Building High-Quality, Research Grade Multi-Source Mortality Databases in a Timely Manner

Kyle William McLean, PhD[1] Kirsty Macaulay, PhD[1]
[1]Veritas Data Research, Delaware, USA

Consolidating complementary mortality data sources materially improves death ascertainment coverage, completeness, timeliness and quality compared with any single source alone

MSR137

Research Question

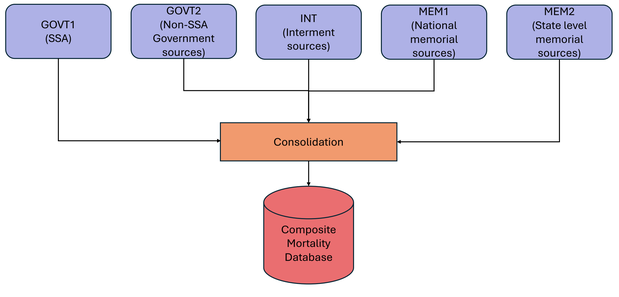
How can composite mortality databases enhance single-source equivalents across coverage, completeness, timeliness and data quality performance?

Why mortality data matters

- Reliable mortality ascertainment underpins real-world evidence across epidemiology, health policy, clinical outcomes, drug safety and health equity research
- Mortality is often treated as a "guaranteed" endpoint; however, **single-source mortality data frequently suffer from incomplete coverage, missing fields, reporting delays and data quality limitations**. Even government sources exhibit systematic gaps that can undermine analyses
- Composite mortality databases, constructed by consolidating multiple complementary sources, offer a pathway to improved performance – but the magnitude and consistency of these gains must be empirically demonstrated

Methods: Building a composite mortality database

- Death records were ingested from multiple government (govt), interment (int) and memorial (mem) data sources
- Records likely representing the same underlying death event were consolidated using internal linkage and deterministic reconciliation logic
- Four performance dimensions were evaluated at the individual death level: **Coverage, Completeness, Timeliness and Quality**
- Source families were incrementally consolidated along a *composition path*, and performance was assessed at each stage
- Results shown are representative of consistent trends observed across all composition paths



Results

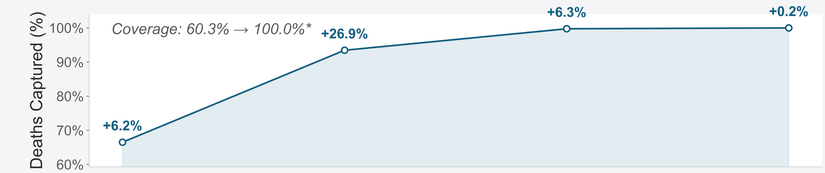
Summary — **Consolidating multiple sources into a single mortality database improves performance vs single source across all four metrics. The result is a research-grade mortality database[1,2]**

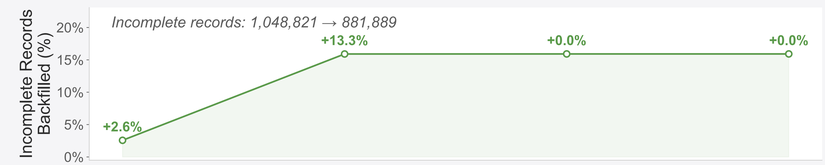
Coverage
Proportion of deaths detected by the source

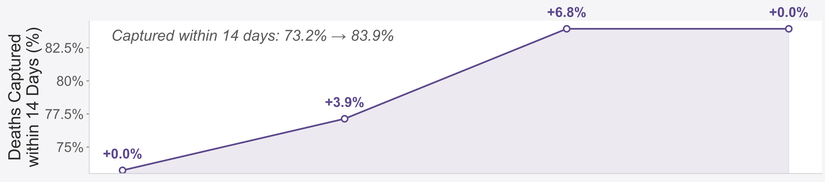
Completeness
Completeness of records pulled from source
Record is considered complete if it contains name, date of birth and date of death

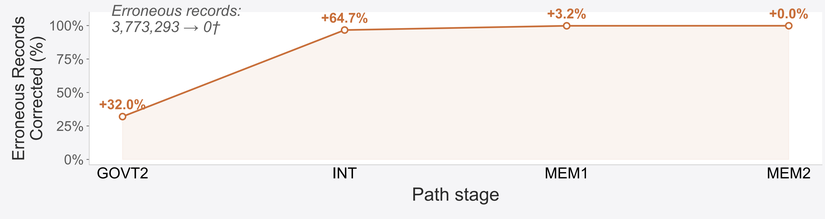
Timeliness
How quickly the source detects a death
In this study, deaths reported within 14 days are considered timely

Quality
Error rate in the death date listed by the source
Errors are identified via Veritas's consolidation engine









Results across four key metrics shown for an illustrative composition path: GOVT1->GOVT2->INT->MEM1->MEM2. Note that the completeness and quality plots are provided as % of records with missing or erroneous data from GOVT1, respectively. Although quantitative values differ, the trends highlighted are independent of composition path.
*Coverage calculated with respect to all deaths present in the final mortality dataset
†Erroneous records are identified via Veritas's consolidation engine, and the number of erroneous records will be 0 when all sources are included

Coverage

65.8% uplift in deaths identified

- Combining complementary sources substantially improves national-level death detection
- The resulting composite dataset **achieves very strong coverage on a national level**; consistent with aggregated CDC counts
- Memorial and interment sources provide the largest marginal gains by filling gaps not covered by administrative data

Completeness

15.9% backfill rate

- **15.9% of incomplete records (n = 1,048,821)** had missing information recovered through linkage with alternative sources
- Interment records were the primary contributors to completeness recovery; mainly due to a bias towards missingness in older deaths

Timeliness

+10% timely death captures

- Reporting delays in government sources result in fewer than **three-quarters of source deaths** being identified within 14 days
- Incremental consolidation with alternative sources substantially increases early detection
- Memorial sources provide the strongest timeliness uplift, often appearing shortly after death occurrence

Quality

> 3M erroneous records corrected

- Date of death discrepancies exist even in authoritative sources
- While government sources exhibit relatively low error rates, **systematic issues persist[3]**
- Cross-source corroboration improves confidence in recorded death dates and reduces reliance on any single source

Key takeaways

- **Single-source mortality databases exhibit meaningful limitations** across coverage, completeness, timeliness, and quality
- **Multi-source fusion consistently enhances performance** across all four dimensions
- Improvement trends are robust to the order in which sources are incrementally combined
- Composite mortality databases provide a **stronger foundation for real-world evidence and health outcomes research**

Limitations

- This study is limited to US mortality data
- Performance metric dependence on death year or geography is not considered
- This study does not seek to evaluate the accuracy of the composite mortality dataset
- Variation due to single-source origin is not shown here but is investigated in the full study (scan QR code)
- Records likely to represent the same underlying death are identified via Veritas's duplication engine; an evaluation of this algorithm is not provided here

References

1. Kapilivsky, J., Islam, F., Roth, E.K., Dow, J., Moran, S., Scherrer, E., Hyun, S.W. and Sangli, C. (2026). Validation of a Composite Mortality End Point in a Large Clinicogenomic Real-World Database of Patients With Advanced Cancer. *JCO Clinical Cancer Informatics*, 10(2)
2. Dong, S., Kansagra, A., Kaur, G., Barcellos, A., Belli, A.J., Fernandes, L.L., Hansen, E., Ambrose, J., Bai, C., Zettler, C.M., He, M. and Wang, C.-K. (2025). Validation of a Composite Real-World Mortality Variable Among Patients With Hematologic Malignancies Treated in the United States. *JCO Clinical Cancer Informatics*, (9)
3. Veritas Data Research (2025, April 11). *Veritas Data Research addresses SSA death data issues*. LinkedIn. *https://www.linkedin.com/posts/veritas-data-research_mortalitydata-veritasdataresearch-dataquality-activity-7316513210679922689--r7g*

Supplementary information

Supplementary materials can be found by scanning the adjacent QR code. This gives further insight into
- Results for all composition paths
- Performance within each evaluation dimension
- Insight into the consolidation engine