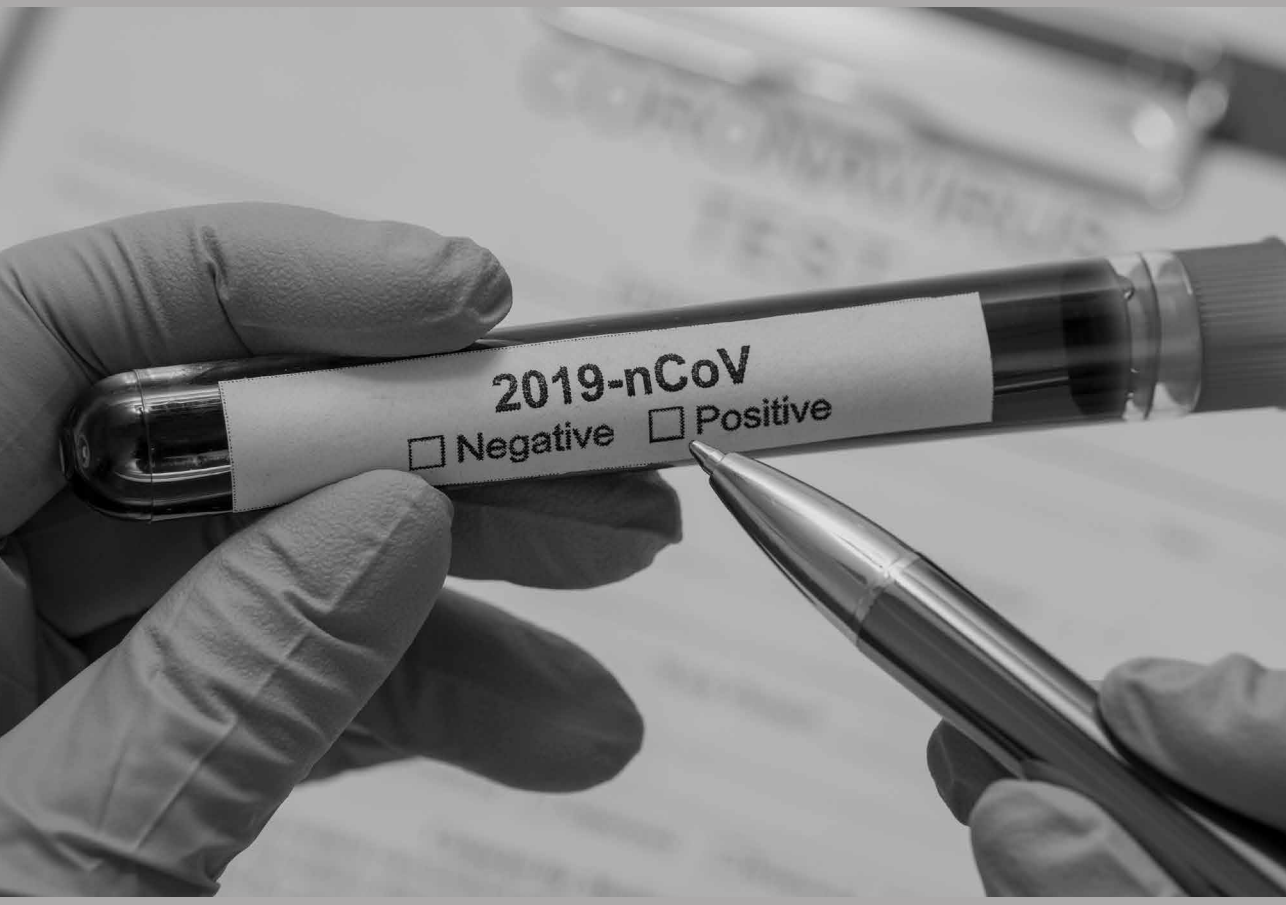
Polski Instytut Ekonomiczny

1/2020

WORKING PAPER



Nadmierna śmiertelność w Polsce podczas pandemii koronawirusa

Stan na dzień 13.05.2020 r.

Warszawa, maj 2020 r.
Autor: Adam Czerwiński
Redakcja: Jakub Nowak, Małgorzata Wieteska
Projekt graficzny: Anna Olczak
Współpraca graficzna: Liliana Gałązka, Tomasz Gałązka, Sebastian Grzybowski
Polski Instytut Ekonomiczny
Al. Jerozolimskie 87
02-001 Warszawa
© Copyright by Polski Instytut Ekonomiczny

ISBN 978-83-66306-79-0

Spis treści

Kluczowe wnioski 4
Streszczenie 5
Wprowadzenie 6
Dane i metodologia 7
Wyniki 13
Dyskusja 16
Bibliografia 18
Załączniki 20
Załącznik 1. Korekta opóźnienia w sprawozdawczości zgonów 20
Załącznik 2. Oczekiwana liczba zgonów w roku *t* na podstawie struktury ludności na koniec roku *t*-1 oraz tablic trwania życia 23

Kluczowe wnioski

- W wielu krajach Europy dotkniętych pandemią koronawirusa zaobserwowano nadmierną śmiertelność, czyli nadwyżkę rzeczywistej liczby zgonów nad tą, której należałoby oczekiwać w normalnej sytuacji (nawet rzędu 75 proc.). Śmiertelności tej nie daje się całkowicie wyjaśnić oficjalną liczbą zgonów związanych z COVID-19.
- W Polsce od 5.03 do 13.05.2020 r. zaobserwowano 80 324 zgony. Oczekiwana liczba zgonów w tym okresie wynosiła (w zależności od oszacowania) od 74 491 do 87 259.
- Na podstawie trendów historycznych nie ma podstaw do stwierdzenia, że w badanym okresie liczba zgonów była wyższa niż należałoby oczekiwać. Bardziej prawdopodobne jest występowanie zaniżonej śmiertelności (1-2 tys. zgonów, 100-200 zgonów tygodniowo).
- Okres pandemii połączony z *lockdownem* cechuje się nasileniem zarówno czynników, które mogą śmiertelność zwiększać, jak i czynników, które powinny ujemnie wpływać na liczbę zgonów (np. niższa aktywność zawodowa, czystsze powietrze). Te dwa zestawy czynników będą się do pewnego stopnia neutralizować. Uwzględniając jednak to, że w innych krajach Europy wykazano nadmierną śmiertelność podczas pandemii oraz to, że zasady *lockdownu* w Polsce i innych krajach są podobne (czyli można spodziewać się zbliżonego, ujemnego wpływu *lockdownu* na śmiertelność) można wnioskować, że niedoszacowanie oficjalnej liczby zgonów związanych z COVID-19 najprawdopodobniej nie jest w Polsce wysokie.

Streszczenie

W wielu krajach podczas pandemii koronawirusa zaobserwowano wyższą liczbę zgonów od tej, której należałoby oczekiwać w normalnej sytuacji. W większości przypadków nadwyżki tej nie daje się wyjaśnić wyłącznie oficjalną liczbą zgonów związanych z COVID-19. W niniejszym opracowaniu przeanalizowano występowanie nadmiernej śmiertelności w Polsce od 5.03 do 13.05.2020 r.

W analizie wykorzystano dane o dziennej liczbie zgonów od 1.01.2010 r. wygenerowane z rejestru PESEL oraz Rejestru Stanu Cywilnego. Na ich podstawie podano rzeczywistą liczbę zgonów w badanym okresie (80 324) i porównano ją z ośmioma oszacowaniami oczekiwanej liczby zgonów (wahającymi się od 74 491 do 87 259). Istotność różnicy między tymi wartościami testowano za pomocą testu *z* dla proporcji, na poziomie istotności 5 proc.

Wyniki testów dla trzech z ośmiu oszacowań wskazują, że w badanym okresie zaobserwowana liczba zgonów była równa oczekiwanej, cztery oszacowania wskazują na występowanie zaniżonej śmiertelności, a jedno na występowanie nadmiernej śmiertelności. Agregując wyniki można stwierdzić, że na podstawie trendów historycznych nie ma podstaw do stwierdzenia, że w badanym okresie liczba zgonów była wyższa niż należałoby oczekiwać. Jedynie wariant minimalny, zakładający, że oczekiwana śmiertelność w okresie *lockdownu* jest zbliżona do śmiertelności z miesięcy letnich wskazuje na występowanie nadmiernej śmiertelności (na poziomie ok 6 tys. zgonów – 600 zgonów tygodniowo). Bardziej prawdopodobną hipotezą niż występowanie nadmiernej śmiertelności jest zmniejszenie liczby zgonów (o około 1-2 tys. zgonów – 100-200 tygodniowo).

Oszacowania nadmiernej śmiertelności podczas pandemii koronawirusa dla krajów Europy Zachodniej są wysokie (sięgają nawet 75 proc.), podczas gdy w Polsce przy zastosowaniu zbliżonej metodologii nie występuje wyraźna różnica między oficjalną a oczekiwaną liczbą zgonów. Można zatem wnioskować, że niedoszacowanie liczby zgonów związanych z COVID-19 w Polsce najprawdopodobniej nie jest wysokie.

Wprowadzenie

Do 19.05.2020 r. na całym świecie oficjalnie potwierdzono blisko 4,7 mln przypadków zarażenia koronawirusem oraz ponad 315 tys. zgonów związanych z COVID-19 (WHO, 2020a). W wielu krajach (m.in. we Włoszech, Portugalii, Hiszpanii, Francji i Wielkiej Brytanii) podczas pandemii koronawirusa zaobserwowano nadmierną śmiertelność (*excess mortality*), czyli śmiertelność powyżej tej, której należałoby oczekiwać w normalnej sytuacji (Checchi, Roberts, 2005). Występowanie nadmiernej śmiertelności podczas pandemii koronawirusa omówiono zarówno w opracowaniach naukowych (Docherty i in., 2020; Felix-Cardoso i in., 2020; Olson i in., 2020; Nogueira i in., 2020; Wetzler, H.P., Wetzler, E.A., 2020), jak i w prasie codziennej (Burn-Murdoch, Romei, Giles 2020; „The Economist", 2020). Wg najlepszej wiedzy autora nie opublikowano do tej pory podobnej analizy dla Polski.

Pierwszy przypadek koronawirusa potwierdzono w Polsce 4.03.2020 r., a pierwszy zgon 12.03. Wg stanu na 19.05 potwierdzono 19 268 przypadków zarażenia koronawirusem oraz zanotowano 948 zgonów (@MZ_GOV_PL, 2020). Liczby te przekładają się na skumulowaną zapadalność 509 przypadków na milion mieszkańców oraz skumulowaną śmiertelność 25 osób na milion mieszkańców. Porównując te wartości z innymi krajami europejskimi należy stwierdzić, że są to liczby stosunkowo niewielkie. Na podstawie danych o pandemii zbieranych przez Uniwersytet Johna Hopkinsa (John Hopkins University, 2020) 19.05.2020 r. tylko cztery kraje UE cechowały się niższą oficjalną skumulowaną zapadalnością (Grecja, Słowacja, Bułgaria, Węgry), a osiem niższą oficjalną skumulowaną śmiertelnością (Słowacja, Łotwa, Malta, Cypr, Grecja, Bułgaria, Litwa, Chorwacja). Pod względem współczynnika śmiertelności na COVID-19 (*case mortality*) 19.05 Polska plasowała się na 14. miejscu w UE.

Niska oficjalna liczba zgonów może wynikać ze skuteczności środków zapobiegawczych, które w Polsce wprowadzono stosunkowo wcześnie (dla przykładu szkoły zamknięto 12.03.2020 r., gdy potwierdzonych przypadków było zaledwie 49). Z drugiej strony, za tym że w Polsce statystyki zgonów mogą być zaniżone może przemawiać kilka kwestii, między innymi stosunkowo niewielka liczba testów (Roser i in., 2020) oraz to, że nie wszystkie zgony powiązane z COVID-19 zgłaszane są do odpowiednich komórek Państwowej Inspekcji Sanitarnej (WSSE, 2020).

Na przykładzie kilku krajów pokazano, że nadmierna śmiertelność w okresie pandemii może przewyższać oficjalne liczby zgonów związanych z COVID-19 nawet o kilkadziesiąt procent (Docherty i in., 2020). Jeśli skala braku odpowiedniego raportowania zgonów związanych z COVID-19 jest wysoka, można spodziewać się, że w Polsce również występuje nadmierna śmiertelność.

W niniejszym opracowaniu dokonano analizy występowania nadmiernej śmiertelności w Polsce w okresie od 5.03 do 13.05.2020 r. Rzeczywistą liczbę zgonów w tym okresie, uzyskaną z Rejestru Stanu Cywilnego, porównano z oczekiwaną liczbą zgonów oszacowaną na podstawie (1) przeskalowanej, historycznej średniej arytmetycznej, (2) analizy kontrfaktycznej i (3) najniższych historycznych współczynników śmiertelności z lat 2010-2019. Występowanie nadmiernej śmiertelności testowano za pomocą testu *z* dla proporcji.

Dane i metodologia

Nadmierna śmiertelność

Nadmierna śmiertelność podczas sytuacji kryzysowej zdefiniowana jest jako śmiertelność powyżej tej, której należałoby oczekiwać w normalnej sytuacji (WHO, 2020b, Checchi, Roberts, 2005). Może być wyrażona w wartościach bezwzględnych jako różnica między rzeczywistą a oczekiwaną liczbą zgonów:

$$\textit{nadmierna liczba zgonów} = \textit{zaobserwowana liczba zgonów} - \textit{oczekiwana liczba zgonów}$$

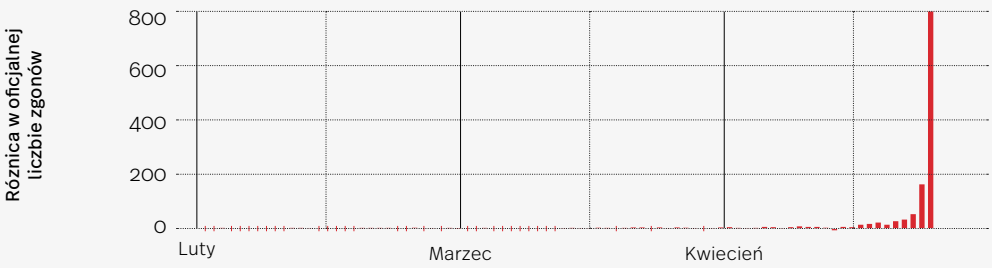
lub w wartościach względnych jako stosunek tych dwóch wartości:

$$\textit{nadmierna smiertelność} = \frac{\textit{zaobserwowana liczba zgonów}}{\textit{oczekiwana liczba zgonów}}$$

Zaobserwowana liczba zgonów

Dane o dobowej liczbie zgonów w okresie 1.01.2010-30.09.2015 otrzymano z Rejestru PESEL, dane z okresu 1.10.2015-15.05.2020 – z Rejestru Stanu Cywilnego. Wygenerowane zbiory danych cechują się opóźnieniem w sprawozdawczości – zgony często odnotowywane są nawet kilka dni po tym, kiedy rzeczywiście miały miejsce. Proces ten można zaobserwować porównując zbiory wygenerowane w dwóch różnych dniach. Na potrzeby niniejszego opracowania dysponowano zbiorami wg stanu na 15.05.2020 r. oraz wg stanu na 24.04.2020 r. Porównanie tych zbiorów w okresie 1.02.2020-24.04.2020 przedstawiono na wykresie 1.

↘ **Wykres 1. Różnica w oficjalnej, dziennej liczbie zgonów między danymi z dnia 15.05.2020 r. a danymi z dnia 24.04.2020 r.**

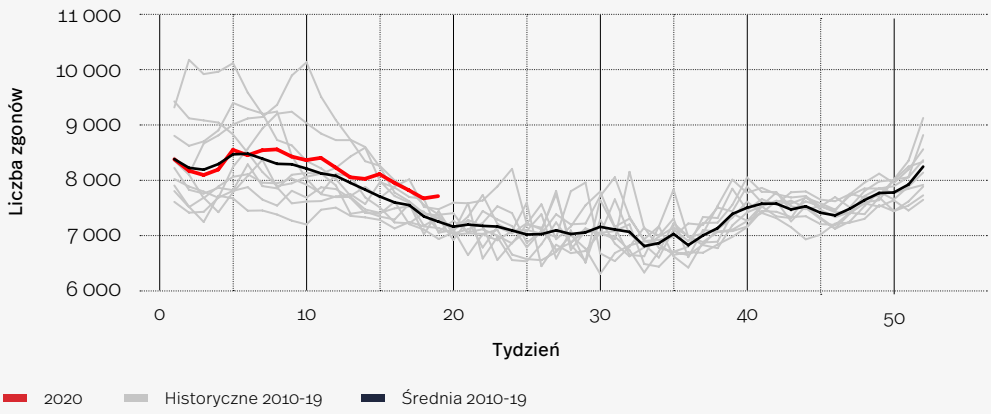


Źródło: opracowanie własne PIE na podstawie Rejestru Stanu Cywilnego.

Istnieje kilka metod korekty opóźnień w raportowaniu (np. Esbjerg, Keiding, Koch-Henriksen, 1994; Lawless, 1994; Midthune i in., 2005; Rosenberg, 1990). W niniejszym opracowaniu korekty dokonano za pomocą uproszczonego, dyskretnego modelu probabilistycznego (załącznik 1).

Zaobserwowaną tygodniową liczbę zgonów w latach 2010-2020 przedstawiono na wykresie 2.

↘ **Wykres 2. Porównanie tygodniowej liczby zgonów w 2020 r. z przebiegiem krzywych zgonów oraz ze średnią liczbą zgonów z lat 2010-2019**



Źródło: opracowanie własne PIE na podstawie rejestru PESEL i Rejestru Stanu Cywilnego.

Oczekiwana liczba zgonów w 2020 r.

Do szacowania oczekiwanej liczby zgonów w badanym okresie wykorzystano oczekiwaną liczbę zgonów w całym 2020 r. Wartość tę oszacowano za pomocą liczby ludności w podziale na płeć i pojedyncze roczniki wieku z 31.12.2019 r. oraz tablic umieralności za 2019 r. Przyjęto, że każdy osobnik w populacji ma 31.12.2019 x+0,5 lat oraz że gęstość zmiennej T określającej długość życia jest stała w przedziałach $[k,k+1)$ dla $k \in Z$ (zgodnie z: Slud, 2001). Przy tych założeniach oczekiwana liczba zgonów dla płci w 2020 r. dana jest przez:

$$\mathbb{E}Z_{2020} = \sum_{x=0}^{100} l_x \cdot \left(1 - \frac{(1-q_x)\left(1-\frac{q_{x+1}}{2}\right)}{1-\frac{q_x}{2}}\right)$$

gdzie:

l_x – liczba osób, które 31.12.2020 r. miały ukończone x lat,

q_x – prawdopodobieństwo, że osoba w wieku x nie dożyje wieku x+1 (załącznik 2).

Zgodnie z powyższą metodą oczekiwana liczba zgonów w 2019 r. wynosi 418 647.

Dla porównania analogiczne obliczenia wykonano dla lat 2011-2019.

Wyniki zaprezentowano w tabeli 1.

↘ Tabela 1. Zaobserwowana i oczekiwana liczba zgonów w latach 2011-2019

Rok	Zaobserwowana liczba zgonów (O_t)	Oczekiwana liczba zgonów (E_t)	Stosunek zaobserwowanej do oczekiwanej liczby zgonów (OE_t)
2011	380 650	381 304	0,9983
2012	388 807	380 808	1,0210
2013	391 590	385 847	1,0149
2014	381 237	387 500	0,9838
2015	400 342	377 500	1,0605
2016	390 634	392 313	0,9957
2017	405 404	388 354	1,0439
2018	414 073	399 201	1,0373
2019	410 959	411 604	0,9984
2020	-	418 647	-
Bezwzględny średni błąd prognozy (MAE): 8 638,55			

Źródło: opracowanie własne PIE na podstawie danych GUS.

Zakładając, że stosunek zaobserwowanych do oczekiwanych zgonów (*OE*) ma pewien rozkład (np. normalny) można wyznaczyć odpowiednie kwantyle, tak by:

$$\mathbb{P}\left(OE_{\frac{\alpha}{2}} \leq OE \leq OE_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

czyli przedziału, w którym stosunek wartości oczekiwanej do zaobserwowanej znajduje się z prawdopodobieństwem 1-α. Traktując *E* jako stałą, a *O* jako zmienną losową możemy przekształcić powyższy wzór tak, by był (1-α) procentowym przedziałem predykcji:

$$\mathbb{P}\left(OE_{\frac{\alpha}{2}} \leq OE \leq OE_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(E_i \cdot OE_{\frac{\alpha}{2}} \leq O_i \leq E_i \cdot OE_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Przy tych założeniach 90-procentowy przedział predykcji dla 2020 r. wynosi [408 063,443 536].

Oczekiwana liczba zgonów w okresie 5.03-13.05.2020

Oczekiwaną liczbę zgonów w okresie 5.03-13.05.2020 oszacowano za pomocą trzech metod:

I. Przeskalowana średnia arytmetyczna tygodniowej liczby zgonów z lat 2010-2019. Rzeczywista liczba zgonów w latach 2010-2019 podlegała znacznym wahaniom (wykres 2, tabela 1). W latach 2010-2019 było ich w Polsce około 380 tys., podczas gdy w latach 2018-2019 już 410-415 tys. Różnica ta wynika przede wszystkim z przemian demograficznych i starzejącego się społeczeństwa. Dlatego też, żeby zapewnić porównywalność między zaobserwowaną a oczekiwaną liczbą zgonów, średnią historyczną przeskalowano (tak, by liczba zgonów w roku była równa oczekiwanej liczbie zgonów dla 2020 r.) (oszacowanie 1a). Porównanie zaobserwowanej liczby zgonów do średniej historycznej jest w dotychczas opublikowanych pracach najpopularniejszą metodą szacowania nadmiernej śmiertelności COVID-19 (m.in. Docherty i in., 2020; Felix-Cardoso i in., 2020; Nogueira i in., 2020). Dla porównania zaobserwowaną liczbę zgonów zestawiono również z 90-procentowym przedziałem predykcji (oszacowania 1b i 1c)

II. Analiza kontrfaktyczna. W latach 2010-2019 zaobserwowano dużą wariancję w częstotliwości zgonów, w szczególności w okresie zimowym (wykres 2). Jednym z podejść, które można zastosować by ograniczyć skalę tego problemu jest analiza kontrfaktyczna przeprowadzona w następujący sposób:

1. Rozkłady zgonów z lat 2010-2019 przeskalowano do liczby zgonów w 2019 r. i oczekiwanej liczby zgonów w 2020 r. Rozkłady te wygładzono za pomocą regresji lokalnej (*locally estimated scatterplot smoothing, LOESS*) (Sharma, Swayna, Obimbo, 2015). Łącznie dla każdego roku obliczono 4 rozkłady porównawcze:
   - → rzeczywisty rozkład – przeskalowany do 2020 r.
   - → wygładzony rozkład – przeskalowany do 2020 r.
   - → rzeczywisty rozkład – przeskalowany do 2019 r.
   - → wygładzony rozkład – przeskalowany do 2019 r.
2. Porównano dane dla stycznia i lutego z punktu 1. z oficjalną liczbą zgonów w styczniu i lutym w 2020 r. (styczeń i luty stanowią odpowiedni zbiór testowy, gdyż można zakładać, że wtedy nie było koronawirusa i zgony powinny występować zgodnie z oczekiwanym rozkładem).
3. Wybrano rozkład cechujący się najniższym błędem średniokwadratowym.
4. Przy założeniu, że zgony w 2020 r. powinny od marca dalej podążać zgodnie z najlepszym rozkładem z punktu 3., porównano oczekiwaną i rzeczywistą liczbę zgonów w badanym okresie.

Spośród badanych rozkładów najlepszym dopasowaniem cechowały się rzeczywisty rozkład z 2014 r. przeskalowany do 2020 r. (oszacowanie 2a) oraz wygładzony rozkład z 2012 r. przeskalowany do 2020 r. (oszacowanie 2b) (tabela 2).

↘ Tabela 2. **Błąd średniokwadratowy w analizie kontrfaktycznej**

Rok	Rozkłady porównawcze			
	skalowanie do 2019 r. – rozkład rzeczywisty	skalowanie do 2019 r. – rozkład wygładzony	skalowanie do 2020 r. – rozkład rzeczywisty	skalowanie do 2020 r. – rozkład wygładzony
2011	94 684	126 409	186 535	233 482
2012	57 796	59 757	49 775	**33 255**
2013	1 105 299	995 381	14 29 155	1 340 198
2014	48 445	51 727	**25 838**	39 758
2015	179 769	57 489	258 138	143 429
2016	48 249	78 903	91 720	134 625
2017	2 375 898	2 103 644	2 950 577	2 611 031
2018	235 782	283 939	394 381	473 162
2019	390 090	355 476	615 949	541 466

Źródło: opracowanie własne PIE na podstawie danych GUS, rejestru PESEL i Rejestru Stanu Cywilnego.

III. Najniższy zaobserwowany współczynnik śmiertelności w latach 2010-2019 w badanym okresie oraz globalnie. Należy oczekiwać, że podczas *lockdownu* śmiertelność powinna być niższa niż w normalnych warunkach. Istnieje kilka czynników, które mogą wpływać na obniżenie śmiertelności: niższa aktywność zawodowa, niższa zapadalność na choroby zakaźne, mniejsza liczba urazów i wypadków (w tym komunikacyjnych), mniejsze zanieczyszczenie powietrza, odwołanie operacji planowych itp. Nogueira i współautorzy (2020) w analizie nadmiernej śmiertelności podczas pandemii koronawirusa w Portugalii sugerują, że umieralność w okresie *lockdownu* może przypominać umieralność obserwowaną w okresie letnim, kiedy część z powyższych czynników również ma zastosowanie (zmniejszona liczba operacji, niższa aktywność zawodowa, niższa zapadalność na choroby zakaźne).

W celu uwzględnienia tego, że śmiertelność podczas *lockdownu* (wyłączając efekt COVID-19 oraz efekt ograniczonego dostępu do usług medycznych) powinna być niższa niż w normalnej sytuacji rozważono trzy dodatkowe oszacowania oczekiwanej liczby zgonów w badanym okresie:

- oszacowanie zakładające, że śmiertelność wg wieku i płci w badanym okre-sie będzie równa śmiertelności z analogicznego okresu z lat 2010-2019, w którym zaobserwowano najniższy standaryzowany wiekiem

i płcią współczynnik śmiertelności (jest to okres 5.03-13.05.2019 r.) (oszacwanie 3a);

- o oszacowanie zakładające, że tygodniowa śmiertelność wg wieku i płci w badanym okresie (10 tygodni) będzie równa śmiertelności wg wieku i płci z analogicznych tygodni z lat 2010-2019, w których zaobserwowano najniższy standaryzowany wiekiem i płcią współczynnik śmiertelności (są to: dla tygodnia 05.03-11.03 rok 2019, dla 12.03-18.03 rok 2017, dla 19.03-25.03 rok 2019, dla 26.03-1.04 rok 2019, dla 2.04-8.04 rok 2017, dla 9.04-15.04 rok 2019, dla 16.04-22.04 rok 2016, dla 23.04-29.04 rok 2016, dla 30.04-6.05 rok 2019, dla 7.05-13.05 rok 2017) (oszacowanie 3b);
- o oszacowanie zakładające, że śmiertelność wg wieku i płci w badanym okresie będzie równa śmiertelności z 10 kolejnych tygodni (tyle czasu upływa od 5.03 do 13.05) z lat 2010-2019, w których zaobserwowano najniższy standaryzowany wiekiem i płcią współczynnik śmiertelności (jest to okres 20.06-27.08.2019 r.) (oszacowanie 3c).

Test hipotezy

Dla każdej z powyższych metod dysponujemy dwiema wartościami: zaobserwowaną liczbą zgonów (O) oraz oczekiwaną liczbą zgonów (E). Przyjmując, że populacja na początku okresu (N) jest równa populacji w dniu 31.12.2019 (czyli, że liczba urodzeń do 4.03 jest równa liczbie zgonów w tym okresie) testowanie hipotezy, że oczekiwana liczba zgonów jest równa zaobserwowanej można sprowadzić do testu z proporcji dla jednej populacji:

$$H_0: p = p_0$$

gdzie estymatorem p jest $\hat{p} = \frac{O}{N}$, a $p_o = \frac{E}{N}$.

Powyższą hipotezę H_o testowano wobec alternatywy: $p > p_o$, czyli równoważnie $O > E$.

Wyniki

Od 5.03 do 13.05 po korekcie (załącznik 1), zaobserwowana liczba zgonów w Polsce wyniosła 80 324. Oszacowania oczekiwanej liczby zgonów oraz p-wartości testów hipotezy, że zaobserwowana śmiertelność była równa oczekiwanej (wobec alternatywy, że zaobserwowana śmiertelność była wyższa) przedstawiono w tabeli 3.

Oszacowanie oczekiwanej liczby zgonów w okresie 5.03-13.05.2020 r. wg metody (1) wyniosło 82 363, a 90-procentowy przedział dla tej predykcji wyniósł [80 280,87 257]. Testy równości między zaobserwowaną a oczekiwaną liczbą zgonów cechowały się p-wartościami równymi ≈1 dla oczekiwanej liczby zgonów (1a), 0,4388 dla dolnej granicy 90-procentowego przedziału dla prognozy (1b) oraz ≈1 dla górnej granicy 90-procentowego przedziału dla prognozy (1c). Wyniki metody (1) zagregowane tygodniowo przedstawiono na wykresie 3.

Oszacowanie oczekiwanej liczby zgonów wg metody (2) wyniosło 82 042 przy założeniu rzeczywistego rozkładu zgonów z 2014 r. przeskalowanego do oczekiwanej liczby zgonów w 2020 r. (2a) oraz 81 632 przy założeniu wygładzonego rozkładu zgonów z 2012 r. przeskalowanego do oczekiwanej liczby zgonów w 2020 r. (2b). Testy równości między zaobserwowaną a oczekiwaną liczbą zgonów cechowały się p-wartościami ≈1 dla obydwu rozkładów. Wyniki metody (2) zagregowane tygodniowo przedstawiono na wykresie 4.

Oszacowanie oczekiwanej liczby zgonów wg metody (3) wyniosło 80 567 (oszacowanie 3a), 79 936 (oszacowanie 3b) oraz 74 491 (oszacowanie 3c). Testy równości między zaobserwowaną a oczekiwaną liczbą zgonów cechowały się p-wartościami 0,8041 (3a), 0,0848 (3b), <0,0001 (3c). Wyniki metody (3) zagregowane tygodniowo przedstawiono na wykresie 5.

Spośród ośmiu oszacowań oczekiwanej liczby zgonów w przypadku siedmiu brak jest podstaw do odrzucenia hipotezy o równości zaobserwowanej i oczekiwanej liczby zgonów wobec alternatywy, że zaobserwowana śmiertelność jest wyższa. Biorąc pod uwagę, że jedynym oszacowaniem, w którym oczekiwana liczba zgonów jest istotnie niższa niż rzeczywista, jest oszacowanie przyrównujące śmiertelność w okresie *lockdownu* do śmiertelności w okresie letnim (oszacowanie 3c), należy stwierdzić, że na podstawie trendów historycznych nie ma podstaw do stwierdzenia, że w badanym okresie liczba zgonów była wyższa niż należałoby się spodziewać. O nadmiernej śmiertelności możemy mówić tylko wtedy, jeśli prawdziwa jest hipoteza Nogueiry i współautorów (2020).

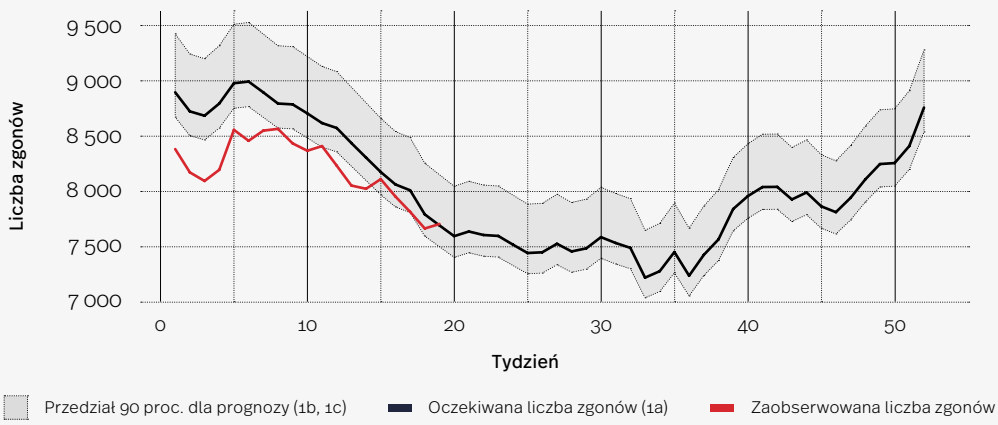
Jeśli przy testowaniu hipotezą alternatywną byłoby $O<E$ to odrzucilibyśmy hipotezę zerową na jej rzecz w przypadku czterech spośród ośmiu oszacowań. Możliwe jest zatem, że podczas pandemii koronawirusa obserwujemy w Polsce mniej zgonów niż należałoby się spodziewać, choć liczba ta jest stosunkowo niewielka (1-2 tys.).

↘ Tabela 3. Oszacowania oczekiwanej liczby zgonów wg metod (1a)-(3c) oraz wyniki testów hipotezy, że zaobserwowana śmiertelność w okresie 5.03.-13.05.2020 r. była równa oczekiwanej

Zaobserwowana liczba zgonów: 80 324			
Metoda		Oczekiwana liczba zgonów	p-wartość dla $H_A: 0 > E$
(1a)	średni rozkład z lat 2010-2019 przeskalowany do oczekiwanej liczby zgonów w 2020 r.	82 363	≈1
(1b)	średni rozkład z lat 2010-2019 przeskalowany do dolnej granicy 90-procentowego przedziału dla prognozy w 2020 r.	80 280	0,4388
(1c)	średni rozkład z lat 2010-2019 przeskalowany do górnej granicy 90-procentowego przedziału dla prognozy w 2020 r.	87 259	≈1
(2a)	rzeczywisty rozkład z 2014 r. przeskalowany do oczekiwanej liczby zgonów w 2020 r.	82 042	≈1
(2b)	wygładzony rozkład z 2012 r. przeskalowany do oczekiwanej liczby zgonów w 2020 r.	81 632	≈1
(3a)	śmiertelność wg wieku i płci równa śmiertelności z analogicznego okresu z lat 2010-2019, w którym zaobserwowano najniższy standaryzowany wiekiem i płcią współczynnik śmiertelności	80 567	0,8041
(3b)	śmiertelność wg wieku i płci równa śmiertelności z analogicznych tygodni z lat 2010-2019, w których zaobserwowano najniższy standaryzowany wiekiem i płcią współczynnik śmiertelności	79 936	0,0848
(3c)	śmiertelność wg wieku i płci równa śmiertelności z 10 kolejnych tygodni z lat 2010-2019, w których zaobserwowano najniższy standaryzowany wiekiem i płcią współczynnik śmiertelności.	74 491	<0,0001

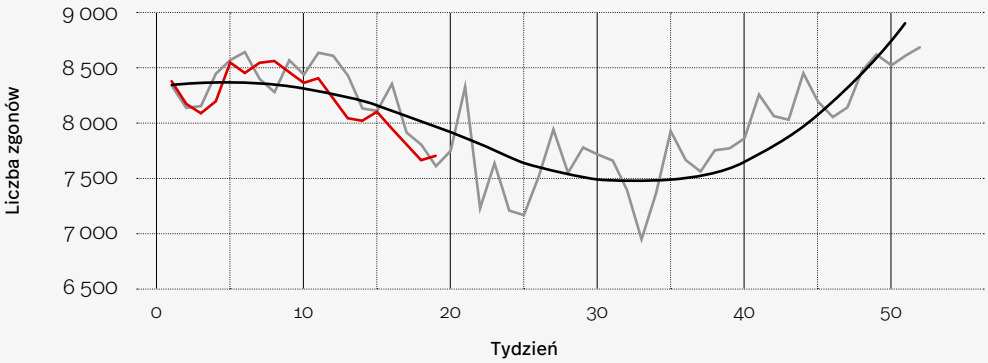
Źródło: jak w tabeli 2.

↘ Wykres 3. Porównanie zaobserwowanej, tygodniowej liczby zgonów w 2020 r. i oczekiwanej liczby zgonów w 2020 r. wg metody (1)



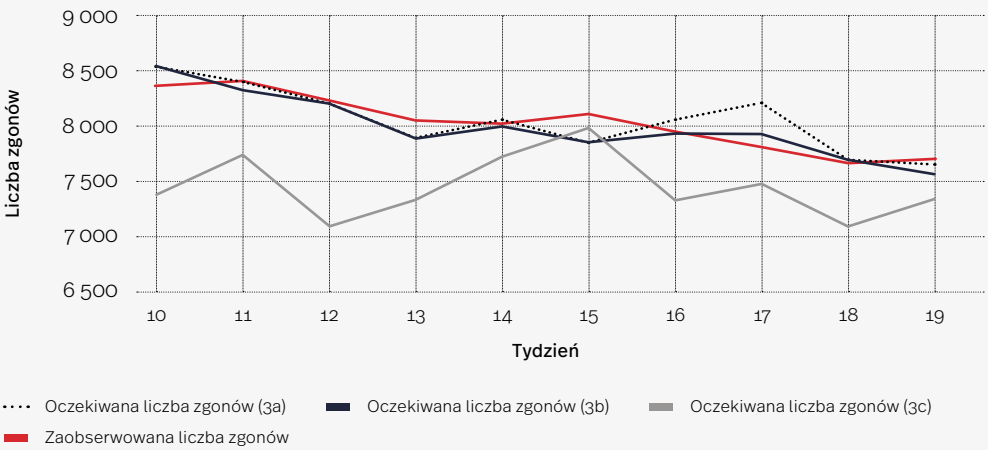
Źródło: opracowanie własne PIE na podstawie danych GUS, rejestru PESEL i Rejestru Stanu Cywilnego.

↘ Wykres 4. Porównanie zaobserwowanej, tygodniowej liczby zgonów w 2020 r. i oczekiwanej liczby zgonów w 2020 r. wg metody (2)



Źródło: jak w wykresie 3.

↘ Wykres 5. Porównanie zaobserwowanej, tygodniowej liczby zgonów w 2020 r. i oczekiwanej liczby zgonów w 2020 r. wg metody (3)



Źródło: jak w wykresie 3.

Dyskusja

Biorąc pod uwagę niewielką oficjalną liczbę zgonów związanych z COVID-19 w Polsce oraz to, że w przypadku innych krajów Europy stwierdzono nadmierną śmiertelność w okresie pandemii, której nie daje się całkowicie wyjaśnić oficjalną liczbą zgonów, w niniejszym opracowaniu sprawdzono występowanie nadmiernej śmiertelności w Polsce w okresie 5.03-13.05.2020 r.

Nadmierną śmiertelność zdefiniowano jako różnicę między zaobserwowaną a oczekiwaną liczbą zgonów. Wyniki analizy wskazują, że nie ma podstaw do stwierdzenia, że w Polsce w okresie 5.03-15.05 2020 r. liczba zgonów była wyższa niż należałoby oczekiwać na podstawie trendów historycznych. Bardziej prawdopodobne jest występowanie odwrotnego zjawiska, tj. zaniżonej śmiertelności (1-2 tys. zgonów, czyli 100-200 zgonów tygodniowo).

Należy zauważyć, że samo to, iż na podstawie trendów historycznych brakuje podstaw do stwierdzenia nadmiernej śmiertelności nie jest jednoznaczne z brakiem występowania w Polsce niedoszacowania oficjalnej liczby zgonów związanych z COVID-19. Powyższe wynika z tego, że pandemia połączona z *lockdownem* cechuje się nasileniem zarówno czynników, które mogą śmiertelność zwiększać, jak i czynników, które mogą na śmiertelność wpływać negatywnie, a przeprowadzona w niniejszym opracowaniu analiza opiera się na liczbach bezwzględnych i nie obejmuje analizy przyczyn zgonów.

Podniesienie liczby zgonów będzie wynikać z samej choroby, ale też z ograniczonego dostępu do opieki medycznej. Z kolei zmniejszenie liczby zgonów może wynikać z szeregu czynników, m.in. z mniejszej liczby wypadków komunikacyjnych, z odwołania operacji planowych, mniejszej liczby urazów, mniejszej aktywności zawodowej, podniesienia czystości powietrza itp. Te dwa zestawy czynników będą się do pewnego stopnia wzajemnie neutralizować i bez dodatkowej analizy przyczyn zgonów nie można jednoznacznie określić skali przyczynowego wpływu COVID-19 na śmiertelność.

Istnieją zasadne argumenty za tym, że liczba zgonów związanych z COVID-19 jest w Polsce wyższa niż oficjalna. Pierwszym z nich jest stosunkowo niewielka liczba wykonanych testów na obecność koronawirusa. Spośród krajów unijnych, dla których dostępne są względnie aktualne dane (czyli wszystkich z pominięciem Cypru, Malty, Francji i Niemiec) mniej testów na 1 tys. mieszkańców wg stanu na 19.05.2020 r. przeprowadzono wyłącznie w Bułgarii, Grecji, Chorwacji, Rumunii i na Węgrzech (Roser i in., 2020). Liczba przeprowadzonych testów jest nieodwracalnie powiązana z liczbą wykrytych przypadków (w obie strony, tj. więcej chorych powoduje, że robi się więcej testów, ale też więcej testów powoduje, że wykrywa się większą liczbę chorych) i niska częstotliwość testowania w Polsce może powodować zaniżenie oficjalnych statystyk, zarówno zgonów, jak i zachorowań.

Po drugie nawet informacje o zgonach osób z potwierdzeniem lub podejrzeniem zarażenia koronawirusem mogą nie znaleźć się w oficjalnych statystykach. Wynika to z braku zgłoszenia zgonu przez podmiot leczniczy do odpowiedniej komórki Państwowej Inspekcji Sanitarnej. Zgodnie z komunikatem Wojewódzkiej Stacji Sanitarno-Epidemiologicznej w Warszawie z 9.04.2020 r. możliwe jest, że nie wszystkie zgony są raportowane w ten sposób (WSSE, 2020).

Biorąc jednak pod uwagę, że w innych krajach Europy wykazano wysoką nadmierną

śmiertelność podczas pandemii koronawirusa oraz to, że zasady *lockdownu* są w Polsce i w tych krajach zbliżone (czyli można spodziewać się zbliżonego, ujemnego wpływu *lockdownu* na śmiertelność) można wnioskować, że niedoszacowanie zgonów związanych z COVID-19 nie jest w Polsce wysokie. Przynajmniej nie na tyle, by było w oczywisty sposób widoczne w analizach statystycznych.

Niniejsze opracowanie ma kilka ograniczeń. Po pierwsze nie uwzględniono w nim przyczyn zgonów, jak wspomniano wyżej.

Po drugie śmiertelność rozważono wyłącznie na zagregowanym poziomie, bez uwzględniania podziału na grupy wiekowe, płeć, miejsce zamieszkania. Nogueira i współautorzy (2020) na przykładzie Portugalii pokazali, że najwyższą nadmierną śmiertelnością podczas pandemii cechowały się najstarsze grupy wiekowe, podczas gdy w młodszych zaobserwowana liczba zgonów była bliska oczekiwanej. Możliwe jest zatem, że brak istotnej różnicy między zaobserwowaną a oczekiwaną liczbą zgonów w Polsce wynika z tego, że w niektórych grupach populacji występuje nadmierna śmiertelność, podczas gdy w innych śmiertelność jest obniżona.

Po trzecie nie uwzględniono bezpośredniego wpływu temperatury i innych sezonowych czynników, które mogą mieć wpływ na zróżnicowanie śmiertelności między latami (choć uwzględniono te czynniki pośrednio, w szczególności w metodzie 2). Chłodniejsze okresy wiążą się z podwyższeniem współczynników śmiertelności (Armstrong, 2006; Huynen i in., 2001; Pattenden, Nikiforov, Armstrong, 2003), a początek 2020 r. w Polsce w porównaniu z poprzednimi latami był względnie ciepły. Dlatego też oszacowanie oczekiwanej śmiertelności na podstawie danych historycznych może cechować się dodatnim obciążeniem.

Bibliografia

@MZ_GOV_PL (Ministerstwo Zdrowia) (2020), *Liczba zakażonych koronawirusem: 19 268/948 (wszystkie pozytywne przypadki/w tym osoby zmarłe)*, Twitter, 19.05.2020, 17:30, https://twitter.com/MZ_GOV_PL/status/1262767586581704706 [dostęp: 19.05.2020].

Armstrong, B. (2006), *Models for the relationship between ambient temperature and daily mortality*, "Epidemiology", No. 17(6).

Burn-Murdoch, J., Romei,V., Giles, C. (2020), *Global coronavirus death toll could be 60% higher than reported, 26 kwietnia*, https://www.ft.com/content/6bd88b7d-3386-4543-b2e9-0d5c6fac846c [dostęp: 20.05.2020].

Checchi, F., Roberts, L. (2005), *Interpreting and using mortality data in humanitarian emergencies*, "Humanitarian Practice Network", No. 52.

Docherty, K., Butt, J., de Boer, R. i in. (2020), *Excess deaths during the Covid-19 pandemic: An international comparison*, "medRxiv", https://doi.org/10.1101/2020.04.21.20073114 [dostęp: 20.05.2020].

Esbjerg, S., Keiding, N., Koch-Henriksen, N. (1999), *Reporting Delay and Corrected Incidence of Multiple Sclerosis*, "Statistics in Medicine", No. 18.

Felix-Cardoso, J., Vasconcelos, H., Pereira Rodrigues, P., Cruz-Correira, R. (2020), *Excess mortality during COVID-19 in five European countries and a critique of mortality data analysis*, "medRxiv", https://doi.org/10.1101/2020.04.28.20083147 [dostęp: 20.05.2020].

Hasell, J., Ortiz-Ospina, E., Mathieu, E. i in. (2020), *Coronavirus (COVID-19) Testing*, https://ourworldindata.org/coronavirus-testing# [dostęp: 19.05.2020].

Huynen, M.M., Martens, P., Schram, D., Weijenberg, M.P., Kunst, A.E. (2001), *The impact of heat waves and cold spells on mortality rates in the Dutch population*, "Environmental Health Perspectives", No. 109(5).

John Hopkins University (2020), *COVID-19 data*, https://github.com/CSSEGISandData/COVID-19 [dostęp: 19.05.2020].

Lawless, J.F. (1994), *Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events*, "The Canadian Journal of Statistics", No. 22(1).

Midthune, D.N., Fay, M.P., Clegg, L.X., Feuer, E.J. (2005), *Modeling Reporting Delays and Reporting Corrections in Cancer Registry Data*, "Journal of the American Statistical Association", No. 100(469).

Nogueira, P.J., Nobre, M.dA., Nicola, P.J., Furtado, C., Vaz Carneiro, A. (2020), *Excess Mortality Estimtion During the COVID-19 Pandemic: Preliminary Data from Portugal*, "Acta Medica Portugesa", No. 33(13).

Olson, D.R., Huynh, M., Fine, A. i in. (2020), *Preliminary Estimate of Excess Mortality During the COVID-19 Outbreak – New York City, March 11-May 2, 2020*, "Centres for Disease Control and Prevention Morbidity and Mortality Weekly Report May 11, 2020", No. 69.

Pattenden, S., Nikiforov, B., Armstrong, B.G. (2003), *Mortality and temperature in Sofia and London*, "Journal of Epidemiology and Community Health", No. 57(8).

Rosenberg, P.S. (1990), *A simple correction of AIDS surveillance data for reporting delays*, "Journal of acquired immune deficiency syndromes", No. 3(1).

Roser, M., Ritchie, H., Ortiz-Ospina, E., Hasell, J. (2020), *Coronavirus Pandemic (COVID-19)*, https://ourworldindata.org/coronavirus [dostęp: 21.05.2020].

Sharma, S., Swayne, D.A., Obimbo, C. (2015), *Automating the Smoothing of Time Series Data*, "Journal of Environmental & Analytical Toxicology", No. 5(5).

Slud, E.V. (2001), *Actuarial mathematics and life-table statistics*, Chapman & Hall/CRC.

"The Economist" (2020), *Tracking COVID-19 excess deaths across countries*, 16 kwietnia, https://www.economist.com/graphic-detail/2020/04/16/tracking-covid-19-excess-deaths-across-countries [dostęp: 20.05.2020].

Wetzler, H.P., Wetzler, E.A. (2020), *COVID-19 Excess deaths in the United States, New York City, and Michigan during April 2020*, "medRxiv", https://doi.org/10.1101/2020.04.02.20051532 [dostęp: 20.05.2020].

WHO (2020a), *Coronavirus disease (COVID-19) Situation Report – 120*, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200519-covid-19-sitrep-120.pdf?sfvrsn=515cabfb_2 [dostęp: 20.05.2020].

WHO (2020b), *Definitions: emergencies*, http://origin.who.int/hac/about/definitions/en/ [dostęp: 24.05.2020].

WSSE (Wojewódzka Stacja Sanitarno-Epidemiologiczna w Warszawie) (2020), *Statystyka zgonów na Mazowszu - odpowiedź Mazowieckiego Państwowego Wojewódzkiego Inspektora Sanitarnego na wypowiedź Prezydenta m.st. Warszawy*, http://wsse.waw.pl/aktualnosci-i-komunikaty/aktualnosci/statystyka-zgonow-na-mazowszu-_-odpowiedz-mazowieckiego-panstwowego--wojewodzkiego-inspektora-sanitarnego-na-wypowiedz-prezydenta [dostęp: 22.05.2020].

Załączniki

Załącznik 1. Korekta opóźnienia w sprawozdawczości zgonów

Korekta – model probabilistyczny

Niech z_t oznacza rzeczywistą liczbę zgonów w dniu t.

Dla każdego dnia t zdefiniujmy $Y_1^t,\ldots, Y_{z_t}^t$, gdzie Y_i^t jest zmienną losową oznaczającą liczbę dni opóźnienia w zarejestrowaniu zgonu i z dnia t. Załóżmy, że wszystkie Y_i^t są niezależne i mają ten sam rozkład:

$$\mathbb{P}(Y_i^t = k) = q_k \text{ dla } k \in \mathbb{N}$$

Niech $T_i^t(j)$ będzie binarną zmienną losową zdefiniowaną jako:

$$T_i^t(j) = \begin{cases} 1 & jeśli\ Y_i^t \le j \\ 0 & jeśli\ Y_i^t > j \end{cases}$$

czyli:

$$\mathbb{P}(T_i^t(j) = 1) = \sum_{k=0}^{j} q_j = 1 - \mathbb{P}(T_i^t(j) = 0)$$

oznaczmy

$$p_j = \sum_{k=0}^{j} q_j$$

gdzie:
$T_i^t(j)$ – zmienna losowa określająca czy zgon i z dnia t został zgłoszony z opóźnieniem maksymalnie j dni,
p_j – prawdopodobieństwo, że zgon zostanie zarejestrowany z maksymalnym opóźnieniem j dni (takie same dla każdego zgonu, co wynika z założenia, że Y_i^t są *IID*).
W danych dotyczących zgonów wg stanu na dzień s obserwujemy wyłącznie wartości $X_t^s(t \le s)$, czyli liczbę zgonów z dnia t zarejestrowaną do dnia s. Zgodnie z wcześniejszymi oznaczeniami X_t^s można zapisać jako:

$$X_t^s = \sum_{k=1}^{z_t} T_k^t(s-t)$$

Otrzymamy zatem:

$$\mathbb{E}(X_t^s) = z_t p_{t-s}$$

Jako że dla każdego t i s dysponujemy tylko jedną obserwacją, dlatego najlepszym estymatorem $z_t p_{t-s}$ jest X_t^s. Możemy przybliżać:

$$X_t^s \approx z_t p_{s-t}$$

Korzystając z centralnego twierdzenia granicznego możemy stwierdzić, że błąd bezwzględny takiego oszacowania będzie co prawda rosnący ze względu na z_t:

$$\begin{aligned}\mathbb{P}(|X_t^s - z_t p_{s-t}| \le c) &= \mathbb{P}(-c \le X_t^s - z_t p_{s-t} \le c) \\ &= \mathbb{P}\left(\frac{-c}{\sqrt{z_t p_{s-t}(1-p_{s-t})}} \le \frac{X_t^s - z_t p_{s-t}}{\sqrt{z_t p_{s-t}(1-p_{s-t})}} \le \frac{c}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) \\ &\approx \Phi\left(\frac{c}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) - \Phi\left(\frac{-c}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) \\ &= 2\Phi\left(\frac{c}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) - 1\end{aligned}$$

ale już błąd względny będzie malejący:

$$\begin{aligned}\mathbb{P}\left(\frac{|X_t^s - z_t p_{s-t}|}{z_t} \le c\right) &= \mathbb{P}(-c \cdot z_t \le X_t^s - z_t p_{s-t} \le c \cdot z_t) \\ &= \mathbb{P}\left(\frac{-c \cdot z_t}{\sqrt{z_t p_{s-t}(1-p_{s-t})}} \le \frac{X_t^s - z_t p_{s-t}}{\sqrt{z_t p_{s-t}(1-p_{s-t})}} \le \frac{c \cdot z_t}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) \\ &\approx \Phi\left(\frac{c\sqrt{z_t}}{\sqrt{p_{s-t}(1-p_{s-t})}}\right) - \Phi\left(\frac{-c\sqrt{z_t}}{\sqrt{z_t p_{s-t}(1-p_{s-t})}}\right) = \\ &= 2\Phi\left(\frac{c \cdot \sqrt{z_t}}{\sqrt{p_{s-t}(1-p_{s-t})}}\right) - 1\end{aligned}$$

co oznacza, że dla dostatecznie dużych z_t, X_t^s jest względnie (w relacji do z_t) dobrym oszacowaniem $z_t p_{s-t}$.

Zakładając, że z_t jest dostatecznie duże dla każdego t oraz dysponując zbiorami danych z dwóch dni: s_1 oraz s_2, gdzie $s_2 > s_1$, możemy dla każdego dnia $t \le s_1$ zapisać następujący układ (przybliżonych) równań:

$$\begin{cases} X_t^{s_2} = z_t p_{s_2 - t} \\ X_t^{s_1} = z_t p_{s_1 - t} \end{cases}$$

Dzieląc stronami otrzymamy:

$$\frac{X_t^{s_2}}{X_t^{s_1}} = \frac{p_{s_2-t}}{p_{s_1-t}} \Rightarrow p_{s_1-t} = p_{s_2-t}\frac{X_t^{s_1}}{X_t^{s_2}}$$

Jeśli do powyższego rozwiązania będziemy po kolei podstawiać możliwe wartości $t \in \{s_1, s_1-1, s_1-2, \ldots\}$, otrzymamy następujący układ równań:

$$U: \begin{cases} p_0 = p_{s_2-s_1} \dfrac{X_{s_1}^{s_1}}{X_{s_1}^{s_2}} \\ p_1 = p_{s_2-s_1+1} \dfrac{X_{s_1-1}^{s_1}}{X_{s_1-1}^{s_2}} \\ p_2 = p_{s_2-s_1+2} \dfrac{X_{s_1-2}^{s_1}}{X_{s_1-2}^{s_2}} \\ \ldots \end{cases}$$

Niezależnie od tego do kiedy sięgają dane, powyższy układ równań ma więcej niewiadomych niż równań i bez dodatkowych założeń nie można uzyskać jednoznacznego rozwiązania. Rozwiązaniem tego problemu może być założenie, że istnieje t_o, dla którego:

$$\forall_{t \le t_0}\, p_{t_0} = 1$$

Założenie to wynika wprost z zależności[1]:

$$p_k \xrightarrow{k \to +\infty} 1$$

co oznacza, że dla dostatecznie dużych k można przybliżać:

$$p_k \approx 1$$

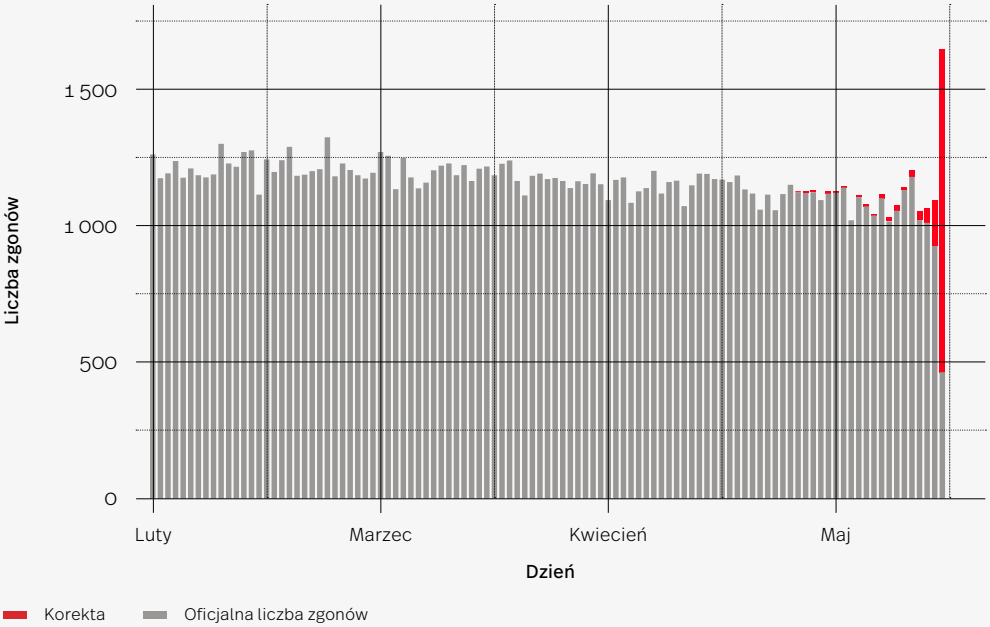
Przy tym założeniu układ U ma już mniejszą liczbę niewiadomych niż równań i pozwala na jednoznaczne wyznaczenie wartości $p_o, p_1, p_2, \ldots$, a w konsekwencji jednoznaczne wyznaczenie wartości z_t.

Korekta – obliczenia

W niniejszym opracowaniu wykorzystane zostały zbiory danych dotyczące zgonów w Polsce wg stanu na dzień 15.05.2020 r. oraz wg stanu na 24.04.2020 r. Dzienne liczby zarejestrowanych zgonów były wysokie (>1000). Przyjęto t_o z modelu probabilistycznego na t_o=90, kierując się przede wszystkim wykresem 1.
Oszacowane dobowe rzeczywiste liczby zgonów (z_t) w okresie 1.02.-15.05.2020 przedstawiono na wykresie A1.

[1] W niniejszym modelu założono, że ostatecznie wszystkie zgony zostaną odnotowane.

↘ **Wykres A1. Oszacowanie rzeczywistej liczby zgonów w Polsce do 15.05.2020 r.**



Korekta zdaje się wyraźnie przeszacowywać dane z dnia wygenerowania drugiego zbioru danych, co może wynikać z tego, że największa część informacji o zgonach spływa niedługo po ich zaistnieniu, a więc nawet różnice w godzinie generowania zbioru na dzień 24.04 i 15.05 mogą powodować znaczne rozbieżności w ostatnim dniu. Korekta wciąż pozwala na dopasowanie danych z kilkunastu poprzedzających dni do postępującego trendu.
W niniejszym opracowaniu wzięto pod uwagę dane do 13.05, czyli pominięto dwa ostatnie dni, co wynikało z przeprowadzenia części analizy na danych zagregowanych tygodniowo, a 13.05 kończy się 19. tydzień roku. Dlatego też przeszacowana korekta z 15.05 nie miała wpływu na ostateczne wyniki.

Załącznik 2. Oczekiwana liczba zgonów w roku t na podstawie struktury ludności na koniec roku $t-1$ oraz tablic trwania życia

Dane o liczbie ludności wg pojedynczych roczników pod względem wieku zawierają informację o liczbie osób, które ukończyły konkretną liczbę lat. Ludzie ujęci w takich statystykach w rzeczywistości są starsi (maksymalnie o rok), dlatego też, aby uniknąć niedoszacowania śmiertelności, należy dokonać dodatkowych założeń. Uśredniając rozkład urodzeń przyjąć należy, że każda osoba ma $x+0{,}5$ lat.

W tablicach trwania życia znajdują się wartości q_x, czyli prawdopodobieństwa, że osoba w wieku x umrze przed osiągnięciem wieku $x+1$. Przy założeniu, że każda osoba jest w wieku $x+0{,}5$ lat do

oszacowania oczekiwanej liczby zgonów w roku potrzebna jest wartość q_x+0,5. Jeśli przyjąć, że gęstość zmiennej T jest stała w przedziałach $[k,k+1)$ dla $k \in Z$, $q_{x+0,5}$ można przedstawić jako funkcję q_x i q_{x+1}:

$$q_x = \mathbb{P}(T < x+1 | T \geq x)$$
$$q_{x+1} = \mathbb{P}(T < x+2 | T \geq x+1)$$

$$q_{x+0,5} = \mathbb{P}(T < x+1,5 | T \geq x+0,5)$$
$$= \frac{\mathbb{P}(T < x+1,5 \wedge T \geq x+0,5)}{\mathbb{P}(T \geq x+0,5)}$$
$$= \frac{\mathbb{P}(x+0,5 \leq T < x+1,5)}{\mathbb{P}(T \geq x+0,5)}$$

$$\mathbb{P}(T \geq x+0,5) = \mathbb{P}(T \geq x+0,5 | T \geq x) \cdot \mathbb{P}(T \geq x)$$

$$\mathbb{P}(x+0,5 \leq T < x+1,5) = \mathbb{P}(T \geq x+0,5) - \mathbb{P}(T \geq x+1,5)$$

$$\mathbb{P}(T \geq x+1,5) = \mathbb{P}(T \geq x+1 | T \geq x) \cdot \mathbb{P}(T \geq x+1,5 | T \geq x+1) \cdot \mathbb{P}(T \geq x)$$

$$q_{x+0,5} = \frac{\mathbb{P}(x+0,5 \leq T < x+1,5)}{\mathbb{P}(T \geq x+0,5)}$$
$$= \frac{\mathbb{P}(T \geq x+0,5) - \mathbb{P}(T \geq x+1,5)}{\mathbb{P}(T \geq x+0,5)}$$
$$= 1 - \frac{\mathbb{P}(T \geq x+1,5)}{\mathbb{P}(T \geq x+0,5)}$$
$$= 1 - \frac{\mathbb{P}(T \geq x+1 | T \geq x) \cdot \mathbb{P}(T \geq x+1,5 | T \geq x+1) \cdot \mathbb{P}(T \geq x)}{\mathbb{P}(T \geq x+0,5 | T \geq x) \cdot \mathbb{P}(T \geq x)}$$
$$= 1 - \frac{(1-q_x)\left(1 - \frac{q_{x+1}}{2}\right)}{1 - \frac{q_x}{2}}$$

Oczekiwana liczba zgonów w roku $t+1$ będzie sumą oczekiwanej liczby zgonów dla wszystkich roczników wieku, czyli:

$$\mathbb{E}Z = \sum_{x=0}^{wiek\ max} l_x \cdot \left(1 - \frac{(1-q_x)\left(1 - \frac{q_{x+1}}{2}\right)}{1 - \frac{q_x}{2}}\right)$$

gdzie:

l_x – liczba osób, które 31.12.2020 r. miały ukończone x lat.

Polski Instytut Ekonomiczny

Polski Instytut Ekonomiczny to publiczny *think tank* gospodarczy, którego historia sięga 1928 roku. Obszary badawcze Polskiego Instytutu Ekonomicznego to przede wszystkim handel zagraniczny, makroekonomia, energetyka i gospodarka cyfrowa oraz analizy strategiczne dotyczące kluczowych obszarów życia społecznego i publicznego Polski. Instytut zajmuje się dostarczaniem analiz i ekspertyz do realizacji Strategii na Rzecz Odpowiedzialnego Rozwoju, a także popularyzacją polskich badań naukowych z zakresu nauk ekonomicznych i społecznych w kraju oraz za granicą.

Polski
Instytut
Ekonomiczny