Supplementary Material 1. Comprehensive definitions of all-cause mortality, expected deaths, excess deaths, and negative excess deaths

Binomial distribution

A binomial distribution can be illustrated with a histogram. A histogram is a series of vertical bars that show the possible values of a probability distribution. The height of each bar represents the probability of each value occurring. This distribution is commonly used in biology to model various phenomena, particularly genetics, ecology, and epidemiology [1]. In epidemiology, the binomial distribution is frequently used to model the spread of diseases within a population. For instance, it can be used to estimate the probability of a certain number of individuals in a population contracting a disease within a specific time frame, given the probability of transmission per contact [2].

The binomial distribution is a versatile tool in biology for modeling random processes involving discrete outcomes, making it invaluable for understanding various biological phenomena and making predictions based on probabilistic reasoning [3]. The key parameters of the binomial distribution are:

n: The number of trials or experiments.
p: The probability of success on each trial or experiments.
k: The number of successes (the random variable).
The probability mass function (PMF) of the binomial distribution is given by:
$P(k;n,p) = (k^{n})p^{k}(1-p)^{n-k}$
$P(k;n,p)$ is the probability of having exactly k successes out of n trials or experiments,
(k^{n}) is the binomial coefficient, equal to n!/k!(n−k)!n!,
p^{k} is the probability of k successes,
$(1-p)^{n-k}$ is the probability of $n-k$ failures.

Correlation coefficient

A correlation coefficient is a statistical method that measures the linear relationship between 2 variables. The correlation coefficient value ranges from -1 to 1. A value of zero means that there is no correlation between x and y (variables). A value of 1 means a perfect correlation between them: when x goes up, y goes up perfectly linearly. This statistical measure helps researchers understand how changes in one variable relate to changes in another. Correlation coefficients play a crucial role in epidemiological studies by quantifying relationships between exposure to risk factors and the occurrence of diseases [4]. It helps epidemiologists identify potential causes or predictors of diseases and inform public health interventions.

In our study, we used Spearman's rank correlation coefficient, denoted by ρ (rho), which is a non-parametric measure of statistical dependence between 2 variables. It assesses the strength and direction of the monotonic relationship between the ranks of the data points of the 2 variables rather than their actual values [5].

Spearman's rank correlation coefficient:

$$\rho = 1 \frac{6 \times \epsilon d_{2}^{i}}{n(n^{2}-1)}$$

ρ is Spearman's rank correlation coefficient.
di represents the difference between the ranks of the corresponding data points.
n is the number of data points

https://doi.org/10.24171/j.phrp.2024.0186

All-causes mortality (ACM)

It is defined as the total mortality (across all reasons of death) occurring in a specified location within a certain period, e.g., year, month, or week. For country *c* and time period *t*, it is expressed by Yc,t [6].

Expected deaths

It is defined as the "counterfactual" or hypothetical total mortality numbers for a country *c* and time period *t* are represented by Ec,t and are forecasted using before the pandemic(historical) mortality data [6].

Excess deaths

It is defined as the disparity in death numbers when comparing pandemic ACM, Yc,t, to the expected, Ec,t, which for country *c* and time period *t*, are expressed as following: $\delta c, t = Yc, t - Ec, t$ [6]

Negative excess deaths

During the pandemic, we calculate ACM. When the pandemic ACM for country *c* over period *t* is inferior or lower than predicted or expected, it leads to a mortality depletion or deficit or "negative" excess and it is described as follows:

$$\delta c, t = Yc, t - Ec, t < 0$$

Where and when it occurs, it is generally due to declines in death numbers for explicit causes during the pandemic duration, e.g., lower-than-expected traffic-related mortality. Nonetheless, it can be an artifact of the model assumptions, e.g., an underestimating ACM or an overestimating expected [6].

Reference

1. Edwards AW. The meaning of binomial distribution. Nature 1960;186:1074.
2. Khan M, Olivier J. Regression to the mean for the bivariate binomial distribution. Stat Med 2019;38:2391–412.
3. DeVeaux M, Kane MJ, Zelterman D. Clinical trial design using a stopped negative binomial distribution. Stat Interface 2018;11:699–707.
4. Zhang J, Yang C, Jin Z, Li J. Dynamics analysis of SIR epidemic model with correlation coefficients and clustering coefficient in networks. J Theor Biol 2018;449:1–13.
5. Abbasi-Ghahramanloo A, Soltani-Kermanshahi M, Mansori K, et al. Comparison of SF-36 and WHOQoL-BREF in measuring quality of life in patients with type 2 diabetes. Int J Gen Med 2020;13:497–506.
6. World Health Organization (WHO) (CH). Global excess deaths associated with COVID-19 (modelled estimates) [Internet]. WHO; 2023 [cited 2024 Mar 11]. Available from: https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates

https://doi.org/10.24171/j.phrp.2024.0186