J Epidemiol Glob Health

https://doi.org/10.1007/s44197-026-00561-8

Article in Press

Understanding Mortality Data: A Step-by-Step Guide to CDC WONDER, Joinpoint Analysis, and Forecasting Models

Received: 17 January 2026

Accepted: 13 April 2026

Published online: 08 May 2026

Cite this article as: Shubietah A., Ruzieh M., Hamed B.M. *et al.* Understanding Mortality Data: A Step-by-Step Guide to CDC WONDER, Joinpoint Analysis, and Forecasting Models. *J Epidemiol Glob Health* (2026). https://doi.org/10.1007/s44197-026-00561-8

Abdalhakim Shubietah, Mohammed Ruzieh, Belal Mohamed Hamed, Sarah Saife, Mohamed Abuelazm, Mohamed S. Elgendy, Ahmed Emara, Ahmed Ibrahim, Ameer Awashra, Ibrahim Gowaily, Mohammed AbuBaha, Ghada Shbeitah, Elsayed Balbaa, Mohamed Saad Rakab & Mohammed Mhanna

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

REVIEW ARTICLE

Title.

Understanding Mortality Data: A Step-by-Step Guide to CDC WONDER, Joinpoint Analysis, and Forecasting Models

Running Title: *CDC WONDER Guide*

Authors. Abdalhakim Shubietah, MD [1]; Mohammed Ruzieh, MD, MSE [2]; Belal Mohamed Hamed, MD [3]; Sarah Saife, MD [4]; Mohamed Abuelazm, MBBCh [5]; Mohamed S. Elgendy, MBBCh [5]; Ahmed Emara, MD [3]; Ahmed Ibrahim, MBBCh [6]; Ameer Awashra, MD [4]; Ibrahim Gowaily, MD [7]; Mohammed AbuBaha, MD [4]; Ghada Shbeitah [8], Elsayed Balbaa, MBBCh [6]; Mohamed Saad Rakab, MBBCh [9]; Mohammed Mhanna, MD, MPH, MSc [10].

Affiliations.

1. Department of Medicine, Advocate Illinois Masonic Medical Center, Chicago, IL, USA.
2. Bluhm Cardiovascular Institute, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.
3. Faculty of Medicine, Al-Azhar University, Cairo, Egypt.
4. Department of Medicine, An-Najah National University, Nablus, Palestine.
5. Faculty of Medicine, Tanta University, Tanta, Egypt.

6. Faculty of Medicine, Alexandria University, Alexandria, Egypt.

7. Department of Medicine, Creighton University School of Medicine, Omaha, Nebraska, USA

8. Language Center, Faculty of Humanities, An-Najah National University, Nablus, West Bank, Palestine

9. Faculty of Medicine, Mansoura University, Mansoura, Egypt.

10. Division of Cardiovascular Medicine, University of Iowa, Iowa, IA, USA.

Corresponding Author. Ameer Awashra, MD. Department of Medicine, An-Najah National University, Nablus, Palestine. **Address:** Old Campus Street 7, P.O. Box 7, Nablus. **Email:** ameer.awashra7@gmail.com.

Abstract

Background: The use of mortality data in public health research has surged with the rise of open-access databases such as CDC WONDER. However, caution is needed when defining the relationship between ICD codes and when transitioning from older to newer versions of the data. This review provides a practical, step-by-step guide to using the CDC WONDER mortality database.

Methods: We outline key functionalities of the CDC WONDER interface, explain mortality rate calculations, and describe best practices for configuring queries using underlying and multiple causes of death. The review further introduces Joinpoint regression to identify temporal trend changes and compares forecasting approaches using traditional ARIMA models and modern deep learning architectures.

Results: Using illustrative examples and visual guides, we demonstrate how data interpretations can vary significantly depending on query configuration, Boolean logic (AND vs. OR), and coding practices. We highlight the strengths and limitations of different analytical strategies and show how misinterpretation can arise from common errors, such as misunderstanding age adjustment or combining ICD codes without appropriate logic.

Conclusion: CDC WONDER is a powerful tool for mortality analysis, but its effective use requires a clear understanding of its data structure, coding logic, and statistical tools. Joinpoint regression and forecasting models complement WONDER data by enabling trend segmentation and future projections. This guide empowers researchers to use these tools accurately, improving the rigor and reproducibility of public health research.

Keywords: CDC WONDER guide, Joinpoint analysis, ARIMA forecasting, deep learning forecasting.

1. Introduction

We write this paper to provide a practical guide for researchers, public health professionals, epidemiologists, clinicians, and medical students to understand what the CDC WONDER is, what mortality data (both age-adjusted and crude) represent, and how to interpret them accurately. We also discuss Joinpoint analysis, which involves dividing mortality trends over the years into segments rather than just one piece to evaluate each group or segment of years together, as well as forecasting analysis, a fancy way of saying we need to see the future mortality trends of our sample based on existing data (what the future looks like given the available data).

With the rising popularity of databases such as CDC WONDER, NHANES, GBD, TriNetX, and MAUDE, among others, journals have become increasingly receptive to publishing research based on these databases. CDC WONDER, in particular, is one of the oldest and reflects data from the US population exclusively. Since its inception, the number of publications using this database has increased significantly. From the PubMed search alone, from 1993 (the database’s inception) to 2010, there were 58 publications, followed by 119 from 2010 to 2020. In recent years, this growth has accelerated even further, as illustrated in **Figure 1.**

- **Insert Figure 1 here.**

2. Overview of the CDC WONDER Database

What is this database, and how is it used?

Essentially, CDC WONDER is a publicly available database published by the Centers for Disease Control and Prevention (CDC) that provides annual data on causes of death in the United States, among other data. An individual may die from a single primary condition that initiated the sequence of events leading to death—referred to as the *underlying cause of death*—but other conditions that contributed to the death may also be recorded. These are known as *multiple* or *contributing* causes of death. Both types of death causes are documented on the death certificate, which is typically completed by the treating physician or a member of the care team. The certificate reflects the physician's clinical judgment about the primary cause of death and any contributing conditions—essentially, what directly caused the death and what other factors played a supporting role (1,2).

The database offers valuable insights into mortality data—both crude and age-adjusted rates—dating back to 1968. Updated annually, it also provides early, incomplete estimates of deaths (counts and often rates) based on death certificates received and coded as of a cutoff date, before the complete set of records is processed. These preliminary figures, referred to as *provisional deaths*, reflect deaths occurring through January 5, 2025, as of January 12, 2026, update (i.e., one week before this inquiry).

3. Mortality Rates: Crude vs. Age-Adjusted

In CDC WONDER, the age-adjusted mortality rate (AAMR) is a method for comparing death rates as if every population group had the same age distribution as a standard chosen population. It does this by calculating death rates within 11 age groups and then combining them using a fixed “standard” age distribution (by default, the year 2000 US standard population). Therefore, the differences you observe are less likely to be driven solely by one group being older than another. The 11 age groups used for standard age-adjustment in WONDER are: <1, 1–4, 5–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, and ≥85 years.

By contrast, the crude mortality rate is the simple “raw” death rate in the population without adjusting for age. Because older people die at higher rates, a population with more older adults will often have a higher crude rate than its age-adjusted rate—but this isn’t guaranteed (it depends on how old the population is compared with the standard population). Both rates are typically reported per 100,000 people in WONDER.

When we say age-adjusted mortality is calculated “as if every group had the same age makeup,” we mean this:

- You first compute age-specific death rates (a rate within each age band).
- Then you re-weight those age-specific rates using the same fixed age distribution (the “standard population”) for everyone you’re comparing—so differences aren’t just because one population is older.

A simple example: Imagine two towns, each with 100,000 people:

Town A (older):

- 80% under 65
- 20% age 65+

Town B (younger):

- 95% under 65
- 5% age 65+

Now, assume both towns have the same age-specific death rates:

- Under 65: 100 deaths per 100,000

- 65+: 2,000 deaths per 100,000

Crude rate (raw rate):

- Town A: 0.80×100 + 0.20×2000 = 80 + 400 = **480 per 100,000**
- Town B: 0.95×100 + 0.05×2000 = 95 + 100 = **195 per 100,000**

So Town A *looks* much worse—but only because it has more older people.

Age-adjusted rate (“as if both towns had the same age makeup”): Pick a single “standard” age mix (for illustration): 90% under 65, 10% age 65+

Apply those same weights to *both* towns:

- 0.90×100 + 0.10×2000 = 90 + 200 = 290 per 100,000 (for both towns)

The two towns in the above example may represent two states (e.g., Florida vs. Utah), two racial groups (e.g., non-Hispanic White vs. Hispanic), or even two time points (e.g., 1999 vs. 2017).

This is exactly the logic used in CDC WONDER: it applies a fixed (standardized) age distribution to ensure fair comparisons of mortality across populations or time periods. **Figure 2** illustrates this age adjustment process.

- **Insert Figure 2 here.**

The platform can be accessed at https://wonder.cdc.gov. Upon entering the site, users are prompted to select the dataset type they wish to explore. The most frequently used datasets are Underlying Cause of Death (UCD), Multiple

Cause of Death (MCD), and Compressed Mortality, which will be the focus of our talk here.

The UCD dataset provides data on the underlying cause of death only, starting in 1999, while the MCD dataset includes data on both underlying and multiple causes of death beginning in 1999. The Compressed Mortality dataset contains older mortality data dating back to 1968, along with some more recent data.

4. ICD Coding Systems and Time Transitions

Still, it is essential to recognize that the International Classification of Diseases (ICD) coding system used in CDC WONDER has evolved over time. ICD-8 was used from 1968 to 1978; ICD-9 from 1979 to 1998; and ICD-10 has been used for US mortality (death certificate) cause-of-death coding, starting with deaths in 1999 and continuing through 2016 and beyond. This can be confusing because the “2015 ICD-10 adoption” refers to a separate transition in clinical billing/claims (ICD-9-CM → ICD-10-CM/PCS) that took effect on October 1, 2015 (3), not the mortality system used in WONDER. When analyzing trends across multiple decades, it is essential to account for these coding transitions to ensure comparability and accurate interpretation.

5. Step-by-Step Guide to Using the CDC WONDER Interface

After selecting the desired dataset, the user accesses a configuration interface to define the parameters of their query. One of the initial steps involves selecting the year or range of years of interest—for example, 1999 to 2020. The user then proceeds to Step 1 in **Figure 3**, where they choose how to group the data. Grouping can be based on demographic or geographic variables such as sex, race, year, census region, or state. It is essential to note that if the underlying or multiple causes of death and demographic categories, such as age groups, selected in Step 3 remain unchanged, the data output will be the same, with differences only in how the results are displayed—for instance, in the Excel file structure. Also, in Step 1, users should check boxes to include age-adjusted mortality rates (AAMR), confidence intervals (CI), standard error (SE), and percentage of total deaths—if desired. It is also important to note that if users choose to group the data by age groups in Step 1, the AAMR will not be available, as it cannot be calculated for the same age groups used in the age adjustment process. In such cases, the AAMR checkbox can be omitted; if it is selected, it will not yield results.

In Step 2, users can narrow the data by location, filtering by state or urbanization level, though these fields are often left at their default settings. Step 3 enables users to define demographic filters such as sex, race, and age group. For analyses focused on adults, the "25 years and over" age group is commonly selected, particularly because data for younger age bands, such as 18–24 years, are often unavailable. WONDER has predefined age strata for the AAMR calculation, which is why the 18-24 age group is unavailable; it’s not used in the AAMR calculation.

Further refinements are available in Step 5, where users can filter by the weekday of death, autopsy status, and place of death. However, these filters are typically not altered unless the research question requires them. The most crucial part of the process is in Steps 6 and 7, where the causes of death are defined using ICD codes. In Step 6, which addresses the underlying cause of death, users can input one or more ICD codes, interpreted as an "OR" condition—meaning any of the selected codes can qualify as the underlying cause. In Step 7, which allows multiple causes of death, users can enter up to 20 ICD codes. These can all be entered in a single box (also interpreted as "OR"), or split between two boxes that use "AND" logic.

For example, one might define ischemic heart disease (IHD) as the underlying cause of death and then add diabetes mellitus type 2 (DM-2), hypertension (HTN), or chronic kidney disease (CKD) in one "OR" box, with stroke listed in the "AND" box. This setup would return data on deaths where IHD was the primary cause, and any of the listed conditions—DM-2, HTN, or CKD—were present, along with stroke as a contributing cause. To do this, you must select the appropriate ICD coding system for ischemic heart disease. Under ICD-8, the codes are 410–413; under ICD-9, the codes are 410–414; and under ICD-10, the codes are I20–I25. Once selected, you can download datasets that include both age-adjusted and crude mortality rates.

Finally, in Step 8, users can configure export settings and specify how the data should be displayed or downloaded. It is recommended to set the number of decimal places to at least 3 or 4, as selecting only 1 decimal place will cause any death rate below that threshold to be displayed as zero. In this case, we selected the maximum value of 9 decimal places to ensure the most precise representation of the data. Generally, Steps 1, 3, 6, and 7—covering data grouping, demographic filters, and cause-of-death coding—are the most essential and frequently used components in most analyses conducted through this interface.

- **Insert Figure 3 here.**

A simplified version of the CDC WONDER query process is illustrated in **Figure 4.**

It is important to note that some data may not be available when certain variables are selected in Step 1, but may become available after filtering in subsequent steps. What does this mean in practice? For example, if you choose to group the data by Weekday in Step 1, the output will include only the number of deaths, without any associated crude death rate or age-adjusted mortality rate (AAMR). This is illustrated in the example for ICD-10 codes C00–D48 (Neoplasms).

However, if you instead go to Step 5 and use Ctrl or Shift to select Monday through Friday, the data output will include both AAMR and crude rates. This is because the system does not provide rate-based data for each day individually, likely due to insufficient sample size. However, when multiple weekdays are combined, the resulting larger sample size enables the calculation of these rates.

It is also important to note that if you use this second method (filtering for weekdays in Step 5) and download the data, the results will represent all selected weekdays combined—without a corresponding comparison group (e.g., weekends). As such, this combined weekday group should not be directly compared to weekend data. Not only is the result age-adjusted only (and not adjusted for other demographic or contextual factors), but it is also not volume-adjusted. In other words, more deaths naturally occur on weekdays than on weekends, so any comparison between the two must be interpreted with caution to avoid misleading conclusions. See **Figure 5** for illustrations.

- **Insert Figure 5 here.**

In addition to mortality data, the CDC WONDER database provides information on environmental changes, cancer incidence and mortality, infant deaths and natality, and the distribution of tuberculosis cases. Refer to **Figure 6** for a detailed overview of the datasets available on the platform. Among these, the datasets on underlying and multiple causes of death are the most widely used for research, which we discussed above.

- **Insert Figure 6 here.**

6. Case Study: CAD and COPD Mortality Queries

Consider the following research topic you want to write up. What is the mortality rate of stable coronary artery disease (CAD) among patients with chronic obstructive pulmonary disease (COPD)?

This title carries four possible interpretations:

First possibility: You designate stable CAD as the *underlying* cause of death and COPD as the *multiple* cause. This means these are patients who died primarily *from stable CAD* and had *COPD recorded as a contributing factor*. In this case, you are not capturing the entire population of people with COPD. Some patients had COPD, but for various reasons, the physician who filled out the death certificate did not judge that COPD contributed to the death, even though the patient had it. This process is very subjective and doesn’t reflect the full sample.

Second possibility: The data was extracted with both stable CAD and COPD listed as *multiple causes*, and the relationship between them was set to *OR (in the same box)*. This includes all patients who had either COPD or stable CAD mentioned anywhere on the death certificate, whether as an underlying or contributing cause. Why? Because if the *underlying cause* field is left blank (by blank, we mean it holds the possibility of any disease to be listed), the main cause of death could be anything. “Anything” could include stable CAD or COPD, but it might also be something else entirely—like stroke, for example. This means the person died and had either COPD or stable CAD or even both listed on their death certificate, without specifying whether it was the primary or a contributing cause. This can make the interpretation very vague.

Third possibility: You selected both stable CAD and COPD as multiple causes, with the relationship set to 'AND' rather than 'OR' (each in a separate box). This means you’re capturing only those death certificates where *both conditions* appear—whether one is listed as underlying and the other as contributing, or both as contributing. This approach is generally more precise than the previous one.

Fourth possibility: The user selected stable CAD and COPD, both as *underlying causes*. In this case, the CDC WONDER system will always treat the relationship as *OR*, because a person can only have one underlying cause of death. This means the individual died *either from stable CAD or from COPD*—but not both. This approach lacks precision. For this, we are against calling it “among patients with COPD” because there are people with COPD who died from stable CAD. Still, the physician did not acknowledge COPD as a contributing factor to death, so it’s better to be revised as “What is the mortality rate of stable CAD when COPD is present as a contributing condition?” See **Table 1 and Figure 7** for details.

- **Insert Table 1 here.**
- **Insert Figure 7 here.**

From our standpoint, WONDER is an effective tool for understanding underlying causes of mortality, such as stroke, atrial fibrillation (AF), or acute coronary syndrome (ACS), over time. Combining earlier data from before 1999 with later data from after 1999 provides a more comprehensive historical perspective. While earlier datasets lack the detailed subgroupings and categories found in more recent datasets (e.g., ethnicity and races beyond Black or White), they still provide valuable insights when analyzing large mortality datasets. This historical view helps illustrate how mortality rates have evolved and raises important questions about the contributing factors. For instance, have public health policies played a role in reducing mortality? Or has the absence of such policies contributed to an increase or plateau in deaths?

Recently, Lippi et al. examined the data from a different angle by focusing on within-year monthly seasonality rather than long-term secular trends, showing that U.S. acute myocardial infarction mortality from 2018 to 2022 consistently peaked in December and January and was lowest during the summer months (4), which provides another useful perspective for examining these data.

Sometimes, an observed rise in mortality may be due to improved detection rather than an actual increase in death rates, as seen in the study examining mortality from amyloidosis over the years, where the increase is likely due to better identification rather than a true rise in cases per se (5). If users choose to use the multiple causes of death in analyses, caution is necessary. The interpretation can vary depending on how Boolean logic is applied in the literature, which may influence the conclusions drawn.

Accordingly, the methodological choices employed when querying CDC WONDER have significant implications for public health. In our CAD and COPD case study, employing a comprehensive 'OR' logic includes a significantly diverse population, in which the actual cause of death may be unrelated to either condition. If researchers depend on this vague question, they could be overestimating how many people die because of the connection between CAD and COPD, which might lead to wrong public health priorities, and could mean that resources for disease prevention are not used as well as they could be (6). On the other hand, using the more exact "AND" logic only looks at the group of people who died with both criteria at the same time. This level of accuracy enables policymakers and doctors to accurately assess the true comorbidity burden, develop integrated care pathways, and develop evidence-based solutions that meet the community's health needs (7).

7. Joinpoint Analysis: Understanding Trend Changes

We now transition to **Joinpoint analysis**.

What is Joinpoint analysis?

In simple terms, Joinpoint analysis is a statistical tool that breaks down a single, long-term trend into smaller, specific time segments. Instead of drawing a single straight line across decades of data, it connects multiple lines to pinpoint exactly when mortality rates changed course, such as the year when a steady decline suddenly turned into an increase.

Consider the following hypothetical example:

The age-adjusted mortality rate (AAMR) for acute myocardial infarction was 0.3 in 1999 and 0.9 in 2020. If you draw a straight line between these two points, you could say that the AAMR increased by 0.6 over 21 years, resulting in an average annual increase of 0.02857 (0.6/21).

While this provides a broad view of the data, it misses important details.

For instance, suppose that between 1999 and 2005, the AAMR actually decreased from 0.3 to 0.1; then from 2005 to 2012, it increased sharply to 1.6; and finally, from 2012 to 2020, it declined again to 0.9. This segmentation provides a much deeper understanding of how the AAMR evolved over the 21 years.

Some of these segmental trends may be statistically significant—for example, a p-value < 0.05 from 1999 to 2005—while others may not (e.g., p-value > 0.05 in subsequent periods).

Joinpoint analysis (or Joinpoint regression) is a method used to examine how mortality trends change over time. Joinpoint models fit multiple connected line segments and identify the years in which the slope changes significantly or non-significantly—these are known as joinpoints. The main objective is to detect the timing of these trend changes and to analyze rate patterns before and after each joinpoint. This approach is well-suited for CDC WONDER data, which often exhibits non-uniform mortality trends.

Joinpoint analysis is commonly conducted using the National Cancer Institute (NCI) Joinpoint Regression Program (8,9), a widely used tool for surveillance analysis of rates. The software allows users to set a minimum and maximum number of joinpoints and then tests whether adding joinpoints significantly improves model fit. The output reports the identified joinpoint years and the slope of each segment, expressed as Annual Percent Change (APC). It also provides an Average Annual Percent Change (AAPC) to summarize overall trends.

For CDC WONDER data, Joinpoint is typically applied to age-adjusted mortality rates (AAMR), which help enable fair comparisons across years despite changes in the population’s age structure. Still, it can also be applied to crude rates, death count, or percentage. The NCI Joinpoint Regression Program includes several models and option settings that are particularly important for analyzing CDC-type mortality data.

The model needs to decide how many “joinpoints” (change-points) to include. One method, the Permutation Test (10), adds joinpoints only when there's strong evidence that the change is real, based on many simulated datasets (e.g., 4499 permutations). This is like running a detailed test to see if adding a joinpoint significantly improves the model. The other method, Weighted BIC (WBIC) (11), skips formal testing and instead picks the number of joinpoints by balancing how well the model fits the data versus how simple it is. WBIC might allow more joinpoints if they slightly improve the model, even without strong statistical proof.

Another difference is in how confidence intervals (CIs) are calculated. These intervals show how certain we are about trend estimates, such as Annual Percent Change (APC). The older Parametric Method assumes that the data follow a specific distribution and tends to yield wider, more cautious intervals. The newer Empirical Quantile (12,13) Method uses resampling (e.g., 5001 model repeats) to directly estimate variability, often resulting in narrower, more accurate intervals. These differences may not matter much in large, smooth datasets. Still, in smaller or noisier subgroups (like by age or race), they can lead to different joinpoint years, different APC values, or even change whether a trend is considered statistically significant. Running both methods can help confirm the robustness of your findings.

Both the Permutation Test and Weighted BIC methods use the same setting for the maximum number of joinpoints—for example, 7—because this is a general rule that limits how complex the model is allowed to get, no matter which selection method you're using. It works like a cap: the software looks at all models from the minimum (usually 0 joinpoints) to the maximum (e.g., 7), and then the chosen method (Permutation or WBIC) picks the “best” one from that range.

However, the number 7 isn't always the default setting. Joinpoint software usually recommends a maximum number of joinpoints based on the amount of data you have. For example, if your data has 27 or more time points (e.g., 27 years), it might suggest a maximum of 5 join points. But you're allowed to set a higher maximum manually (up to 9) if you’re using a more detailed search, called a grid search, though this can take longer to run.

It is important to note that the software accepts a grouping variable (such as ethnicity, race, or gender) alongside the year as the independent variable, and requires the count or rate to be analyzed—such as age-adjusted mortality rate (AAMR), crude rate, or death count—along with its standard error (SE). Confidence intervals (CIs) are not required. We advise using the Empirical Quantile method for confidence intervals and the Weighted BIC for model selection, as shown in **Figure 8.**

For a simple explanation of Joinpoint analysis, please see **Figure 9.**

- **Insert Figure 8 here.**

8. Forecasting Mortality Trends

Forecasting uses historical data to predict future mortality trends, assuming no major changes in clinical care or public health policies occur. This section includes two main approaches: traditional statistical models (such as ARIMA), which are excellent for straightforward, historical patterns, and modern machine learning models (such as deep learning), which are better suited for massive, complex datasets with multiple variables.

Two commonly used approaches are ARIMA (AutoRegressive Integrated Moving Average) and deep learning (DL) models, both of which fall under the broader umbrella of machine learning. However, ARIMA is often considered a traditional statistical method, while deep learning reflects modern computational advances.

These models generally assume that underlying mortality trends remain stable, meaning that there are no significant changes in treatment practices, diagnostic criteria, or external interventions over time. In other words, they attempt to answer the question:

"If we make no changes to clinical care or public health interventions, what will mortality look like in 5, 10, or 20 years?"

ARIMA models are most effective for small to moderately sized datasets—typically when working with monthly or annual data spanning 10–30 years and involving relatively few variables (14). These models are highly interpretable and well-suited to data with consistent seasonal or linear patterns. They rely on previous observations and errors to capture time dependence and are widely used because of their simplicity and ease of interpretation (15,16). In practice, ARIMA models are commonly implemented using R packages such as "forecast", or Python packages such as "pmdarima" and "statsmodels".

On the other hand, deep learning models, such as LSTM, GRU, or Transformer-based architectures, are more powerful for handling larger and more complex datasets, including high-frequency data (e.g., daily or hourly), longer time spans, or multivariate time series with many features (e.g., clinical, demographic, or policy variables). While these models require more data and computational power, they can capture non-linear trends and complex dependencies that ARIMA models cannot (17). They are typically implemented in R using Keras, or in Python using TensorFlow/Keras or PyTorch, and perform best when applied to large or complex datasets.

For a simple explanation of forecasting models, please see **Figure 9.**

- **Insert Figure 9 here.**

9. Examples from the Literature Where CDC WONDER's Interpretation Matters.

It's important to understand that studies based on CDC WONDER are cross-sectional in nature, involving population-level death-certificate data of deaths from a given condition or conditions in the United States. There is no comparative arm, and therefore, these studies cannot be classified as retrospective cohort studies. However, they can appropriately be described as retrospective cross-sectional studies.

In the study by Billion et al.(18), the authors examined trends and disparities in lung cancer mortality, but incorrectly described their work as a retrospective cohort study. The same issue applies to the study by Maniya et al. (19), which analyzed atherosclerotic cardiovascular disease–related mortality among older adults, but also misclassified the study design as a retrospective cohort study. In another study by Nabi et al. (20), the authors used “AND” logic to link both obesity and hypertension. ICD codes for one condition were entered into the first box of the MCD section, while codes for the other condition were entered into the second box. The UCD was left at the default setting of “All Causes of Death,” meaning the primary cause could be obesity, hypertension, or another condition entirely. The authors appropriately defined and applied their logic.

Additionally, in Dort et al. (21), the authors titled their work *“Trends in pulmonary embolism mortality in cancer patients,”* suggesting that the study included all cancer patients whose primary cause of death was pulmonary embolism (PE). However, the authors used AND logic, similar to Nabi et al., meaning that both PE and cancer were listed somewhere on the death certificate—either as an MCD or UCD—but not necessarily that PE was the direct cause of death in cancer patients. In this configuration, the actual UCD could be PE, cancer, or another unrelated condition. A more accurate and transparent title for their study would be: *“Mortality trends related to PE and cancer patients.”* In another study by the same lead author (Nabi et al.) (22), they used acute myocardial infarction (AMI) as the underlying cause and chronic kidney disease (CKD) as a contributing cause of death, titling the study *“Mortality Trends from Acute MI with Underlying CKD.”* While it is true that if CKD is listed as an MCD, the patient had it at the time of death, the approach does not capture all AMI cases in which CKD was recorded as an MCD, as this is entirely dependent on how the death certificate was completed. Instead, it identifies only a subset of patients in whom AMI was the UCD and CKD was explicitly documented as a contributing cause.

10. Adaptation of Analytical Methods to Global Mortality Databases

While the CDC WONDER database provides comprehensive mortality data for the US only, the analytical principles and statistical techniques discussed in this guide can be translated to global health research. Researchers investigating international mortality trends can apply similar workflows, such as differentiating between underlying and multiple causes of death, adjusting for age to allow cross-population comparisons, and accounting for ICD coding transitions when using global repositories.

Prominent global resources include the World Health Organization (WHO) Mortality Database, which provides cause-of-death statistics based on ICD classifications (23) (24), and the Global Burden of Disease (GBD) studies, which provide comprehensive demographic and epidemiological estimates across 204 countries and territories (25) (26). When applying tools like Joinpoint regression and forecasting models to these international datasets, investigators must remain vigilant about database-specific limitations. For example, the quality, completeness, and reporting of death certificates vary significantly between low- and high-income countries (27) (28). Also, global databases often aggregate data using different standard populations for age-adjustment (29). However, by adapting the rigorous query configurations and statistical modeling approaches outlined in this guide, researchers can effectively leverage global databases to track disease burden, evaluate international public health interventions, and forecast future epidemiological challenges on a global scale.

11. Limitations of CDC WONDER and Cautions in Interpretation

The CDC WONDER mortality databases have several key limitations. First, they are based on death certificates rather than clinical registries and therefore offer limited clinical detail and sparse demographic data. They do not include laboratory results, imaging, medications, procedures, or adjudicated diagnoses. Instead, a single designated certifier makes the medical judgment, and the underlying cause is determined afterward through standardized coding rules rather than through a multiperson adjudication process. Prior literature has shown that certifiers may record terminal mechanisms such as “cardiopulmonary arrest” or “cardiac arrest” instead of the actual disease process causing death, even though these terms reflect mechanisms of dying rather than valid underlying causes of death. As a result, mortality data may be influenced not only by disease burden and survival, but also by certification practices and coding behavior. Additionally, WONDER cannot distinguish whether mortality trends reflect changes in disease occurrence, survival, treatment, or coding practices; thus, causal interpretation requires caution. Suppression rules also present a practical challenge: counts under 10 are suppressed and cannot be published, which distorts analyses of rare outcomes or small populations. In addition, low-count rates (for example, <20) may be flagged as unreliable. Finally, changes in coding systems, such as transitions from ICD-8 to ICD-9 and ICD-10, can introduce discontinuities in trends, necessitating the use of comparability ratios (30–32).

Misclassification of race/ethnicity—particularly among American Indian/Alaska Native and Hispanic groups—can bias disparity analyses, while inaccuracies in recording education and occupation (reported after 2020) can distort trends and analyses related to socioeconomic and occupational factors. Provisional mortality files are affected by reporting delays, increased suppression, and flagged unreliability, making them more suitable for surveillance than for detailed subgroup analysis. Finally, analytic decisions about whether to use only underlying causes or include multiple causes can substantially affect results, as multiple-cause data may reflect comorbidity rather than true causation.

12. Strategies for Overcoming Data Limitations.

While the inherent limitations of CDC WONDER cannot be avoided, researchers can employ specific methodological strategies to mitigate their impact.

Regarding suppressed data: To protect confidentiality, CDC WONDER suppresses death counts of fewer than 10, which disproportionately affects rural areas and rare causes of death. The most straightforward strategy to overcome this is temporal or spatial aggregation, such as pooling 3- to 5-year periods into rolling averages or grouping individual counties into larger geographic regions to surpass the reporting threshold. For more advanced geospatial analyses, researchers can apply Bayesian spatial smoothing or multiple imputation techniques that weight population data and regional averages to estimate suppressed values without discarding incomplete datasets (33,34).

Additionally, to correct for misclassification, researchers evaluating demographic disparities should apply the misclassification correction ratios routinely published by the National Center for Health Statistics (NCHS) to adjust their age-adjusted mortality rates (35). Alternatively, if detailed demographic specifics are not a primary aim of the study, researchers may limit comparative analyses to non-Hispanic White and non-Hispanic Black cohorts. These groups have historically exhibited superior validity and minimal misclassification in vital statistics records (36).

13. Conclusion

CDC WONDER is a valuable tool for analyzing US mortality trends. Still, its effective use depends on a clear understanding of its data structure, including the distinction between crude and age-adjusted mortality rates, ICD coding transitions, and the logic applied when selecting underlying and multiple causes of death. This review provides practical guidance on navigating the platform, configuring accurate queries, and applying advanced analytical methods such as Joinpoint regression and forecasting models. When used thoughtfully, CDC WONDER, paired with robust statistical tools, can offer powerful insights into temporal mortality patterns and inform evidence-based public health decisions.

Statements

Data Availability Statement

All data supporting the findings of this study are included in the manuscript or its supplementary materials.

Author Contributions

Abdalhakim Shubietah was responsible for conceptualization, project leadership, and writing the initial manuscript draft. Mohammed Ruzieh supervised the project and provided critical insights and revisions to the initial draft. Belal Mohamed Hamed contributed to the writing of the sections on Joinpoint analysis and forecasting models. Sarah Saife designed the figures and wrote their interpretations. Mohamed Abuelazm, Mohamed S. Elgendy, Ahmed Emara, Ahmed Ibrahim, Ameer Awashra, Ibrahim Gowaily, Elsayed Balbaa, Mohamed Saad Rakab, and Mohammed AbuBaha contributed to the writing of the second version of the manuscript and the literature review. Ghada Shbeitah contributed critical language editing and revised the final manuscript for clarity, coherence, and academic style. Mohammed Mhanna supervised the project and provided critical insights throughout the research process.

Funding

No funding was received for conducting this study.

Conflict of Interest

The authors declare no competing interests relevant to the content of this article.

Ethics Approval

Not applicable.

Declaration of AI use

Figure 9 was created using the AI tool Image Generator, while all other figures were created using BioRender.

Consent to Publish

Not applicable.

Consent to Participate

Not applicable.

References:

1. CDC WONDER [Internet]. [cited 2026 Jan 16]. Available from: https://wonder.cdc.gov/
2. Friede A, Freedman MA, Paul JE, Rizzo NP, Pawate VI, Turczyn KM. DATA2000: CDC WONDER information system linking Healthy People 2000 objectives to data sets. Am J Prev Med. 1994;10(4):230–4. PubMed PMID: 7803066.

3. Kusnoor SV, Blasingame MN, Williams AM, DesAutels SJ, Su J, Giuse NB. A narrative review of the impact of the transition to ICD-10 and ICD-10-CM/PCS. JAMIA Open. 2020 Apr 1;3(1):126–31. doi:10.1093/jamiaopen/ooz066

4. Lippi G, Sanchis-Gomar F, Lavie CJ. The Recent (2018-2022) US Monthly Mortality for Acute Myocardial Infarction Still Peaks in December and January. Am J Med. 2025 Jun;138(6):1038–41. doi:10.1016/j.amjmed.2025.01.006

5. Shubietah A, Owda D, Rakab MS, Dawoud Y, Khraiwesh A, Elgendy MS, et al. Nationwide patterns of cardiac-related mortality in amyloidosis cases: an epidemiologic study. BMC Cardiovasc Disord. 2025 Dec 30;25(1):907. doi:10.1186/s12872-025-05353-9

6. Tungbani B, Nhlapo MS. Accuracy of Cause-of-Death Classification in Ghana: Evidence from the 2023 District Health Information Management System (DHIMS) II. J Health Inform Afr. 2025;12(2):69–82. doi:10.12856/JHIA-2025-v12-i2-613

7. Iburg KM, Mikkelsen L, Adair T, Lopez AD. Are cause of death data fit for purpose? evidence from 20 countries at different levels of socio-economic development. PLOS ONE. 2020 Aug 24;15(8):e0237539. doi:10.1371/journal.pone.0237539

8. Joinpoint Regression Program [Internet]. [cited 2026 Jan 16]. Available from: https://surveillance.cancer.gov/joinpoint/

9. Kim HJ, Chen HS, Midthune D, Wheeler B, Buckman DW, Green D, et al. Data-driven choice of a model selection method in joinpoint regression. J Appl Stat. 2023 Jul 4;50(9):1992–2013. doi:10.1080/02664763.2022.2063265

10. Joinpoint Help System [Internet]. [cited 2026 Jan 16]. Permutation Test Details. Available from: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/model-selection-method/permutation-tests/permutation-test-details

11. Joinpoint Help System [Internet]. [cited 2026 Jan 16]. Weighted BIC (WBIC). Available from: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/model-selection-method/weighted-bic-wbic

12. Joinpoint Help System [Internet]. [cited 2026 Jan 16]. Empirical Quantile Confidence Interval. Available from:

https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/apc-aapc-tau-confidence-intervals/empirical-quantile

13. Kim HJ, Luo J, Chen HS, Green D, Buckman D, Byrne J, et al. Improved confidence interval for average annual percent change in trend analysis. Stat Med. 2017 Aug 30;36(19):3059–74. doi:10.1002/sim.7344 PubMed PMID: 28585245; PubMed Central PMCID: PMC5652079.

14. Grøntved S, Jørgine Kirkeby M, Paaske Johnsen S, Mainz J, Brink Valentin J, Mohr Jensen C. Towards reliable forecasting of healthcare capacity needs: A scoping review and evidence mapping. Int J Med Inf. 2024 Sep;189:105527. doi:10.1016/j.ijmedinf.2024.105527

15. Forecasting: Principles and Practice (3rd ed) [Internet]. [cited 2026 Jan 17]. Available from: https://otexts.com/fpp3/

16. Jiang H, Rehm J, Tran A, Lange S. Interrupted Time Series Design and Analyses in Health Policy Assessment. medRxiv. 2024 Aug 2;2024.08.01.24311280. doi:10.1101/2024.08.01.24311280 PubMed PMID: 39132471; PubMed Central PMCID: PMC11312671.

17. Kong X, Chen Z, Liu W, Ning K, Zhang L, Muhammad Marier S, et al. Deep learning for time series forecasting: a survey. Int J Mach Learn Cybern. 2025 Aug 1;16(7):5079–112. doi:10.1007/s13042-025-02560-w

18. Billion T, Verheyen L, Abdul Jabbar AB, Asghar N, Mirza M, Mirza M, et al. Regional trends and disparities of lung cancer mortality in the United States, 1999–2022: a retrospective cohort study. Respir Res. 2025 Oct 1;26(1):280. doi:10.1186/s12931-025-03344-0

19. Maniya MT, Siddiqi AK, Shahzad M, Ali KM, Chachar MA, Amin SB, et al. Rising atherosclerotic cardiovascular disease–related mortality among older adults in the United States: a 22-year retrospective cohort study using the CDC WONDER database. Hellenic J Cardiol. 2025 Jul;S1109966625001435. doi:10.1016/j.hjc.2025.06.001

20. Nabi R, Zanub A, Akhtar M, Chaudhry SAA, Awais AR, Farooqi HA, et al. Concomitant mortality trends due to obesity and hypertension in the U.S.: a 20-year retrospective analysis of the CDC WONDER database. BMC Cardiovasc Disord. 2025 Jul 7;25(1):496. doi:10.1186/s12872-025-04909-z

21. Dort E, Rud H, Billion T, Tauseef A. Trends in pulmonary embolism mortality in cancer patients in the United States from 1999-2022– A CDC Wonder database study. Respir Res. 2025 Jul 16;26(1):248. doi:10.1186/s12931-025-03319-1

22. Nabi R, Mansoor MH, Shahbaz MH, Kharal M, Fatima N, Munir ST, et al. Mortality Trends from Acute MI with Underlying CKD in the US from 1999 to 2020: A Cross-Sectional Analysis of the CDC WONDER Database. Int Urol Nephrol. 2025 Aug 25. doi:10.1007/s11255-025-04720-x

23. Hua J, Li L, Ning P, Schwebel D, He J, Rao Z, et al. Road traffic death coding quality in the WHO Mortality Database. Bull World Health Organ. 2023 Oct 1;101(10):637–48. doi:10.2471/BLT.23.289683

24. WHO Mortality Database - WHO [Internet]. [cited 2026 Mar 26]. Available from: https://www.who.int/data/data-collection-tools/who-mortality-database

25. Global Burden of Disease (GBD) [Internet]. [cited 2026 Mar 6]. Available from: https://www.healthdata.org/research-analysis/gbd

26. Naghavi M, Ong KL, Aali A, Ababneh HS, Abate YH, Abbafati C, et al. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024 May;403(10440):2100–32. doi:10.1016/S0140-6736(24)00367-2

27. International completeness of death registration (Volume 50 - Article 38 | Pages 1151–1170) - Demographic Research [Internet]. [cited 2026 Mar 6]. Available from: https://www.demographic-research.org/articles/volume/50/38

28. Johnson SC, Cunningham M, Dippenaar IN, Sharara F, Wool EE, Agesa KM, et al. Public health utility of cause of death data: applying empirical algorithms to improve data quality. BMC Med Inform Decis Mak. 2021 Dec;21(1):175. doi:10.1186/s12911-021-01501-1

29. Thurber KA, Thandrayen J, Maddox R, Barrett EM, Walker J, Priest N, et al. Reflection on modern methods: statistical, policy and ethical implications of using age-standardized health indicators to quantify inequities. Int J Epidemiol. 2021 Jul 5;51(1):324–33. doi:10.1093/ije/dyab132 PubMed PMID: 34223891; PubMed Central PMCID: PMC8855998.

30. Schuppener LM, Olson K, Brooks EG. Death Certification: Errors and Interventions. Clin Med Res. 2020 Mar;18(1):21–6. doi:10.3121/cmr.2019.1496 PubMed PMID: 31597655; PubMed Central PMCID: PMC7153801.

31. Maharjan L, Shah A, Shrestha KB, Shrestha G. Errors in cause-of-death statement on death certificates in intensive care unit of

Kathmandu, Nepal. BMC Health Serv Res. 2015 Nov 12;15:507. doi:10.1186/s12913-015-1168-6 PubMed PMID: 26563325; PubMed Central PMCID: PMC4643506.

32. Brooks EG, Reed KD. Principles and Pitfalls: a Guide to Death Certification. Clin Med Res. 2015 Jun;13(2):74–82; quiz 83–4. doi:10.3121/cmr.2015.1276 PubMed PMID: 26185270; PubMed Central PMCID: PMC4504663.

33. Quick H. Estimating County-Level Mortality Rates Using Highly Censored Data From CDC WONDER. Prev Chronic Dis. 2019;16. doi:10.5888/pcd16.180441

34. Tiwari C, Beyer K, Rushton G. The Impact of Data Suppression on Local Mortality Rates: The Case of CDC WONDER. Am J Public Health. 2014 Aug;104(8):1386–8. doi:10.2105/AJPH.2014.301900

35. Arias E, Heron M, National Center for Health Statistics, Hakes J, US Census Bureau. The Validity of Race and Hispanic-origin Reporting on Death Certificates in the United States: An Update. Vital Health Stat 2. 2016 Aug 1;(172):1–21. PubMed PMID: 28436642.

36. Johnson J, Moore B, Hwang E, Hickner A, Yeo H. The accuracy of race & ethnicity data in US based healthcare databases: A systematic review. Am J Surg. 2023 May 1;226. doi:10.1016/j.amjsurg.2023.05.011

Tables

Table 1. Interpretation of CDC WONDER Query Combinations for Stable CAD and COPD in Mortality Data.

Scenario	**WONDER Configuration**	**Logic Applied**	**What This Query Captures**	**Implication for Research**
1. Primary & Contributory	Underlying Cause: CAD Multiple Cause: COPD	AND	Patients whose primary cause of death was CAD, and who had COPD recorded specifically as a contributing factor.	Too Restrictive: Misses COPD patients where the physician failed to list it as a contributor (subjective error).
2. Broad Inclusion (OR)	Multiple Cause (Box 1): CAD, COPD	OR	Patients who had either CAD or COPD were mentioned anywhere on the certificate. The actual cause of death could be unrelated (e.g., Stroke).	Too Vague: Captures an overly broad population; difficult to attribute mortality specifically to the interaction of these diseases.
3. Co-Occurrence (AND)	Multiple Cause (Box 1): CAD Multiple Cause (Box 2): COPD	AND	Patients where both conditions appear on the death certificate, regardless of which was primary or contributory.	Recommended: The most precise method for analyzing the mortality burden of co-occurring conditions.
4. Underlying Exclusive	Underlying Cause: CAD, COPD	OR	Patients who died either primarily from CAD or primarily from COPD. A person cannot have two underlying causes.	Misleading: This creates two separate cohorts (one dying of CAD, one of COPD) rather than a cohort with both conditions.

*By CAD, we refer to stable CAD.

Figures Legend

Figure 1. Rising use of the CDC WONDER database in published research over time.

Figure 2. Calculation and interpretation of crude versus age-adjusted mortality rates.

Figure 3. Step-by-step configuration interface for querying mortality data in CDC WONDER (Multiple Cause of Death, 1999–2020).

Figure 4. Simplified overview of the CDC WONDER query workflow.

Figure 5. Comparison of weekday filtering in Step 5 (right) versus grouping by weekday in Step 1 (left) using the CDC WONDER Multiple Cause of Death database (1999–2020). Filtering weekdays (e.g., Monday to Friday) in Step 5 yields full output including deaths, crude death rates, age-adjusted mortality rates (AAMR), and their respective confidence intervals and standard errors. In contrast, grouping by weekday in Step 1 provides only death counts, with all rate-based measures labeled "Not Applicable." This visual comparison highlights how the choice of step affects data availability and interpretability.

Figure 6. CDC WONDER system interface displaying the range of available public health data sources.

Figure 7. Common methodological pitfalls in CDC WONDER–based mortality analyses.

Figure 8. Joinpoint Regression Program setup for trend analysis of age-adjusted mortality rates.

Top Panel (Step 1 – Variable Selection):

Loads the dataset containing age-adjusted mortality rates (AAMR), year, subgroup (e.g., race), and standard error. The year is set as the independent variable, and AAMR is selected as the dependent variable. Log transformation is applied to linearize exponential trends.

Middle Panel (Step 2 – Joinpoint Configuration):

The model defaults to the number of joinpoints based on the loaded data. Confidence intervals are calculated using the Empirical Quantile Method with 5,001 resamples for increased precision, and model selection is based on the Weighted Bayesian Information Criterion (WBIC).

Bottom Panel (Step 3 – Model Selection & Execution):

The model switches from BIC-based methods to the Permutation Test, which uses 4,499 permutations to determine whether each joinpoint significantly improves model fit. Trends are summarized over the full time range (1968–2023) using the Average Annual Percent Change (AAPC), with confidence intervals calculated using parametric methods.

Figure 9. Overview of the analytical framework for mortality trend research, showing a two-stage workflow: joinpoint regression to identify changes in temporal trends and estimate APC/AAPC, followed by forecasting models such as ARIMA and deep learning to project future mortality patterns.

WONDER Online Databases

- AIDS Public Use Data
- Births
- Cancer Statistics

Deaths:

All Ages:

- Underlying Cause of Death
- Multiple Cause of Death
- U.S. - Mexico Border Area Mortality
- Compressed Mortality
- Fetal Deaths
- Infant Deaths

Environment:

- Heat Wave Days May-September
- Daily Air Temperatures & Heat Index
- Daily Land Surface Temperatures
- Daily Fine Particulate Matter
- Daily Sunlight
- Daily Precipitation
- Online Tuberculosis Information System

Population:

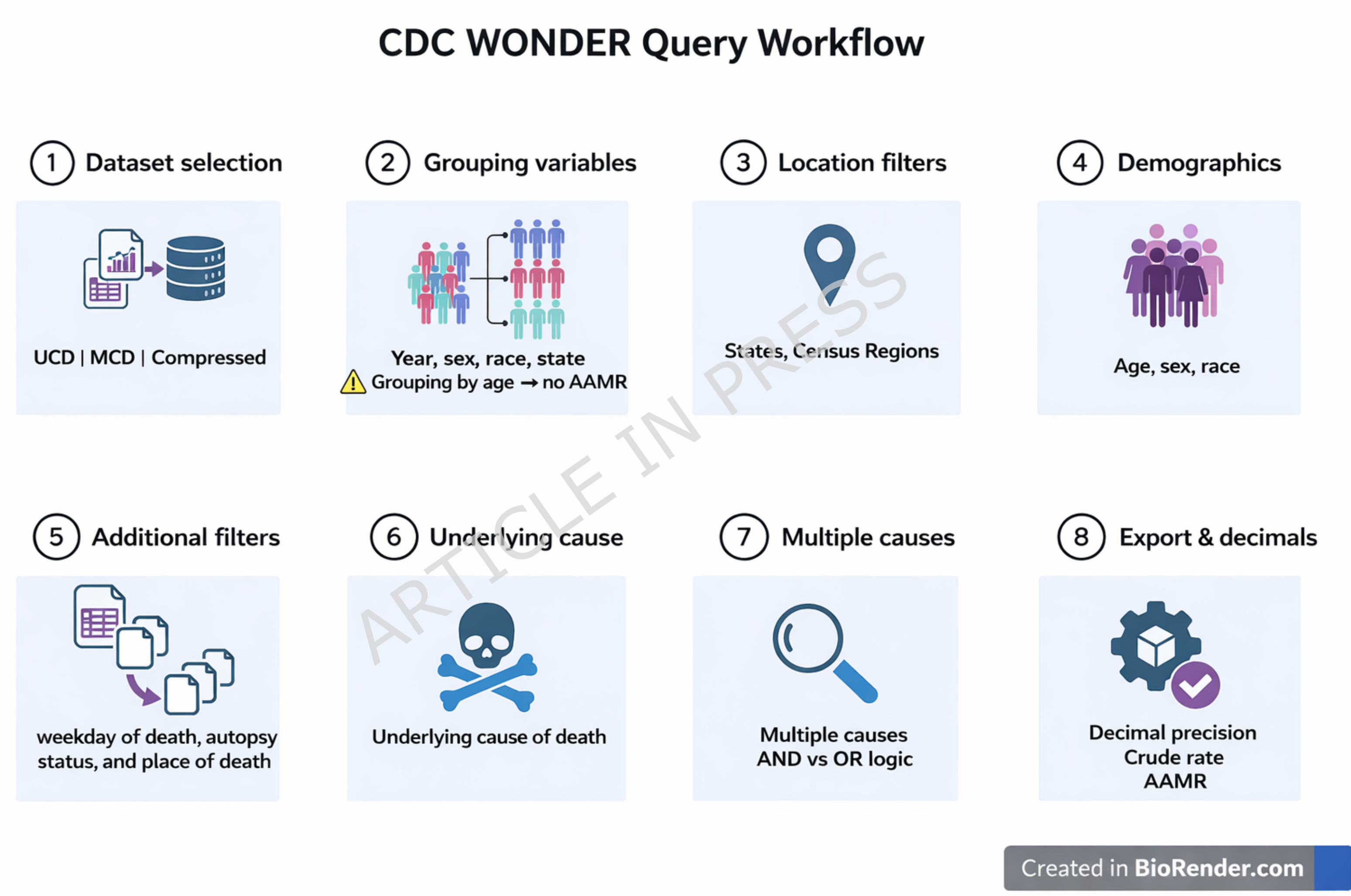
- Bridged-Race Population
- Single-Race Population (from Census)
- Population Projections (from Census)
- Puerto Rico Population Estimates (from Census)
- Sexually Transmitted Disease Morbidity
- Vaccine Adverse Event Reporting

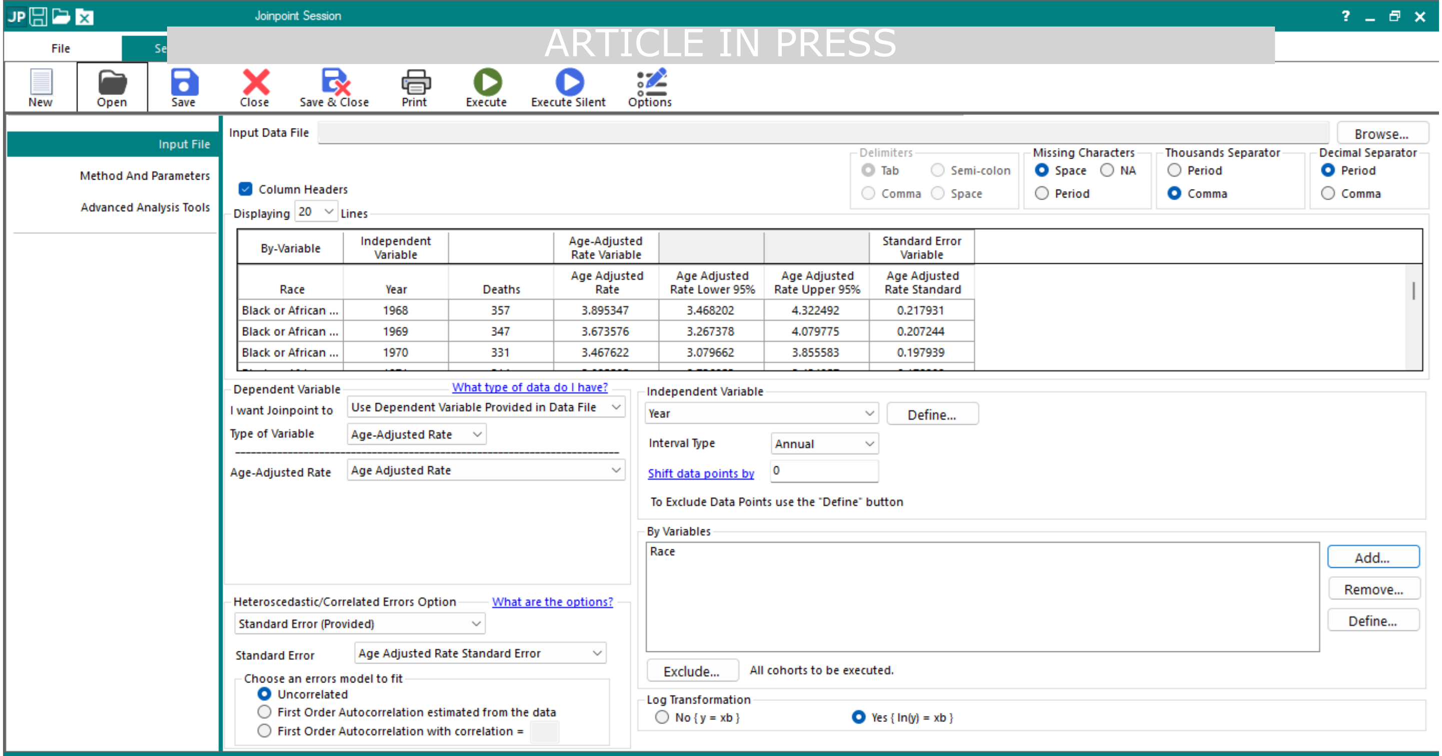
National Notifiable Conditions

- NNDSS Annual Summary Data Query
- NNDSS Annual Tables
- NNDSS Weekly Tables

Other Query Systems

- Healthy People


CDC WONDER Query Workflow
1 Dataset selection
UCD | MCD | Compressed
2 Grouping variables
Year, sex, race, state
⚠ Grouping by age → no AAMR
3 Location filters
States, Census Regions
4 Demographics
Age, sex, race
5 Additional filters
weekday of death, autopsy status, and place of death
6 Underlying cause
Underlying cause of death
7 Multiple causes
Multiple causes
AND vs OR logic
8 Export & decimals
Decimal precision
Crude rate
AAMR
Created in BioRender.com



Input Data File

Delimiters: ○ Tab ○ Semi-colon ○ Comma ○ Space

Missing Characters: ● Space ○ NA ○ Period

Thousands Separator: ○ Period ● Comma

Decimal Separator: ● Period ○ Comma

☑ Column Headers

Displaying 20 Lines

By-Variable	Independent Variable		Age-Adjusted Rate Variable			Standard Error Variable
Race	Year	Deaths	Age Adjusted Rate	Age Adjusted Rate Lower 95%	Age Adjusted Rate Upper 95%	Age Adjusted Rate Standard
Black or African ...	1968	357	3.895347	3.468202	4.322492	0.217931
Black or African ...	1969	347	3.673576	3.267378	4.079775	0.207244
Black or African ...	1970	331	3.467622	3.079662	3.855583	0.197939

Dependent Variable — What type of data do I have?

I want Joinpoint to: Use Dependent Variable Provided in Data File

Type of Variable: Age-Adjusted Rate

Age-Adjusted Rate: Age Adjusted Rate

Independent Variable: Year

Interval Type: Annual

Shift data points by: 0

To Exclude Data Points use the "Define" button

By Variables: Race

All cohorts to be executed.

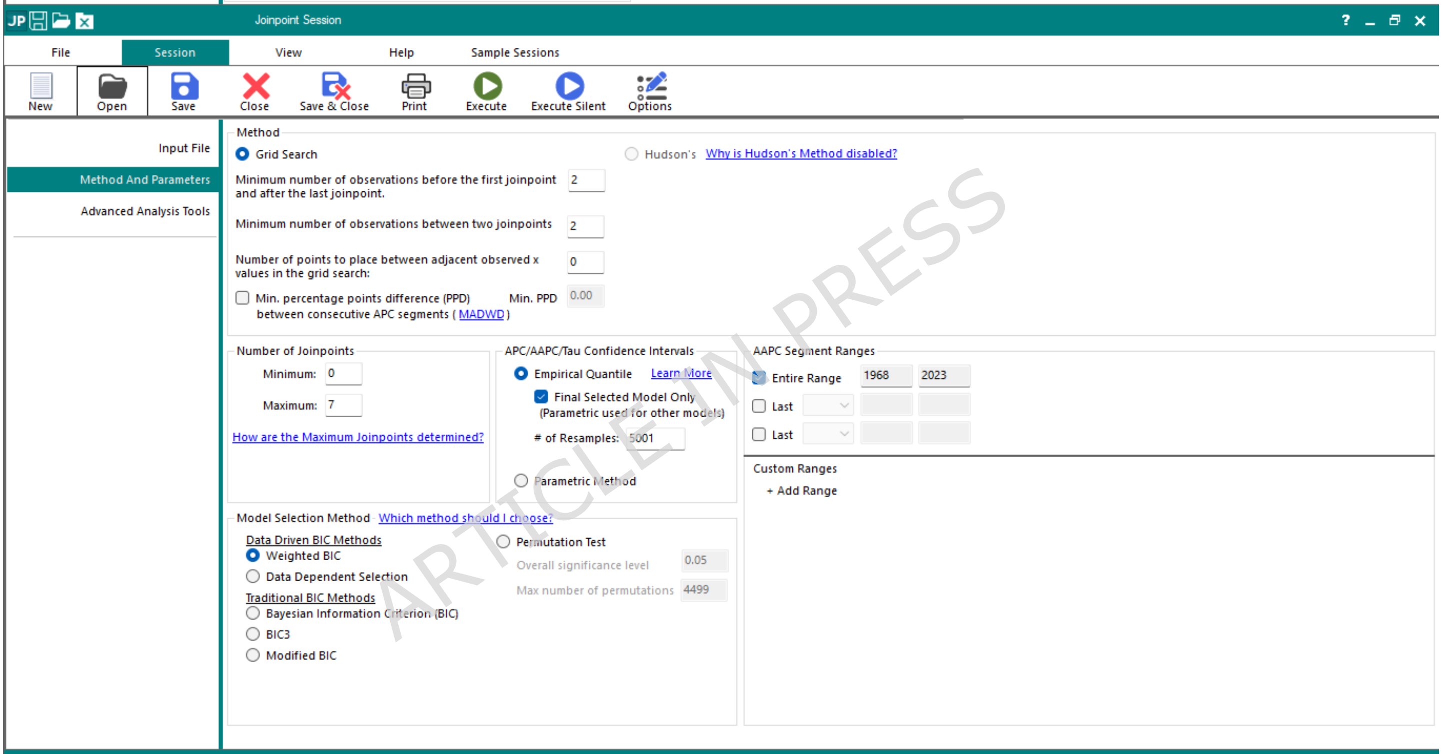
Heteroscedastic/Correlated Errors Option — What are the options?

Standard Error (Provided)

Standard Error: Age Adjusted Rate Standard Error

Choose an errors model to fit
● Uncorrelated
○ First Order Autocorrelation estimated from the data
○ First Order Autocorrelation with correlation =

Log Transformation: ○ No { y = xb } ● Yes { ln(y) = xb }



Method
● Grid Search ○ Hudson's — Why is Hudson's Method disabled?

Minimum number of observations before the first joinpoint and after the last joinpoint. 2

Minimum number of observations between two joinpoints 2

Number of points to place between adjacent observed x values in the grid search: 0

☐ Min. percentage points difference (PPD) between consecutive APC segments (MADWD) Min. PPD 0.00

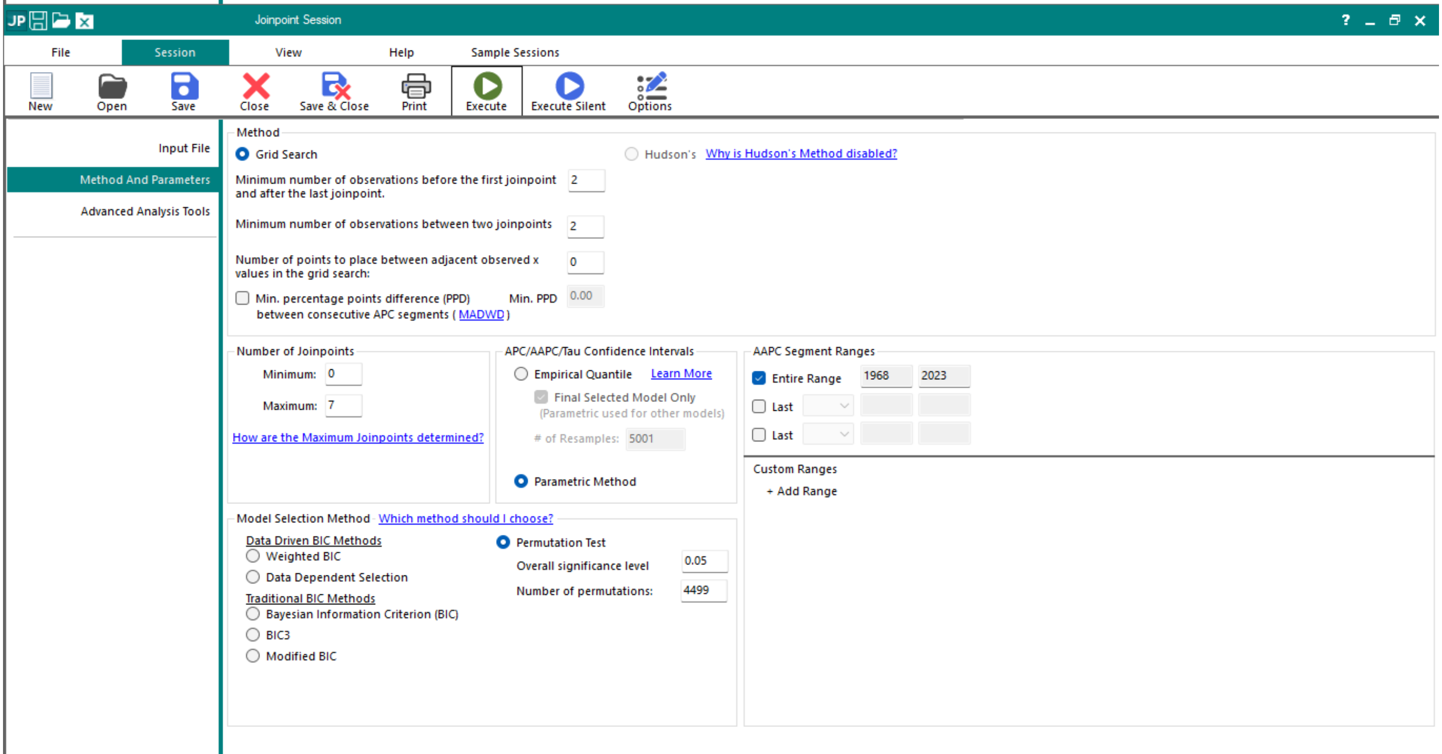
Number of Joinpoints
Minimum: 0
Maximum: 7
How are the Maximum Joinpoints determined?

APC/AAPC/Tau Confidence Intervals
● Empirical Quantile — Learn More
☑ Final Selected Model Only (Parametric used for other models)
\# of Resamples: 5001
○ Parametric Method

AAPC Segment Ranges
☑ Entire Range 1968 2023
☐ Last
☐ Last

Custom Ranges

Model Selection Method — Which method should I choose?
Data Driven BIC Methods
● Weighted BIC
○ Data Dependent Selection
Traditional BIC Methods
○ Bayesian Information Criterion (BIC)
○ BIC3
○ Modified BIC
○ Permutation Test
Overall significance level 0.05
Max number of permutations 4499



Method
● Grid Search ○ Hudson's — Why is Hudson's Method disabled?

Minimum number of observations before the first joinpoint and after the last joinpoint. 2

Minimum number of observations between two joinpoints 2

Number of points to place between adjacent observed x values in the grid search: 0

☐ Min. percentage points difference (PPD) between consecutive APC segments (MADWD) Min. PPD 0.00

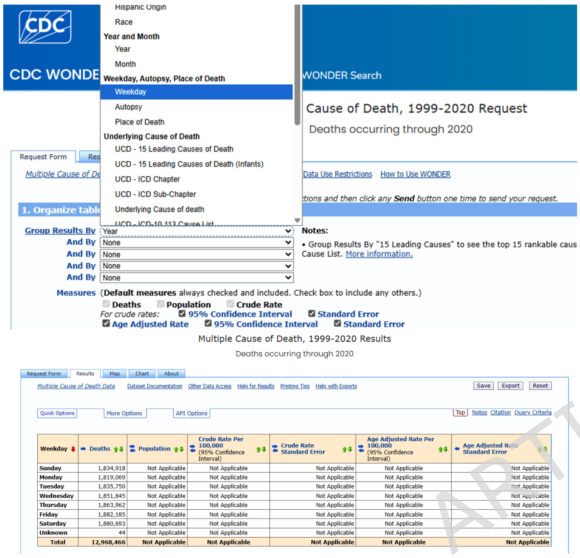
Number of Joinpoints
Minimum: 0
Maximum: 7
How are the Maximum Joinpoints determined?

APC/AAPC/Tau Confidence Intervals
○ Empirical Quantile — Learn More
☑ Final Selected Model Only (Parametric used for other models)
\# of Resamples: 5001
● Parametric Method

AAPC Segment Ranges
☑ Entire Range 1968 2023
☐ Last
☐ Last

Custom Ranges

Model Selection Method — Which method should I choose?
Data Driven BIC Methods
○ Weighted BIC
○ Data Dependent Selection
Traditional BIC Methods
○ Bayesian Information Criterion (BIC)
○ BIC3
○ Modified BIC
● Permutation Test
Overall significance level 0.05
Number of permutations: 4499



Cause of Death, 1999-2020 Request

Deaths occurring through 2020

tions and then click any **Send** button one time to send your request.

Group Results By Year
And By None
And By None
And By None
And By None

Notes:

- Group Results By "15 Leading Causes" to see the top 15 rankable caus Cause List. More information.

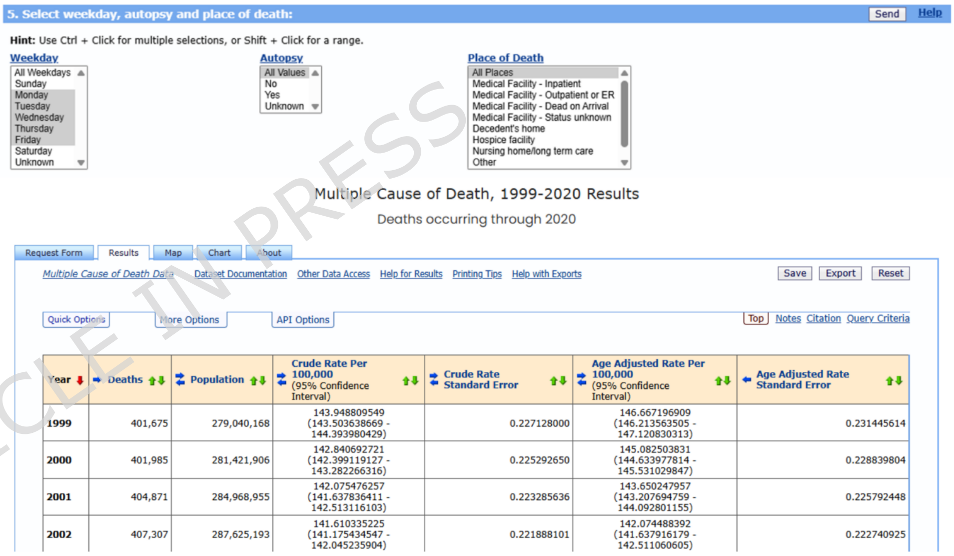
Measures (**Default measures** always checked and included. Check box to include any others.)

☑ **Deaths** ☑ **Population** ☑ **Crude Rate**
For crude rates: ☑ **95% Confidence Interval** ☑ **Standard Error**
☑ **Age Adjusted Rate** ☑ **95% Confidence Interval** ☑ **Standard Error**

Multiple Cause of Death, 1999-2020 Results

Deaths occurring through 2020

Weekday	Deaths	Population	Crude Rate Per 100,000 (95% Confidence Interval)	Crude Rate Standard Error	Age Adjusted Rate Per 100,000 (95% Confidence Interval)	Age Adjusted Rate Standard Error
Sunday	1,834,918	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Monday	1,819,069	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Tuesday	1,835,750	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Wednesday	1,851,845	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Thursday	1,863,962	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Friday	1,882,185	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Saturday	1,880,693	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Unknown	44	Not Applicable	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Total	**12,968,466**	**Not Applicable**	**Not Applicable**	**Not Applicable**	**Not Applicable**	**Not Applicable**



5. Select weekday, autopsy and place of death:

Hint: Use Ctrl + Click for multiple selections, or Shift + Click for a range.

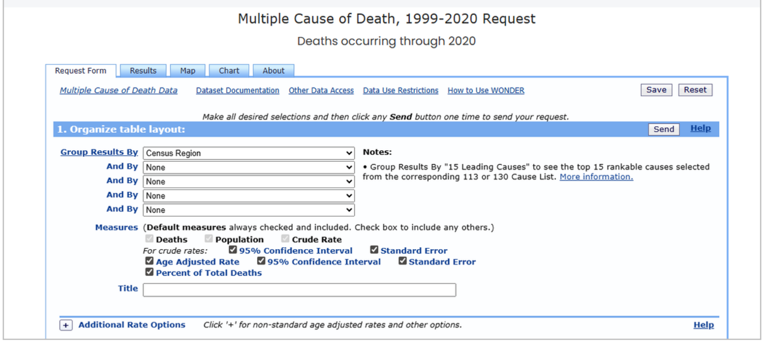
Weekday	Autopsy	Place of Death
All Weekdays Sunday Monday Tuesday Wednesday Thursday Friday Saturday Unknown	All Values No Yes Unknown	All Places Medical Facility - Inpatient Medical Facility - Outpatient or ER Medical Facility - Dead on Arrival Medical Facility - Status unknown Decedent's home Hospice facility Nursing home/long term care Other

Multiple Cause of Death, 1999-2020 Results

Deaths occurring through 2020

Year	Deaths	Population	Crude Rate Per 100,000 (95% Confidence Interval)	Crude Rate Standard Error	Age Adjusted Rate Per 100,000 (95% Confidence Interval)	Age Adjusted Rate Standard Error
1999	401,675	279,040,168	143.948809549 (143.503638669 - 144.393980429)	0.227128000	146.667196909 (146.213563505 - 147.120830313)	0.231445614
2000	401,985	281,421,906	142.840692721 (142.399119127 - 143.282266316)	0.225292650	145.082503831 (144.633977814 - 145.531029847)	0.228839804
2001	404,871	284,968,955	142.075476257 (141.637836411 - 142.513116103)	0.223285636	143.650247957 (143.207694759 - 144.092801155)	0.225792448
2002	407,307	287,625,193	141.610335225 (141.175434547 - 142.045235904)	0.221888101	142.074488392 (141.637916179 - 142.511060605)	0.222740925

Step 1. Organize table layout



Multiple Cause of Death, 1999-2020 Request

Deaths occurring through 2020

*Make all desired selections and then click any **Send** button one time to send your request.*

1. Organize table layout: Send Help

Group Results By Census Region
And By None
And By None
And By None
And By None

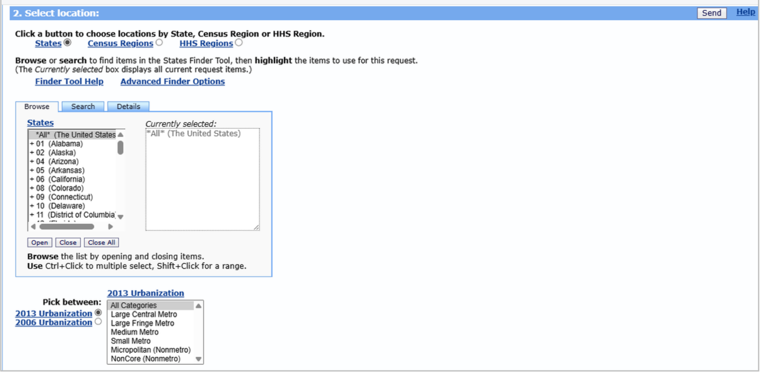
Notes:
- Group Results By "15 Leading Causes" to see the top 15 rankable causes selected from the corresponding 113 or 130 Cause List. More information.

Measures (**Default measures** always checked and included. Check box to include any others.)
☑ **Deaths** ☑ **Population** ☑ **Crude Rate**
For crude rates: ☑ **95% Confidence Interval** ☑ **Standard Error**
☑ **Age Adjusted Rate** ☑ **95% Confidence Interval** ☑ **Standard Error**
☑ **Percent of Total Deaths**

Title

+ **Additional Rate Options** *Click '+' for non-standard age adjusted rates and other options.* Help

Step 2. Select location



2. Select location: Send Help

Click a button to choose locations by State, Census Region or HHS Region.
States ● Census Regions ○ HHS Regions ○

Browse or **search** to find items in the States Finder Tool, then **highlight** the items to use for this request.
(The *Currently selected* box displays all current request items.)

States
All (The United States
+ 01 (Alabama)
+ 02 (Alaska)
+ 04 (Arizona)
+ 05 (Arkansas)
+ 06 (California)
+ 08 (Colorado)
+ 09 (Connecticut)
+ 10 (Delaware)
+ 11 (District of Columbia

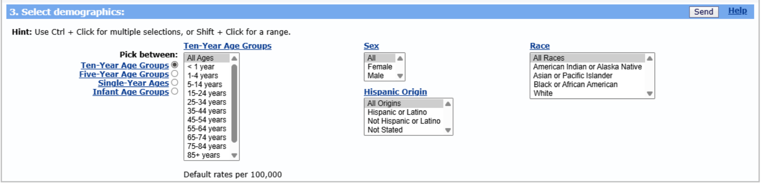
Currently selected:
All (The United States)

Browse the list by opening and closing items.
Use Ctrl+Click to multiple select, Shift+Click for a range.

Pick between:
2013 Urbanization ●
2006 Urbanization ○

2013 Urbanization
All Categories
Large Central Metro
Large Fringe Metro
Medium Metro
Small Metro
Micropolitan (Nonmetro)
NonCore (Nonmetro)

Step 3. Select demographics



3. Select demographics: Send Help

Hint: Use Ctrl + Click for multiple selections, or Shift + Click for a range.

Pick between:
Ten-Year Age Groups ●
Five-Year Age Groups ○
Single-Year Ages ○
Infant Age Groups ○

Ten-Year Age Groups
All Ages
< 1 year
1-4 years
5-14 years
15-24 years
25-34 years
35-44 years
45-54 years
55-64 years
65-74 years
75-84 years
85+ years

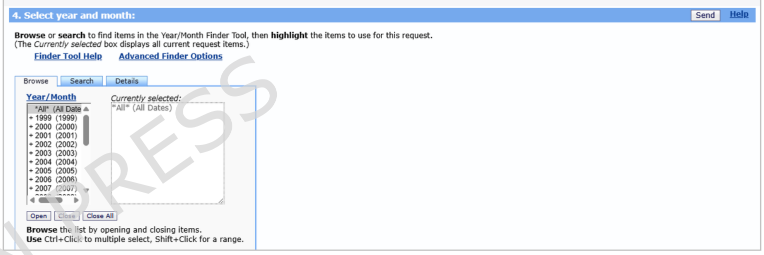
Default rates per 100,000

Sex
All
Female
Male

Hispanic Origin
All Origins
Hispanic or Latino
Not Hispanic or Latino
Not Stated

Race
All Races
American Indian or Alaska Native
Asian or Pacific Islander
Black or African American
White

Step 4. Select year and month



4. Select year and month: Send Help

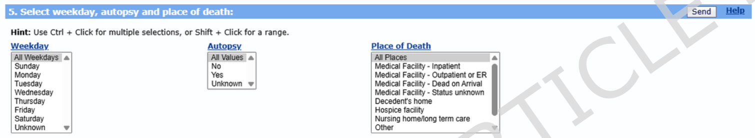
Browse or **search** to find items in the Year/Month Finder Tool, then **highlight** the items to use for this request.
(The *Currently selected* box displays all current request items.)

Year/Month
All (All Date
+ 1999 (1999)
+ 2000 (2000)
+ 2001 (2001)
+ 2002 (2002)
+ 2003 (2003)
+ 2004 (2004)
+ 2005 (2005)
+ 2006 (2006)
+ 2007 (2007)

Currently selected:
All (All Dates)

Browse the list by opening and closing items.
Use Ctrl+Click to multiple select, Shift+Click for a range.

Step 5. Weekday, autopsy, and place of death



5. Select weekday, autopsy and place of death: Send Help

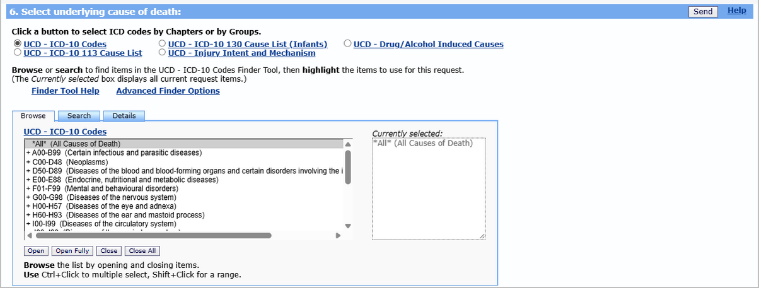
Hint: Use Ctrl + Click for multiple selections, or Shift + Click for a range.

Weekday
All Weekdays
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Unknown

Autopsy
All Values
No
Yes
Unknown

Place of Death
All Places
Medical Facility - Inpatient
Medical Facility - Outpatient or ER
Medical Facility - Dead on Arrival
Medical Facility - Status unknown
Decedent's home
Hospice facility
Nursing home/long term care
Other

Step 6. Select underlying cause of death



6. Select underlying cause of death: Send Help

Click a button to select ICD codes by Chapters or by Groups.
● UCD - ICD-10 Codes ○ UCD - ICD-10 130 Cause List (Infants) ○ UCD - Drug/Alcohol Induced Causes
○ UCD - ICD-10 113 Cause List ○ UCD - Injury Intent and Mechanism

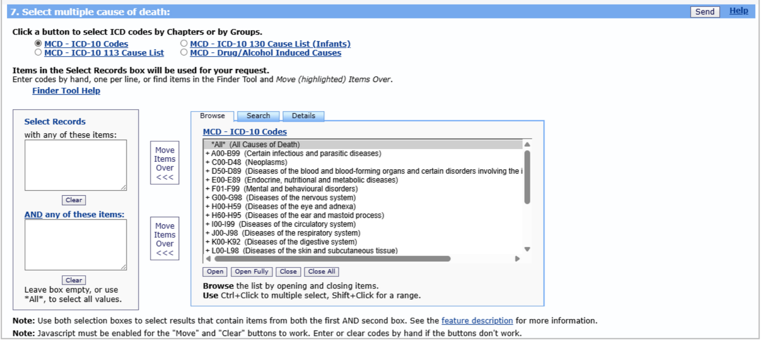
Browse or **search** to find items in the UCD - ICD-10 Codes Finder Tool, then **highlight** the items to use for this request.
(The *Currently selected* box displays all current request items.)

UCD - ICD-10 Codes
All (All Causes of Death)
+ A00-B99 (Certain infectious and parasitic diseases)
+ C00-D48 (Neoplasms)
+ D50-D89 (Diseases of the blood and blood-forming organs and certain disorders involving the i
+ E00-E88 (Endocrine, nutritional and metabolic diseases)
+ F01-F99 (Mental and behavioural disorders)
+ G00-G98 (Diseases of the nervous system)
+ H00-H57 (Diseases of the eye and adnexa)
+ H60-H93 (Diseases of the ear and mastoid process)
+ I00-I99 (Diseases of the circulatory system)

Currently selected:
All (All Causes of Death)

Browse the list by opening and closing items.
Use Ctrl+Click to multiple select, Shift+Click for a range.

Step 7. Select multiple cause of death



7. Select multiple cause of death: Send Help

Click a button to select ICD codes by Chapters or by Groups.
● MCD - ICD-10 Codes ○ MCD - ICD-10 130 Cause List (Infants)
○ MCD - ICD-10 113 Cause List ○ MCD - Drug/Alcohol Induced Causes

Items in the Select Records box will be used for your request.
Enter codes by hand, one per line, or find items in the Finder Tool and *Move (highlighted) Items Over*.

Select Records
with any of these items:

AND any of these items:

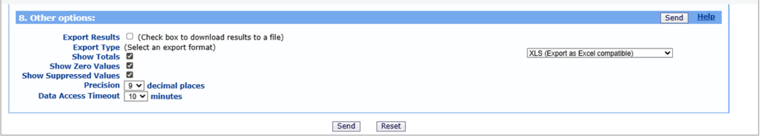
Leave box empty, or use *All*, to select all values.

MCD - ICD-10 Codes
All (All Causes of Death)
+ A00-B99 (Certain infectious and parasitic diseases)
+ C00-D48 (Neoplasms)
+ D50-D89 (Diseases of the blood and blood-forming organs and certain disorders involving the i
+ E00-E89 (Endocrine, nutritional and metabolic diseases)
+ F01-F99 (Mental and behavioural disorders)
+ G00-G98 (Diseases of the nervous system)
+ H00-H59 (Diseases of the eye and adnexa)
+ H60-H95 (Diseases of the ear and mastoid process)
+ I00-I99 (Diseases of the circulatory system)
+ J00-J98 (Diseases of the respiratory system)
+ K00-K92 (Diseases of the digestive system)
+ L00-L98 (Diseases of the skin and subcutaneous tissue)

Browse the list by opening and closing items.
Use Ctrl+Click to multiple select, Shift+Click for a range.

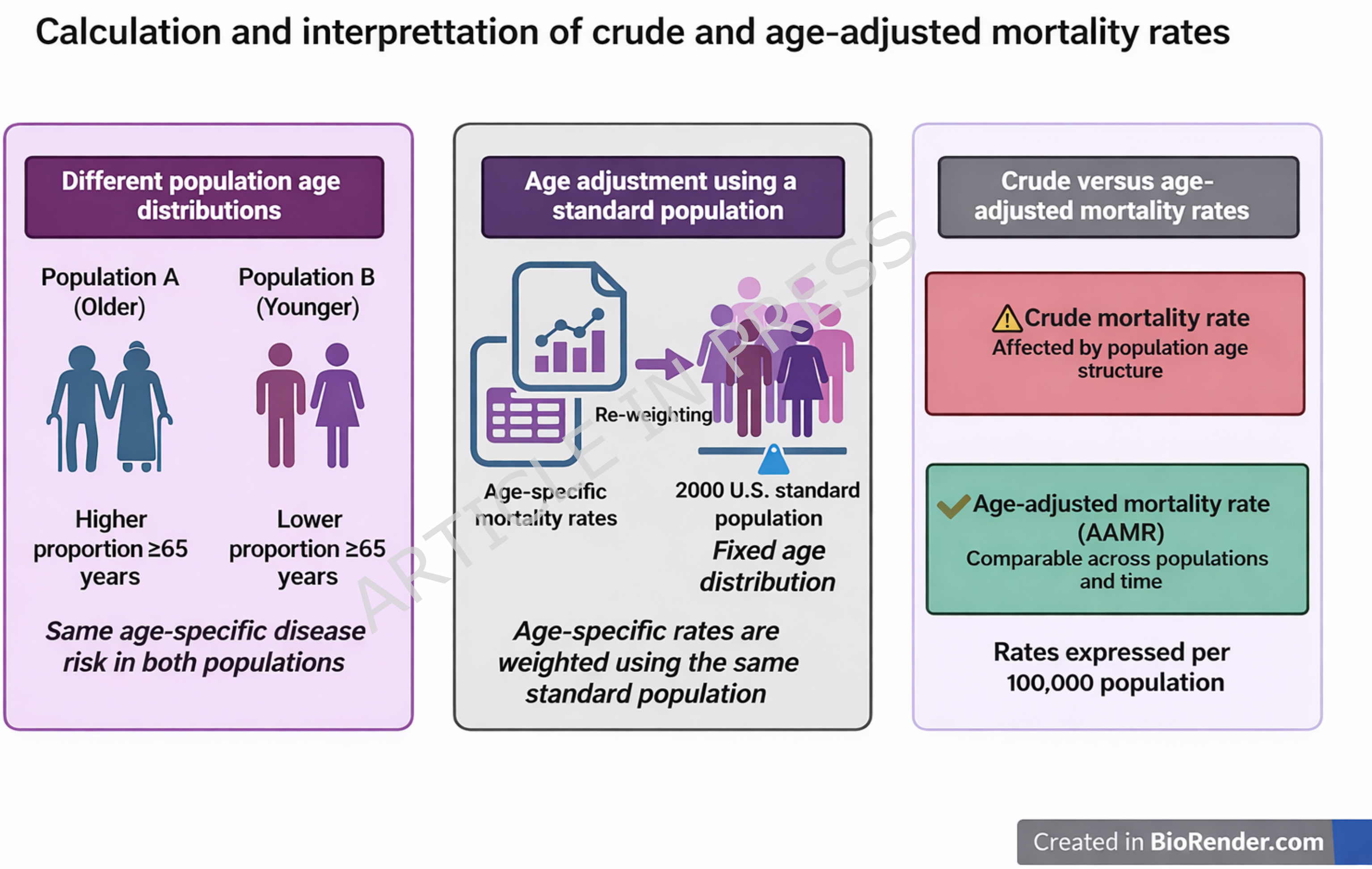
Note: Use both selection boxes to select results that contain items from both the first AND second box. See the feature description for more information.
Note: Javascript must be enabled for the "Move" and "Clear" buttons to work. Enter or clear codes by hand if the buttons don't work.

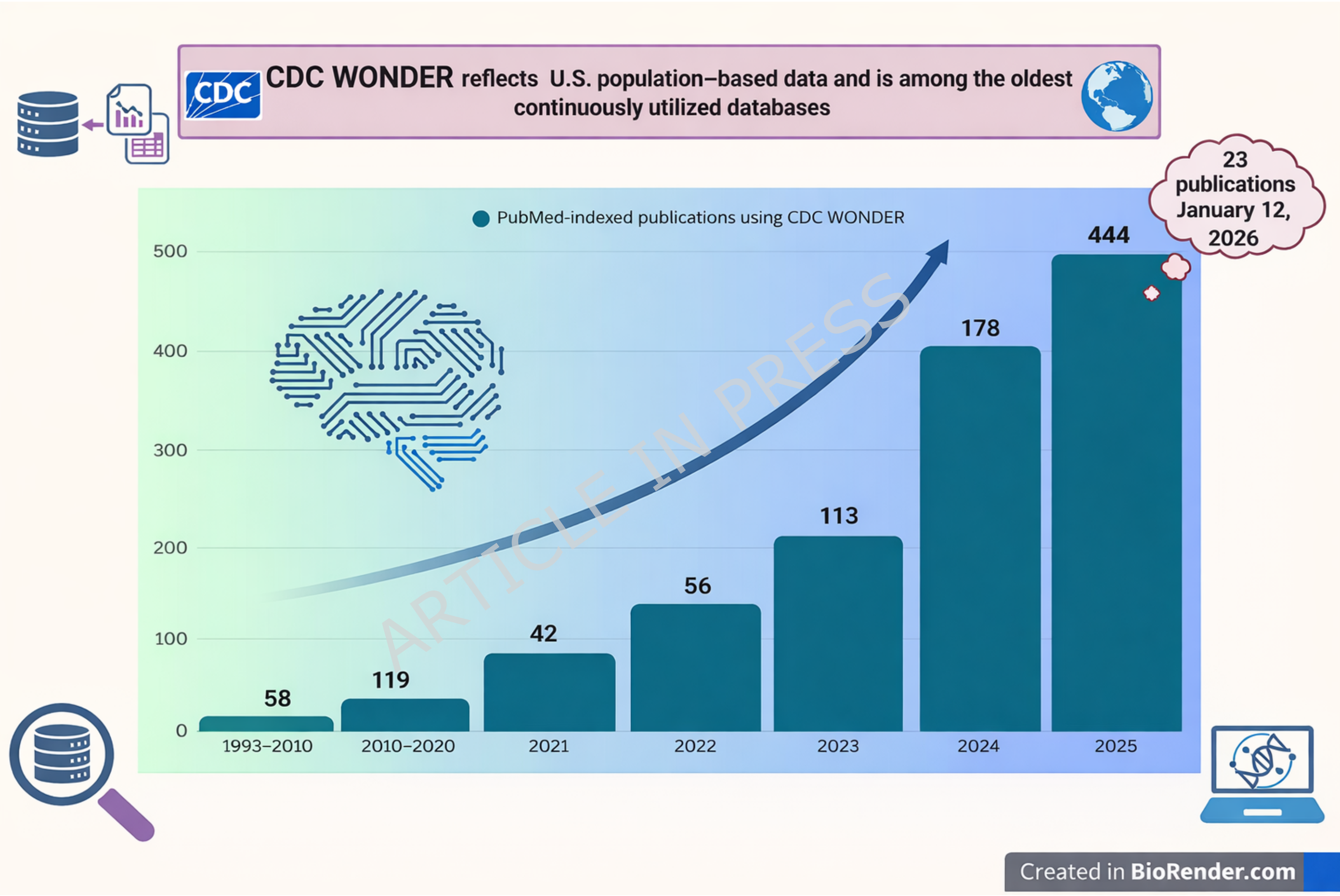
Step 8. Other options / export settings



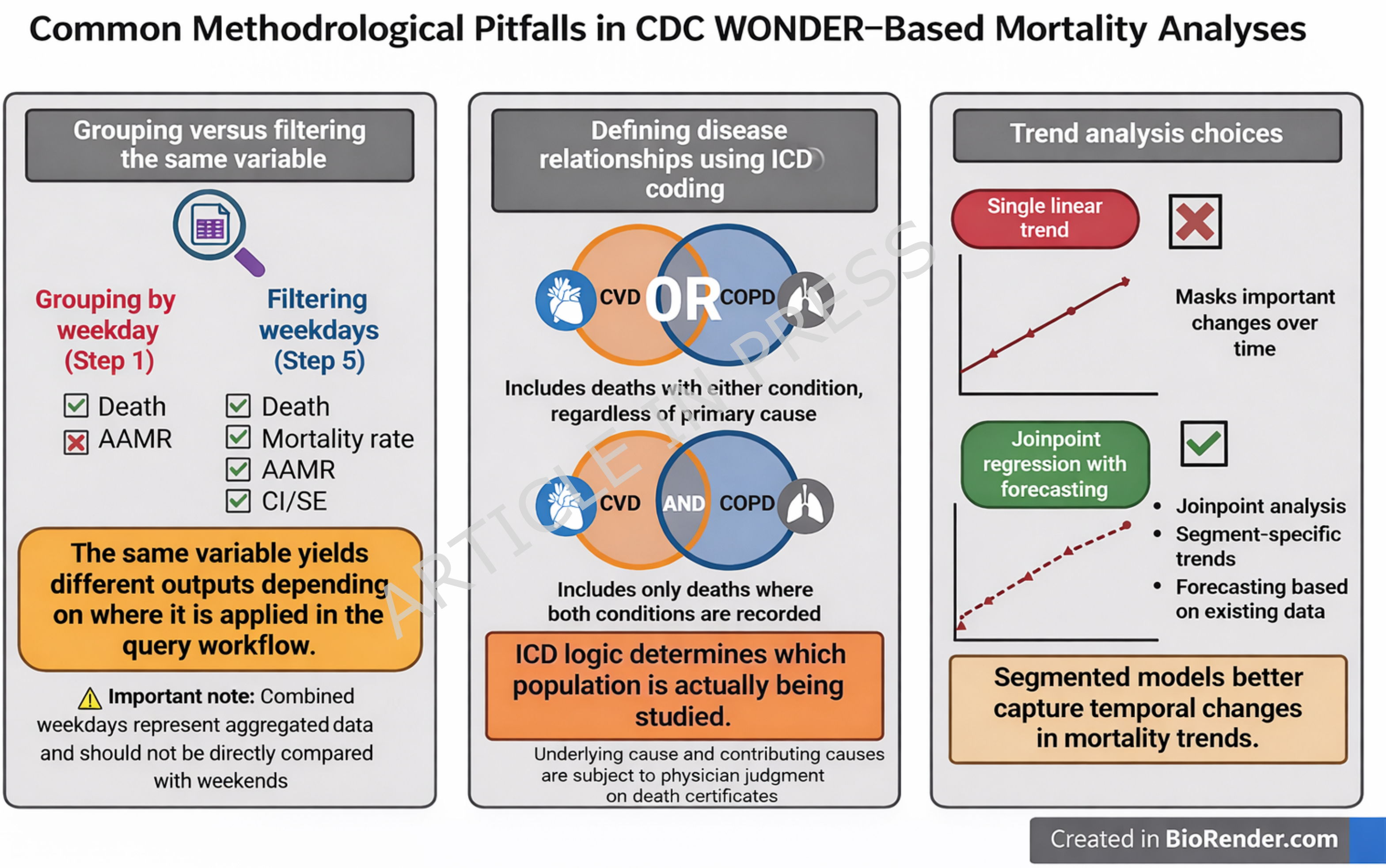
8. Other options: Send Help

Export Results ☐ (Check box to download results to a file)
Export Type (Select an export format) XLS (Export as Excel compatible)
Show Totals ☑
Show Zero Values ☑
Show Suppressed Values ☑
Precision 9 **decimal places**
Data Access Timeout 10 **minutes**


Calculation and interprettation of crude and age-adjusted mortality rates
Different population age distributions
Population A (Older)
Population B (Younger)
Higher proportion ≥65 years
Lower proportion ≥65 years
Same age-specific disease risk in both populations
Age adjustment using a standard population
Re-weighting
Age-specific mortality rates
2000 U.S. standard population
Fixed age distribution
Age-specific rates are weighted using the same standard population
Crude versus age-adjusted mortality rates
⚠ Crude mortality rate
Affected by population age structure
✓ Age-adjusted mortality rate (AAMR)
Comparable across populations and time
Rates expressed per 100,000 population
Created in BioRender.com



CDC
CDC WONDER reflects U.S. population–based data and is among the oldest continuously utilized databases
PubMed-indexed publications using CDC WONDER
23 publications January 12, 2026
500
400
300
200
100
0
58
119
42
56
113
178
444
1993–2010
2010–2020
2021
2022
2023
2024
2025
Created in BioRender.com



Common Methodological Pitfalls in CDC WONDER–Based Mortality Analyses

Grouping versus filtering the same variable

Grouping by weekday (Step 1)

☑ Death
☒ AAMR

Filtering weekdays (Step 5)

☑ Death
☑ Mortality rate
☑ AAMR
☑ CI/SE

The same variable yields different outputs depending on where it is applied in the query workflow.

⚠ **Important note:** Combined weekdays represent aggregated data and should not be directly compared with weekends

Defining disease relationships using ICD coding

Includes deaths with either condition, regardless of primary cause

Includes only deaths where both conditions are recorded

ICD logic determines which population is actually being studied.

Underlying cause and contributing causes are subject to physician judgment on death certificates

Trend analysis choices

Single linear trend ☒

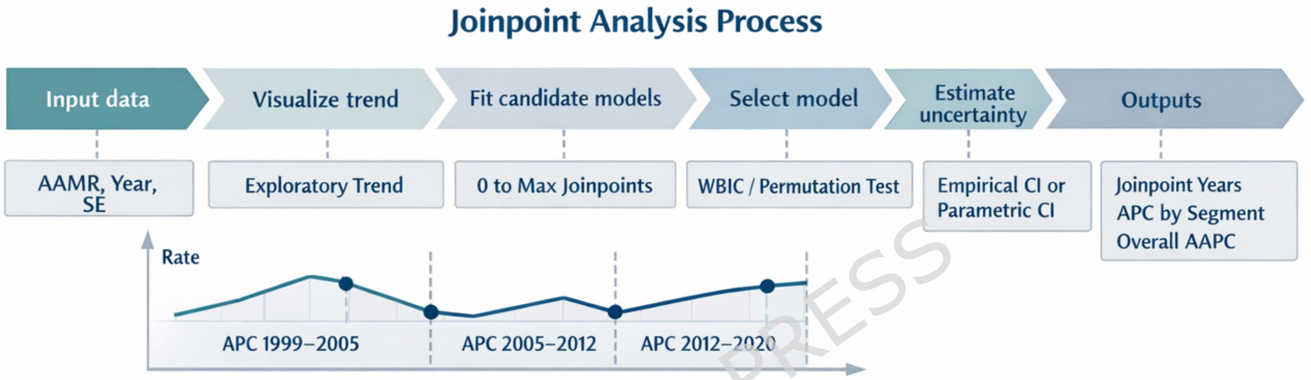
Masks important changes over time

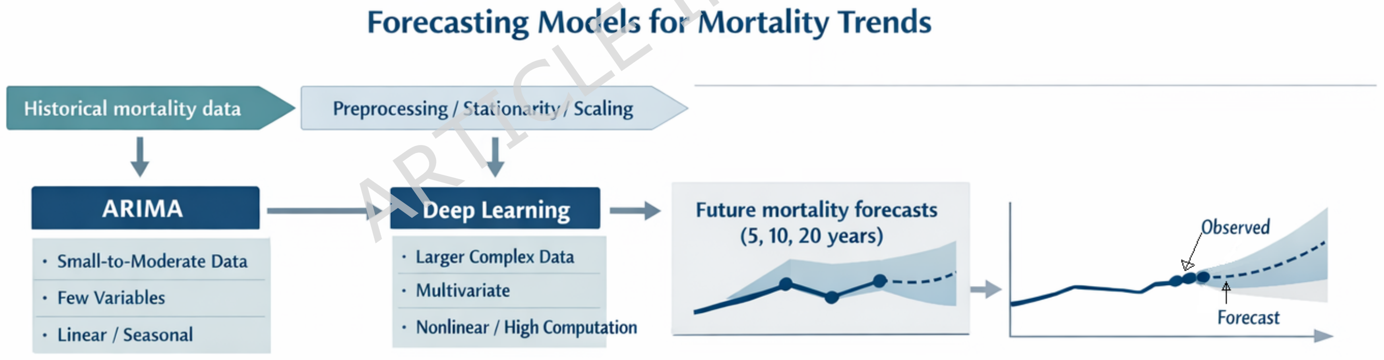
Joinpoint regression with forecasting ☑

- Joinpoint analysis
- Segment-specific trends
- Forecasting based on existing data

Segmented models better capture temporal changes in mortality trends.

Created in **BioRender.com**


Joinpoint Analysis Process
Input data
Visualize trend
Fit candidate models
Select model
Estimate uncertainty
Outputs
AAMR, Year, SE
Exploratory Trend
0 to Max Joinpoints
WBIC / Permutation Test
Empirical CI or Parametric CI
Joinpoint Years
APC by Segment
Overall AAPC
Rate
APC 1999–2005
APC 2005–2012
APC 2012–2020


Forecasting Models for Mortality Trends
Historical mortality data
Preprocessing / Stationarity / Scaling
ARIMA
Small-to-Moderate Data
Few Variables
Linear / Seasonal
Deep Learning
Larger Complex Data
Multivariate
Nonlinear / High Computation
Future mortality forecasts
(5, 10, 20 years)
Observed
Forecast