

Mirosław Szreder*

Rejestry administracyjne – konkurencja czy uzupełnienie badań statystycznych?

Obserwowany w ostatnich dekadach dynamiczny rozwój technologii informacyjnych przyczynia się do rosnącego zakresu zmian w organizacji i przebiegu badań statystycznych niezależnie od ich tematyki. Zmiany te są szczególnie widoczne w badaniach tzw. oficjalnej statystyki, nazywanej w Polsce statystyką publiczną. Nie znaczy to jednak, że nie występują one w innego rodzaju badaniach statystycznych, np. prowadzonych dla celów *stricte* naukowych. Wykorzystanie elektronicznych formularzy

ankiet, elektroniczne panele respondentów, liczne internetowe bazy danych (służące m.in. do tworzenia operatu losowania), a także wszelkie komputerowe narzędzia przetwarzania i analizowania danych są tylko przykładami postępującej w tym zakresie ewolucji. Równie istotna – chociaż może mniej spektakularna zmiana – dokonuje się w zakresie tworzenia i obsługi elektronicznych rejestrów urzędowych, zwanych dalej rejestrami administracyjnymi. Rośnie zarówno ich liczba, jak i zakres oraz najczęściej także jakość gromadzonych przez te rejestry informacji. Niemal każdy z nas ma poczucie, że różnego rodzaju instytucje

* Prof. dr hab. Mirosław Szreder – Uniwersytet Gdański.

będące dysponentami tych informacji wiedzą o nas, o indywidualnych konsumentach, społeczeństwie i gospodarce, coraz więcej. W Austrii uznano, że na tyle dużo, iż powszechny spis ludności stał się w swojej tradycyjnej formie zbędny – możliwy do zastąpienia informacjami pobranymi z rejestrów urzędowych. *Austria jest jedynym dużym państwem, które w czasie jednej dekady zdecydowało się na kompleksową zmianę metodologii spisu z w pełni tradycyjnego na całkowicie oparty na rejestrach* – stwierdza Gołata [2018, s. 67]. Tendencja do uzupełniania lub zastępowania badań statystycznych, zwłaszcza wyczerpujących (obejmujących pomiarem wszystkie jednostki badanej zbiorowości), informacjami z rejestrów jest widoczna także w wielu innych dziedzinach życia gospodarczego i społecznego. Uzasadnione wydaje się więc pytanie: czy statystyka oparta na rejestrach (ang. *register-based statistics*) zmierza stopniowo do wyparcia, a ostatecznie zastąpienia tradycyjnych badań statystycznych, czy pozostanie jedynie ich uzupełnieniem?

Rejestry administracyjne – szanse i wyzwania

Być może ekonomistów najłatwiej byłoby przekonać do korzyści płynących z wykorzystania rejestrów administracyjnych przykładem odnoszącym się do badań wysokości wynagrodzeń w gospodarce. Podawana raz w miesiącu przez GUS informacja o średnim wynagrodzeniu w sektorze przedsiębiorstw (zatrudniających 10 osób i więcej) budzi u wielu osób niedosyt. Średnia arytmetyczna jest bowiem miarą tendencji centralnej, która jest wrażliwa na duże (nawet gdy nieliczne) wartości występujące w analizowanym szeregu wynagrodzeń. W gospodarce polskiej wynagrodzenie średnie lub wyższe otrzymuje jedynie 1/3 zatrudnionych, stąd oczekiwanie, że obok średniego wynagrodzenia GUS poda medianę wynagrodzeń, a być może także którąś z miar zróżnicowania (np. odchylenie standardowe). Tego jednak statystycy GUS-u nie są w stanie dokonać, bo w przeciwieństwie do obliczeń średniej arytmetycznej – będącej ilorzem dwóch agregatów: sumy funduszu płac wykorzystanego w danym miesiącu w przedsiębiorstwach i liczby zatrudnionych w nich osób – inne miary przeciętne oraz miary dyspersji wymagają znajomości indywidualnych wysokości wynagrodzeń w badanej zbiorowości. Aby wysokości te poznać, wykonywane jest co dwa lata reprezentacyjne badanie struktury wynagrodzeń na próbie prawie 35 tys. podmiotów zatrudniających co najmniej 10 osób. Wyniki tego badania stanowią szczegółową i rzetelną informację o strukturze wynagrodzeń według zawodów, a ponadto dostarczają informacji o zróżnicowaniu wynagrodzeń m.in. ze względu na płeć, poziom wykształcenia i wiek (por. [GUS, 2022]).

Badanie to ma jednak jedną podstawową słabość. Jest ono czasochłonne, co powoduje, że pełne jego wyniki publikowane są najczęściej dopiero po ponad roku (zwykle po 15 miesiącach). W chwili ogłoszenia tych wyników uwaga ekonomistów skierowana jest już na inny, bieżący moment czasu, zwłaszcza gdy z powodu wysokiej inflacji dynamika wynagrodzeń bywa także duża. Oczywiście świadomość dużego odstępu czasu między momentem badania a dniem opublikowania jego pełnych wyników istniała w środowisku ekonomistów i statystyków od dawna. Z tego powodu w Głównym Urzędzie Statystycznym prowadzone są od kilku lat zaawansowane prace nad zastąpieniem wspomnianego wyżej reprezentacyjnego badania struktury wynagrodzeń danymi pochodzącymi z rejestrów Zakładu Ubezpieczeń Społecznych (ZUS) oraz uzupełniająco z rejestrów Ministerstwa Finansów i Bazy Jednostek Statystycznych GUS. Pozwoliłoby to na wzbogacenie zakresu danych o wynagrodzeniach w gospodarce, znaczne przyśpieszenie ich publikacji, a dodatkowo istotnie obniżyłoby koszty badania i uwolniło przedstawicieli przedsiębiorstw uczestniczących w dotychczasowych badaniach od absorbującej pracy sprawozdawczej. Byłby to swego rodzaju przełom w badaniu wysokości i struktury wynagrodzeń w Polsce. Zaawansowanie tych prac i obiecujące wyniki omówione są w publikacji GUS [2023].

Rejestry administracyjne stają się coraz częściej ważnym źródłem danych dla statystyków w badaniach z różnych obszarów życia gospodarczego i społecznego. Niekiedy stanowią jedno z kilku źródeł informacji, jak na przykład w badaniach z zakresu produkcji rolnej, gdzie dodatkowo korzysta się z danych satelitarnych, szacunków rzeczoznawców oraz wyników badań biometrycznych. Częściej jednak rejestry traktowane są jako uzupełnienie badań reprezentacyjnych. Na przykład rejestr pod nazwą Centralna Baza Ofert Pracy (CBOP) służy łącznie z reprezentacyjnym badaniem pod nazwą „Popyt na pracę” do oszacowania liczby wolnych miejsc pracy w Polsce. Podobnie jest w badaniach GUS na temat transformacji polskiej gospodarki w kierunku gospodarki troszczącej się o stan środowiska naturalnego (zielonej gospodarki). Rośnie natomiast liczba tych badań statystycznych, w których dane z rejestrów – podobnie jak w omówionym wcześniej przypadku wynagrodzeń – w całości zastępują tradycyjne badania statystyczne. Przykładem może być estymacja liczby cudzoziemców na krajowym rynku pracy w ujęciu regionalnym, wymagająca wykorzystania kilku rejestrów łącznie, czyli danych: PESEL, KRUS, ZUS oraz Urzędu do Spraw Cudzoziemców.

W wielu innych badaniach statystyki publicznej informacje pochodzące z rejestrów służą do kontroli jakości danych z badań statystycznych, a także do

imputacji brakujących danych. Nie należy nie doceniać tego ostatniego zastosowania rejestrów. Braki w odpowiedziach respondentów stanowią bowiem coraz bardziej uciążliwy problem w większości badań reprezentacyjnych. Bez wiarygodnego, wysokiej jakości alternatywnego źródła informacji (np. z rejestru) konsekwencje błędów spowodowanych brakami w odpowiedziach (ang. *non-response errors*) mogą być bowiem bardzo poważne.

Trudności i wyzwań związanych z wykorzystaniem rejestrów administracyjnych w badaniach statystycznych jest niemało. Przede wszystkim trzeba sobie zdawać sprawę z tego, że sam rejestr nie jest jeszcze zbiorem danych statystycznych, czyli nie zawsze ma strukturę i uporządkowanie, jakich oczekuje się po zbiorach statystycznych. Dodatkowymi trudnościami w bezpośrednim wykorzystaniu rejestrów mogą być różne definicje podmiotów, których dotyczą dane, a także różne definicje zmiennych i kategorii używanych w rejestrach i w badaniach statystycznych. Różne mogą być w nich także klasyfikacje i grupowania. Wszystkie tego typu wyzwania są na ogół warte podjęcia, gdyż tendencja do coraz szerszego wykorzystania danych administracyjnych przez statystyków jest raczej trwała. Obszernie o tym zjawisku traktuje ubiegłoroczna publikacja Polskiej Agencji Rozwoju Przedsiębiorczości *Wtórne wykorzystanie danych administracyjnych w statystyce publicznej – obecnie i w przyszłości*, PARP [2022].

Czego rejestry nie ujmują?

We współczesnej, dynamicznie rozwijającej się sferze wymiany informacji można ulec złudzeniu, że niedługo tradycyjne badania statystyczne odejdą do lamusa – zastąpią je coraz doskonalsze rejestry oraz zbiory i narzędzia big data. Wydaje się jednak, że droga do pełnego zastąpienia badań statystycznych, w szczególności badań reprezentacyjnych, innymi zbiorami i narzędziami jest jeszcze dość długa. Warto tym razem sięgnąć do innego przykładu, który ilustruje i w pewnym stopniu uzasadnia tę tezę. Ścisłej rzecz biorąc prezentuje on sytuację, w której za uprawnione, a nawet konieczne uznać należy równoległe funkcjonowanie w określonej problematyce rejestrów administracyjnych i statystycznych badań sondażowych.

Jeden z takich przypadków omawia w swojej książce poświęconej statystycznemu pomiarowi przestępczości w Stanach Zjednoczonych Sharon L. Lohr [2019]. W USA – podobnie jak w wielu innych krajach – rozwijane i doskonalone są policyjne rejestry przestępstw. Do roku 2021 policyjne statystyki przestępstw funkcjonowały w ramach The Uniform Crime Reporting (UCR) Program, który po rozbudowie przekształcony został w The National Incident-Based Reporting System (NIBRS). W rejestrach tych

notowane są wszystkie zgłoszone policji przestępstwa oraz te, o których policja zdobyła wiedzę w inny sposób. Wydawać by się mogło, że policyjne statystyki dobrze opisują liczby i rodzaje przestępstw, a przez to wiarygodnie informują o bezpieczeństwie w danym kraju lub na jego określonym obszarze. Tymczasem, na co zwraca uwagę nie tylko autorka wspomnianej książki, w Stanach Zjednoczonych szacuje się, że jedynie 55% wszystkich najpoważniejszych przestępstw oraz tylko 35% przestępstw przeciwko mieniu zostaje odnotowanych w policyjnych rejestrach (dane za rok 2015, na podst. [Sharon, 2019, s. 37]). Wyjątkiem w tej drugiej grupie przestępstw są kradzieże pojazdów, dla których wskaźnik ten jest znacznie wyższy i wynosi ponad 75%. Wiąże się to oczywiście z wymogami zakładów ubezpieczeń, które przed wypłaceniem odszkodowania za kradzież żądają zwykle potwierdzenia zgłoszenia tego faktu policji. Autorka wskazuje na szereg przyczyn, z powodu których nie wszystkie przestępstwa, zarówno przeciwko życiu i zdrowiu człowieka, jak i przeciwko mieniu, są zgłaszane policji. Do najważniejszych z nich należą: nieświadomość niektórych osób, że stały się ofiarą przestępstwa (np. wykorzystane seksualnie dziecko), niechęć do zgłoszenia czynu przestępczego (np. w przypadkach przemocy domowej), poczucie zawstydzenia, że padło się ofiarą przestępstwa, przekonanie, że zdarzenie może być zbyt błahе, by traktować je jako przestępstwo, niechęć do kontaktu z policją (np. osób o nieuregulowanym statusie imigracyjnym). Ogół tych przyczyn sprawia, że policyjne statystyki – zapewne nie tylko w USA – są w dużym stopniu niepełne, czyli nie przedstawiają faktycznego stanu i struktury przestępczości. Istotnym uzupełnieniem tego rejestru jest prowadzone w USA od prawie 50 lat coroczne badanie reprezentacyjne gospodarstw domowych na temat wszelkiego rodzaju przestępstw (z wyłączeniem zabójstw), których w minionym roku doświadczyli członkowie tworzący dane gospodarstwo domowe. W badaniu tym, o nazwie The National Crime Victimization Survey (NCVS), bierze udział ok. 240 tys. respondentów z losowo wybranych 150 tys. gospodarstw domowych. Informacje zbierane w tym badaniu obejmują szereg aspektów przestępstw, spośród których za najważniejsze w kwestionariuszu badania należy uznać: charakterystykę przestępcy, szczegółowy opis samego zdarzenia, stwierdzenie, czy przestępstwo zostało zgłoszone policji, a jeżeli nie, to z jakich powodów, straty ofiary będące wynikiem przestępstwa (m.in. finansowe i zdrowotne). Wskaźnik odmów wśród wylosowanych do próby gospodarstw domowych nie jest duży – w 2021 r. nie przekroczył 33%. Uzyskane z tego badania wyniki są podstawą wnioskowania statystycznego o liczbie i strukturze przestępstw na całym obszarze Stanów Zjednoczonych. Dopiero oba te źródła – policyjny rejestr (NIBRS) oraz wyniki ankiety badania

gospodarstw domowych (NCVS) – pozwalają nie tylko dobrze scharakteryzować przestępczość, ale także trafniej zidentyfikować ułomności każdego z tych źródeł i wskazać właściwe sposoby ich doskonalenia.

Powyższy przykład pokazuje, że nawet dobrze zaprojektowany i zorganizowany rejestr administracyjny może wymagać dodatkowego wsparcia tradycyjnym badaniem statystycznym. Dlatego we współczesnych badaniach statystycznych na świecie dąży się przede wszystkim do łączenia kilku źródeł danych, z których najważniejszymi pozostają badania statystyczne i rejestry urzędowe. Na ogół sam rejestr nie stanowi wystarczającego źródła informacji, które mogłoby stanowić satysfakcjonującą alternatywę dla badania statystycznego. Ale i to się zmienia, gdyż kombinacja rejestrów administracyjnych i technik pozyskiwania

danych z internetu oraz sieci społecznościowych (big data) może stopniowo wypierać w wielu obszarach tradycyjne badania statystyczne.

Bibliografia

- Gołata E. [2018]. *Koniec ery tradycyjnych spisów ludności*. Poznań: Wydawnictwo UEP.
- GUS [2022]. *Struktura wynagrodzeń według zawodów w październiku 2020 r.*, Główny Urząd Statystyczny.
- GUS [2023]. *Zeszyt metodologiczny. Wynagrodzenia w gospodarce narodowej na podstawie źródeł administracyjnych*. Główny Urząd Statystyczny.
- Lohr S.L. [2019]. *Measuring Crime. Behind the Statistics*, CRS Press.
- PARP [2022]. *Wtórne wykorzystanie danych administracyjnych w statystyce publicznej – obecnie i w przyszłości*. Polska Agencja Rozwoju Przedsiębiorczości.