AUGMENTATION OF MORTALITY DATA USING DISPARATE REAL-WORLD DATA SOURCES: DOES MORE DATA = BETTER VALIDITY?

Authors: Elaine Liu[1], David Diakun[1], Nicole Princic[1], Ryan Ross[1], Liisa Palmer[1]
[1]MarketScan by Merative, Real World Data Research & Analytics, Ann Arbor, MI, USA

MarketScan by MERATIVE

Study Summary

Study Question: How can different real-world data sources reliably be used together to identify death information?

Study Design

Random sample, 6,000 patients in the MarketScan Commercial Claims and Encounters Database, age 18+

↓

Identify deaths in MarketScan (MSN) data

Identify deaths in Death Master File (DMF)

Identify deaths in National Death Index (NDI)

Evaluate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) for MSN, DMF identified deaths vs. NDI

Study Results: Accuracy for MSN-, DMF- and MSN+DMF- identified Deaths vs. NDI

	MSN	DMF	MSN + DMF
NPV (95% C.I.)	86% (85%, 87%)	80% (78%, 81%)	91% (90%, 92%)
PPV (95% C.I.)	93% (91%, 94%)	96% (94%, 98%)	93% (92%, 95%)
Specificity (95% C.I.)	99% (98%, 99%)	99.6% (99.4%, 99,8%)	98% (98%, 99%)
Sensitivity (95% C.I.)	54% (51%, 56%)	23% (21%, 26%)	71% (69%, 74%)
Date concordance (+/- 1 day)	97%	98%	98%

Conclusion: Using different real-world data sources together can improve reliable death ascertainment for real-world evidence (RWE) generation.

Background and Objective

- Mortality is an important outcome in pharmacoepidemiologic research yet is often missing or incomplete. Identifying reliable death data in real-world data (RWD) sources is difficult due to variability in quality, completeness and interoperability among different sources.[1,2]
- The objective of this study was to evaluate the added value in using disparate RWD sources to accurately identify presence and date of death.

Methods

Data Source

- This retrospective analysis utilized death information coming from inpatient discharge status on administrative claims data from the Merative[TM] MarketScan® Commercial and Medicare Database (MSN), employer-sourced death data from MSN and the Social Security Administration Death Master File (DMF).
- Patients were linked to the National Center for Health Statistics at the Center of Disease Control and Prevention's National Death Index (NDI), which holds death data from death certificates from vital records offices.

Study Design and Outcomes

- A random sample of 6,000 patients (age 18+) in MSN whose employer-sponsored insurance benefits ended in 2019 or 2022 (to optimize potential death identification and exclude active pandemic years) were identified.
- Patients with either an inpatient discharge status of "died" or death reported from the employer and the associated date of death were identified in MSN. Presence and date of death was identified in DMF.
- All patients were linked to the NDI for death ascertainment.
- Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were analyzed, using the NDI as the 'gold standard'.
- Date concordance, based on NDI reported death date, was evaluated.

Results

- Of the 6,000 randomly selected patients in the study sample, 1,150 deaths were identified using either MSN or DMF data.
  - 782 (68.0%) patients were identified in MSN only, 278 (24.2%) patients had a death documented in DMF only, and 90 (7.8%) patients were identified in both MSN and DMF (Figure 1).
- 1,504 deaths were identified in the NDI using the same patient sample. Of these, 808 (53.7%) were also identified using MSN data only while 1,073 (71.3%) were identified when MSN was augmented with DMF data (Table 1).

Figure 1. Proportions of deaths recorded in MSN, DMF, both datasets

- DMF Death Only
- MSN Death Only
- DMF & MSN Death
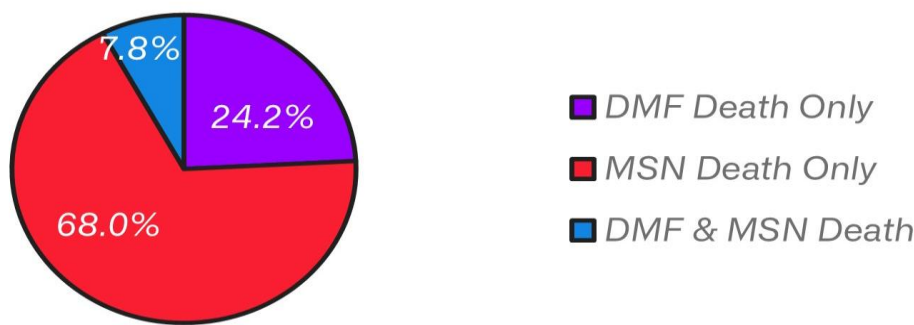
DMF: Death Master File; MSN: MarketScan; NDI: National Death Index

- When DMF and MSN death data were combined, accuracy remained similar or improved in comparison with either data source used alone when comparing with NDI-derived death data (Figure 2)

Table 1. MSN, MSN+DMF deaths linked to NDI deaths, 2019, 2022

	Death in NDI		
	Yes	No	Total
Death in MSN Only			
Yes	808	64	872
No	4,432	4,419	5,128
Total	5,240	696	6,000
Death in MSN or DMF			
Yes	1,073	77	1,150
No	431	4,419	4,850
Total	1,504	4,496	6,000

DMF: Death Master File; MSN: MarketScan; NDI: National Death Index

Figure 2. Accuracy, MSN, MSN+DMF deaths vs. NDI deaths, 2019, 2022
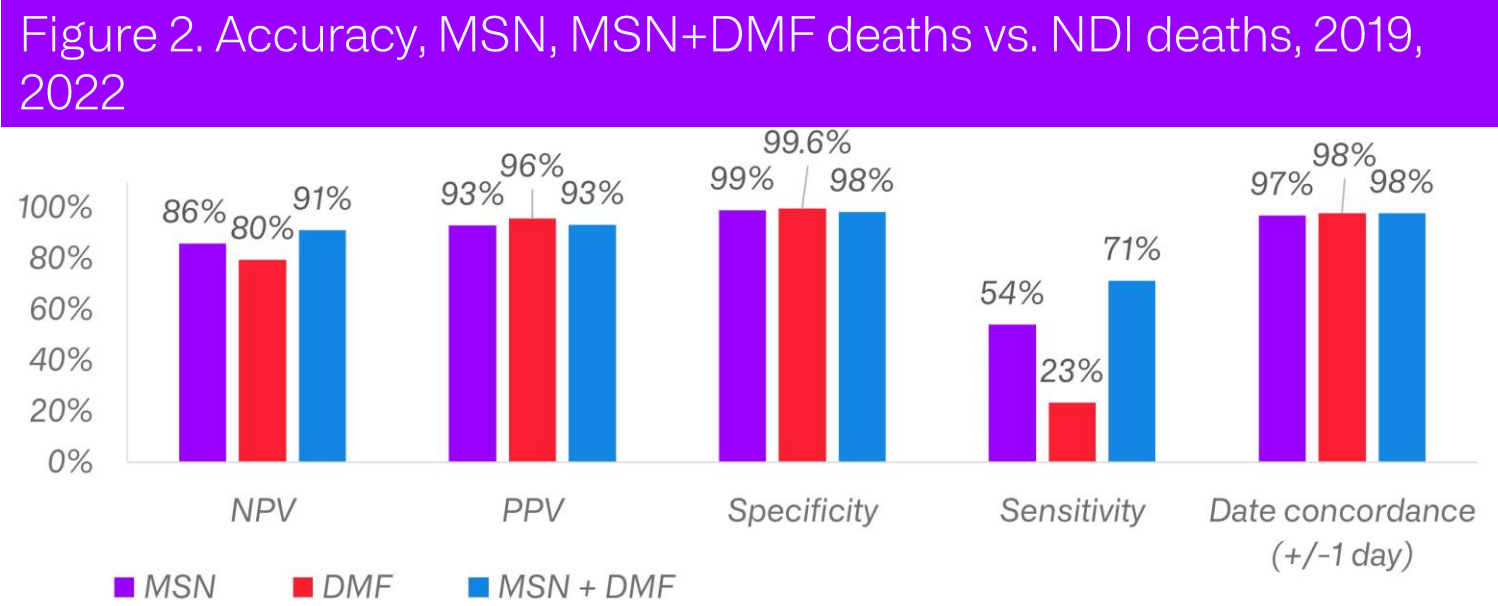
DMF: Death Master File; MSN: MarketScan; NDI: National Death Index, NPV: Negative predictive value; PPV: Positive predictive value

Limitations

- This study was based on a random sample of patients with commercial health coverage and with death data available from various resources, therefore it may not be generalizable to patients with other types of insurance or with no insurance.

Conclusions

- Although death information coming exclusively from employer data feeds may under capture death, death information captured (presence of death and specific date of death) are very robust. The inclusion of secondary death data sources, including the Social Security Death Master file, may be considered to increase the comprehensive capture of death for RWE generation.

References

1. Levin MA, et al. Health Serv Res. 2019 Jun;54(3):722.
2. Sholle ET, et al. AMIA Annu Symp Proc. 2024 Jan 11:2023:634-640. eCollection 2023.

MERATIVE