

# Oczekiwana długość życia a wskaźnik porodów wśród nastolatków w wybranych państwach na podstawie Rocznego Raportu WHO 2022

Agata Juda, Oliwia Makuch

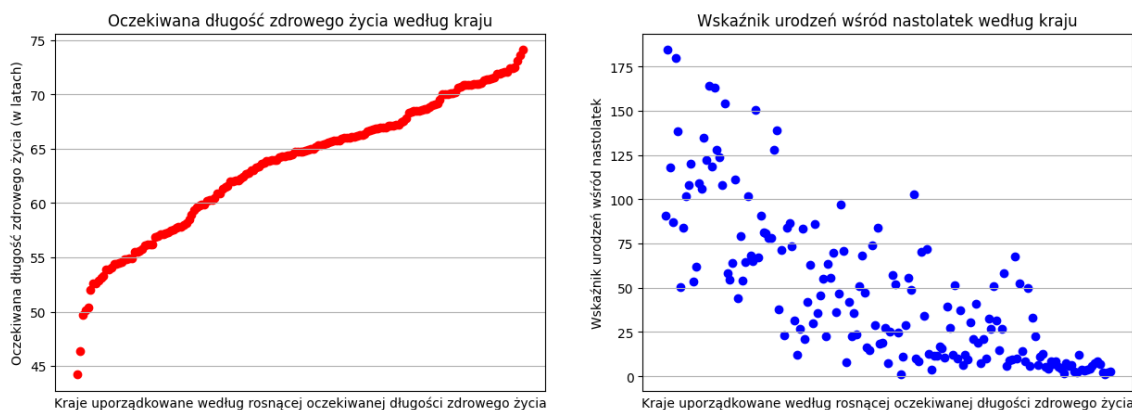
5 maja 2024

## 1 Wstęp

Z roku na rok medycyna rozwija się coraz szybciej, a wraz z tym nasza świadomość tego, czy coś jest dla nas dobre. W tym raporcie, korzystając z danych z Oficjalnego Raportu Rocznego WHO z roku 2022<sup>[1]</sup>, zanalizujemy oraz zbadamy zależność między wskaźnikiem porodów na 1000 nastolatków w wieku 15-19 lat<sup>[2]</sup> oraz oczekiwaną długością życia w zdrowiu dla obu płci w chwili urodzenia podawaną w latach<sup>[3]</sup>. Dane zostały podzielone przez WHO według państw uznawanych przez Organizację Narodów Zjednoczonych. Z porównania wyłączyliśmy państwa, w przypadku których jedna lub obie dane były nieznane. Celem naszej pracy jest sprawdzenie jak duży wpływ na oczekiwaną długość życia w zdrowiu ma częstotliwość zachodzenia w ciążę wśród nastolatków. Już z samych wykresów danych można zauważyć pewną korelację danych, ponieważ nastoletnie ciążę mogą powodować większe ryzyko chorób u dziecka, jak i zagrożenie dla zdrowia matki.

Dane dotyczące oczekiwanej długości zdrowego życia oraz wskaźnika porodów wśród nastolatków uporządkowane względem rosnącej oczekiwanej długości zdrowego życia zostały przedstawione na Wykresie 1. Na Wykresie 2 natomiast przedstawiona została zależność pomiędzy danymi.

Ten problem i jego pochodne były już poruszane m.in. w publikacjach organizacji takich jak The World Bank<sup>[4]</sup>, The Lancet<sup>[5]</sup>, United Nations Department of Economic and Social Affairs<sup>[6]</sup>, BMJ Global Health<sup>[7]</sup> czy National Library of Medicine<sup>[8]</sup>.



Wykres 1: Wykresy oczekiwanej długości zdrowego życia w zdrowiu dla obu płci w chwili urodzenia (po lewej) oraz wskaźnika porodów wśród nastolatków (po prawej) dla wybranych państw. Państwa uporządkowane według rosnącej oczekiwanej długości życia



Wykres 2: Wykres zależności pomiędzy oczekiwaną długością zdrowego życia a wskaźnikiem urodzeń wśród nastolatków

## 2 Podstawowe statystyki

Do obliczenia statystyk korzystamy z funkcji wbudowanych python oraz funkcji bibliotek *numpy* i *scipy*. Długość próby dla obu statystyk jest ta sama i wynosi 173 (liczba rozważanych państw).

### 2.1 Statystyki dla oczekiwanej długości zdrowego życia

Statystyka	Wartość
Średnia arytmetyczna	63,50
Średnia geometryczna	63,18
Średnia harmoniczna	62,85
Średnia winsorowska (dla $k = 17$ )	63,66
Mediana	64,70
Współczynnik skośności	-0,55
Kwartyl pierwszy	58,5
Kwartyl trzeci	68,4
Rozstęp międzykwartylowy	9,9
Rozstęp z próby	29,9
Wariancja	38,73
Odchylenie standardowe	6,22
Kurtoza	-0,37

Tabela 1: Podstawowe statystyki dla oczekiwanej długości zdrowego życia

Jedną z podstawowych statystyk są średnie. Średnie obliczymy z następujących wzorów:

- średnia arytmetyczna:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- średnia geometryczna:

$$\text{śr}_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

- średnia harmoniczna:

$$\text{śr}_h = n \left( \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

- średnia winsorowska:

$$\bar{x}_{win} = \frac{1}{n} \left[ (k+1) x_{(k+1)} + \sum_{i=k+1}^{n-k-1} x_{(i)} + (k+1) \cdot x_{(n-k)} \right]$$

gdzie  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  - uporządkowane elementy próby.

W Tabeli 1 przedstawione zostały m.in. wyliczone średnie dla oczekiwanej długości życia w zdrowiu. Wyniki zostały podane w zaokrągleniu do dwóch liczb po przecinku.

Analizując te wyniki możemy dojść do kilku wniosków. Każda średnia oscyluje wokół liczby 63. Średnia winsorowska (po odrzuceniu 10% wartości brzegowych) i średnia geometryczna (która jest mniej wrażliwa na skrajne wartości) są zbliżone do średniej arytmetycznej, co wskazuje na to, że eliminacja wartości skrajnych nie ma dużego wpływu na centralną tendencję danych. Średnia harmoniczna natomiast, bardziej wrażliwa na wartości niskie, jest najniższa, co oznacza, że wśród danych są państwa o oczekiwaniach znacząco niższych od średniej.

Medianę dla próby liczymy ze wzoru:

$$x_{med} = \begin{cases} x_{((n+1)/2)}, & \text{gdy } n \text{ jest parzyste} \\ \frac{1}{2} (x_{(n/2)} + x_{(n/2+1)}), & \text{gdy } n \text{ jest nieparzyste} \end{cases}$$

Dla długości życia mediana wynosi 64,7. Ponieważ  $\bar{x} < x_{med}$ , próba jest skośna w lewo, lecz nie jest to znacząca różnica, co oznacza, że próbie blisko do symetrii. Fakt ten potwierdzić można, korzystając ze współczynnika skośności, obliczanego według wzoru:

$$\alpha = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}}$$

Dla tego przewidywania  $\alpha$  jest równe w przybliżeniu -0,55, zatem rzeczywiście pojawia się niewielka lewoskośność.

Kwartyle pierwszy (Q1) oraz trzeci (Q3) to kolejno mediana obserwacji mniejszych od  $x_{med}$  i mediana obserwacji większych od  $x_{med}$ . Q1 próby wynosi 58,5, a Q3 - 68,4, co oznacza że rozstęp międzykwartylowy wyrażany wzorem:

$$IQR = Q3 - Q1$$

jest równy 9,9. Rozstęp ten oznacza że, połowa danych zamyka się w przedziale o tej długości.

Rozstęp z próby obliczany jest ze wzoru:

$$R = x_n - x_1$$

i oznacza różnicę między największym a najmniejszym wynikiem próby. Dla wartości oczekiwanej długości zdrowego życia z naszych danych rozstęp wynosi w przybliżeniu 29,9 lat.

Wariancja została obliczona ze wzoru:

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

i w przybliżeniu wynosi 38,73, a odchylenie standardowe, liczone ze wzoru:

$$\sigma = \sqrt{\text{Var}(X)}$$

wynosi około 6,22. Z różnicy między tymi dwiema wartościami możemy wysunąć wniosek o sporym rozproszeniu danych na świecie.

Jako ostatnią obliczamy kurtozę, czyli miarę spłaszczenia. Liczona jest ze wzoru

$$K = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma^4} - 3$$

Miara ta dla próby wynosi w zaokrągleniu -0,37, co oznacza, że próba ma rozkład platykurtyczny, tzn. jej ogony są "lżejsze" niż rozkładu normalnego - ekstremalne wartości są mniej prawdopodobne niż w rozkładzie normalnym.

Podstawowe statystyki dla oczekiwanej długości zdrowego życia zostały zebrane w Tabeli 1.

## 2.2 Statystyki dla wskaźnika porodów wśród nastolatków

Podobna analiza została przeprowadzona dla wskaźnika porodów na 1000 nastolatków w wieku 15-19 lat. Wzór na obliczenie jego wartości został podany na stronie WHO i jest postaci:

$$\text{Wskaźnik porodów} = \frac{\text{Liczba porodów wśród kobiet w wieku 15-19 lat}}{\text{Populacja kobiet w wieku 15-19 lat w połowie roku}}$$

Średnie wartości dla wskaźnika porodów wśród nastolatków zostały zaokrąglone do dwóch miejsc po przecinku i zebrane w Tabeli 2.

Statystyka	Wartość
Średnia arytmetyczna	47,19
Średnia geometryczna	27,29
Średnia harmoniczna	12,20
Średnia winsorowska (dla $k = 17$ )	44,64
Mediana	34,0
Współczynnik skośności	1,06
Kwartył pierwszy	10,8
Kwartył trzeci	71,0
Rozstęp międzykwartyłowy	60,2
Rozstęp z próby	183,5
Wariancja	1828,67
Odchylenie standardowe	42,76
Kurtoza	0,5

Tabela 2: Podstawowe statystyki dla wskaźnika porodów wśród nastolatków

Tak jak przy latach przyjmijmy średnią arytmetyczną za odnośnik. Średnia winsorowska jest bardzo zbliżona co oznacza że skrajne wartości są małe lub wystarczająco się równoważą. Biorąc jednak pod uwagę średnią geometryczną zauważamy, iż wartości skrajne mają duży wpływ, gdyż odbiega ona znacznie od dwóch wcześniej porównanych. Najbardziej odbiega jednak średnia harmoniczna, która wskazuje nam na to, że w zestawieniu występują znacząco mniejsze od średniej arytmetycznej ilości ciąż na 1000 nastolatków w wieku 15-19, gdyż średnia ta, jak opisałyśmy wyżej, jest mocno wrażliwa na niskie wartości.

Mediana dla tej próby wyniosła 34,0. Ponieważ  $\bar{x} > x_{med}$ , w przypadku wskaźnika porodów mamy do czynienia z rozkładem skośnym w prawo i bardziej asymetrycznym niż predykcja długości życia. W celu potwierdzenia, obliczony został też współczynnik skośności. Jest on równy 1,06, co potwierdza, iż rozkład jest prawoskośny.

Większe zróżnicowanie możemy też zauważyć patrząc na kwartyle, ponieważ  $Q1=10,8$  i  $Q3=71,0$ , co oznacza że  $IQR=60,2$ . Oznacza to że dane są mocno rozrzucone oraz często wstępują wartości skrajne, co można zauważyć na Wykresie 1. Rozstęp z próby również jest bardzo duży, ponieważ między najniższym a najwyższym współczynnikiem jest różnica aż 183,5. Oznacza to, że pomiędzy krajem, w którym zjawisko nastoletniej ciąży jest najczęstsze oraz krajem, w którym jest ono najrzadsze, jest aż 18 punktów procentowych różnicy w statystyce opisującej część kobiet zostających matkami w wieku nastoletnim.

Wariancja dla próby to w przybliżeniu aż 1828,67, co wynika z przypadków mocno wyższych niż średnia arytmetyczna co można zauważyć na Wykresie 1. Powoduje to że odchylenie standardowe to aż 42,76. Potwierdza to tym bardziej informację o dużym rozproszeniu na całym świecie ilości nastoletnich ciąż.

Miara spłaszczenia (kurtoza) dla próby wynosi około 0,5, co oznacza że ogony są "cięższe" niż w rozkładzie normalnym - ekstremalne wartości są bardziej prawdopodobne.

## 2.3 Korelacja oczekiwanej długości zdrowego życia i wskaźnika porodów wśród nastolatek

Dla pełnej analizy postanowiliśmy obliczyć dwie różne korelacje. Pierwszą z nich jest podstawowa korelacja, czyli korelacja Pearsona liczona ze wzoru:

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}X} \cdot \sqrt{\text{Var}Y}} = \frac{\sum x_i y_i - n \cdot \bar{X} \cdot \bar{Y}}{(n-1) \cdot \sigma_x \cdot \sigma_y}$$

Dla naszych zmiennych współczynnik ten jest równy około -0,8. Inną korelacją jest korelacja Spearmana, czyli współczynnik Pearsona liniowy dla rangi zmiennych zamiast surowych wartości, gdzie ranga jest równa pozycji w rosnącym porządku. Liczony jest on ze wzoru:

$$r_i = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ gdzie } d_i = R_{x_i} - R_{y_i}, \text{ czyli różnica rang.}$$

Wartość współczynnik korelacji Spearmana jest bardzo zbliżona, bo wynosi -0,82. Z analizy wynika zatem, że zmienne są mocno skorelowane tak, że gdy jedna z nich rośnie - druga maleje.

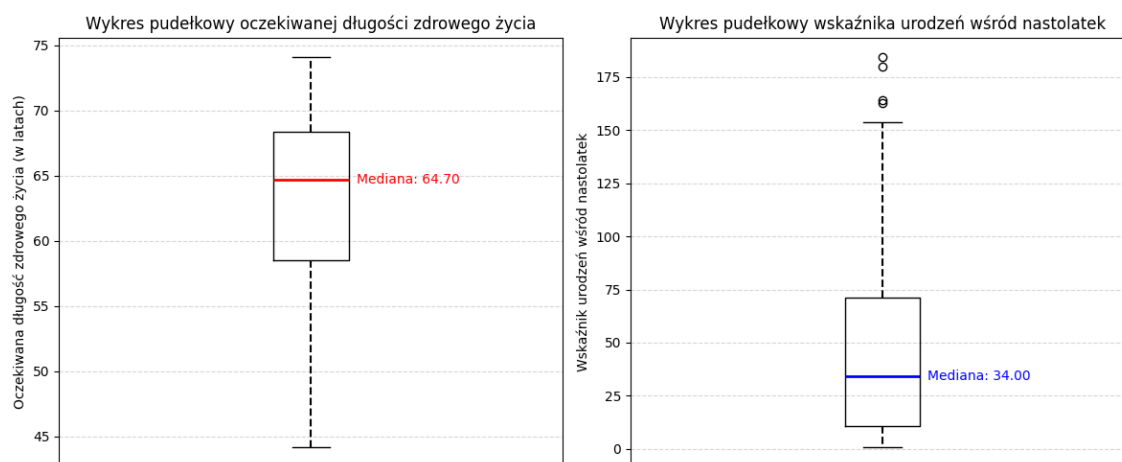
## 3 Wizualizacja danych

Aby zwizualizować dane z raportu WHO przedstawiliśmy oczekiwane długości zdrowego życia oraz wskaźniki porodów wśród nastolatek na wykresach pudełkowych (Wykres 3).

Z wykresów odczytać można medianę zestawów danych oraz zauważyć wyliczone wcześniej lewoskośność (dla oczekiwanej długości życia) oraz prawoskośność (dla wskaźnika urodzeń wśród nastolatek). Ciekawą obserwacją jest również to, że w przypadku oczekiwanej długości zdrowego życia wszystkie wartości mieszczą się na przedziale pomiędzy "wysięgnikami", tj. żadne z danych nie są szczególnie odstające. W przypadku wskaźnika urodzeń wśród nastolatek pojawiają się wartości znacząco większe od średniej, co wpływa na bardzo duże odchylenie standardowe zestawu danych. W praktyce oznacza to, że istnieją państwa, w których liczba nastolatek zachodzących w ciążę jest znacznie wyższa niż w pozostałych. Z analizy arkusza z Raportem Rocznym WHO wynika, że są to m.in. Republika Środkowoafrykańska (wskaźnik urodzeń na poziomie 184,4), Mozambik (180,0) i Mali (164,0).

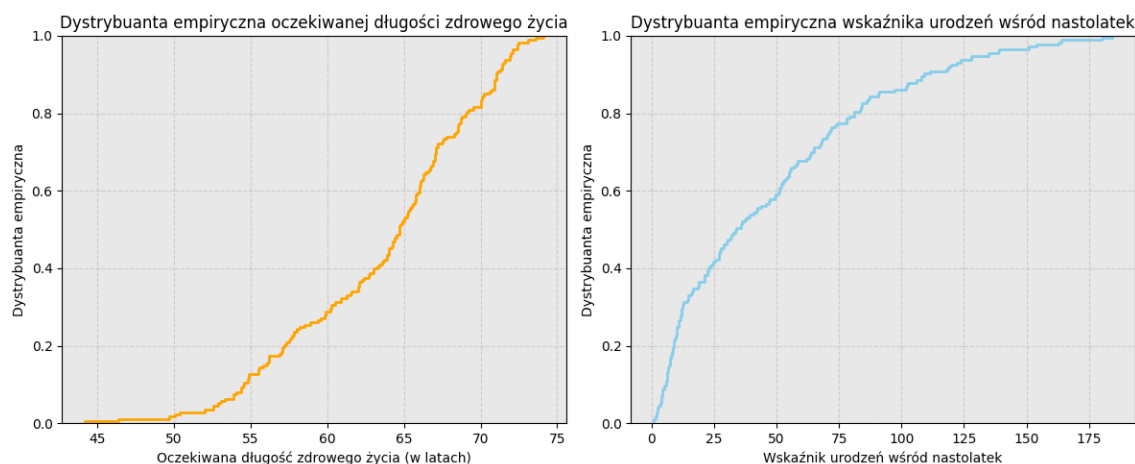
Na Wykresie 4 przedstawiono dystrybuanty empiryczne oczekiwanej długości zdrowego życia oraz wskaźnika urodzeń wśród nastolatek. Z definicji dystrybuanty empirycznej:

$$F(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$$



Wykres 3: Wykresy pudełkowe oczekiwanej długości zdrowego życia w zdrowiu dla obu płci w chwili urodzenia (po lewej) oraz wskaźnika porodów wśród nastolatek (po prawej)

oznacza to, że funkcja rośnie w najszybszym tempie wtedy, gdy w danym przedziale wartości koncentruje się większa liczba obserwacji. Możemy zauważyć, że dystrybuanta oczekiwanej długości zdrowego życia rośnie najszybciej w okolicy średniej, natomiast przy krańcach przedziału wypłaszcza się. Wypłaszczenie jest bardziej widoczne dla małych wartości, jednak najszybszy wzrost funkcji zachodzi dla wartości większych od mediany. Tłumaczy to niewielką lewoskośność funkcji mimo tego, że średnia winsorowska jest nieznacznie większa od arytmetycznej. Wartości skrajne i odstające nie mają tak dużego wpływu na średnią, jak np. w przypadku wskaźnika urodzeń. Wykres drugiej dystrybuanty jest wyraźnie wklęsły, z czego wynika, że dane są znacznie bardziej rozproszone i niekonsekwentne. Możemy zauważyć także znacznie bardziej wyraźne wypłaszczanie się wykresu dla większych wartości - wykres wypłaszcza się na stosunkowo długim fragmencie, z czego wynika występowanie odstających danych.



Wykres 4: Wykresy dystrybant empirycznych oczekiwanej długości zdrowego życia w zdrowiu dla obu płci w chwili urodzenia (po lewej) oraz wskaźnika porodów wśród nastolatek (po prawej)

Podobne wnioski wyciągnąć można na podstawie wykresów gęstości empirycznych oraz unormowanych histogramów dla obu danych (Wykres 5). Po wykresach gęstości widać wyraźnie, że dane dotyczące oczekiwanej długości zdrowego życia są bardziej skoncentrowane niż te dotyczące wskaźnika urodzeń.



Wykres 5: Wykresy gęstości empirycznych oraz unormowanych histogramów oczekiwanej długości zdrowego życia w zdrowiu dla obu płci w chwili urodzenia (po lewej) oraz wskaźnika porodów wśród nastolatek (po prawej)

## 4 Podsumowanie

Analiza danych dotyczących oczekiwanej długości zdrowego życia oraz wskaźnika porodów wśród nastolatek przyniosła kilka istotnych wniosków. Pierwszym spostrzeżeniem jest silna negatywna korelacja między tymi dwiema zmiennymi. Oznacza to, że kraje, w których częstość ciąży wśród nastolatek jest wysoka, charakteryzują się zazwyczaj niższą oczekiwaną długością życia.

Analiza statystyczna pokazała, że średnia oczekiwana długość życia w zdrowiu dla obu płci wynosi około 63 lat, z odchyleniem standardowym wynoszącym 6,22. Natomiast średni wskaźnik porodów wśród nastolatek wynosi około 47 porodów na 1000 kobiet w wieku 15-19 lat, z bardzo dużym odchyleniem standardowym wynoszącym 42,76. To sugeruje, że dane dotyczące wskaźnika porodów są bardziej rozproszone i mniej stabilne niż dane dotyczące oczekiwanej długości życia.

Dodatkowo, analiza wykresów pudełkowych, dystrybuant empirycznych oraz gęstości empirycznych potwierdziła obserwacje wynikające z analizy statystycznej. Wykresy te pokazały, że dane dotyczące oczekiwanej długości życia są bardziej skoncentrowane wokół średniej wartości, podczas gdy dane dotyczące wskaźnika porodów wykazują dużą zmienność i obecność wartości odstających.

Wnioski te są zgodne z wynikami analiz przeprowadzonych we wspomnianych publikacjach. W 65-stronowym raporcie "Płodność nastolatków od czasu Międzynarodowej Konferencji na temat Ludności i Rozwoju (ICPD) w Kairze" [6] opublikowanym przez ONZ możemy znaleźć następującą konkluzję:

**"Istnieją duże obawy dotyczące rodzenia dzieci przez nastoletnie kobiety i jego związku ze zwiększonym ryzykiem dla zdrowia matek i ich dzieci, niższym poziomem wykształcenia i utrzymującym się ubóstwem wśród kobiet, które zostają matkami we wczesnym okresie dojrzewania"** ("There is much concern about adolescent childbearing and its association with heightened health risks for mothers and their infants and lower educational attainment and persistent poverty among women who become mothers during early adolescence")

Jest to wstępna analiza, która otwiera drogę do dalszych badań nad tym złożonym problemem. Istnieje wiele innych czynników, takich jak dostęp do opieki medycznej czy edukacja seksualna, które mogą wpływać zarówno na wskaźnik porodów wśród nastolatek, jak i oczekiwaną długość życia. Zbadanie tych czynników bardziej szczegółowo może dostarczyć głębszego zrozumienia złożonej dynamiki między nimi oraz pomóc w identyfikacji skutecznych strategii działania.

## Literatura

- [1] [World health statistics 2022: monitoring health for the SDGs, sustainable development goals](#), World Health Organisation; Data, Analytics & Delivery for impact (DDI), Maj 2022.
- [2] [Adolescent birth rate \(per 1000 women\)](#), World Health Organisation.
- [3] [Healthy life expectancy at birth \(years\)](#), World Health Organisation.
- [4] [The Social and Educational Consequences of Adolescent Childbearing](#), The World Bank, Luty 2022.
- [5] [Maternal mortality in adolescents compared with women of other ages: evidence from 144 countries](#), Dr Andrea Nove, PhD, Prof Zoë Matthews, PhD, Sarah Neal, PhD, Alma Virginia Camacho, MD; Styczeń 2014.
- [6] [Adolescent Fertility since the International Conference on Population and Development \(ICPD\) in Cairo](#), United Nations; Grudzień 2013.
- [7] [The Effect of Adolescent Pregnancy on Child Mortality in 46 Low- and Middle-Income Countries](#), Navideh Noori, Joshua L Proctor, Yvette Efevbera, Assaf P Oron; Styczeń 2023.
- [8] [Adolescent fertility and risky environments: a population-level perspective across the lifespan](#), Caitlyn D. Placek and Robert J. Quinlan; Lipiec 2012.