

Raport wykonany na potrzeby przedmiotu:  
Mikroekonometria

*Determinanty przeżycia choroby COVID-19*

Opracowali:

Oskar Furmańczuk 81794

Mateusz Koch 72158

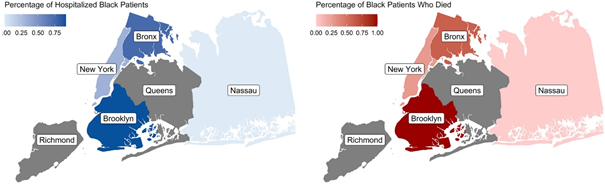
*Warszawa 2020*

**Wprowadzenie**

Rok 2020 z pewnością zostanie zapisany na kartach historii jako rok rozpoczęcia jednej z największych pandemii jaka dotknęła naszą cywilizację. Wirus SARS-CoV-2, wywołujący chorobę COVID-19, pozbawił życia milionów ludzi w każdym zakątku naszej planety. Jak wskazują ostatnie informacje nie istnieje grupa etniczna, rasa, płeć ani grupa wiekowa, która byłaby zupełnie odporna na działanie koronawirusa. Zestawienie tych cech może natomiast dać pewne przybliżenie szans na przetrwanie choroby COVID-19.

**Rasa jako determinanta przebiegu choroby COVID-19 u pacjentów z wykrytym nowotworem**

Jednym z projektów prowadzonych na rzecz opisu przebiegu COVID-19 u pacjentów pochodzących z różnych grup rasowych jest „Racial disparities in patients with coronavirus disease 2019 infection and gynecologic malignancy”. Badanie to zostało przeprowadzone w Stanach Zjednoczonych na podstawie danych zebranych przez nowojorskie szpitale. Grupę badanych stanowili pacjenci powyżej osiemnastego roku życia u których stwierdzono nowotwór ginekologiczny, a także w okresie terapii onkologicznego wykryto SARS-CoV-2. Pacjentów podzielono na odpowiadające im grupy rasowe lecz ze względu na bardzo wysoki procentowy udział białych i czarnych pacjentów do zbiorczego zestawienia przyjęto 3 grupy rasowe: biała, czarna i inna.



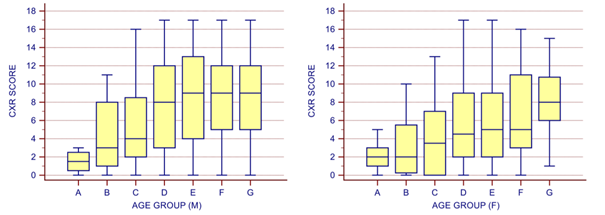
*Rysunek 1. Procent hospitalizowanych oraz procent zgonów wśród czarnych pacjentów*

Rezultatem badań był wniosek o istotnej przewadze ilości ciężkich przypadków u pacjentów rasy czarnej. Pomimo, iż stanowili oni zaledwie 1/3 badanej grupy to tworzyli ponad 45% hospitalizowanych oraz ponad 40% śmiertelnych przypadków. Ponadto, młodzi czarni pacjenci byli 5 razy bardziej narażeni na hospitalizację niż pozostałe osoby w tej samej kategorii wiekowej w badanej grupie.

Źródłem relatywnie cięższego przebiegu COVID-19 u czarnych pacjentów może być między innymi większa ilość współistniejących chorób. Te z kolei, zdaniem autorów, miałby być pośrednio rezultatem niskich średnich dochodów wśród czarnych społeczności oraz wynikiem rasizmu systemowego.[[1]](#footnote-2)

**Wpływ wieku i płci na przebieg COVID-19**

Jedno z badań wpływu wieku oraz płci pacjentów na przebieg choroby wywołanej SARS-CoV-2 zostało przeprowadzone w pierwszym kwartale 2020 roku w mieście Brescia (Włochy). W projekcie brało udział 786 pacjentów z czego 523 stanowili mężczyźni a 251 kobiety. Wszystkim badanym wykonano badanie rentgenowskiego klatki piersiowej, a następnie każdemu z pacjentów przyznano odpowiedni wynik w skali CXR (system oceny zmian płucnych). Następnie zestawiono wyniki różnych kategorii wiekowych z uwzględnieniem płci pacjentów. Zmienna „płeć” przyjęła dwie wartości: „M” dla mężczyzn oraz „F” dla kobiet. Grupy wiekowe opisano literami od A do G, gdzie A to najmłodsza grupa (20-29 lat), a G najstarsza (powyżej 80 lat).

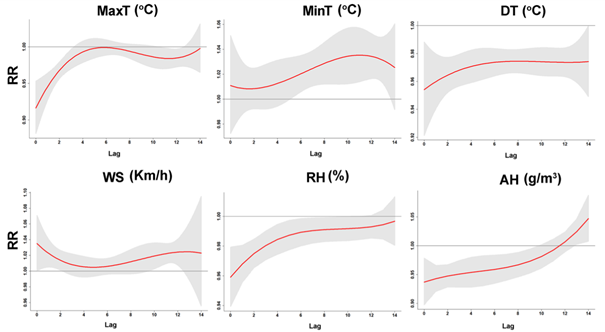
*Rysunek 2. Wskaźnik CXR dla mężczyzn (lewa strona) i kobiet (prawa strona) w różnych grupach wiekowych*

Rozpatrując wyniki pomiędzy kategoriami wiekowymi obserwowalna jest silna dodatnia korelacja pomiędzy stopniem uszkodzenia płuc a wiekiem. Wniosek ten jest zgodny z danymi o przypadkach COVID-19, które systematycznie dostarczane są przez władzę krajów Unii Europejskiej (w Polsce szansa przeżycia osoby po 75 rok roku życia jest ponad 10 razy mniejsza niż u osoby przed 35 rokiem życia).

Kolejnym wnioskiem z analizy wyników jest obserwacja znacząco większych zmian płucnych u mężczyzn z najstarszych grup wiekowych (powyżej 50 lat) w zestawieniu z wynikami kobiet z tego samego przedziału wiekowego. Natomiast w grupie kobiet największy spadek wskaźnika CXR miał miejsce dla pacjentek powyżej 80 roku życia. [[2]](#footnote-3)

**Pora roku oraz zmienne** **meteorologiczne te jako determinanty częstotliwości zachorowań na COVID-19**

Opublikowany w wrześniu 2020 artykuł „Effect of meteorological factors on COVID-19 cases in Bangladesh” koncertuje się na opisie wpływu warunków pogodowych na możliwość zarażenia się koronawirusem. Badanie na temat którego powstał wspomniany artykuł powstał na terytorium Bangladeszu. Do jego przeprowadzenia zebrano dane z 43 stacji meteorologicznych z których wyodrębniono informacje na temat kilkunastu parametrów atmosferycznych (maksymalna temperatura dobowa (MaxT), minimalna temperatura dobowa (MinT), siła wiatru (WS), wilgotność relatywna (RH), wilgotność absolutna (AH) itp.) dla odpowiadającym im regionów. Na podstawie danych o przypadkach zachorowani na COVID-19 z odpowiadających im obszarów ustalono wskaźnik relatywnego ryzyka (RR), który oceniał prawdopodobieństwo zarażenia się SARS-CoV-2 w określonym dniu.

*Rysunek 3. Wskaźnik RR dla wybranych zmiennych meteorologicznych (opóźnienie 14 dniowe)*

Wyniki wykazały, że minimalna (MinT) i średnia temperatura, prędkość wiatru (WS), wilgotność względna (RH) i wilgotność bezwzględna (AH) miały znaczącą dodatnią korelację z liczbą potwierdzonych przypadków COVID-19. Najwyższe wartości RR odnotowano m.in. dla wartości MinT = 18 °C, WS = 21 km/h oraz RH = 46%, AH = 23 g/m3. Ogólnym wnioskiem płynącym z badania jest fakt o głębokim wpływie czynników meteorologicznych na transmisję SARS-CoV-2. [[3]](#footnote-4)

Głęboką zależność pomiędzy czynnikami meteorologicznymi a liczbą zachorowani na COVID-19 można wnioskować również biorąc pod uwagę dane z terenu Polski. Podczas okresu letniego, z przeważającą ilością dni z temperaturą maksymalną powyżej 20 °C, odnotowanych zostało znacznie mniej nowych dziennych zachorowani na COVID-19 w porównaniu z okresem jesienno-zimowym.[[4]](#footnote-5)

**Analiza danych**

Dane użyte w dalszej części raportu pochodzą z serwisu internetowego stworzonego przez Centers for Disease Control and Prevention (CDC), instytucję publiczną wchodzącą w skład Departamentu Zdrowia Stanów Zjednoczonych. Zawieją one informację na temat wszystkich odnotowanych przypadków COVID-19 na terenie USA, w sumie ponad 8 milionów obserwacji. Dane w całości pochodzą z roku 2020.

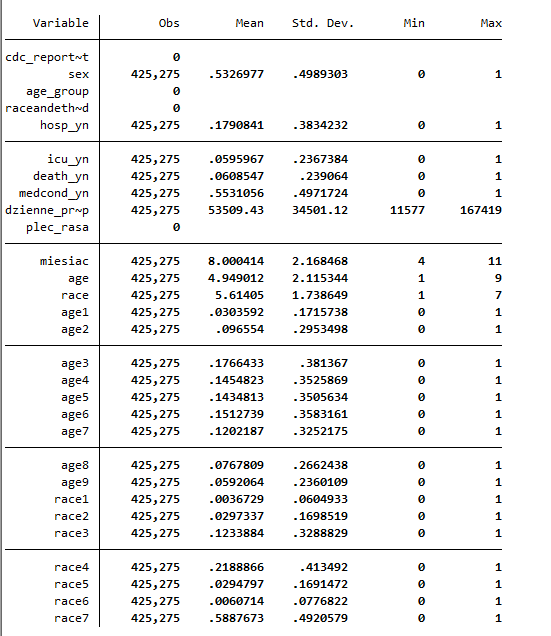
By zapewnić szybszą kalkulację obliczeń oraz klarowność w formułowaniu wniosków postanowiono odrzucić obserwacje posiadające brakujące/zagubione dane. W wyniku takiej eliminacji otrzymaliśmy dane dla ponad 425 tysięcy obserwacji. Taka ilość umożliwia prowadzenie dalszych analiz nad wybranym zbiorem danych.

Zgodnie z przytoczoną na wstępie literaturą wybrano 8 zmiennych objaśniających możliwy przebieg choroby COVID-19 (zakończenie wyzdrowieniem lub zgonem). Ich opis oraz przyjmowane wartości znajdują się w poniższej tabeli:

*Tabela 1. Wykorzystane zmienne wraz z ich opisem*



# *Tabela 2. Podsumowanie zmiennych*



Do zilustrowania informacji na temat pacjentów znajdujących się w badanej próbie wykorzystano poniższy wykres oraz tabele:

*Rysunek 4. Ilość pacjentów z wyróżnieniem płci w różnych grupach wiekowych*

*Tabela 3. Ilość pacjentów z wyróżnieniem płci w grupach rasowo-etnicznych*



Najliczniejszą grupę pacjentów stanowią osoby z białym kolorem skóry będące między 30 a 60 rokiem życia. Grupą etniczną, dla której uzyskano najmniej danych stanowią natywni amerykanie. Liczba kobiet i mężczyzn jest porównywalna, choć z przewagą liczby kobiet.

# Wybór modelu

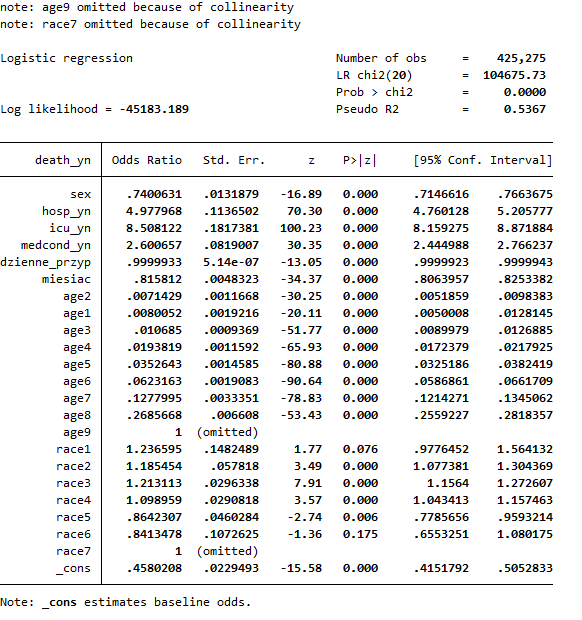
Ze względu na dwumianową zmienną objaśnianą odrzucone zostały modele liniowe   
w których wynik estymacji w takich przypadkach nie nadaje się dobrze do reprezentowania związku między zmienną objaśnianą a wielkością prawdopodobieństwa. Ze względu   
na minimalnie lepsze miary dopasowania w pierwszych (roboczych) dla modelu logitowego względem probitowego do dalszej estymacji oraz analizy wybrany został model logitowy.

# Prezentacja wyników i ich interpretacja

## Estymacja

Po przeprowadzonej estymacji uzyskano poniższe wyniki. Jak można zauważyć zmienne age9 oznaczająca przedział wiekowy 80+ lat oraz race7 oznaczający rasę: white/ non-hispanic zostały pominięte ze względu na współliniowość zmiennych. Wynika to z zmiany zmiennych jakościowych o wielu kategoriach na zmienne zero-jedynkowe. Gdy zmienna nie należy do żadnej z pierwszych kategorii to na pewno należy do ostatniej. Dlatego współliniowość w tym przypadku nie wpływa negatywnie na estymacje modelu.

Interpretacja wyników została przeprowadzona w późniejszym etapie. Najpierw została przeprowadzona analiza poprawności budowy modelu aby ocenić czy metoda estymacji jest poprawna oraz czy model w sposób zadawalający przewiduje wyniki.

*Tabela 4. Wynik estymacji modelu*

## Poprawność modelu

Pierwszą wartością jaką będziemy sprawdzać w modelu (na podstawie tabeli 4) jest pseudo R2 czyli dopasowanie zmiennych do modelu. Wartość 0,54 jest wartością dość wysoką jak na model mikroekonometryczny. Może to wynikać z dość szczegółowych danych jakie były użyte do wyjaśnienie zgonów na COVID-19. Wysoki R2 upewnia nas, pokazuje, że pod względem dopasowania model jest poprawny.

Istotność całego modelu możemy odczytać z zmiennej Prob >chi2. Na tej podstawie można określić, że występuje brak podstaw do odrzucenia hipotezy zerowej mówiącej o nieistotności modelu a co za tym idzie możemy uznać, że model jest istotny na poziomie 95%.

Analizując statystyczną istotność poszczególnych zmiennych zauważyć można, że wszystkie zmienne poza zmiennymi: race1 oraz race6 są istotne statystycznie na poziomie istotności 95%. Można przypuszczać, że brak istotności zmiennych race1 oraz race6 może wynikać z tego, że te rasy czyli kolejno: American Indian/Alaska Native, Non-Hispanic, Native Hawaiian/Other Pacific Islander, Non-Hispanic są bardzo małą populacją w próbie zaledwie 0,003% oraz 0,006%. Jednak ze względu, na niosące informacje oraz mały wpływ na wyniki estymacji (z powodu małej populacji w próbie) nie wyrzucamy tych zmiennych z modelu (mając na względzie ostatnie zalecenia American Statistical Association).

## Tablice trafności

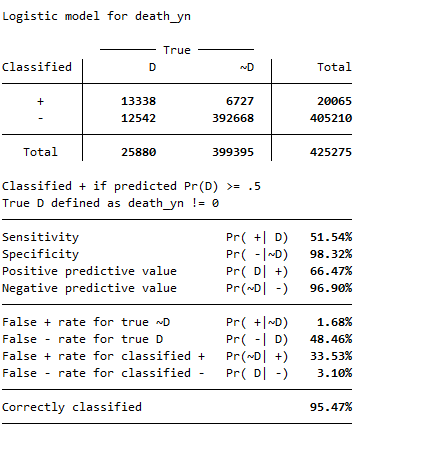
Ze względu na brak zbilansowania próby skonstruowano dwie tablice trafności. Jedną standardową (punkt odcięcia 0.5) (Tabela 5) a drugą według zasady Cramera (Tabela 6). Punkt odcięcia dla naszych danych wyniósł: 0.0608547. Pierwsza tablica poprawnie prognozuje około 95% wyników druga, ta utworzona przy pomocy metody Cramera, mniej i jest to około 90%. Jest dość duża różnica jednak jeszcze bardziej znaczące różnice pojawiają się przy jakości prognozowania poszczególnych stanów:

* prognoza ins=1 (sensitivity): 51% oraz 93%
* prognoza ins=0 (specificity): 98% oraz 89%.

Zauważyć można, że tablica wykorzystująca zasadę Cramera znacznie lepiej prognozuje zgony a tablica pierwsza nieznacznie lepiej prognozuje brak zgonu. Ze względu na specyfikę zmiennej zgon dla której poprawna estymacja dla wyniku pozytywnego (zgonu) jest ważniejsza niż w przypadku wyniku negatywnego (braku zgonu) należy uznać drugą tablicę za poprawniejszą w naszym przypadku.

Przyjmując za poprawniejszą drugą tablicę trafności prognozy możemy przyjąć, że nasz model poprawnie przewiduje 90% wyników. Jest to wynik zadowalający dla naszej prognozy dlatego możemy przejść do interpretacji modelu.

# *Tabela 5. Tablica trafności (standardowa)* *Tabela 6. Tablica trafności (metoda Cramera)*



## Interpretacja modelu

Po upewnieniu się, że skonstruowany model jest poprawny, można przejść do interpretacji wyników modelu.

W programie stata wygenerowano efekty krańcowe (Tabela 7) dla poszczególnych zmiennych obliczone dla wartości średnich. Razem z ilorazem szans (Tabela 4) posłużą one   
do interpretacji otrzymanych wyników. Wszystkie interpretacje w swym założeniu mówią osobach mających identyczne wartości zmiennych objaśniających za wyjątkiem tej jednej, która objaśniamy (zasada ceteris paribus).

Dane pokazują, że mężczyźni mają o 26% większą szansę na zgon spowodowany Covid-19. Biorąc po uwagę badania z Brescii omówione w wstępie można przypuszczać,  
 że jest to spowodowane większym uszkodzeniem płuc u mężczyzn.

Hospitalizacja pacjenta jak i przebywanie na intensywnej terapii zwiększa o kolejno 0,016 i 0,035 (średnio) prawdopodobieństwo zgonu pacjenta. Jest to fakt zgodny  
 z oczekiwaniami, jednak szczególnie pobyt na intensywnej terapii, wiążę się z dużym wzrostem prawdopodobieństwa zgonu. Wywnioskować można, że pacjenci, którzy trafiają do szpitala, są w gorszym stanie niż ci do niego nie trafiający.

Często powtarzane w mediach choroby współistniejące zwiększają prawdopodobieństwo zgonu o 0,005. Jest do dość wysoka wartość, jednak wpływ tej zmiennej jest niższy niż np. wiek pacjenta lub jego hospitalizacja.

Dzienny przyrost zakażeń w dniu zarejestrowania pacjenta praktycznie nie ma wpływu na prawdopodobieństwo zgonu. Na tej podstawie można wnioskować, że służba zdrowia  
 w USA poprawnie radzi sobie z pandemią i nawet przy dużej ilości zakażeń jest w stanie pomóc wszystkim pacjentom. Z ogólnie dostępnych danych wiemy, że niestety nie wszystkie kraje poradziły sobie pod tym kątem poprawnie, szczególnie w pierwszej fazie rozwoju pandemii.

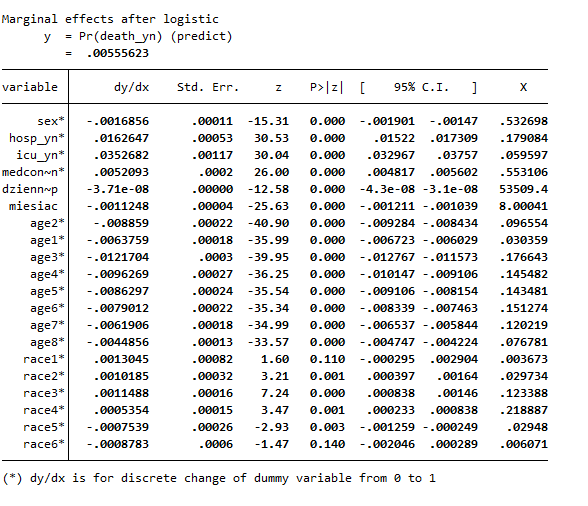
Zachorowanie w każdym kolejnym miesiącu trwania pandemii zmniejsza szanse na zgon o 27,5%. Prawdopodobnie jest to związane z tym, że z czasem system zdrowia lepiej zaczął radzić sobie z wirusem i wprowadzono kilka względnie skutecznych terapii. Jednak, może to również wiązać się z mniejszą ilością wykrytych pozytywnych zakażeń co sprawia, że wykrywane zaczęły być tylko te cięższe przypadki. Z trzeciej strony może to również wiązać się z mutacją sprawiającą mniejszą „zjadliwość” wirusa. Na podstawie zebranych danych nie jesteśmy w stanie, jednoznacznie zinterpretować tego wyniku, należałoby to zrobić w bardziej wyspecjalizowanych badaniach.

Zmienne age oraz race będziemy interpretować, z powodu specyfiki modelu, względem pominiętych zmiennych czyli age9 (wiek 80+) oraz race7 (rasa biała).

Model wykazał, że wraz z wiekiem szanse na zgon pacjentów rosną. Jest to zgodne   
z ogólnodostępnymi danymi. Największy skok szansy zgonu widać miedzy grupą age9 i age8. Pacjenci z grupy age8 mają około 73% mniejsze szanse na zgon w stosunku do grupy z age9. Potem kolejno od najstarszej grupy: 87%, 94%, 97%, 98%, 99%, 99,3%, 99,2%. Jak widać   
u osób młodych szansa na zgon w stosunku do osób najstarszych jest bardzo niska.

Wpływ rasy na szanse zgonu jest zauważalny. Szczególnie dużą różnicę można zauważyć dla ras American Indian, Asian i Black gdzie szansa zgonu jest o około 20% większa niż u osób rasy białej. Z kolei dla ras określonych jako Multiple/Other oraz Native Hawaiian szanse na zgon z powodu zakażenia są o około 15% mniejsze w stosunku do rasy białej. Jednak z powodu małej populacji tych ras w próbie należałoby przeprowadzić dokładniejsze badania. Zauważyć należy, że wyższa szansa zgonu w grupie pacjentów czarnych jest zgodna   
z badaniami: „Racial disparities in patients with coronavirus disease 2019 infection and gynecologic malignancy”.

*Tabela 7. Efekty krańcowe zmiennych*



**Zakończenie**

Przeprowadzona estymacja daje wiele informacji na temat zgonów z powodu Covid-19. Wyniki potwierdzają część wcześniejszych badań i niosą dodatkowe informacje.

Na podstawie wyników zalecić można, przeprowadzenie ściślejszych badań na temat wpływu śmiertelności Covid-19 ze względu na rasę. Model wykazał dość dużą rozbieżność  
 w śmiertelności w zależności od rasy. Mogła mieć na to wpływ niepoprawna próbka statystyczna jednak warto przeprowadzić dokładniejsze badania, szczególnie w grupie pacjentów o rasie Multiple/Other oraz Native Hawaiian. Dane pokazują, że mogą mieć oni większą odporność na wirusa co może pomóc odnalezieniu skutecznego leku na wirusa.

Dodatkową zależnością wartą zbadania jest powód, dla którego u pacjentów zakażonych wirusem w późniejszych miesiącach pandemii śmiertelność jest niższa niż u tych zakażonych wcześniej.

Pandemia Covid-19 pomimo wynalezienia szczepionki na pewno potrwa jeszcze wiele miesięcy. Dane zebrane podczas naszej analizy mogą posłużyć do zmniejszania ilości zgonów oraz powiększenia naszej wiedzy na temat samej choroby i jej przebiegu.

# Bibliografia

1. Olivia D. Lara MD Maria J. Smith BS Yuyan Wang PhD, *Racial disparities in patients with coronavirus disease 2019 infection and gynecologic malignancy,* American Cancer Society (2020)
2. Andrea Borghesi, Angelo Zigliani, Roberto Masciullo, Salvatore Golemi, *Radiographic severity index in COVID-19 pneumonia: relationship to age and sex in 783 Italian patients.* La radiologia medica (2020)
3. Abu Reza Md. Towfiqul Islam, Md. Hasanuzzaman, Md. Abul Kalam Azad*, Effect of meteorological factors on COVID-19 cases in Bangladesh,* Environment, Development and Sustainability (2020)
4. *Raport zakażeń koronawirusem SARS-CoV-2*, <https://www.gov.pl/web/koronawirus/>
5. Valentin Amrhein, Sander Greenland, Blake McShane, *Scientists rise up against statistical significance*. Nature (2019)

1. Olivia D. Lara MD Maria J. Smith BS Yuyan Wang PhD, *Racial disparities in patients with coronavirus disease 2019 infection and gynecologic malignancy,,* American Cancer Society (2020) [↑](#footnote-ref-2)
2. Andrea Borghesi, Angelo Zigliani, Roberto Masciullo, Salvatore Golemi, *Radiographic severity index in COVID-19 pneumonia: relationship to age and sex in 783 Italian patients.* La radiologia medica (2020) [↑](#footnote-ref-3)
3. Abu Reza Md. Towfiqul Islam, Md. Hasanuzzaman, Md. Abul Kalam Azad, *Effect of meteorological factors on COVID-19 cases in Bangladesh*, Environment, Development and Sustainability (2020) [↑](#footnote-ref-4)
4. *Raport zakażeń koronawirusem (SARS-CoV-2),* https://www.gov.pl/web/koronawirus/ [↑](#footnote-ref-5)