

Anna Smaga<sup>1</sup>, Anna Tymicka<sup>1</sup>, Marta Niedźwiedzka-Stadnik<sup>2</sup>, Bartosz Napiórkowski<sup>3</sup>,  
Tomasz Stachurski<sup>3</sup>, Magdalena Rosińska<sup>2</sup>

## DEVELOPMENT OF AN HIV CASE DEFINITION USING INSURANCE CLAIMS DATA IN POLAND: A STEP TOWARD COMPLEMENTING NATIONAL SURVEILLANCE WITH NATIONAL HEALTH FUND DATA

### OPRACOWANIE DEFINICJI PRZYPADKU HIV Z WYKORZYSTANIEM DANYCH ROZLICZENIOWYCH Z SYSTEMU FINANSOWANIA ŚWIADCZEŃ ZDROWOTNYCH W POLSCE: KROK W KIERUNKU UZUPEŁNIENIA KRAJOWEGO NADZORU O DANE NARODOWEGO FUNDUSZU ZDROWIA

<sup>1</sup>Department of Population Health Monitoring and Analysis,  
National Institute of Public Health NIH – National Research Institute, Poland  
Zakład Monitorowania i Analiz Stanu Zdrowia Ludności,  
Narodowy Instytut Zdrowia Publicznego PZH – Państwowy Instytut Badawczy

<sup>2</sup>Department of Epidemiology of Infectious Diseases and Surveillance,  
National Institute of Public Health NIH – National Research Institute, Poland  
Zakład Epidemiologii Chorób Zakaźnych i Nadzoru,  
Narodowy Instytut Zdrowia Publicznego PZH – Państwowy Instytut Badawczy

<sup>3</sup>Department of Quality Monitoring Analysis and Service Optimization, National Health Fund  
Departament Analiz Monitorowania Jakości i Optymalizacji Świadczeń, Narodowy Fundusz Zdrowia

#### ABSTRACT

**BACKGROUND.** Electronic health records and reimbursement claims data are increasingly recognized as valuable resources for enhancing health surveillance, particularly in evaluating continuity of care. However, their use is often complicated by coding errors.

**OBJECTIVE.** The aim of the study was to develop case definitions for a person living with HIV in Poland and for a newly diagnosed HIV case, based on claims data.

**MATERIAL AND METHODS.** Data on services and prescriptions reimbursed by the National Health Fund (NHF) between January 2010 and June 2022, in Poland, were used to construct candidate definitions for a person with diagnosed HIV (PHIV-D). Based on the selected definition and analysis of the time intervals between registered services, the size of the PHIV-D population residing in Poland was estimated. To evaluate the validity of the approach, the results were compared with national HIV surveillance data using mean squared error and visual trend analysis. Percentage growth was applied to assess the reporting delay in NHF data.

**RESULTS.** According to the selected definition a PHIV-D was defined as an individual who had used at least 2 services funded by NHF with an HIV-related ICD-10 code listed as either main or concurrent cause. PHIV-D who did not use any NHF-funded services or prescriptions for consecutive 24 months, and had no recorded death were considered likely to have left the country. Based on this definition, the estimated PHIV-D population residing in Poland increased from 8.9 thousand in 2013 to 15.7 thousand in 2021, with 10,993 new diagnoses recorded during this period.

**CONCLUSIONS.** The developed definition allows to estimate the size of the PHIV-D population in Poland, which is challenging for existing surveillance system. It provides a valuable tool for monitoring and planning healthcare services for individuals living with HIV.

**Keywords:** *surveillance, HIV infection, claims-based surveillance, Poland*

\* The authors had equal contribution in the preparation of the article / Autorzy w równym stopniu przyczynili się do powstania artykułu

## STRESZCZENIE

**WPROWADZENIE.** Elektroniczna dokumentacja medyczna i dane z wniosków o refundację są coraz częściej uznawane za cenne zasoby wzmacniające nadzór epidemiologiczny, szczególnie w ocenie ciągłości opieki. Jednak ich wykorzystanie często utrudniają błędy w kodowaniu.

**CEL.** Celem badania było opracowanie definicji przypadku osoby żyjącej ze zdiagnozowanym zakażeniem HIV w Polsce oraz nowego rozpoznania HIV, na podstawie danych dotyczących zrealizowanych świadczeń.

**MATERIAŁ I METODY.** Dane dotyczące świadczeń i recept refundowanych przez Narodowy Fundusz Zdrowia (NFZ) w okresie od stycznia 2010 r. do czerwca 2022 r. w Polsce posłużyły do skonstruowania kandydatów na definicje osoby ze zdiagnozowanym zakażeniem HIV (PHIV-D). Na podstawie wybranej definicji i analizy przerw między zarejestrowanymi świadczeniami, oszacowano liczebność populacji PHIV-D zamieszkałej w Polsce. Aby ocenić trafność tego podejścia, wyniki porównano z krajowymi danymi z nadzoru epidemiologicznego nad HIV, stosując średni błąd kwadratowy i wizualną analizę trendów. Do oceny opóźnienia raportowania w danych NFZ zastosowano wskaźnik wzrostu procentowego.

**WYNIKI.** Zgodnie z wybraną definicją, PHIV-D została zdefiniowana jako osoba, która skorzystała z co najmniej 2 świadczeń finansowanych przez NFZ, z kodem ICD-10 związanym z HIV wymienionym jako przyczyna główna lub współistniejąca. Osoby z PHIV-D, które nie korzystały z żadnych świadczeń ani recept finansowanych przez NFZ przez kolejne 24 miesiące i u których nie odnotowano zgonu, uznawano za osoby, które prawdopodobnie wyemigrowały. Na podstawie tej definicji, szacunkowa populacja osób z PHIV-D zamieszkujących w Polsce wzrosła z 8,9 tys. w 2013 r. do 15,7 tys. w 2021 r., przy czym w tym okresie odnotowano 10 993 nowe rozpoznania.

**WNIOSKI.** Opracowane definicje pozwalają oszacować liczebność populacji PHIV-D w Polsce, co stanowi wyzwanie dla istniejącego systemu nadzoru. Są one cennym elementem wspierającym monitorowanie i planowanie świadczeń zdrowotnych dla osób żyjących z HIV.

**Słowa kluczowe:** zakażenie HIV, nadzór, nadzór oparty na danych o świadczeniach, Polska

## INTRODUCTION

Estimating the size of the population of people with diagnosed HIV (PHIV-D) residing in the country, at any given time is crucial for monitoring the development of the epidemic. In Poland, new diagnoses of HIV infection and AIDS incidence are registered under the Law on Prevention and Control of Infections and Infectious Diseases in Humans (Journal of Laws 2008 No. 234 item 1570, as amended). As part of epidemiological surveillance, physicians and laboratory diagnosticians are required to report new diagnoses, but monitoring continuity of care remains a challenge (1). In addition, there are no data on the migration of PHIV-D (from and to Poland) and deaths of PHIV-D people that occurred from causes other than AIDS (2).

Ongoing monitoring of the epidemic's development in Poland is especially critical given the recent change in migration patterns. Since 2022, there has been a significant influx of war refugees from Ukraine (approximately 1 million in temporary protection at the beginning of 2023), along with a growing number of immigrants from other countries (3). Another contributing factor potentially influencing the dynamics of the HIV epidemic in Poland is the occurrence of risky sexual behaviour among Polish

## WSTĘP

Oszacowanie liczebności populacji osób z rozpoznaniem zakażeniem HIV (PHIV-D) zamieszkujących w kraju w danym momencie jest kluczowe dla monitorowania rozwoju epidemii. W Polsce nowe rozpoznania zakażenia HIV i zachorowania na AIDS są rejestrowane na podstawie ustawy o zapobieganiu oraz zwalczaniu zakażeń i chorób zakaźnych u ludzi (Dz. U. 2008 Nr 234 poz. 1570 z późn. zm.). W ramach nadzoru epidemiologicznego lekarze i diagnosty laboratoryjni są zobowiązani do zgłaszania nowych rozpoznań, ale monitorowanie ciągłości opieki pozostaje wyzwaniem (1). Ponadto brak jest danych na temat migracji PHIV-D (z i do Polski) oraz zgonów PHIV-D, które nastąpiły z przyczyn innych niż AIDS (2).

Ciągły monitoring rozwoju epidemii w Polsce jest szczególnie istotny, biorąc pod uwagę ostatnie zmiany wzorców migracyjnych. Od 2022 r. obserwuje się znaczący napływ uchodźców wojennych z Ukrainy (około 1 miliona osób objętych ochroną tymczasową na początku 2023 r.), a także rosnącą liczbę imigrantów z innych krajów (3). Innym czynnikiem potencjalnie wpływającym na dynamikę epidemii HIV w Polsce jest występowanie ryzykownych zachowań seksualnych wśród polskich emigrantów, którzy mogą podejmować takie działania za granicą i utrzymywać

out-migrants, who may engage in such activities abroad and maintain contact with residents in Poland either during temporary visits or upon return (4). Estimating the number of out-migrating and possibly re-entering population is not feasible through current surveillance mechanisms. To address these challenges, we intend to use data from the National Health Fund (NHF), Poland's sole public healthcare payer.

The electronic health records (EHR) and also reimbursement claims are increasingly recognised as a source of data that could be used for informing public health decisions. Despite multiple challenges such as the poor data quality in EHRs, difficulty in analysing, cleaning and accessing unstructured data, data privacy and security and the lack of interoperability standards the EHR-based systems are considered a way forward to modernising surveillance (5,6). A subset of EHR is usually reported to the payer to obtain reimbursement for services. Such claims data, depending on the context may provide useful information, in particular, when the cross-field questions are investigated requiring integration of data from different providers (7,8). Usually, in order to efficiently use these data, it is necessary to understand how to efficiently identify the cohort of interest based on the reported claims. Prior research in the U.S.A. indicates that this is also the case for identification of PHIV-D in the claims data (9,10). We therefore aimed at exploring whether the claims data could be used for HIV monitoring in Poland. We developed an algorithm to identify PHIV-D individuals residing in Poland and newly diagnosed HIV case, using NHF data on medical services related to HIV care – as well as unrelated services – and data on deaths from all causes. This data source and the algorithm enable regular monitoring of the PHIV-D population size, with updates available on a monthly basis.

## MATERIALS AND METHODS

**Study design.** We have conducted secondary data analysis basing on reimbursement claims data. The following steps were undertaken in the study: 1) identification of relevant ICD-10 codes and medical service types, 2) development of candidate definitions, 3) data extraction and sensitivity analysis, and 4) data triangulation with other data sources.

Since HIV treatment is conducted in specialised centres (11) and severe opportunistic infections are typically treated in infectious disease clinics, the data on services performed in specialised clinics and hospitals was considered potentially more reliable than those reported from general practice. In addition, since HIV treatment requires regular monitoring, it was considered that the expected number of services

kontakt z mieszkańcami w Polsce podczas wizyt tymczasowych lub po powrocie (4). Oszacowanie liczebności populacji migrującej i potencjalnie powracającej nie jest możliwe za pomocą obecnych mechanizmów nadzoru. Możliwym rozwiązaniem jest wykorzystanie danych rozliczeniowych zbieranych przez Narodowy Fundusz Zdrowia (NFZ), jedyne publicznego płatnika opieki zdrowotnej w Polsce.

Elektroniczna dokumentacja medyczna (EHR) oraz wnioski o zwrot kosztów zrealizowanych świadczeń są coraz częściej uznawane za źródła danych, które mogą być wykorzystane do podejmowania decyzji w zakresie zdrowia publicznego. Pomimo licznych wyzwań, takich jak niska jakość danych w EHR, trudności w analizie, czyszczeniu i dostępie do nieustrukturyzowanych danych, prywatność i bezpieczeństwo danych oraz brak standardów interoperacyjności, systemy oparte na EHR są uważane za sposób na usprawnienie nadzoru (5,6). Podzbiór EHR jest zazwyczaj zgłaszany płatnikowi w celu rozliczenia świadczeń. Dane rozliczeniowe, w zależności od kontekstu, mogą dostarczyć użytecznych informacji, w szczególności w przypadku badania kwestii międzydziedzinowych wymagających integracji danych pochodzących z różnych placówek medycznych (7,8). Zazwyczaj, aby efektywnie wykorzystać te dane, konieczne jest zrozumienie, jak skutecznie zidentyfikować interesującą kohortę na podstawie rozliczonych świadczeń. Wcześniejsze badania w USA wskazują, że dotyczy to również identyfikacji PHIV-D w tego typu danych (9,10). Stąd naszym celem było zbadanie, czy dane dotyczące rozliczonych świadczeń zdrowotnych mogłyby być wykorzystane do monitorowania HIV w Polsce. Opracowaliśmy algorytm identyfikujący PHIV-D mieszkające w Polsce oraz nowo zdiagnozowane przypadki HIV, wykorzystując dane NFZ dotyczące usług medycznych – zarówno tych bezpośrednio związanych z opieką nad osobami żyjącymi z HIV, jak i pozostałych – a także dane dotyczące zgonów z wszystkich przyczyn. To źródło danych i zaproponowany przez nas algorytm umożliwiają regularne monitorowanie liczebności populacji PHIV-D, z miesięcznymi aktualizacjami.

## MATERIAŁ I METODY

**Schemat badania.** Przeprowadziliśmy analizę danych wtórnych w oparciu o dane z wniosków o zwrot kosztów zrealizowanych świadczeń. W badaniu podjęto następujące kroki: 1) identyfikacja odpowiednich kodów ICD-10 i rodzajów usług medycznych, 2) opracowanie kilku wariantów definicji, 3) ekstrakcja danych i analiza wrażliwości oraz 4) triangulacja danych z innymi źródłami danych.

with ICD-10 code indicative of HIV infection would be high. It was also assumed that individuals with no recorded services or reimbursed prescription, over an extended period and not identified as deceased, were likely to have left the country, permanently or temporarily. We were not able to use the antiretroviral medication prescriptions as the ART is provided via treatment programme coordinated by the National AIDS Centre, and the prescriptions are not reported to the NHF.

Candidate definitions for person with diagnosed HIV were therefore developed based on the number of specific types of services, such as specialist care, hospitalizations, or primary care services, and the impact of each parameter was investigated in the sensitivity analysis.

All analyses requiring processing of personal data, including pseudonymized data, were performed by the NHF analytical team. Only fully anonymized data were released with the granularity level of demographic characteristics ensuring that at least 5 cases fell into each of demographic categories (e.g. using age group instead of age, residence type category instead of exact place of residence). Datasets used for data triangulation included epidemiological surveillance dataset (case-based reporting) and data on the number of individuals in antiretroviral treatment (ARV).

**Data source.** In order to define the PHIV-D population in Poland, individuals with at least one HIV related service were identified from NHF between January 2010 and June 2022. NHF is the single public payer in Poland financing public health services based on universal national insurance. The NHF data include information on outpatient visits, hospitalisations (service data) by provider, as well as information on prescriptions.

The services were assigned to patients using personal identification numbers, allowing analysis of longitudinal patient data over time. The period between 2010 and 2022 was chosen for analysis, as centralized health service database was not available before 2010, and the substantial migration influx in 2022 could introduce differences in patterns of service use compared to resident Polish population. In the National Health Fund (NHF) data services (i.e. outpatient visit or hospitalization) are assigned ICD-10 codes corresponding to the diagnoses relevant to this particular service. In order to identify the HIV-related services we used a set of ICD-10 codes as specified in the Table 1.

Visits to outpatient specialised care or hospitalisation in infectious disease treatment facilities and those providing addiction therapy were considered as specialised services. To define PHIV-D residing in Poland, prescription data from NHF were also used.

Ponieważ leczenie HIV jest prowadzone w wyspecjalizowanych ośrodkach (11), a ciężkie zakażenia oportunistyczne są zazwyczaj leczone w klinikach chorób zakaźnych, dane dotyczące świadczeń udzielanych w specjalistycznych klinikach i szpitalach uznano za potencjalnie bardziej wiarygodne niż dane zgłaszane przez lekarzy ogólnych. Ponadto, ponieważ leczenie HIV wymaga regularnego monitorowania, uznano, że oczekiwana liczba świadczeń oznaczonych kodem ICD-10 wskazującym na zakażenie HIV będzie wysoka. Założono również, że osoby, które nie odnotowały świadczeń ani nie otrzymały refundowanych recept przez dłuższy czas, prawdopodobnie opuściły kraj na stałe lub tymczasowo. Nie byliśmy w stanie wykorzystać danych o przepisanych lekach antyretrowirusowych, z uwagi na fakt, że terapia ta jest realizowana w ramach programu leczenia koordynowanego przez Krajowe Centrum ds. AIDS, a recepty nie są zgłaszane do NFZ.

Warianty definicji osoby ze zdiagnozowanym zakażeniem HIV opracowano zatem na podstawie liczby określonych rodzajów usług, takich jak opieka specjalistyczna, hospitalizacje lub podstawowa opieka zdrowotna, a wpływ każdego parametru zbadano w analizie wrażliwości.

Wszystkie analizy wymagające przetwarzania danych osobowych, w tym danych pseudonimizowanych, zostały przeprowadzone przez zespół analityczny NFZ. Udostępniono wyłącznie w pełni zanonimizowane dane, charakteryzujące się szczegółowością cech demograficznych, zapewniającej, że co najmniej 5 przypadków mieściło się w każdej kategorii demograficznej (np. grupa wiekowa zamiast wieku, kategoria miejsca zamieszkania zamiast dokładnego miejsca zamieszkania). Zbiory danych wykorzystane do triangulacji obejmowały zbiór danych pochodzący z nadzoru epidemiologicznego (zgłoszenia nowo rozpoznanych przypadków) oraz dane dotyczące liczby osób objętych leczeniem antyretrowirusowym (ARV).

**Źródło danych.** Aby zdefiniować populację PHIV-D w Polsce, zidentyfikowano osoby, które skorzystały z co najmniej jednej usługi związanej z HIV w ramach NFZ w okresie od stycznia 2010 r. do czerwca 2022 r. NFZ jest jedynym płatnikiem publicznym w Polsce, finansującym publiczne świadczenia zdrowotne w oparciu o powszechne ubezpieczenie społeczne. Dane NFZ obejmują informacje o wizytach ambulatoryjnych, hospitalizacjach (dane dotyczące usług) według świadczeniodawcy, a także informacje o receptach.

Świadczenia przypisano pacjentom za pomocą numerów identyfikacyjnych, co umożliwiło analizę wzdłużną danych pacjentów w czasie. Do analizy wybrano okres między 2010 a 2022 rokiem, ponieważ scentralizowana baza danych usług zdrowotnych nie była dostępna przed 2010 rokiem, a znaczny napływ

Table 1. ICD-10 codes used for the identification of HIV cases  
Tabela 1. Kody ICD-10 zastosowane dla identyfikacji osób zakażonych HIV

HIV	B20.0, B20.1, B20.2, B20.3, B20.4, B20.5, B20.6, B20.7, B20.8, B20.9, B21.0, B21.1, B21.2, B21.3, B21.7, B21.8, B21.9, B22.0, B22.1, B22.2, B22.7, B23.0, B23.1, B23.2, B23.8, B24, Z21, F02.4, R75
-----	---

**Development of definition of person with diagnosed HIV.** To build the definition of a person living with a diagnosis of HIV, three conditions were considered: 1) the occurrence of any HIV-related service; 2) the receipt of an HIV-related service within specialized care; and 3) the occurrence of an HIV-related hospitalization. In addition, the number of occurrences of these conditions for each individual was considered. The candidate definitions were then constructed by combining these conditions.

Assessment criteria for selecting the optimal definition:

1. Comparability to other existing data

The selection was made by comparing the number of new diagnoses of HIV infection according to the tested definition to epidemiological surveillance data. Mean squared error and visual trend analysis were used on annual count of new diagnoses. In the NHF data, a new diagnosis was defined as the date of the first recorded health service with an HIV-related code.

Due to a possible underestimation for 2020-2021 in the surveillance data (delays related to reporting during the COVID-19 pandemic years), the comparison was limited to 2013-2019. The year 2013 was chosen as a starting point to minimise the influence of previously diagnosed cases (before 2010), which – due to the characteristics of the database – may have appeared in the early years of observation period.

2. Simplicity

Preference was given to definitions that minimized complexity while maintaining sufficient accuracy and clinical relevance. The simplicity was understood as the application of an algorithm with the smallest possible number of conditions.

3. Reducing the risk of delayed inclusion as meeting the definition

Definitions are constructed in a way that subsequent data updates can affect historical results. If the condition requires more than one HIV-related service, the patient's inclusion may be delayed. To evaluate the extent of this effect, we analyzed the number of cases retrospectively added to previous years across three database updates. Based on percentage growth in the following year, we assessed whether these updates significantly influenced epidemiological trends and the overall estimated population size.

**Identifying the population of people diagnosed with HIV residing in Poland.** In order to calculate the number of people living with HIV in Poland we

migracyjny w 2022 roku mógł wprowadzić różnice w schematach korzystania z usług w porównaniu z rezydentami polskiej populacji.

W danych Narodowego Funduszu Zdrowia (NFZ) do świadczeń (np. wizyta ambulatoryjna lub hospitalizacja) są przypisane kody ICD-10 odpowiadające rozpoznaniom związanym z danym świadczeniem. Do identyfikacji świadczeń związanych z HIV wykorzystano zestaw kodów ICD-10, jak określono w Tabeli 1.

Wizyty w ambulatoryjnej opiece specjalistycznej lub hospitalizacje w placówkach leczenia zakaźnego oraz w placówkach prowadzących terapię uzależnień uznano za usługi specjalistyczne. Do zdefiniowania statusu PHIV-D zamieszkałego w Polsce wykorzystano również dane dotyczące recept z NFZ.

**Opracowanie definicji osoby z rozpoznaniem zakażeniem HIV.** Aby zbudować definicję osoby żyjącej z rozpoznaniem zakażenia HIV, wzięto pod uwagę trzy warunki: 1) wystąpienie jakiegokolwiek usługi związanej z HIV; 2) otrzymanie usługi związanej z HIV w ramach opieki specjalistycznej; oraz 3) wystąpienie hospitalizacji związanej z HIV. Dodatkowo wzięto pod uwagę liczbę wystąpień tych warunków u każdej osoby. Następnie, poprzez połączenie tych warunków, skonstruowano 3 warianty definicji.

Kryteria oceny służące do wyboru optymalnej definicji:

1. Porównanie z innymi istniejącymi danymi

Wyboru dokonano poprzez porównanie liczby nowych rozpoznań zakażenia HIV zgodnie z testowaną definicją z danymi z nadzoru epidemiologicznego. Do porównania rocznej liczby nowych rozpoznań wykorzystano analizę średniego błędę kwadratowego i wizualną analizę trendów. W danych NFZ nowe rozpoznanie zdefiniowano jako datę pierwszego zarejestrowanego świadczenia z kodem związanym z zakażeniem HIV.

Ze względu na możliwe niedoszacowanie danych z nadzoru za lata 2020-2021 (opóźnienia związane z raportowaniem w latach pandemii COVID-19), porównanie ograniczono do lat 2013-2019. Rok 2013 wybrano jako punkt wyjścia, aby zminimalizować wpływ przypadków zdiagnozowanych wcześniej (przed 2010 r.), które – ze względu na specyfikę bazy danych – mogły pojawić się we wczesnych latach okresu obserwacji.

2. Prostota

Preferowano definicje minimalizujące złożoność, przy jednoczesnym zachowaniu wystarczającej do-

need to account for out-migration of people with HIV diagnosis. Information about leaving Poland (temporary or permanent) is not recorded in the NHF data or other data sources of public statistics linked to NHF data. In order to address this issue, we hypothesize that people who reside in the country use health care services. In fact, OECD estimates that an average rate of outpatient consultations across countries was 6.8 consultations per person per year, with most countries reporting between four and ten. In Poland, the rate was 7.8 consultations per person per year. Reported statistics in USA, Australia and the UK demonstrate that 80% – 90% of individuals visits general practitioner at least once yearly (12–15). We expect that this percent may be even larger for PHIV-D. This is a consequence of the characteristics of HIV disease, which, if left untreated for a long time, can develop into AIDS. In order to establish the cut-off for time-since-last-service/prescription, above which the person will be considered to have out-migrated we analysed the history of individual patients meeting the selected case definition. Specifically, we examined the distribution of time intervals between services for any reason.

For analytical purposes, each interval was classified into one of three types:

1. gap – an interval between two consecutive services, where both are observed during the study period;
2. last – an interval after the last recorded service or prescription, with no subsequent service observed before the end of follow-up;
3. death – an interval ending in a registered date of death.

Subsequently, we described the distribution of the length of time intervals between consecutive registered services, focusing on those of the “gap” type, for all individuals meeting the selected PHIV-D definition.

The calibration of the definition based on intervals was performed using first dataset (up to end of 2021), before data for the first half of 2022 became available. The analysis covered all services financed by the NHF, regardless of the associated diagnosis, including outpatient medical appointments, hospital admissions and drug prescriptions. For services lasting more than one day, they are treated as occurring in all months from start to finish.

In addition, we distinguished in the analysis the last intervals recorded in patients:

- Interval from last recorded service for a patient with no death recorded – number of months to 31/12/2021 including the month 12/2021 (e.g. if the last service was 13/11/2021 then the number of months with no service is 1). Interval type – ‘last’.
- Gap from last recorded service for a patient with a death – number of months to date of death

kładności i trafności klinicznej. Prostotę rozumiano jako zastosowanie algorytmu z najmniejszą możliwą liczbą warunków.

3. Zmniejszenie ryzyka opóźnionego włączenia w przypadku spełnienia definicji

Definicje są konstruowane w taki sposób, że kolejne aktualizacje danych mogą wpływać na wyniki historyczne. Jeśli definicja wymaga więcej niż jednej usługi związanej z HIV, włączenie pacjenta do badania może zostać opóźnione. Aby ocenić skalę tego efektu, przeanalizowaliśmy liczbę przypadków retrospektywnie dodanych do danych z poprzednich lat w ramach trzech aktualizacji bazy danych. Na podstawie procentowego wzrostu w kolejnym roku oceniliśmy, czy te aktualizacje miały istotny wpływ na trendy epidemiologiczne i ogólną szacowaną wielkość populacji.

**Identyfikacja populacji osób zdiagnozowanych z HIV zamieszkujących w Polsce.** Aby obliczyć liczbę osób żyjących z HIV w Polsce, musieliśmy uwzględnić migrację osób z rozpoznaniem HIV. Informacje o opuszczeniu Polski (czasowym lub stałym) nie są rejestrowane w danych NFZ ani w innych źródłach danych statystyki publicznej powiązanych z danymi NFZ. Aby rozwiązać ten problem, stawiamy hipotezę, że osoby mieszkające w kraju korzystają z usług opieki zdrowotnej. W rzeczywistości, OECD szacuje, że średni wskaźnik wizyt ambulatoryjnych w różnych krajach wynosił 6,8 wizyty na osobę rocznie, przy czym większość krajów zgłasza od czterech do dziesięciu. W Polsce, wskaźnik ten wynosi 7,8 wizyt na osobę w ciągu jednego roku. Statystyki raportowane w USA, Australii i Wielkiej Brytanii pokazują, że 80% – 90% osób odwiedza lekarza rodzinnego co najmniej raz w roku (12-15). Oczekujemy, że ten odsetek może być jeszcze wyższy w przypadku PHIV-D. Jest to konsekwencja charakterystyki choroby HIV, która, jeśli będzie przez długi czas nieleczone, może rozwinąć się w AIDS. Aby ustalić granicę czasu od ostatniej zrealizowanej usługi/recepty, powyżej której dana osoba zostanie uznana za osobę, która wyemigrowała, przeanalizowaliśmy historię poszczególnych pacjentów spełniających wybraną definicję osoby zdiagnozowanej. W szczególności zbadaliśmy rozkład odstępów czasu między dowolnymi świadczeniami.

Dla celów analitycznych każdy interwał został zaklasyfikowany do jednego z trzech typów:

1. przerwa – interwał pomiędzy dwoma kolejnymi świadczeniami, przy czym oba są obserwowane w okresie badania;
2. ostatni – okres po ostatnim odnotowanym świadczeniu lub recepcie, bez zaobserwowania żadnego kolejnego świadczenia przed końcem okresu obserwacji;

including the month in which the death occurred. (e.g. if the last service was 13/11/2020 and the death was 15/01/2021 then the number of months without services is 1). Interval type - 'death'.

## RESULTS

**HIV population definition.** At least 1 HIV-related service occurred in 39 459 people alive in June 2022. Table 2 shows the breakdown of this population by the number of registered HIV-related services in total and those provided in specialised services.

Drawing on the data presented in the Table 2, and focusing on cells with a substantial number of cases, three definitions were identified for further analysis:

1. Definition 1: At least 2 HIV-related services – the patient received at least 2 services funded by the National Health Fund between 2010 and 2021 with an ICD-10 code related to HIV infection as the main or concurrent diagnosis.
2. Definition 2: At least 2 HIV-related services, and 1 specialised HIV-related service – the patient between 2010 and 2021 met the Definition 1 and registered a service with an HIV-related diagnosis at a facility specializing in infectious diseases.
3. Definition 3: At least 2 HIV-related services or 1 specialised service – the patient between 2010 and 2021 has met the Definition 1 or has registered a service with an HIV-related diagnosis in a facility specializing in infectious diseases.

The total number of PHIV-D was the highest based the Definition 3 (22 432 individuals) and the lowest based on the Definition 2 (17 772 individuals) (Table 2). The size of cohort identified with the Definition 1 exceed the number identified by the Definition 2 by 4% (18 476 vs. 17 772). A substantial proportion of individuals identified by the Definition 3 (3 956, 18%) had only a single HIV-related visit.

Regardless of the definition adopted, the date of diagnosis is defined as the date of the first HIV-related

3. śmierć – okres kończący się zarejestrowaną datą zgonu.

Następnie opisaliśmy rozkład długości odstępów czasu pomiędzy kolejnymi zarejestrowanymi usługami, ze szczególnym uwzględnieniem usług typu „przerwa” dla wszystkich osób spełniających wybraną definicję PHIV-D.

Kalibrację definicji opartej na okresach bez korzystania z usług przeprowadzono przy użyciu pierwszego zestawu danych (do końca 2021 r.), zanim zostały udostępnione dane za pierwszą połowę 2022 r. Analiza objęła wszystkie świadczenia finansowane przez NFZ, niezależnie od diagnozy, w tym wizyty ambulatoryjne, hospitalizacje i recepty na leki. W przypadku świadczeń trwających dłużej niż jeden dzień, uznaje się je za realizowane we wszystkich miesiącach od początku do końca trwania świadczenia.

Dodatkowo w analizie wyróżniliśmy ostatnie interwały w korzystaniu z usług zarejestrowane u pacjentów:

- Odstęp czasu od ostatniej zarejestrowanej usługi dla pacjenta, dla którego nie odnotowano zgonu – liczba miesięcy do 31.12.2021 r., wliczając miesiąc 12.2021 r. (np. jeśli ostatnia usługa miała miejsce 13.11.2021 r., to liczba miesięcy bez usługi wynosi 1). Typ interwału – „ostatni”.
- Odstęp od ostatniej zarejestrowanej usługi dla pacjenta, który zmarł – liczba miesięcy do daty zgonu, wliczając miesiąc, w którym nastąpił zgon. (np. jeśli ostatnia usługa miała miejsce 13.11.2020, a zgon nastąpił 15.01.2021, wówczas liczba miesięcy bez usług wynosi 1). Typ interwału – „zgon”.

## WYNIKI

**Definicja populacji żyjącej ze zdiagnozowanym zakażeniem HIV.** Stan na czerwiec 2022 r. wskazuje, że 39 459 osób miało odnotowane w systemie co najmniej jedno świadczenie związane z HIV. Tabela 2 przedstawia podział tej populacji według liczby zare-

Table 2. Number of HIV individuals alive in June 2022, by the conditions considered in the case definition

Tabela 2. Liczba osób żyjących z HIV w czerwcu 2022 r. według warunków uwzględnionych w definicji przypadku

Number of patients for a given combination		Number of HIV-related services				Sum
		1	2	3	>=4	
Number of HIV-related services recorded by specialised facility	0	17 027	517	92	95	17 731
	1	3 956	95	17	20	4 088
	2		1 085	60	28	1 173
	>=3			575	15 892	16 467
	Total	20 983	1 697	744	16 035	39 459
Definition 1			1 697	744	16 035	18 476
Definition 2			1 180	652	15 940	17 772
Definition 3		3 956	1 697	744	16 035	22 432

health service. Accordingly, a person is regarded as newly diagnosed with HIV in a given year if their first HIV-related service occurred in that year.

Definition 3 (at least 2 HIV services or 1 specialised service) identifies the largest annual numbers of newly diagnosed patients, that significantly exceeds the number of new HIV diagnoses recorded in 2013-2018 according to surveillance data (Fig. 1). The deviations are much larger than the other two approaches, so this definition was rejected at the outset. The mean squared error of this solution with respect to surveillance data was 74 817.

The other definitions give visually similar results, but a lower mean squared error (9 098 vs. 15 412) on the analysed period 2013-2019 compared to the surveillance data is given by Definition 1 (at least 2 HIV services). This definition is also favoured by the simplicity criterion.

The authors therefore chose the Definition 1 for further analysis. This definition requires at least two HIV-related services for a patient to be included in PHIV-D population but takes the time of occurrence of the first service as the moment of diagnosis. It may lead to an underestimation of the new diagnoses particularly for the latest years of data. An estimate of the magnitude of this underestimation was made by comparing the annual numbers of new diagnoses in the adult population calculated from the datasets taken in 03/2022, 06/2023 and 12/2023 (Table 3, Fig. 2). The number of new diagnoses in the latest year changed only marginally from data extracted 6 month and 12 months after the end of the year.

jestrowanych świadczeń związanych z wirusem HIV ogółem oraz usług świadczonych w placówkach specjalistycznych.

W oparciu o dane przedstawione w Tabeli 2, skupiając się na komórkach zawierających znaczną liczbę przypadków, sformułowano trzy definicje do dalszej analizy:

1. Definicja 1. Co najmniej 2 świadczenia związane z zakażeniem HIV – pacjent w latach 2010-2021 zrealizował co najmniej 2 świadczenia finansowane przez Narodowy Fundusz Zdrowia z kodem ICD-10 związanym z zakażeniem HIV na miejscu rozpoznania głównego lub współistniejącego.
2. Definicja 2. Co najmniej 2 świadczenia związane z HIV, w tym co najmniej 1 zrealizowane w placówce specjalistycznej – pacjent w latach 2010-2021 spełniał definicję 1. i zarejestrował się w placówce specjalizującej się w chorobach zakaźnych, w której zdiagnozowano zakażenie HIV.
3. Definicja 3. Co najmniej 2 świadczenia związane z HIV lub 1 świadczenie zrealizowane w placówce specjalistycznej – pacjent w latach 2010-2021 spełnił Definicję 1. lub zarejestrował się w placówce specjalizującej się w chorobach zakaźnych, w której zdiagnozowano zakażenie HIV.

Łączna liczba PHIV-D była najwyższa w oparciu o definicję 3. (22 432 osoby), a najniższa w oparciu o definicję 2. (17 772 osoby) (Tabela 2). Liczebność kohorty zidentyfikowanej w oparciu o definicję 1. przewyższała liczbę osób zidentyfikowanych w opar-

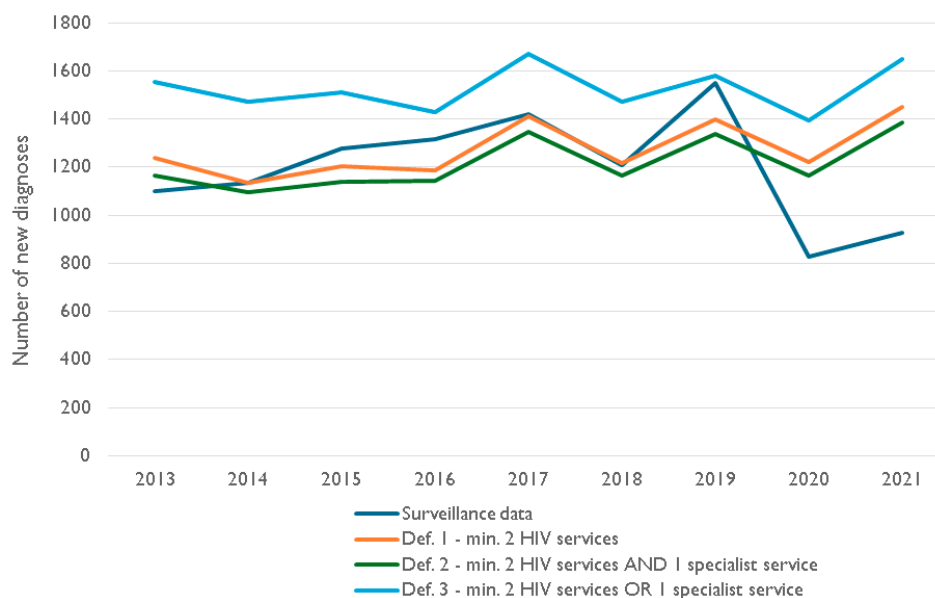


Figure 1. Number of new HIV diagnoses according to the 3 definition options compared to surveillance data in Poland between 2013-2021 years

Rycina 1. Liczba nowo rozpoznanych zakażeń HIV według 3 opcji definicji przypadku w porównaniu do danych z nadzoru epidemiologicznego w Polsce w latach 2013-2021

### Population of people diagnosed with HIV residing in Poland, based on “gaps” analysis.

The number of intervals between services (“gaps”) identified after diagnosis of HIV infection was 3 332 435 (‘last’ and ‘death’ intervals were omitted). The number of gaps lasting at least 1 month was 516 209, while the number of gaps lasting at most 12 months was 509 276. Of the gaps lasting at least 1 month, less than 0.5 per cent were longer than 24 months (Table 4).

Based on this analysis we identified PHIV-D who likely out-migrated and those who would be included in the PHIV-D living in Poland, to be considered in the cascade of care (see below).

1. A patient is considered to have resided in Poland in a given month if they used at least one health service (of any type) financed by the National Health Fund (including reimbursed prescriptions) in the given month or within the 24 months preceding it.
2. A patient is considered to have resided in Poland in a given year if they met the above monthly criterion for at least one month during that year.

ciu o definicję 2. o 4% (18 476 w porównaniu do 17 772). Znaczna część osób zidentyfikowanych w oparciu o definicję 3. (3956, 18%) miała tylko jedną wizytę związaną z zakażeniem HIV.

Niezależnie od przyjętej definicji, data diagnozy jest definiowana jako data pierwszej wizyty związanej z zakażeniem HIV. W związku z tym, osobę uznaje się za nowo zdiagnozowaną w danym roku, jeśli jej pierwsza wizyta związana z zakażeniem HIV miała miejsce w tym roku.

Definicja 3. (co najmniej 2 świadczenia związane z HIV lub 1 świadczenie zrealizowane w placówce specjalistycznej) wskazuje na największą roczną liczbę nowo zdiagnozowanych pacjentów, która znacznie przewyższa liczbę nowych diagnoz HIV odnotowanych w latach 2013- 2018 w danych nadzorowych (Ryc. 1). Odchylenia są znacznie większe niż w przypadku dwóch pozostałych podejść, dlatego definicja ta została odrzucona na wstępie. Średni kwadratowy błąd tego rozwiązania w odniesieniu do danych z nadzoru wyniósł 74 817.

Table 3. Change in the number of new diagnoses calculated from subsequent updates of downloaded data

Tabela 3. Zmiana liczby nowych diagnoz obliczona na podstawie kolejnych aktualizacji pobranych baz danych

	Number of new diagnoses in NHF data		
	2021	2022	2023
Date extracted 03/2022	1408		
Date extracted 06/2023	1440	1360	
Data extracted 12/2023	1442	1361	1396
Change 12/2023 vs. 03/2022 extraction	2,4%		
Change 12/2023 vs. 06/2023 extraction	0,1%	0,1%	

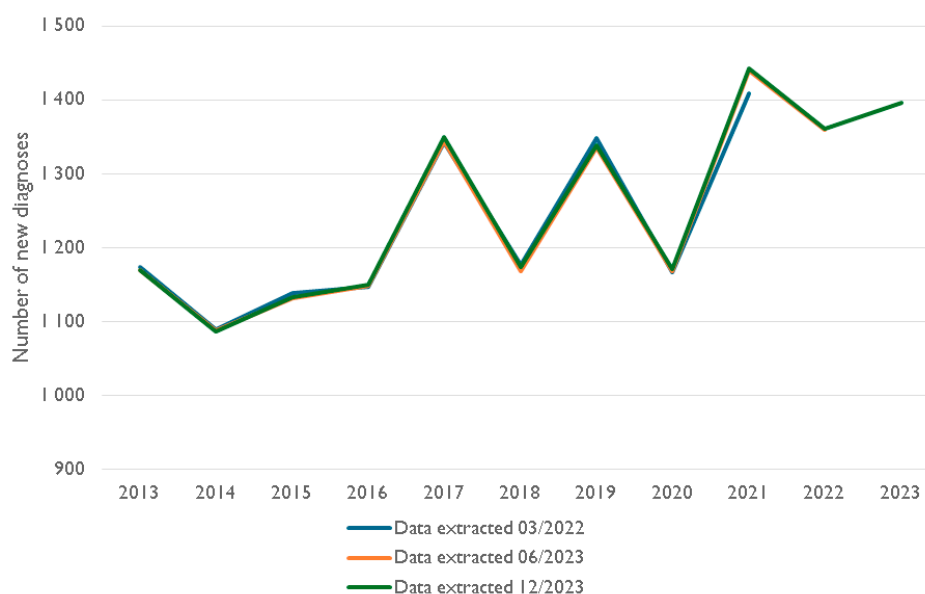


Figure 2. Number of new diagnoses per year estimated from data extracted in 03/2022, 06/2023 and 12/2023

Rycina 2. Liczba nowych rozpoznań HIV w ciągu roku w zależności od roku rozpoznania w oparciu o dane pobrane z okresów: 03/2022, 06/2023 i 12/2023

Table 4. Distribution of the time interval ( $\geq 1$  month) between 2 consecutive registered services among PHIV-D  
Tabela 4. Rozkład przerw ( $\geq 1$  miesiąc) pomiędzy zarejestrowanymi usługami u osób zdiagnozowanych żyjących z HIV

Interval length (months)	No. of intervals	%
<12	509 276	98.66
12-23	4 541	0.88
24-35	1 241	0.24
36-47	492	0.10
48-59	294	0.06
60-71	129	0.02
72-83	108	0.02
84-95	52	0.01
>95	76	0.01
Total	516 209	100.00

Otherwise, a patient with an HIV diagnosis is considered to be outside Poland. The status can of course change in subsequent calendar years.

The results of applying the described algorithm – dividing the PHIV-D population into those residing in Poland and abroad in a given year – are presented in the Figure 3A. In 2021, there were 15.7 thousand people (aged 18 and above) living with diagnosed HIV in Poland, which is almost twice as much as in 2013 (Fig. 3A).

Pozostałe definicje dają wizualnie zbliżone wyniki, ale definicja 1. (co najmniej 2 świadczenia związane z HIV) daje niższy średni błąd kwadratowy (9098 vs 15412) w analizowanym okresie 2013–2019 w porównaniu z danymi z nadzoru. Definicja ta jest również preferowana ze względu na kryterium prostoty.

Autorzy, do dalszych analiz, wybrali zatem definicję pierwszą. Definicja ta wymaga co najmniej dwóch świadczeń związanych z HIV, aby pacjent został włączony do populacji PHIV-D, ale za moment diagnozy przyjmuje się moment wystąpienia pierwszego świadczenia. Może to prowadzić do niedoszacowania nowych diagnoz, szczególnie w przypadku danych z ostatnich lat obserwacji. Oszacowanie skali tego niedoszacowania wykonano porównując roczną liczbę nowych diagnoz w populacji dorosłych obliczoną na podstawie zestawów danych pobranych w miesiącach 03/2022, 06/2023 i 12/2023 (Tab. 3, Ryc. 2). Liczba nowych diagnoz w ostatnim roku obserwacji zmieniła się nieznacznie w porównaniu z danymi wyodrębnionymi 6 i 12 miesięcy po zakończeniu roku.

**Liczba osób żyjących ze zdiagnozowanym zakażeniem HIV, zamieszkujących w Polsce, bazując na analizie „interwałów”.** Liczba interwałów między świadczeniami („przerw”) zidentyfikowanych po zdiagnozowaniu zakażenia HIV wyniosła 3 332 435 (pominięto interwały „ostatnie” i „zgony”). Liczba przerw trwających co najmniej 1 miesiąc wyniosła 516 209, natomiast liczba przerw trwających maksymalnie 12 miesięcy wyniosła 509 276. Spośród przerw trwających co najmniej 1 miesiąc, mniej niż 0,5% trwały dłużej niż 24 miesiące (Tab. 4).

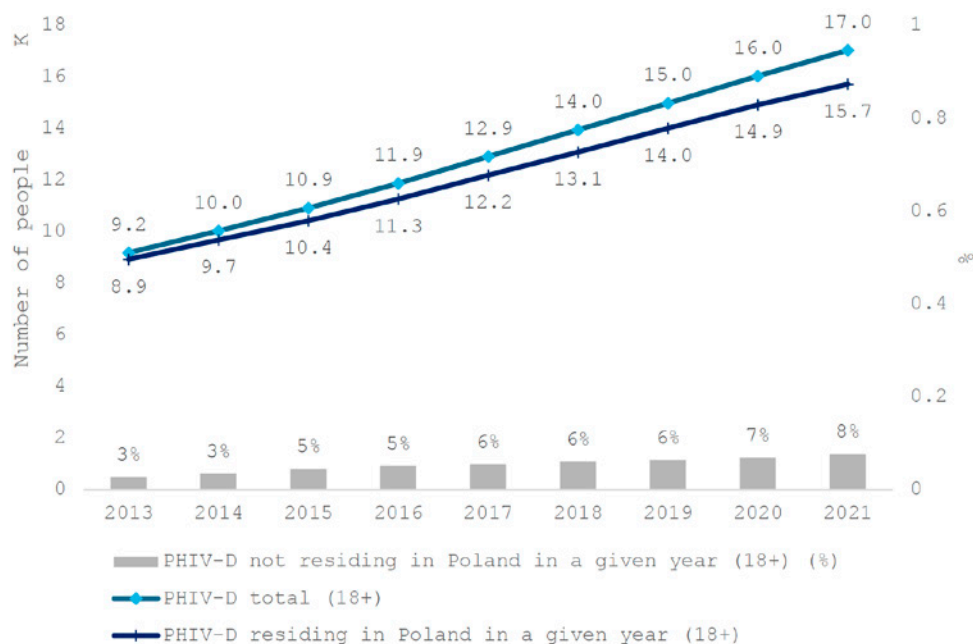


Figure 3A. Overall trend of the number of PHIV-D in total and residing in Poland in 2013-2021  
Rycina 3A. Obserwowany trend liczby osób zdiagnozowanych żyjących z HIV ogółem i u osób zamieszkałych w Polsce w latach 2013-2021

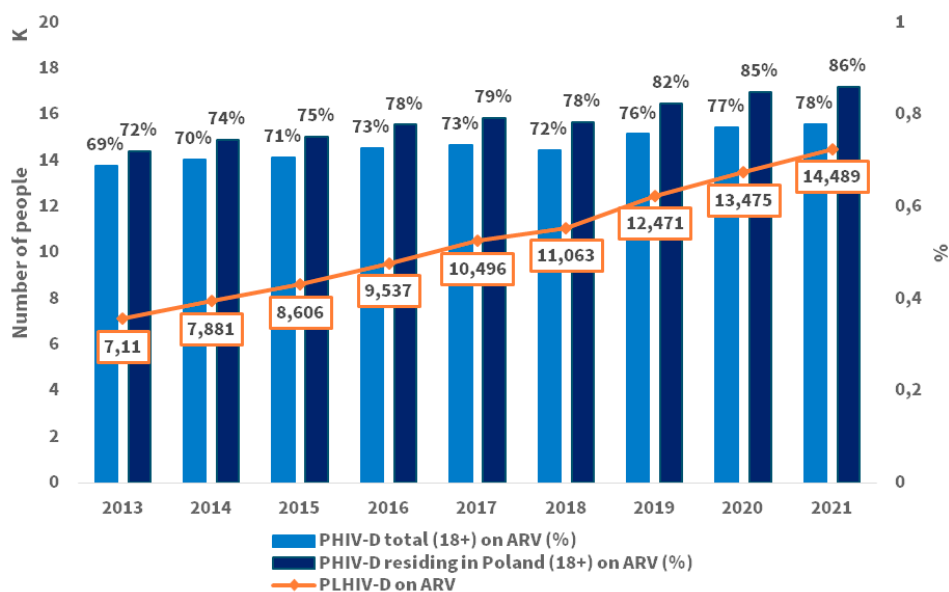


Figure 3B. ARV treatment cascade: number of treated individuals and percentage treated among total and residing in Poland, 2013–2021

Rycina 3B. Kaskada leczenia ARV: liczba osób leczonych oraz odsetek leczonych wśród osób zdiagnozowanych żyjących z HIV ogółem i u osób zamieszkałych w Polsce, 2013-2021

When these results are compared with treatment data from the National AIDS Centre, Poland (Fig. 3B), a clear improvement in treatment coverage is observed. Between 2013 and 2021, the proportion of people receiving antiretroviral therapy (ARV) among PHIV-D increased from 69% to 78%. When the cascade is limited to those residing in Poland, the percentage of treated individuals rose even more markedly, from 72% in 2013 to 86% in 2021.

The predominant group among PHIV-D population are men. The proportion of women has declined over the analyzed years, from 24% in 2013 to 21% in 2021. The age distribution of females and males living with diagnosed HIV in Poland has several similar characteristics. The largest age group among both genders are individuals aged 35-44 (Table 5). Moreover, there has been a discernible rise in the proportion of people of both sexes older than 45 years.

## DISCUSSION

The developed method leveraging the NHF databases makes it possible to estimate the number of PHIV-D residing in Poland accounting for the number of new HIV diagnoses, migration and deaths of PHIV-D. The method can be used in the framework of monitoring the development of the HIV epidemic in Poland due to the fact that the complete NHF dataset is available for analytical purposes in the mode of monthly updates with a lag of 3 months, the estimation can be carried out in a reproducible manner, and the

Na podstawie tej analizy zidentyfikowaliśmy PHIV-D, które prawdopodobnie wyemigrowały za granicę, oraz te, które zostaną włączone do PHIV-D mieszkających w Polsce i które należy uwzględnić w kaskadzie opieki (patrz poniżej).

Pacjenta uważa się za zamieszkałego w Polsce w danym miesiącu, jeżeli skorzystał z co najmniej jednego świadczenia zdrowotnego (dowolnego rodzaju) finansowanego przez Narodowy Fundusz Zdrowia (w tym z recept refundowanych) w danym miesiącu lub w ciągu 24 miesięcy poprzedzających go.

Pacjenta uważa się za zamieszkałego w Polsce w danym roku, jeżeli spełnił powyższe kryterium miesięczne przez co najmniej jeden miesiąc w danym roku.

W przeciwnym razie pacjent z rozpoznaniem HIV jest uznawany za osobę przebywającą poza granicami Polski. Status ten może oczywiście ulec zmianie w kolejnych latach kalendarzowych.

Wyniki zastosowania opisanego algorytmu, dzielącego populację PHIV-D na zamieszkałą w danym roku w Polsce i za granicą – przedstawiono na rysunku Ryc. 3A. W 2021 r. w Polsce żyło już 15,7 tys. osób (w wieku 18 lat i więcej) z rozpoznaniem zakażeniem HIV, co stanowi prawie dwukrotność liczby z 2013 r. (Ryc. 3A).

Porównanie tych wyników z danymi dotyczącymi leczenia pochodzącymi z Krajowego Centrum ds. AIDS w Polsce (Ryc. 3B) wyraźnie pokazuje poprawę w zakresie osób objętych leczeniem. W latach 2013–2021 odsetek osób otrzymujących terapię antyretrowirusową (ARV) wśród pacjentów z grupy PHIV-D

Table 5. Sex and age structure (aged 18 and above) of people living with diagnosed HIV in Poland in 2013-2021

Tabela 5. Charakterystyka osób z rozpoznaniem zakażeniem HIV żyjących w Polsce w 2013-2021 roku z podziałem na płeć i wiek (wiek powyżej 18 lat)

		2013	2014	2015	2016	2017	2018	2019	2020	2021
	F	24%	24%	23%	22%	22%	21%	21%	21%	21%
	M	76%	76%	77%	78%	78%	79%	79%	79%	79%
	Total	8 945	9 696	10 438	11 276	12 210	13 104	14 028	14 927	15 732
F	18-24	4%	4%	4%	4%	4%	4%	3%	4%	3%
	25-34	31%	28%	25%	22%	21%	19%	17%	15%	14%
	35-44	38%	39%	41%	40%	39%	39%	39%	38%	38%
	45-54	15%	16%	17%	19%	21%	23%	25%	26%	27%
	55-64	7%	7%	8%	9%	9%	9%	10%	10%	10%
	65+	4%	5%	5%	5%	6%	6%	7%	7%	7%
	Total	2 183	2 309	2 420	2 517	2 665	2 778	2 915	3 085	3 231
M	18-24	4%	4%	4%	4%	4%	3%	3%	3%	3%
	25-34	30%	29%	28%	27%	27%	26%	25%	24%	23%
	35-44	38%	38%	38%	38%	38%	38%	38%	38%	38%
	45-54	20%	20%	20%	21%	21%	22%	22%	23%	23%
	55-64	7%	7%	7%	7%	7%	8%	8%	9%	10%
	65+	2%	2%	3%	3%	3%	3%	3%	3%	4%
	Total	6 762	7 387	8 018	8 759	9 545	10 326	11 113	11 842	12 501

results are available within a relatively short time after the analysed period.

As compared to the routine case-based surveillance in place in Poland our method has several advantages. Firstly, it allows to analyse the patient's pathway, including visits following the diagnosis, which can help to study retention in care as well as to identify people who stay in the country or who likely moved out. It is also possible to monitor mortality. Migration, retention in care and mortality are challenging to monitor through routine surveillance. Yet, these are necessary to be able to reliably estimate the number of people living with diagnosed HIV in the country and, effectively, the treatment cascade. Moreover, the system based on NHF data is less prone to disturbances due to public health crisis such as COVID-19 pandemic, engaging the public health personnel responsible for surveillance. On the other hand, the NHF data lack epidemiological variables such as transmission routes or prior testing and the clinical information is very limited. These should therefore remain the focus of case-base surveillance activities.

Our definition to identify HIV cases was similar to the algorithm used by EHR systems from Florida (9), which to identify HIV cases requires data on patient with at least one diagnostic code for HIV and meet one of the following criteria: have 1+ positive HIV laboratory, have been prescribed with HIV medications, or have 3+ visits with HIV diagnostic codes (sensitivity – 98.9%; specificity – 97.6%). In

wzrósł z 69% do 78%. W przypadku ograniczenia kaskady do osób zamieszkałych w Polsce, odsetek leczonych wzrósł jeszcze bardziej – z 72% w 2013 r. do 86% w 2021 r.

Dominującą grupą wśród populacji PHIV-D są mężczyźni. Odsetek kobiet zmniejszył się w analizowanych latach, z 24% w 2013 r. do 21% w 2021 r. Struktura wiekowa kobiet i mężczyzn żyjących z rozpoznaniem zakażeniem HIV w Polsce ma kilka podobnych cech. Najliczniejszą grupę wiekową wśród obu płci stanowią osoby w wieku 35-44 lat (Tab. 5). Ponadto zauważalny jest wzrost odsetka osób obu płci powyżej 45 roku życia.

## DYSKUSJA

Opracowana metoda wykorzystująca bazę danych NFZ umożliwia oszacowanie liczby osób z grupy PHIV-D przebywających w Polsce, uwzględniając liczbę nowych rozpoznań HIV, migracji i zgonów. Metoda ta może być wykorzystana w celu monitorowania rozwoju epidemii HIV w Polsce, ponieważ pełny zbiór danych NFZ jest dostępny do celów analitycznych w trybie comiesięcznych aktualizacji z opóźnieniem 3 miesiące, estymacja może być przeprowadzana w sposób powtarzalny, a wyniki są dostępne w stosunkowo krótkim czasie po analizowanym okresie.

W porównaniu z rutynowym nadzorem opartym na zgłaszanych przypadkach, stosowanym w Polsce, nasza metoda cechuje się kilkoma istotnymi atutami. Po

other coded study (10), among 272 women enrollees of a U.S. health plan, with  $\geq 1$  day with an HIV diagnosis code was analysed, and 227 women were included as patients with HIV infection based also on medical record-abstracted clinician's note stating that the patient had a diagnosis of HIV, or in the absence of a clinician's note, laboratory evidence. But also in this study, it was proved that using an algorithm requiring  $\geq 1$  HIV diagnosis code the HIV infection was confirmed for 80.5% women and this number increased appreciably when  $\geq 2$  HIV diagnosis codes were required or more (93.9%). This conclusion is in line with our definition of the classification of HIV patients diagnosed on the basis of the NHF database in Poland.

As the HIV treatment program is provided through targeted governmental budget, the HIV test results and HIV medications are not recorded in the claims database. We note however, that similarly to the cited results, excluding cases with only a single service with HIV codes, especially outside of the specialized HIV clinics, greatly improved the coherence of the results with the surveillance data. The presence of individual medical services identified with the HIV code indicates frequent errors in data coding and requires validation and caution in using such data. Similar problems have been noted in the coding of other diseases, including diabetes, in the NHF data in Poland (16).

In addition, the age and gender distribution of the study participants is comparable to the data from routine HIV epidemiological surveillance in Poland (1) or in other European countries (17). In ECDC report from 2024 year, men had higher age-specific HIV diagnosis rates than women, in all age groups, except people below 15 years old. The highest overall age-specific rate was observed in the age group 30-39 years (13.1 per 100 000 population). Among men, the highest rate was in the age group 25-29 years (19.6 per 100 000), while for women, it was in the age group 30-39 years (7.2 per 100 000). In Poland in 2019 (before the disruptions related to the pandemic) there were 1 551 cases reported to national surveillance, and the distribution of male to female, was 85% in males to 15% to females, with 64% of all cases diagnosed in the age 20-39 years, so results obtained from the NHF data presented in the study are consistent with the expected results. Other data could be also used for triangulation with our results including the census data for HIV-related deaths as published by the Statistics Poland and the antiretroviral treatment programme data available from National AIDS Centre. However, such characteristics as the number of deaths and the number of patients receiving antiretroviral therapy (ART) cannot be directly compared across data sources. In NHF data we are able to study only all-cause mortality, which without individual linkage could be only

pierwsze, pozwala ona na analizę ścieżek pacjentów, w tym wizyt po diagnozie, co może pomóc w badaniu ciągłości opieki, a także identyfikacji osób, które pozostają w kraju, lub z dużym prawdopodobieństwem się wyprowadziły. Możliwe jest również monitorowanie umieralności. Monitorowanie migracji, ciągłości opieki i umieralności stanowi wyzwanie w ramach nadzoru epidemiologicznego. Jest to jednak niezbędne do wiarygodnego oszacowania liczby osób żyjących z zdiagnozowanym zakażeniem HIV w kraju oraz kaskady leczenia. Ponadto, system oparty na danych NFZ jest mniej podatny na zakłócenia spowodowane kryzysami zdrowia publicznego, takimi jak pandemia COVID-19, które angażują personel zdrowia publicznego odpowiedzialny za nadzór. Z drugiej strony, dane NFZ nie uwzględniają istotnych zmiennych epidemiologicznych, takich jak drogi zakażenia, czy wcześniejsze badania w kierunku HIV, a informacje kliniczne są bardzo ograniczone. Dlatego, powinny one pozostać kluczowym przedmiotem działań w ramach nadzoru epidemiologicznego.

Nasza definicja służąca do identyfikacji przypadków HIV jest zbliżona do algorytmu stosowanego przez systemy EHR na Florydzie (9), który do identyfikacji przypadków HIV wymaga danych pacjenta potwierdzających wystąpienie co najmniej jednego kodu diagnostycznego HIV i spełniających jedno z następujących kryteriów: wystąpienie co najmniej 1 pozytywnego wyniku badania laboratoryjnego na HIV, otrzymanie przepisanych leków na HIV lub odbycie co najmniej 3 wizyt z kodami diagnostycznymi HIV (czułość – 98,9%; swoistość – 97,6%). W innym badaniu, wykorzystującym kody ICD-10 (10) przeanalizowano 272 kobiety objęte amerykańskim planem zdrowotnym, z  $\geq 1$  dniem z kodem diagnostycznym HIV. Spośród nich 227 kobiet zaklasyfikowano jako pacjentki z zakażeniem HIV na podstawie adnotacji w dokumentacji medycznej potwierdzającej rozpoznanie HIV lub w przypadku braku takiej notatki, obecność wirusa stwierdzono na podstawie wyników badań laboratoryjnych. W badaniu tym wykazano również, że przy użyciu algorytmu wymagającego  $\geq 1$  kodu diagnostycznego HIV, zakażenie HIV potwierdzono u 80,5% kobiet, a odsetek ten znacząco wzrósł, gdy wymagane były  $\geq 2$  lub więcej kodów diagnostycznych HIV (93,9%). Wniosek ten jest zgodny z naszą definicją klasyfikacji pacjentów ze zdiagnozowanym HIV na podstawie bazy danych NFZ w Polsce.

Ponieważ program leczenia HIV jest finansowany z budżetu państwa, wyniki testów na HIV i leki na HIV nie są rejestrowane w bazie danych pochodzących z wniosków o zwrot kosztów zrealizowanych świadczeń. Zauważamy jednak, że podobnie jak w przypadku cytowanych wyników, wyłączenie przypadków z tylko jednym świadczeniem z przypisanym kodem

compared to the number of deaths classified as due to AIDS in the Statistics Poland report. Likewise, the number of patients on ART reported by the National AIDS Centre does not represent all people living with HIV, as not all infected individuals are currently on treatment. We note that individual linkage especially between the ART program and NHF databases, if legally possible would highly strengthen our conclusions.

However, the method has several limitations, primarily related to the characteristics of the NHF data and the purpose, for which they are collected. Although the NHF's status as the only public payer allows for near-comprehensive national coverage, NHF data do not include care financed privately – either by patients themselves or by employers. The lack of data on privately funded services could introduce uncertainty into the estimates. However, it is less of an issue in case of HIV patients, due to simplified access of these patients to publicly funded services (18,19). The financing of HIV treatment in Poland is regulated by the Act on Healthcare Services Financed from Public Funds and by the Government Health Policy Program “Antiretroviral Treatment of People Living with HIV in Poland for 2022-2026”. According to these regulations, antiretroviral (ARV) treatment of people living with HIV is free of charge for the patient and financed from public funds.

Another limitation is the potential impact of legislative or organizational changes on the way HIV-related services are recorded and reimbursed. Continuous monitoring of such changes is necessary to maintain consistency in data interpretation over time. Additionally, there is currently no linkage between antiretroviral treatment (ARV) data collected by the National AIDS Centre and NHF claims data, which constrains the ability to validate and refine the estimates.

Moreover, the arrival of the individuals who were previously diagnosed with HIV in their countries of origin will be considered by our algorithm new diagnoses, which may complicate the interpretation of the date in the context of epidemic development. Further research would be necessary to account for this issue. The accuracy of estimates of the PHIV-D immigrant population is dependent on the scope of access to publicly funded services.

Another issue worth exploring in future research is the group of individuals who have been registered with HIV-related service but remain undiagnosed for a prolonged period. Such analysis could provide valuable insight into potential missed opportunities for earlier HIV detection.

Despite limitations, the NHF data remains useful for infectious disease surveillance purposes. They

HIV, zwłaszcza poza specjalistycznymi placówkami HIV, znacznie poprawiło spójność wyników z danymi z nadzoru. Obecność pojedynczych świadczeń oznaczonych kodem HIV wskazuje na częste błędy w kodowaniu danych i wymaga walidacji oraz ostrożności przy ich wykorzystywaniu. Podobne problemy odnotowano w kodowaniu innych chorób, w tym cukrzycy, w danych NFZ w Polsce (16).

Ponadto, rozkład wieku i płci uczestników badania jest porównywalny z danymi pochodzącymi z rutynowego nadzoru epidemiologicznego HIV w Polsce (1) lub w innych krajach europejskich (17). W raporcie ECDC z 2024 roku mężczyźni we wszystkich grupach wiekowych, z wyjątkiem osób poniżej 15. roku życia mieli wyższe wskaźniki zachorowalności na HIV niż kobiety. Najwyższy ogólny wskaźnik w zależności od wieku zaobserwowano w grupie 30-39 lat (13,1 na 100 000 ludności). Wśród mężczyzn najwyższy wskaźnik był w grupie wiekowej 25-29 lat (19,6 na 100 000), natomiast wśród kobiet w grupie wiekowej 30-39 lat (7,2 na 100 000). W Polsce w 2019 roku (przed wystąpieniem zakłóceń spowodowanych pandemią) zgłoszono 1 551 przypadków do krajowego nadzoru, a rozkład ze względu na płeć wynosił 85% mężczyzn i 15% kobiet, przy czym 64% wszystkich przypadków zdiagnozowano w grupie 20-39 lat, więc wyniki uzyskane z danych NFZ przedstawionych w badaniu są zgodne z oczekiwanymi wynikami. Do triangulacji z naszymi wynikami można by również wykorzystać inne dane, w tym dane z rejestru zgonów, w zakresie zgonów związanych z HIV, publikowane przez Główny Urząd Statystyczny oraz dane dotyczące realizacji programu leczenia antyretrowirusowego dostępne w Krajowym Centrum ds. AIDS. Jednakże, nie jest możliwe bezpośrednie porównanie takich charakterystyk jak liczba zgonów, czy liczba pacjentów otrzymujących terapię antyretrowirusową (ART) między różnymi źródłami danych. W danych NFZ możliwe jest badanie umieralności ogółem, bez wyróżnienia przyczyn, która bez indywidualnego powiązania danych, może być porównywana jedynie z liczbą zgonów z powodu AIDS raportowaną w danych Głównego Urzędu Statystycznego. Podobnie, liczba pacjentów korzystających z ART zgłaszana przez Krajowe Centrum ds. AIDS nie odzwierciedla wszystkich osób żyjących z rozpoznaniem zakażeniem HIV, ponieważ nie wszystkie zakażone osoby są objęte leczeniem. Należy zauważyć, że powiązanie danych dotyczących tej samej osoby, zwłaszcza między bazą danych programu leczenia ART i bazami danych NFZ, jeśli będzie to prawnie możliwe, mogłoby znacząco wzmocnić wnioski niniejszego badania.

Opracowana metoda ta ma jednak kilka ograniczeń, związanych przede wszystkim z charakterystyką danych NFZ i celem, dla którego są gromadzone.

partly fulfil the functions envisioned for electronic health record (EHR)-based systems—namely, reducing reporting burdens while improving the timeliness and completeness of infectious disease data (6).

## CONCLUSION

The developed method for estimating the number of people living with diagnosed HIV (PHIV-D) constitutes a valuable element of the national HIV surveillance system in Poland. Given its utility and potential, the authors consider it appropriate to pursue further development of the approach – specifically, by designing an algorithm capable of identifying cases of late diagnosis, as well as a variant tailored to account for the epidemiological and healthcare access characteristics of immigrant populations. A particularly important enhancement would be the integration of antiretroviral treatment data from the National AIDS Centre. This addition would enable monitoring not only of the size of the PLHIV-D population but also of key stages of the HIV care cascade.

## REFERENCES

1. Niedźwiedzka-Stadnik M, Nowakowska-Radziwonka E, Marzec-Bogusławska A. HIV infections and AIDS in Poland in 2022. *Przegl Epidemiol* 2024; 78(4):459-478; doi: 10.32394/pe/197855;
2. Rosinska M, Grzeszuk A, Ankiersztejn-Bartczak M, Kowalska J. Nadzór epidemiologiczny nad HIV/AIDS. W: *Zasady opieki nad osobami żyjącymi z HIV Zalecenia PTN AIDS 2024*; Warszawa - Szczecin: Polskie Towarzystwo Naukowe AIDS; 2024, s. 30–34.; <https://ptnaids.pl/category/rekomendacje/>
3. Główny Urząd Statystyczny. Migracje zagraniczne ludności; <https://stat.gov.pl/obszary-tematyczne/ludnosc/migracje-zagraniczne-ludnosci/>
4. Ganczak M, Czubińska G, Korzeń M, Szych Z. A Cross-Sectional Study on Selected Correlates of High risk Sexual Behavior in Polish Migrants Resident in the United Kingdom. *Int J Environ Res Public Health*; 2017;14(4):422; doi:10.3390/ijerph14040422
5. Aliabadi A, Sheikhtaheri A, Ansari H. Electronic health record–based disease surveillance systems: A systematic literature review on challenges and solutions. *J Am Med Inform Assoc*; 2020;27(12):1977–86; doi: 10.1093/jamia/ocaa186
6. European Centre for Disease Prevention and Control. Assessment of electronic health records for infectious disease surveillance: final mapping exercise report. Stockholm: ECDC; 2021.

Chociaż status NFZ jako jedyne płatnika publicznego pozwala na niemal kompleksowe pokrycie całego kraju, dane NFZ nie obejmują opieki finansowanej prywatnie – ani przez samych pacjentów, ani przez pracodawców. Brak danych o usługach finansowanych ze środków prywatnych może wprowadzać niepewność do szacunków. Problem ten jest jednak mniejszy w przypadku pacjentów z HIV ze względu na uproszczony dostęp tych pacjentów do usług finansowanych ze środków publicznych (18,19). Finansowanie leczenia HIV w Polsce jest regulowane przez Ustawę o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych oraz przez Rządowy Program Polityki Zdrowotnej „Leczenie antyretrowirusowe osób żyjących z wirusem HIV w Polsce na lata 2022-2026”. Zgodnie z tymi przepisami leczenie antyretrowirusowe (ARV) osób żyjących z HIV jest bezpłatne dla pacjenta i finansowane ze środków publicznych.

Kolejnym ograniczeniem jest potencjalny wpływ zmian legislacyjnych lub organizacyjnych na sposób rejestrowania i refundacji usług związanych z HIV. Ciągłe monitorowanie takich zmian jest niezbędne dla zachowania spójności interpretacji danych w czasie. Ponadto obecnie nie ma powiązania między danymi dotyczącymi leczenia antyretrowirusowego (ARV) gromadzonymi przez Narodowe Centrum ds. AIDS a danymi dotyczącymi wniosków o świadczenia w ramach Narodowego Funduszu Zdrowia (NFZ), co ogranicza możliwość weryfikacji i udoskonalania szacunków.

Dodatkowo, przybycie migrantów, u których wcześniej zdiagnozowano zakażenie wirusem HIV w ich krajach pochodzenia, będzie przez nasz algorytm uwzględniane jako nowa diagnoza HIV, co może utrudniać interpretację daty diagnozy w kontekście rozwoju epidemii. Konieczne będą dalsze badania, które pozwoliłyby rozwiązać ten problem. Dokładność szacunków populacji imigrantów PHIV-D zależy od zakresu dostępu do usług finansowanych ze środków publicznych.

Kolejnym zagadnieniem wartym zaadresowania w przyszłych badaniach jest grupa osób, u których zostało zarejestrowane świadczenie z kodem HIV, ale przez dłuższy czas nie miały kolejnego świadczenia. Taka analiza mogłaby dostarczyć cennych informacji na temat potencjalnie utraconych szans na wcześniejsze objęcie opieką osób żyjących z HIV.

Pomimo ograniczeń dane NFZ pozostają użyteczne do celów nadzoru nad chorobami zakaźnymi. Częściowo spełniają one funkcje przewidziane dla systemów opartych na elektronicznej dokumentacji medycznej (EHR), a mianowicie zmniejszają obciążenia związane ze sprawozdawczością, a jednocześnie poprawiają aktualność i kompletność danych dotyczących chorób zakaźnych (6).

Available from: <https://www.ecdc.europa.eu/en/publications-data/assessment-electronic-health-records-infectious-disease-surveillance>

7. Jung J, Im JH, Ko YJ, Huh K, Yoon C gyo, Rhee C, et al. Complementing conventional infectious disease surveillance with national health insurance claims data in the Republic of Korea. *Sci Rep.* 2019;9(1):8750; doi: 10.1038/s41598-019-45409-3
8. Bazazi AR, Thomas M, Vittinghoff E, Walkup J, Hermida R, Arnold EA, et al. Schizophrenia and Retention in HIV Care Among Adults Insured Through Medicaid in the United States: A Population-Based Retrospective Cohort Study. *JAIDS J Acquir Immune Defic Syndr.* 99(1):20–30; doi: 10.1097/QAI.00000000000003606
9. Liu Y, Siddiqi KA, Cook RL, Bian J, Squires PJ, Shenkman EA, et al. Optimizing Identification of People Living with HIV from Electronic Medical Records: Computable Phenotype Development and Validation. *Methods Inf Med.* 2021;60(03/04):84–94; doi: 10.1055/s-0041-1735619
10. Pocobelli G, Oliver M, Albertson-Junkans L, Gundersen G, Kamineni A. Validation of human immunodeficiency virus diagnosis codes among women enrollees of a U.S. health plan. *BMC Health Serv Res.* 2024;24(1):234; doi: 10.1186/s12913-024-10685-x
11. Minister Zdrowia, Krajowe Centrum ds. AIDS. Harmonogram realizacji Krajowego Programu Zapobiegania Zakażeniom HIV i Zwalczenia AIDS opracowany na lata 2022 – 2026. Warszawa; 2022.
12. OECD. Health at a Glance 2021: OECD Indicators; 2021; [https://www.oecd.org/en/publications/health-at-a-glance-2021\\_ae3016b9-en/full-report.html](https://www.oecd.org/en/publications/health-at-a-glance-2021_ae3016b9-en/full-report.html)
13. CDC. National Center for Health Statistics. Ambulatory Care Use and Physician office visits.; <https://www.cdc.gov/nchs/fastats/physician-visits.htm>
14. Australian Bureau of Statistics. Patient Experiences; <https://www.abs.gov.au/statistics/health/health-services/patient-experiences/2021-22>
15. NHS. GP patient surveys; <https://gp-patient.co.uk/surveysandreports>
16. Bogusławski S, Wojciechowska A, Smaga A, Strojek K, Wróbel K, Wojtyniak B, et al. Analysis of diabetes-related complications and the cost of diabetes in Poland. *Eur J Transl Clin Med.* 2023;6(1):31–40; doi: 10.31373/ejtc/m/162189
17. WHO Regional Office for Europe, European Centre for Disease Prevention and Control. WNIOSKI  
Opracowana metoda szacowania liczby osób żyjących z rozpoznaniem zakażeniem HIV (PHIV-D) stanowi cenne uzupełnienie krajowego systemu nadzoru epidemiologicznego nad HIV w Polsce. Biorąc pod uwagę jej użyteczność i potencjał, autorzy uważają za zasadne dalsze rozwijanie tego podejścia – w szczególności poprzez opracowanie algorytmu umożliwiającego identyfikację przypadków późnego rozpoznania, a także wariantu uwzględniającego specyfikę epidemiologiczną i dostęp do opieki zdrowotnej w populacji imigrantów. Szczególnie istotnym usprawnieniem byłaby integracja danych dotyczących leczenia antyretrowirusowego z Krajowego Centrum ds. AIDS. Umożliwiłoby to monitorowanie nie tylko liczebności populacji PHIV-D, ale także kluczowych elementów kaskady opieki nad osobami z HIV.
- HIV/AIDS surveillance in Europe 2024 – 2023 data; Copenhagen: WHO Regional Office for Europe; 2024; <https://www.ecdc.europa.eu/en/publications-data/hiv-aids-surveillance-europe-2024-2023-data>
18. Minister Zdrowia. Rządowy Program Polityki Zdrowotnej leczenie antyretrowirusowe osób żyjących z wirusem HIV w Polsce na lata 2022-2026; <https://www.gov.pl/web/zdrowie/rzadowy-program-polityki-zdrowotnej-leczenie-antyretrowirusowe-osob-zyjacych-z-wirusem-hiv-w-polsce-na-lata-2022-2026>
19. Ustawa z dnia 27 sierpnia 2004 r. o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych. Dz.U. 2004 nr 210 poz. 2135; <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=wdu20042102135>

**Received:** 10.07.2025

**Accepted for publication:** 26.11.2025

Otrzymano: 10.07.2025 r.

Zaakceptowano do druku: 26.11.2025 r.

**Address for correspondence:**

Adres do korespondencji:

Anna Tymicka

Zakład Monitorowania i Analiz Stanu Zdrowia

Ludności,

Narodowy Instytut Zdrowia Publicznego PZH – PIB

ul. Chocimska 24, 00-791 Warszawa

email: [atymicka@pzh.gov.pl](mailto:atymicka@pzh.gov.pl)