjne</i> | √                                              | v                                                   |
| „Obserwacyjne” — <i>dane eksperymentalne</i> | v                                              | √                                                   |

U w a g a. √ — dominujące (rekomendowane); v — umożliwione (jako hybrydowe); • — dopuszczalne (warunkowo).

Ź r ó d ł o: opracowanie własne.

Pozwala to na ukazanie kierunków możliwości rozszerzania ujęć analitycznych bądź jako przede wszystkim danych obserwacyjnych, do których obok analiz sondażowych można (w określonym zakresie) stosować modele analiz przyczynowych, bądź jako zasadniczo danych eksperymentalnych pozwalających zarazem na generalizację zależności przyczynowych na populację.

### *Analiza sondażowa („eksperymentalne” dane obserwacyjne)*

Za przykład pierwszego typu ujęć posłużyć może badanie eksperymentalne przeprowadzone w ramach Europejskiego Sondażu Społecznego nad międzykulturową porównywalnością wybranych pojęć kluczowych (kategorii klasyfikacyjnych) przez respondentów (np. odnoszenie do siebie stosowanych w analizach stratyfikacji społecznej kategorii „klasa średnia”, „klasa wyższa”, „klasa pracownicza” za pomocą techniki winietowej, w wersji winiet zakotwiczonych — Hopkins, King, 2010).

Innym przykładem są zastosowania posługujące się techniką winiet zakotwiczonych konstruowanych w pozycjach opisujących hipotetyczne obiekty i sytuacje, do których badany ustosunkowuje się poprzez wskazanie winiety mu najbliższej, czyli badania nad wyznacznikami preferencji i wyborów miejsca zamieszkania (Bruch, Mare, 2012). Na podstawie winiet przedstawiających kombinację 14 elementów otoczenia rasowego wokół danego miejsca — *od wszystkie domy dookoła są zamieszkałe przez etnicznie obcych do wszystkie domy są zamieszkałe przez naszych* — 5 kombinacji/wzorów pokazywano respondentom 3 różnych grup.

Jeszcze innego przykładu analizy sondażowej wzbogaconej o elementy przyczynowe dostarcza ES stosowany w celu identyfikacji występowania wspomnianego już efektu kontekstowego sondażu w sensie relacji nadrzędności/podrzędności pomiędzy tzw. pytaniami kontekstowymi i celowymi (Tourangeau i in., 2009). Jak wykazał J. Buckley (2008), efekt kontekstu występuje nie tylko w zwykłych, ale też w winietowych formach prezentacji pytań. Stąd sugestia autora dotycząca dokonywania randomizacji w zakresie wszystkich trzech aspektów kontekstu: miejsca pozycji służących do odpowiedzi (samooceniań), porządku stosowania winiet oraz porządku przedstawiania zestawów łącznych (pozycje oceniania i winiety).

Efekt kontekstu sondażu występuje w różnego typu badaniach sondażowych, pozwalających na identyfikację związanego z nim obciążenia, np. przy posłużeniu się techniką wizualną, jak pokazanie zdjęcia osoby zdrowej albo chorej, własna ocena zdrowia przez respondenta zależy wyraźnie od tego, jakie zdjęcie zostało mu pokazane (niższa w pierwszym i wyższa w drugim przypadku — Couper i in., 2007).

### ***Analiza eksperymentalna („sondażowe” dane eksperymentalne)***

Analizy wykorzystujące ES skoncentrowane na zależnościach przyczynowych stosunkowo najwcześniej i najszerzej rozwinęły się w badaniach nad procesami formowania opinii i postaw. Bezpośrednim ich przedmiotem był efekt bodźca i hipotezy stwierdzające jego przyczynowy wpływ — określenie kierunku i wielkości — na rejestrowane wyniki. Przykładem mogą być badania nad postawami wobec tzw. akcji afirmatywnej w Stanach Zjednoczonych, przy równoczesnej kontroli, czy i w jakim stopniu ingeruje w ten proces zjawisko stereotypów (jak zakładano w wielu konceptualizacjach/hipotezach na ten temat — Druckman, Leeper, 2010).

Obecnie coraz większy udział mają badania prowadzone metodą ES w oszacowywaniu efektów programów (w tym publicznych) podejmowanych przede wszystkim w ramach badań ewaluacyjnych. W dużym stopniu wynika to z powszechnie znanej rozbieżności w ocenie polityki/programów rządowych dokonywanych na podstawie danych sondażowych oraz danych z eksperymentów naturalnych (Gaines i in., 2006). Użycie techniki łączeniowej (*matching, propensity score*) czy dwustopniowej regresji może redukować obciążenie, ale go nie eliminuje.

### ***ES w badaniach ewaluacyjnych — efektywność programu a obciążenia postaw wobec polityki redystrybucyjnej***

Problem obciążenia opinii (w szczególności braku poparcia) wobec programów redystrybucji dochodów w zależności od „grupy odniesienia”, z jaką identyfikuje się respondent, sygnalizowany w literaturze jako ważny dla efektywnego wdrażania programów redystrybucyjnych, podjęli Cruces i in. (2013). W celu określenia kierunku i rozmiaru tego obciążenia odwołali się do ES. Zbierane w sondażu szczegółowe dane na temat samoidentyfikacji z grupą dochodową — wykazujące zazwyczaj tendencję do zawyżania oceny swojej pozycji w porównaniu do innych — uzupełnili o eksperyment polegający na losowym przyporządkowaniu uczestniczących w sondażu respondentów do grup informowanych o przynależeniu do danej kategorii w rankingu dochodów. Po zarejestrowaniu odpowiedzi na pytania dotyczące wielkości i charakteru redystrybucji poprzez świadczenia społeczne i poprzez podatki, respondenci byli informowani o rzeczywistym miejscu zajmowanym na skali dochodów. Pytani powtórnie o stosunek do programów redystrybucyjnych, badani z reguły zmieniali swoje opinie w kierunku spójnym z pozycją faktycznie zajmowaną w rankingu dochodów — im większa rozbieżność pomiędzy oryginalną (przeszacowaną) pozycją własną, tym większy wzrost poparcia dla programów redystrybucji. Dodać można, że jest to zarazem zgodne z koncepcją „zasłony niewiedzy” w teorii Rawlsa podziału sprawiedliwego, nakazującej w konsekwencji maksymalizować pomoc dla znajdujących się najniżej w uporządkowaniu społecznym (Lissowski i in., 1991).

## ***ES w kontekście nierówności i redystrybucji***

Testowanie hipotez głoszących zależność postaw wobec programów redystrybucyjnych od postrzeganej nierówności rozkładów dochodów jest również coraz częściej podejmowane na podstawie danych generowanych za pomocą ES. Przykładem są badania nad wpływem informacji na temat nierówności na stosunek do redystrybucji poprzez system podatkowy, w których do realizacji ES wykorzystano wspomniany serwis *Amazon* (Kuziemko i in., 2013). Zbadano 5000 respondentów losowo przyporządkowanych do bodźców polegających na podaniu informacji o nierówności dochodów w Stanach Zjednoczonych oraz zależności pomiędzy opodatkowaniem wysokich dochodów i wzrostem gospodarczym, uzyskując znaczący wpływ informacji na postrzeganie nierówności przez badanych, ale bez konsekwencji w postaci poparcia dla programów redystrybucji (ani poprzez podatki, ani poprzez świadczenia).

Wyznaczniki postaw wobec innego typu programów redystrybucyjnych — w postaci np. protekcyjizmu wobec określonego sektora gospodarki — mogą być trafnie zidentyfikowane dzięki zastosowaniu ES, jak ilustrują to badania Naoi i Kume (2011). Dążąc do wyjaśnienia powszechnej w Japonii akceptacji wysokich cen żywności, przy równoczesnym ograniczaniu jej importu (nawet w czasie kryzysu, jak w 2008 r.), autorzy zidentyfikowali projekcję własnego zagrożenia pracy respondentów spoza sektora związanego z produkcją żywności na sytuację farmerów jako mechanizm odpowiedzialny za pozornie sprzeczne z ich interesem, jako konsumentów, popieranie polityki utrzymywania wysokich cen żywności.

## ***Badania ewaluacyjne oraz studia nad praktykami i procedurami ewaluacyjnymi***

W badaniach ewaluacyjnych ES podlegają wyznaczaniu w sposób pozwalający na ocenę efektywności interwencji poprzez rygorystyczne, losowe przyporządkowanie badanych do grup doświadczalnych i kontrolnych. Z tego punktu widzenia dyskutowany jest, postulowany często przez sponsorów badań ewaluacyjnych, schemat losowej kontroli prób *Random Control Trial* (RCT). Jednakże z rezerwą do schematu RCT jako do złotego środka — stosowanego w tej roli w ewaluacyjnych badaniach klinicznych — podchodzi Europejskie Stowarzyszenie Ewaluacyjne (*European Evaluation Society* — EES), nie rekomendując w efekcie żadnej konkretnej procedury.

Istotną kwestią w toczącej się dyskusji na ten temat jest m.in. pytanie, czy dysponowanie grupą kontrolną wystarcza do skonstruowania kontrfaktycznej wersji interwencji, albowiem wyznaczanie na podstawie losowania i kontroli (jak w RCT) nie może według ekspertów ewaluacji być uznane za najlepsze podejście do ustalania zależności przyczynowo-skutkowej. Przykładowo M. Scriven (2008) uważa, że ES oraz eksperyment naturalny są z metodologicznego punktu widzenia

lepszym i poprawniejszym sposobem ustalania takich zależności w odniesieniu do przedsięwzięć (programów) w sferze społecznej, m.in. z racji ich większej autentyczności. Jest to spowodowane np. niemożnością wykluczenia korelacji pozornej w ramach RCT, gdzie trudno o spełnienie warunku tzw. podwójnego zaślepienia (nieświadomości bycia wybranym do grupy poddanej lub niepoddanej działaniu bodźca), dlatego nawet najpoprawniejsze badanie w sensie RCT powinno — według Scrivena — być uzupełnione o „test polowy”.

### ***ES w badaniach metodologicznych***

Przykłady coraz liczniejszego stosowania ES w badaniach tematycznych nie dorównują popularności ich występowania w badaniach metodologicznych, przede wszystkim dotyczących problematyki badań sondażowych, wśród których badania tą metodą stanowią, zdaniem ekspertów, ok. 3/4 wszystkich badań (Guterbock, 2010). Czynniki decydującymi o tym są:

- usprawnienia w zakresie obniżenia kosztów badania, np. poprzez wspomniane dostosowanie techniki zbierania danych w sposób uwzględniający dostępność i chęć uczestniczenia różnych kategorii respondentów;
- szybkość dostarczania wyników, m.in. dzięki łatwości porównań rezultatów w ramach kontynuowanego sondażu;
- zwiększenie ogólnej użyteczności badania, m.in. z racji uwzględnienia zależności przyczynowych oraz elastyczności w odnoszeniu wyników do różnych dziedzin badań (np. uwzględnienie czynników zmiany postaw na skuteczność programu w badaniach ewaluacyjnych);
- zmniejszenie całkowitego błędu sondażowego, zarówno błędu próbkowego jak i błędu pokrycia, np. porównanie losowania adresów z losowaniem jednostek terytorialnych czy losowania systemem RDD (*Random Digit Dial*) z wyborem z listy adresowej mieszkań itd.;
- usprawnienie prowadzenia wywiadu i ogólnej organizacji procesu zbierania danych;
- podniesienie jakości danych, m.in. przez zmniejszenie zakresu niedoszacowania dochodów czy uczestniczenia w programach socjalnych, w szczególności przez redukcję błędu związanego z brakami danych, jak zwiększenie uczestnictwa, np. za pomocą strategii wyznaczania całkowitego Dillmana (*Total Design Survey Method* — TDM) (Lavrakas, 2008) oraz z brakami odpowiedzi (np. techniki wizualizacji, systemy pominięć, inne ułatwienia).

We wspomnianym sondażowym badaniu postaw konsumenckich (Luiten, Schouten, 2013) losowo dobrane 3000 gospodarstw domowych objęto eksperymentalnym badaniem za pomocą sekwencyjnej kombinacji techniki komputerowo-internetowo-pocztowo-telefonicznej. Celem było dostosowanie sposobu realizacji sondażu do grup respondentów mniej i bardziej skłonnych do uczestniczenia w badaniu, opierając się na paradanych w postaci informacji z wcześ-

niejszych badań nt. przyczyn braku danych. Zanotowano znaczący wzrost uczestnictwa i reprezentatywności (przy nieznacznym wzroście kosztów z racji preferowania pocztowego kontaktu nad internetowym, ale mniej kosztownego niż CATI, o który należało uzupełnić wersję pocztową dla uzyskania wzrostu odpowiedzi). Do podniesienia jakości danych przyczynia się też redukcja błędu pomiarowego związanego z narzędziem (poprzez odpowiedni stopień ustrukturyzowania kwestionariusza, sformułowania i sekwencję pytań, etykiety w skalach itp.) oraz związanego z ankierem (sposoby raportowania, wywiady poznawcze itp.). ES stanowi najlepszą formę identyfikacji możliwych usprawnień w tych dziedzinach.

Cechy ES ujawniają się w szczególności w badaniach panelowych typu systematycznego oraz w eksperymentach złożonych (wieloczynnikowych). Ze względu na schemat wyznaczania badania metodą ES można podzielić na wyznaczanie proste — gdy porównuje się np. odmienne wersje kwestionariusza (efekt konstrukcji, zmiany porządku pytań itp.) oraz złożone (czynniki) — gdy kontroluje się efekt wielu zmiennych w jednym badaniu (wielodoświadczeniowym), np. układ pytań, treść pytań, sposób realizacji kwestionariusza oraz efekt interakcji pomiędzy czynnikami. W ostatnim typie wyznaczania występuje jednak ryzyko „eksplozji kombinatorycznej”, np. dla 3 zmiennych o 3 wariantach każda istnieje aż 27 możliwych doświadczeń. W badaniach z 10 czynnikami, z których każdy mógł występować w 2 do 9 wariantów, otrzymano 270 tys. możliwych kombinacji winietowych (Nock, Guterbock, 2010). Stąd potrzeba ograniczenia poprzez stosowanie wyznaczania frakcyjnego (tylko część spośród  $2^k$  komórek jest realizowana w badaniu, ale w ich wyborze powinna być zachowana proporcjonalność).

### ***Redukcja braków danych w badaniach longitudinalnych***

Badania podłużne, w tym sondaże panelowe, mimo ich szczególnych zalet z racji możliwości powtarzania pomiarów i porównywania wyników w czasie narażone są na swoiste dla nich — obok trapiących także sondaże przekrojowe — błędy związane z ubytkiem w kolejnych rundach (efekt ścierania się próby). Ubytki te bowiem z reguły nie są losowe i systematyczne zniekształcenia próby nakładają się na siebie, wymagając złożonych metod ich identyfikacji i uwzględniania w fazie analizy danych. ES stanowią operacyjne podejście do redukcji takich błędów poprzez odkrywanie powodów ubytków i sposobów radzenia sobie z nimi w fazie terenowej. Przykładu dostarczają badania za pomocą dwukrotnego zastosowania ES w Brytyjskim Sondażu Panelowym Gospodarstw Domowych (Fumagalli i in., 2013). W pierwszym eksperymencie porównywano strategię dotyczącą sposobu kontaktowania się i informowania badanych. Skłonność do uczestniczenia w kolejnej rundzie badania okazała się wzrastać dzięki międzyrundowej wysyłce karty pocztowej z prośbą o informację nt. ewentualnej zmiany adresu (wysyłanie karty o potwierdzenie adresu, tak samo jak brak wysyłki karty, nie podnosił udziału). W drugim eksperymencie porównywano stan-

daryzowane i dostosowane do respondentów formularze dla ludzi młodych oraz dla osób bardzo zajętych. Wprawdzie w obu grupach odnotowano nieznaczny efekt pozytywny, ale opcja w postaci krótkiego wywiadu telefonicznego znacząco polepszała uczestnictwo w stosunku do wywiadu personalnego (Fumagalli i in., 2013).

### ***Problemy związane ze stosowaniem ES*** **— zakłócenia wielokrotnego eksperymentu**

ES realizowane w ramach tego samego sondażu populacji — w szczególności typu panelowego — podejmowane dla wyjaśnienia, dlaczego jakiś program jest (lub nie jest) efektywny, są narażone na obciążenia wynikające ze wzajemnego oddziaływania różnych ich aplikacji (dyfuzji i *spillover*, gdy doświadczenia wcześniejsze wpływają na późniejsze wyniki eksperymentu). Przykładem są badania Gainesa i in. (2006), którzy zastosowali losowe przyporządkowanie podmiotów do grup kontrolnej i eksperymentalnych (bodźcowanych) dwukrotnie. W pierwszej rundzie uczestników eksperymentu pytano o to, czy ludzie obcy etnicznie są leniwi. Jednocześnie każdy z uczestników miał 50% szansy otrzymania przed pytaniem informacji na temat tzw. akcji afirmatywnej (AA) dotyczącej preferowania tych ludzi w różnego typu systemach rekrutacji. Paralelnie (drugie doświadczenie) pytano o brak odpowiedzialności ludzi obcych etnicznie, z taką samą szansą otrzymania informacji na temat urodzeń pozamałżeńskich wśród nich. Grupa kontrolna nie otrzymywała żadnej informacji dodatkowej. Każdy uczestnik miał następnie określone prawdopodobieństwo odzwierciedlające jego skłonności do stereotypowania ( $S$ ), przy warunku znalezienia się w sytuacji eksperymentalnej ( $e$ ) bądź realnej ( $r$ ), mogąc formalnie być zaliczonym do grupy eksperymentalnej ( $e$ ) i poddanym działaniu bodźca ( $T$ ) lub nie ( $C$ ), czyli w jednym z czterech stanów:

$$\begin{array}{ll} P(S|C_rC_e)=b & P(S|C_rC_e)=b+p \\ P(S|T_rC_e)=b+p & P(S|T_rT_e)=b+p+pd \quad 0 \leq d \leq 1 \end{array}$$

gdzie:

- $b$  — prawdopodobieństwo (początkowe) podzielania stereotypu,
- $p$  — dodatkowa wielkość, o którą prawdopodobieństwo może być zwiększone w wyniku obu doświadczeń,
- $d$  — reprezentuje zmianę w  $b$  i  $p$  jako wynik wcześniejszego (przed badaniem) wpływu na odpowiedzi badanych, tzn. na ich skłonność do stereotypów ( $d=1$  oznacza jednakowy wzrost takich skłonności ze strony wszystkich oddziaływań;  $d=0$ , gdy tylko jednorazowe oddziaływanie zwiększa skłonność do stereotypowania).

W odróżnieniu od porównania różnic pomiędzy wielkościami w grupach eksperymentalnych i kontrolnej można rozpatrywać łączne prawdopodobieństwo dla oszacowania efektu *spillover*, przyjmując jako za  $y_k$  proporcję dzielących

stereotyp *k*. Efekt możliwego oddziaływania pierwszego doświadczenia na wynik zarejestrowany w doświadczeniu drugim jest tym większy, im większe odstępstwo od losowości w wyznaczaniu uczestników do doświadczenia drugiego.

W badaniach nad innowacyjnością problem *spillover* jest także obecny, np. w postaci dyfuzji innowacji w przestrzeni.

## Podsumowanie

Metoda ES wydaje się odgrywać coraz większą rolę w badaniach statystyki publicznej, ale generalnie jest ona ograniczona do określonej klasy badań, tj. raczej do badań metodologicznych niż tematycznych. Na wprowadzenie do głównego nurtu statystyki publicznej pierwszego typu badań, realizowanych za pomocą ES, zasługują przede wszystkim badania w zakresie redukcji obciążenia respondentów, przy równoczesnym wzroście ich udziału w badaniach nad wyznaczaniem sondaży dostosowanych do wyróżnionych grup oraz nad podniesieniem jakości danych poprzez możliwie (dzięki ES) lepsze wyznaczanie próby w kolejnych rundach badań stałych.

Spośród badań tematycznych stosunkowo najbardziej obiecujące wydają się badania za pomocą ES realizowane w ramach niezależnie prowadzonych badań ogólnokrajowych i ewaluacyjnych (jak badanie budżetów gospodarstw domowych czy badanie siły roboczej itp.). Chociaż trudno postulować (np. ze względów etycznych), aby określone grupy gospodarstw czy osoby były wykluczane z jakiegoś podległego ewaluacji programu, np. typu socjalnego, to jako uzupełniające badanie postaw i opinii na wybranych (celowych) grupach może przyczynić się ono do usprawnienia lub eliminacji efektów będących wynikiem zakłóceń pochodzenia zewnętrznego wobec samego programu. Przykładem są badania nad wyborami w zakresie (poparcia/odrzućenia) programów lokalnych, włączające ich uczestników (w celu identyfikacji wyznaczników preferencji i znalezienia sposobów perswazji) w procesy decyzyjne dotyczące w szczególności dóbr publicznych.

Możliwości takie otwierają szerokie pole do współpracy ośrodków akademickich i administracji publicznej — jako stron najbardziej zainteresowanych tego typu badaniami, a także dla statystyki publicznej — jako dostarczyciela platformy dla takiej współpracy.

---

**dr hab. Włodzimierz Okrasa** — GUS, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

## LITERATURA

- Barabas J., Jerit J. (2010), *Are Survey Experiments Externally Valid?*, „American Political Science Review”, Vol. 104, No. 2, May 2010
- Berinsky A. J., Huber G. A., Lenz G. S. (2012), *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, „Political Analysis”, Vol. 20, No. 3

- Bruch E., Mare R. (2012), *Methodological Issues in the Analysis of Residential Preferences and Residential Mobility*, „Sociological Methodology”, Vol. 42
- Buckley J. (2008), *Survey Context Effects in Anchoring Vignettes*, <http://polmeth.wustl.edu/workingpapers.php>
- Cassino D., Erisen C. (2008), *Priming Bush and Iraq in 2008: A Survey Experiment*, „American Politics Research”, Vol. 38, No. 2
- Clark A. E., Frijters P., Shields M. (2007), *Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles*, „Journal of Economic Literature”, Vol. 46(1)
- Couper M. P., Conrad F. G., Tourangeau R. (2007), *Visual Context Effects in Web Surveys*, „Public Opinion Quarterly”, Vol. 71, No. 4
- Cruces G., Perez-Truglia R., Tetaz M. (2013), *Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment*, „Journal of Public Economics”, Vol. 98
- Druckman J. N., Green D. P., Kuklinski J. H., Lupia A. (2006), *The Growth and Development of Experimental Research in Political Science*, „American Political Science Review”, Vol. 100, No. 4
- Druckman J., Leeper T. (2010), *Bias in Political Communication Experiments*, Presentation at the 2010 Annual Conference, Harvard Program in Survey Research
- Esterling K. M., Neblo M. A., Lazer D. M. J. (2008), *Estimating Treatment Effects in the Presence of Noncompliance and Nonresponse: The Generalized Endogenous Treatment Model. Experiments in Political Science 2008 Conference Paper*, <http://ssrn.com/abstract=1269485>
- Freese J. (2010), *TESS: Overview, Status, and Future*. Sociology & Institute for Policy Research. Northwestern University. [tess.experimentcentral.org](http://tess.experimentcentral.org)
- Fumagalli L., Laurie H., Lynn P. (2013), *Experiments with methods to reduce attrition in longitudinal surveys*, „Journal of the Royal Statistical Society Series A”, Royal Statistical Society, Vol. 176(2)
- Gaines B. J., Kuklinski J. H. (2011), *Experimental Estimation of Heterogeneous Treatment Effects Related to Self-Selection*, „American Journal of Political Science”
- Gaines B. J., Kuklinski J. H., Quirk P. J. (2006), *The Logic of the Survey Experiment Reexamined*, „Political Analysis”, Vol. 15, No. 1
- Groves R. M., Couper M. P., Presser S., Singer R., Tourangeau R., Acosta G. P., Nelson L. (2006), *Experiments in Producing Nonresponse Bias*, „Public Opinion Quarterly”, Vol. 70, No. 5
- Guidelines on Measuring Subjective Well-being* (2013), OECD Publishing, <http://dx.doi.org/10.1787/9789264191655-en>
- Guterbock T. M. (2010), *Survey Experiments: Past, Present, Future*, Presentation at the 2010 Annual Conference, Harvard Program in Survey Research (October)
- Hainmueller J., Hiscox M. J. (2010), *Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment*, „American Political Science Review”, Vol. 104, No. 1 (February)
- Hall L., Wallack R. Z., Eggers F. (2012), *Fair Market Rent Survey: Results of Methodological Experiment*, Redstone Research, LLC (mimeo)
- Heckman J. J., Ichimura H., Todd P. (1998), *Matching as an Econometric Evaluation Estimator*, „Review of Economic Studies”, Vol. 65
- Heckman J. J., Smith J. A. (1995), *Assessing the Case for Social Experiments*, „Journal of Economic Perspectives”, Vol. 9(2)
- Heckman J. J., Vytlacil E. J. (2007), *Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation*, [in:] J. J. Heckman, E. E. Leamer, *Handbook of Econometrics*, Elsevier
- Holland P. W. (1986), *Statistics and Causal Inference*, „Journal of the American Statistical Association”, Vol. 81, No. 396

- Hopkins D., King G. (2010), *A Method of Automated Nonparametric Content Analysis for Social Science*, „American Journal of Political Science”, Vol. 54, No. 1
- King G., Koehane R. O., Verba S. (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton University Press
- Kropf M. E., Blair J. (2005), *Eliciting Survey Cooperation. Incentives, Self-Interest, and Norms of Cooperation*, „Evaluation Review”, Vol. 29, No. 6
- Kuziemko I., Norton M. I., Saez E., Stantcheva S. (2013), *How elastic are preferences for redistribution? Evidence from randomized survey experiments*, National Bureau of Economic Research, Working Paper 18865, <http://www.nber.org/papers/w18865>
- Ladd J. M. (2010), *Why Americans Hate the Media and How It Matters*, Princeton University Press
- Lavrakas P. J. (2008), *Total Design Method (TDM)*, Encyclopedia of Survey Research Methods, <http://dx.doi.org/10.4135/9781412963947>
- Lissowski G., Tyszka T., Okrasa W. (1991), *Principles of Distributive Justice: Experiments on Ethical Preferences in Poland and America*, „Journal of Conflict Resolution”, Vol. 35, No. 1
- Luiten A., Schouten B. (2013), *Tailored fieldwork design to increase representative household survey response: an experiment in the Survey of Consumer Satisfaction*, „Journal of the Royal Statistical Society A”, Vol. 176, No. 1
- Lyall J., Blair G., Imai K. (2013), *Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan*, „American Political Science Review”, Vol. 107, No. 4
- Mutz D. C. (2011), *Population-Based Survey Experiments*, Princeton University Press, Princeton i Oxford
- Naoi M., Kume I. (2011), *Explaining Mass Support for Agricultural Protectionism: Evidence from a Survey Experiment During the Global Recession* International Organization, Vol. 65, No. 4
- Nock S. L., Guterbock T. M. (2010), *Survey Experiments*, [in:] J. Wright, P. Marsden (red.), *Handbook of Survey Research*, Second Edition, Wiley Interscience
- Okrasa W. (2012), *Statistics and Sociology: The mutually-supportive development from the perspective of interdisciplinaryization of social research*, „Statistics in Transition new series”, „Journal of Polish Statistical Association”, Vol. 13, No. 2
- Okrasa W., Scott K. (1998), *Analysis of Latvia Diary Experiment*, Mimeo, Development Research Group, World Bank, Washington
- Rosenbaum P. R. (2002), *Observational Studies*, Springer-Verlag, New York, second edition
- Rosenbaum P. R., Rubin D. B. (1983), *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, „Biometrika”, Vol. 70, No. 1
- Rubin D. B. (1974), *Estimating causal effects of treatments in randomized and nonrandomized studies*, „Journal of Educational Psychology”, Vol. 66, No. 5
- Scheubel B., Schunck D., Winter J. (2013), *Strategic Responses: A Survey Experiment on Opposition to Pension Reforms*, „The Scandinavian Journal of Economics”, Vol. 115, No. 2
- Schlueter E., Schmidt P. (2010), *Special Issue: Survey Experiments*, „European Journal of Research Methods for the Behavioral and Social Sciences”, Vol. 6, No. 3
- Scriven M. (2008), *A Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research*, „Journal of MultiDisciplinary Evaluation”, Vol. 5, No. 9
- Smith T. W. (1986), *Conditional order effects*, „Technical Report GSS Methodological Research Report”, No. 55, National Opinion Research Center, Chicago
- Sniderman P. M., Hagendoorn L., Prior M. (2004), *Predisposing Factors and Situational Triggers: Exclusionary Reactions to Immigrant Minorities*, „American Political Science Review”, Vol. 98, No. 1
- Sterngold A., Warland R. H., Herrmann R. O. (1994), *Do Surveys Overstate Public Concerns!*, „Public Opinion Quarterly”, Vol. 58, No. 2

- Tourangeau R., Rips L. J., Rasinski K. (2000), *The Psychology of Survey Response*, Cambridge University Press, Cambridge
- Tourangeau R., Groves R. M., Kennedy C., Ting Yan (2009), *The Presentation of a Web Survey, Nonresponse and Measurement Error among Members of Web Panel*, „Journal of Official Statistics”, Vol. 25, No. 3
- Transue J. E., Lee D. J., Aldrich J. H. (2009), *Treatment Spillover Effects across Survey Experiments*, „Political Analysis”, Vol. 17, No. 2

## SUMMARY

*The article has a dual purpose — besides a general introduction to the method of the survey experiment, and related research opportunities there are formulated research problems to which it can be recommended in the studies of public statistics, with particular emphasis on research aimed to evaluation of public programs. Its structure consists of three parts: a general description of the method and its potency, examples of the most important applications in the literature and a summary from the perspective of the modernization of social research in official statistics.*

## РЕЗЮМЕ

*Статья имеет двойную цель — рядом с общим введением в метод зондажного эксперимента и связанных с ним исследовательских возможностей, определяются исследовательские проблемы, к которым может он рекомендоваться в обследованиях официальной статистики, с особым учетом обследований направленных на оценку государственных программ. Его структуру составляют три части: общая характеристика метода и его потенциал, примеры важнейшего использования в литературе и подведение итогов с точки зрения модернизации социальных обследований в официальной статистике.*