

Validation of Mortality Data Sources Compared to the National Death Index in the Healthcare Integrated Research Database

Aziza Jamal-Allial¹, Todd Sponholtz¹, Shiva K Vojjala², Mark Paullin², Anahit Papazian¹, Biruk Eshete², Seyed Hamidreza Mahmoudpour³, Patrice Verpillat³, Daniel C Beachler²

¹Safety & Epidemiology, Carelon Research, Newton, MA, USA; ²Safety & Epidemiology, Carelon Research, Wilmington, DE, USA; ³Merck Healthcare KGaA, Darmstadt, Germany

Correspondence: Daniel C Beachler, Carelon Research, 123 Justison Street, Suite 200, Wilmington, DE, 19801, USA, Email Daniel.beachler@carelon.com

Background: The National Death Index (NDI) is the gold standard for mortality data in the United States (US) but has a time lag and can be operationally intensive. This validation study assesses the accuracy of various mortality data sources with the NDI.

Methods: This validation study is a secondary analysis of an advanced cancer cohort in the US between January 2010 and December 2018, with an established NDI linkage. Mortality data sources, inpatient discharge, disenrollment, death master file (DMF), Center for Medicare and Medicaid Services (CMS), Utilization management data (U.M.), and online obituary data were compared to NDI.

Results: Among 40,692 patients, 25,761 (63.3%) had a death date using NDI; the composite algorithm had a sensitivity of 88.9% (95% CI = 88.5%, 89.3%), specificity was 89.1% (95% CI = 88.6%, 89.6%). At the same time, positive predictive value (PPV) was 93.4% (95% CI = 93.1%, 93.7%), negative predictive value (NPV) was 82.3% (95% CI = 81.7%, 82.9%), and when comparing each individual source, each had a high PPV but limited sensitivity.

Conclusion: The composite algorithm was demonstrated to be a sensitive and precise measure of mortality, while individual database sources were accurate but had limited sensitivity.

Plain Language Summary:

What is known on this topic:

- The National Death Index (NDI) is the gold standard for ascertaining death in the US, with several limitations, such as time lag and population restrictions.

What this study adds:

- A composite mortality algorithm of several sources of death was shown to be a sensitive and precise in death ascertainment compared to the NDI.
- Individual mortality sources, such as the Social Security Death Master File and online obituaries, had limited sensitivity on their own compared to the NDI.

Keywords: mortality, validation, National Death Index, real word evidence, obituary, claims

Introduction

Generating evidence using real-world evidence depends heavily on the underlying quality and information diversity data sources. Thus, using such data, eg, administrative claims or electronic medical records, which was not explicitly designed for research is challenging. Therefore, important information, eg, mortality data, is often only available with additional efforts, such as linkage. Several mortality data sources are utilized for research purposes, such as the United States (US) National Death Index (NDI) contains over 100 million death records in the US since 1979.¹ NDI is the gold standard of

death ascertainment with high sensitivity and specificity available for research with the ability to match other databases.^{1–6}

NDI is a robust resource due to its completeness; however, NDI has a data lag of up to two years, which limits its utility for research projects. In addition, obtaining NDI linkage is impossible for data sources unable to share personal identifiers and can be operationally intensive, particularly for large studies.^{1–5,7} Other sources of death date and cause are available, eg, the Social Security Administration's Death Master File (DMF) from the National Technical Information Service Social Security Administration (NTIS).⁸ However, a policy change resulted in rigorous and limited disclosure of the state death record data with limited exceptions. Historically, DMF contained 89 million records, and the policy change resulted in the removal of approximately 4.2 million records, lowered the annual records, and decreased the addition of new records by 1 million annually.⁷ Similarly, the Center for Medicare and Medicaid Services Vital Status files have also been used. However, the limitation of only including Medicare members makes it suitable as a stand-alone source for only a portion of the population.

Hospital discharge claims are another source of mortality data that captures in-hospital deaths precisely and does not capture other death cases;^{9–12} however, the decline in the number of deaths occurring during hospitalization¹³ makes this data source unable to capture an increasing proportion of deaths. Obituaries have been used for several decades as a mortality data source, for example, the field of occupational epidemiology has utilized this source, and it has been reported to have a high sensitivity.^{14,15} With obituaries becoming available online, their use increased as one of the sources for ascertaining mortality data and was shown to be a valid and reliable source,^{16,17} though it may be limited in sensitivity. We hypothesized that combining each of these six mortality data sources in a composite algorithm could provide a sensitive and specific measure for mortality that could be a reliable approach when NDI linkage is not possible. The main objective of our study is to assess the validity and completeness of six mortality data sources and a composite algorithm within the Healthcare Integrated Research Database (HIRD) compared to NDI data, ie, the gold standard, in the advanced-stage cancer patient population from 2010 to 2018.¹⁸

Methods

This study was performed using data from the HIRD, which contains medical and pharmacy claims data from health plan members across the US. The HIRD is approximately 88 million unique individuals with comprehensive medical and pharmacy claims since 2006, representing the commercial and Medicare advantage-insured US population. This is a secondary analysis of de-identified data from a cohort of individuals in the HIRD diagnosed with advanced cancer from 2010 to 2018. More recent data was not available from the NDI at the time of the analysis.

Several sources of death status (Yes/No) and death date (if available) were available, including:

- DMF: National Technical Information Service, Social Security Administration (NTIS) known as the Social Security Death Master File¹⁹
- Discharge: Hospital discharge status on medical claims (Discharge status codes, 20, 40, 41, 42) available through medical claims data from Elevance Health insurance plans
- Disenrollment: Health plan reason-for-disenrollment data is available through administrative data from Elevance Health insurance plans
- CMS: Center for Medicare and Medicaid Services
- U.M.: Utilization of management data
- Obituary: Online obituary data obtained through natural language processing^{17,20,21}

Dates obtained from a subset of online sources included only the month and year of death and were assumed to occur on the 15th of the month. Death data from these six available sources in the HIRD are updated monthly. NDI data were obtained for this study. Patients identified with death in 2018 were matched to NDI early-release data, which may need to be completed. The study protocol was reviewed and approved by the WCG Institutional Review Board (IRB) and the

National Center for Health Statistics (NCHS) overseeing the NDI. The time trends in the completeness and validity of the mortality sources compared to the NDI are described in a separate manuscript.²²

Data Analysis

For each of the six death sources, the difference in days between the death dates from a death source and the NDI was calculated among members with a death date.

In addition, the mean and standard deviation (SD) of the difference in dates and the proportion of members in different categories of difference between the dates from death source and NDI was calculated.

The mean, standard deviation, and proportion of participants within categories of sex, age (<18, 18–44, 45–64, ≥65 years), geographic region (Northeast, Midwest, South, West), and plan type (commercial, Medicaid advantage) for all and by death status (death source only, NDI death only, death source, and NDI death) were calculated.

Validation metrics were calculated to assess the performance of internally available death sources.^{23,24}

- Se = proportion of NDI death with an identifiable death in a death source (True positives/(True positives + false negatives))
- Sp = proportion of patients without an NDI death and no identifiable death in a death source (True negatives/(True negatives + false positives))
- PPV = proportion of patients with an identifiable death in a death source and NDI death (True positives/(True positives + false positives))
- NPV = proportion of patients without death in the death source and no NDI death (True negatives/(True negatives + false negatives))

The 95% confidence intervals (95% CI) were calculated using the binomial method. Death dates from death sources were a match with an NDI if they occurred 60 days before or 30 days after the NDI date.

We created a composite death data algorithm that assigned a death if a death was recorded in any of the six death sources. The death date of the composite algorithm was assigned using the following hierarchy of death sources, based on a) the highest proportion of internally sourced dates with NDI dates for members who had dates from death source and NDI, b) the proportion of members with internally sourced dates lacking an NDI date, and c) the proportion of members with an NDI date where the internally sourced date occurred within ±30 days. The same methods were followed for each death source. The hierarchy was defined as the DMF > Discharge > Disenrollment > CMS > U.M. > Online Obituary. If death was recorded in one data source (eg discharge) and not the others, the individual was recorded as a death and the death date was defined as the date from the source. When a death was recorded in multiple data sources, the death date was defined as the date recorded in the source highest in the hierarchy.

To assess the impact of the gap between the internally sourced data and NDI, different allowable gaps were used for sensitivity analysis: ±60, ±30, ±15, ±7 days of the NDI date. Analyses were performed using SAS[®] Enterprise Guide[®] 8.3; Cary, NC: SAS Institute Inc.

Results

Among a cohort of advanced cancer patients from 2010 to 2018, 27,396 (67.3%) had at least one death date. The mean age was 67.2 years with an SD of 13.5 years; 54.3% were women, and 56.1% were commercially insured (Table 1). Compared to patients without a death, those who had a death date in at least one death source were older (mean age 69.1 vs 63.1 years) and more likely to be male (49.7% vs 37.4%) and have a Medicare plan (49.8% vs 32.5%). Among all patients with death in NDI and death source, the mean difference in days was 1.83 (95% CI = 1.25–2.40). Death dates were available in 53.4% of obituaries, compared to 33.2% and 35.2% of discharge and DMF data sources (data is not shown).

Table 2 shows the distribution of the various death gaps between the composite mortality, six internal death sources, and NDI death date. For composite mortality, 84.7% of captured death dates were precisely matched to the NDI death date. In contrast, among all six sources, CMS and U.M. had the highest exact match death date, 99.3% and 91.5%,

Table 1 Baseline Characteristics for the Study Population

	No Observed Death Date	Observed Death Date	Overall
N(%)	13,296 (32.7%)	27,383 (67.3%)	40,679 (100.0%) ^a
Sex			
Female	8,321 (62.6%)	13,777 (50.3%)	22,098 (54.3%)
Male	4,975 (37.4%)	13,606 (49.7%)	18,581 (45.7%)
Age (years)			
Mean (SD)	63.1 (15.0)	69.1 (12.2)	67.2 (13.5)
<18	177 (1.3%)	14 (0.1%)	191 (0.5%)
18–44	966 (7.3%)	684 (2.5%)	1,650 (4.1%)
45–64	6,104 (45.9%)	9,052 (33.1%)	15,156 (37.3%)
65+	6,049 (45.5%)	17,633 (64.4%)	23,682 (58.2%)
Region ^b			
Midwest	3,780 (31.0%)	8,822 (37.6%)	12,602 (35.3%)
Northeast	2,154 (17.7%)	3,910 (16.7%)	6,064 (17.0%)
South	2,912 (23.9%)	6,155 (26.2%)	9,067 (25.4%)
West	3,356 (27.5%)	4,572 (19.5%)	7,928 (22.2%)
Insurance Type			
Commercial	8,238 (67.5%)	11,777 (50.2%)	20,015 (56.1%)
Medicare	3,971 (32.5%)	11,685 (49.8%)	15,656 (43.9%)

Notes: ^aAll baseline characteristics are missing for 11 members. ^bN missing: No observed death, 1094; Observed death date, 3924; Overall, 5018. ^cN missing: No observed death, 1087; Observed death date, 3921; Overall, 5008.

Table 2 The Difference in Days Between Each Death Date Source and the NDI Death Date

The Difference in Dates (Days)	Composite Death N (%)	DMF N (%)	Obituary N (%)	Discharge N (%)	Disenrollment N (%)	CMS N (%)	U.M. N (%)
> -60	29 (0.1%)	21 (0.2%)	61 (0.4%)	0 (0.0)	15 (0.4%)	0 (0.0)	0 (0.0)
-60 to -31	26 (0.1%)	< 5 (N/A)	< 5 (N/A)	0 (0.0)	59 (1.5%)	< 5 (N/A)	< 5 (N/A)
-30 to -16	179 (0.8%)	< 5 (N/A)	15 (0.11%)	< 5 (N/A)	969 (25.2%)	< 5 (N/A)	0 (0.0)
-15 to -8	692 (3.0%)	531 (5.9%)	969 (7.0%)	< 5 (N/A)	625 (16.3%)	< 5 (N/A)	0 (0.0)
-7 to -1	1109 (4.8%)	525 (5.8%)	980 (7.1%)	72 (0.8%)	590 (15.4%)	21 (0.3%)	114 (6.0%)
0	19,478 (84.7%)	7808 (86.1%)	9686 (70.1%)	4641 (54.3%)	1488 (38.8%)	6350 (99.3%)	1746 (91.5%)
1 to 7	620 (2.7%)	107 (1.2%)	969 (7.0%)	2375 (27.8%)	79 (2.1%)	5 (0.1%)	35 (1.9%)
8 to 15	623 (2.7%)	55 (0.6%)	1019 (7.4%)	915 (10.7%)	< 5 (N/A)	< 5 (N/A)	< 5 (N/A)
16 to 30	167 (0.7%)	8 (0.1%)	96 (0.7%)	452 (5.3%)	< 5 (N/A)	< 5 (N/A)	< 5 (N/A)
31 to 60	15 (0.1%)	< 5 (N/A)	< 5 (N/A)	85 (1.0%)	< 5 (N/A)	0 (0.0)	< 5 (N/A)
> 60	68 (0.3%)	10 (0.1%)	11 (0.1%)	6 (0.07%)	10 (0.3%)	11 (0.2%)	< 5 (N/A)

Abbreviations: NDI, National Death Index; DMF, Death master file; CMS, Center for Medicare and Medicaid Services; U.M., Utilization Management data.

respectively, followed by DMF and obituary, 86.1% and 70.1%, respectively. For discharge, 54.3% were an exact match, and 27.8% were one to seven days after the NDI date. For deaths obtained from disenrollment data, only 38.8% were an exact match, and 31.7% were reported one to fifteen days earlier than the NDI death date.

Table 3 shows the performance of the composite mortality database and the individual death sources. The composite mortality had sensitivity at 88.9% (95% CI = 88.5%, 89.3%), specificity at 89.1% (95% CI = 88.6%, 89.6%), PPV with 93.4% (95% CI = 93.1%, 93.7%), and NPV at 82.3% (95% CI = 81.7%, 82.9%). When comparing sources, their sensitivity and NPV varied dramatically; obituary data had the highest sensitivity and NPV, 53.4% (95% CI = 52.8%, 54.0%) and 54.0% (95% CI = 53.4%, 54.6%), respectively. While U.M. and disenrollment had the lowest sensitivity, 7.4% (95% CI = 7.1%, 7.8%) and 14.9% (95% CI = 14.4%, 15.3%), respectively, and the lowest NPV, 38.4% (95% CI = 37.9%, 38.9%), and 40.2% (95% CI = 39.7%, 40.7%), respectively. Regarding specificity, the U.M. and disenrollment death databases had the highest specificity at 99.3% (95% CI = 99.2%, 99.4%) and 98.8% (95% CI = 98.6%, 99.0%),

Table 3 Validation Performance Metrics of Death Data Sources Compared to the NDI

	Composite Death	DMF	Obituary	Discharge	Disenrollment	CMS	U.M.
Sensitivity	88.9% (88.5%, 89.3%)	35.2% (34.6%, 35.7%)	53.4% (52.8%, 54.0%)	32.7% (32.2%, 33.3%)	14.9% (14.4%, 15.3%)	24.8% (24.3%, 25.4%)	7.4% (7.1%, 7.8%)
Specificity	89.1% (88.6%, 89.6%)	97.7% (97.5%, 98.0%)	94.3% (93.9%, 94.7%)	97.5% (97.3%, 97.8%)	98.8% (98.6%, 99.0%)	97.7% (97.5%, 98.0%)	99.3% (99.2%, 99.4%)
PPV	93.4% (93.1%, 93.7%)	96.4% (96.0%, 96.7%)	94.2% (93.8%, 94.6%)	95.8% (95.4%, 96.2%)	95.4% (94.7%, 96.0%)	94.9% (94.4%, 95.4%)	94.6% (93.6%, 95.6%)
NPV	82.3% (81.7%, 82.9%)	46.6% (46.1%, 47.2%)	54.0% (53.4%, 54.6%)	45.7% (45.1%, 46.2%)	40.2% (39.7%, 40.7%)	43.0% (42.5%, 43.5%)	38.4% (37.9%, 38.9%)

Abbreviations: NDI, National death index; DMF, Death Master file; CMS, Center for Medicare and Medicaid Services; U.M., Utilization management data; PPV, Positive predictive value; NPV, Negative predictive value.

respectively. In addition, DMF and discharge had the highest PPV 96.4% (95% CI = 96.0%, 96.7%) and 95.4% (95% CI = 95.4%, 96.2%), respectively, while the obituary data had lower specificity and PPV 94.3% (95% CI = 93.9%, 94.7%) and 94.2% (95% CI = 93.8%, 94.6%), respectively.

[Supplemental Table 1](#) shows the performance of the death date gap segments between the composite mortality database and NDI. The highest sensitivity and NPV were observed for ± 60 days gap 89.0% (95% CI = 88.6%, 89.4%) and 82.4% (95% CI = 81.8%, 83.0%), respectively; and the lowest sensitivity and NPV were observed among the ± 7 days gap, 82.4% (95% CI = 81.4%, 82.8%) and 74.5% (95% CI = 73.9%, 75.2%), respectively. While specificity and PPV remained consistent across all death day gaps, 89.1% (95% CI = 88.6%, 89.6%) and 93.4% (95% CI = 93.1%, 93.7%), respectively.

Discussion

In our secondary analysis of cancer patients with an established NDI linkage between January 2010 and December 2018, the composite mortality database was highly sensitive and accurate in identifying deaths, 88.8% (95% CI = 88.4%, 89.2%) and 93.3% (95% CI = 93.0%, 93.6%), respectively, compared to NDI death dates. When we compared the composite mortality database to the individual death data sources, all had greater than PPV at or greater than 94.0%; however, each of their sensitivities was less than 54%. Thus, a composite of death data sources of the DMF, obituary, discharge, disenrollment, CMS, and U.M. could provide an accurate and more sensitive death assessment than a single source.

Mortality ascertainment is challenging because it is only partially ascertained directly in administrative claims or EMR. The two most populated death data sources were online obituary and DMF, but both still had limited sensitivities on their own. Our study suggests that using several mortality sources, including linkages to other sources, is more robust and accurate for the death assessment events than relying on a single death source alone. Historically, the DMF death database was previously considered a reliable and complete source of death information. However, with the 2011 policy change surrounding sharing state death registries, its sensitivity declined over time. Many reports noted that DMF has become unreliable as a single source of death data for research and that there is a need to find additional sources.^{3,8}

As previously reported, NDI is considered the gold standard for mortality ascertainment in the US¹⁻⁶ However, this source is time-lagged, and privacy restrictions of some populations prevent NDI linkage. Historically, occupational epidemiology utilized obituaries from industry-specific journals as a source of mortality.^{14,15} With advanced technology, such as natural language processing, the use of online obituaries for research is surging.^{16,18,20} However, other studies suggested that the under-capture of women and racial and ethnic minorities is a limitation in addition to its limited availability before 2015.²⁰ Even though NDI is the gold standard for death ascertainment, it has several limitations, such as using probabilistic matching,¹⁸ which could involve errors in matching the right individuals.

In addition, the time lag of NDI could create hurdles in utilizing the NDI data. For our study, the ascertainment of death dates in 2018 was impacted by the availability of NDI data as the early-release NDI death data was the only available data, which likely impacted the sensitivity of the composite death database. Other limitations of the NDI data

include its privacy restriction on some populations. In addition, the performance of the NDI varies in different populations, including women and minorities, which has been noted for the death sources.¹⁸

Though our study included a large population of patients with linked NDI data, one limitation is its population type: commercially and Medicare advantage-insured, advanced cancer older adults in the US. Thus, the performance of the composite mortality database and the various death databases might differ in other populations. This study is also limited in that it does not provide death data after 2018, so further research in assessing the performance of the composite mortality algorithm, especially in the COVID-19 era, is warranted.

Conclusion

Our study demonstrates that creating an algorithm combining DMF with additional death sources, such as discharge, disenrollment, and online obituary, provides an accurate and more sensitive assessment of death outcomes among advanced cancer subjects. The algorithm offers high sensitivity and accuracy compared to NDI and suggests that using a composite algorithm is particularly important when estimating the absolute risk of death in real-world database studies. Further research is needed to understand the performance of this algorithm in other populations and the COVID-19 era.

Ethics Approval Statement

This study was reviewed and approved by WCG IRB (formerly New England Institutional Review Board) review. The Board found that this research meets the requirements for a waiver of consent under 45 CFR 46.116(f) [2018 Requirements] 45 CFR 46.116(d). The National Center for Health Statistics (NCHS), a US Centers for Disease Control division overseeing the NDI, reviewed and approved this study.

Funding

This study was conducted and funded by a subsidiary of Elevance Health.

Disclosure

Dr Seyed Mahmoudpour is an employee of Merck KgaA, Darmstadt, Germany, outside the submitted work. Patrice Verpillat is an employee of Merck Group, during the conduct of the study. Dr Daniel Beachler is employed by Celeron Research, a subsidiary of Elevance Health. The authors report no other conflicts of interest in this work. The following manuscript is also available on pre-print servers (Authorea and ResearchGate).

References

1. Skopp NA, Smolenski DJ, Schwesinger DA, Johnson CJ, Metzger-Abamukong MJ, Reger MA. Evaluation of a methodology to validate National Death Index retrieval results among a cohort of U.S. service members. *Ann Epidemiol.* 2017;27(6):397–400. doi:10.1016/j.annepidem.2017.05.004
2. LaVeist TA, Diala C, Torres M, Jackson JS. Vital status in the National Panel Survey of Black Americans: a test of the National Death Index among African Americans. *J Natl Med Assoc.* 1996;88(8):501–505.
3. Navar AM, Peterson ED, Steen DL, et al. Evaluation of mortality data from the social security administration death master file for clinical research. *JAMA Cardiol.* 2019;4(4):375–379. doi:10.1001/jamacardio.2019.0198
4. Chase HC. Report on a national death index--pros and cons. *Am J Public Health.* 1972;62(5):719–723. doi:10.2105/AJPH.62.5.719
5. Lash TL, Silliman RA. A comparison of the National Death Index and Social Security Administration databases to ascertain vital status. *Epidemiology.* 2001;12(2):259–261. doi:10.1097/00001648-200103000-00021
6. Calle EE, Terrell DD. Utility of the National Death Index for ascertainment of mortality among cancer prevention study II participants. *Am J Epidemiol.* 1993;137(2):235–241. doi:10.1093/oxfordjournals.aje.a116664
7. Levin MA, Lin HM, Prabhakar G, McCormick PJ, Egorova NN. Alive or dead: validity of the social security administration death master file after 2011. *Health Serv Res.* 2019;54(1):24–33. doi:10.1111/1475-6773.13069
8. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the Death Master File. *Circ Cardiovasc Qual Outcomes.* 2013;6(1):124–128. doi:10.1161/CIRCOUTCOMES.112.968826
9. Conway RBN, Armistead MG, Denney MJ, Smith GS. Validating the matching of patients in the linkage of a large hospital system's EHR with state and national death databases. *Appl Clin Inform.* 2021;12(1):82–89. doi:10.1055/s-0040-1722220
10. Garry EM, Weckstein AR, Quinto K, et al. Categorization of COVID-19 severity to determine mortality risk. *Pharmacoepidemiol Drug Saf.* 2022;31(7):721–728. doi:10.1002/pds.5436
11. Rhee C, Dantes R, Epstein L, et al. Incidence and Trends of Sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA.* 2017;318(13):1241–1249. doi:10.1001/jama.2017.13836

12. (CDC-NCHS) CfDCNCfHS. In-hospital mortality among hospital confirmed COVID-19 encounters by week from selected hospitals. 2023. Available from: <https://www.cdc.gov/nchs/covid19/nchs/hospital-mortality-by-week.htm>. Accessed April 1, 2023.
13. QuickStats. Percentage of deaths, by place of death — national vital statistics system, United States, 2000–2018. *MMWR*. 2020;69:611. doi:10.15585/mmwr.mm6919a4
14. McDonald JA, Li FP, Mehta CR. Cancer mortality among beekeepers. *J Occup Med*. 1979;21(12):811–813.
15. Samkoff JS, Hockenberry S, Simon LJ, Jones RL. Mortality of young physicians in the United States, 1980–1988. *Acad Med*. 1995;70(3):242–244. doi:10.1097/00001888-199503000-00018
16. Buonanno P, Puca M. Using newspaper obituaries to “nowcast” daily mortality: evidence from the Italian COVID-19 hot-spots. *Health Policy*. 2021;125(4):535–540. doi:10.1016/j.healthpol.2021.01.006
17. Soowamber ML, Granton JT, Bavaghar-Zaeimi F, Johnson SR. Online obituaries are a reliable and valid source of mortality data. *J Clin Epidemiol*. 2016;79:167–168. doi:10.1016/j.jclinepi.2016.05.012
18. National Death Index user’s guide. 2013. Available from: https://www.cdc.gov/nchs/data/ndi/ndi_users_guide.pdf. Accessed February 4, 2025.
19. Service NTI. Death Master File. In.
20. Lerman MH, Holmes B, St Hilaire D, et al. Validation of a mortality composite score in the real-world setting: overcoming source-specific disparities and biases. *JCO Clin Cancer Inform*. 2021;5:401–413. doi:10.1200/CCI.20.00143
21. Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *J Affect Disord*. 2021;279:1–8. doi:10.1016/j.jad.2020.09.113
22. Sponholtz T, Jamal-Allial A, Vojjala SK, et al. Trends in the completeness and validity of sources of death data against the national death index from 2010 to 2018. *Pharmacoepidemiol Drug Saf*. 2024;33(8):e5870. doi:10.1002/pds.5870
23. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliability, and pitfalls in research and practice. *Front Public Health*. 2017;5:307. doi:10.3389/fpubh.2017.00307
24. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45–50. doi:10.4103/0301-4738.37595

Pragmatic and Observational Research

Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>

Dovepress
Taylor & Francis Group