Development and validation of a high-quality composite real-world mortality endpoint

Melissa D. Curtis, MSc[1]; Sandra D. Griffith, PhD[1]; Melisa Tucker[1]; Michael D. Taylor, PharmD, PhD[2]; William B. Capra, PhD[2]; Gillis Carrigan, PhD[2]; Ben Holzman[1]; Aracelis Z. Torres, PhD, MPH[1]; Paul You, MPH[1]; Brandon Arnieri[2]; Amy P. Abernethy, MD, PhD[1]

[1]Flatiron Health, New York, NY, USA. [2]Genentech, South San Francisco, CA, USA.

Research objective

- Evidence based on real-world data derived from electronic health records (EHRs) can be leveraged for retrospective and prospective research[1].
- For many research questions, mortality data are critical for determining outcomes.
- Sources of mortality data are EHRs and publicly available national datasets, but both have drawbacks:
  - In EHRs, mortality data are frequently incomplete because EHRs are largely designed for managing patient care and tracking billing, and death date collection often is not part of clinic workflows.
  - Available national mortality datasets traditionally used for research are inadequate for many real-world evidence use cases due to recency and accessibility challenges.
- **We sought to develop a recent, linkable, high-quality EHR-derived mortality dataset supplemented with multiple data sources to fill gaps in structured data, and benchmark it to a gold standard[2].**

Study design

- The mortality dataset was generated by sequentially adding new sources of death information to resolve gaps in data. Structured EHR death data from the Flatiron Health Network were supplemented with:
  - external commercial data (Commercial Death Data 1; CDD1),
  - publicly available US Social Security Death Index (SSDI) data, and
  - abstracted information from unstructured documents within EHRs (ABS).
- Quality was assessed at each step of dataset development by measuring sensitivity, specificity, positive and negative predictive value (PPV and NPV), and date agreement, benchmarked against the most complete US mortality data, the National Death Index (NDI).

Table 1: Validation metrics for mortality data.

		NDI data		
		Deceased	**Alive**	
EHR-derived composite data	**Deceased**	True positives (A)[1]	False positives (B)[2]	**PPV** = A / (A + B)
	Alive	False negatives (C)[3]	True negatives (D)[4]	**NPV** = D / (C + D)
		Sensitivity = A / (A + C)	**Specificity** = D / (B + D)	

[1]True positives: all individuals with a death date in both the composite dataset and the NDI. [2]False positives: all individuals with a death date in the composite dataset but not in the NDI. [3]False negatives: individuals without a death date in the composite death dataset but with a death date in the NDI. [4]True negatives: all individuals who did not have a death date in the composite death dataset or in the NDI. **Sensitivity** indicated the percent of deaths in the NDI that were correctly recorded in the composite dataset, computed as the proportion of true positives among all the positives in the NDI gold standard [A/(A+C)]. **Specificity** indicated the percent of individuals without a death date in the NDI who were also not recorded as deceased in the composite dataset, computed as the proportion of true negatives among all the negatives in the NDI gold standard [D/(B+D)]. **PPV** indicated the percent of individuals with a death date in the composite dataset who were also considered dead in the NDI gold standard dataset [A/(A+B)]. **NPV** indicated the percent of individuals without a date of death in the composite dataset who were also not recorded as deceased in the NDI gold standard [D/(C+D)].

- The framework was developed for an advanced non-small cell lung cancer (advNSCLC) cohort.
- Subsequently the framework was extended to advanced melanoma (advMelanoma), metastatic colorectal cancer (mCRC), and metastatic breast cancer (mBC) cohorts, and sensitivity analyses were performed.
- The impact of mortality variable completeness on overall survival estimates was measured:
  - Using datasets of varying completeness created in the process of developing the final composite mortality dataset, overall survival was calculated to determine how it was impacted by quality of the mortality data.
  - For comparison, overall survival was also calculated for NDI data only, which in this study was the benchmark, and therefore was assumed to have 100% completeness.

Population studied

- Cohort inclusion criteria were:
  1. cancer diagnosis, as documented by ICD-9 or -10 code and confirmed in unstructured EHR data
  2. confirmation of advanced or metastatic disease by trained chart abstractors
  3. >1 oncologist visit on or after January 1, 2011
  4. diagnosis of advanced or metastatic disease on or after January 1, 2011 through December 31, 2015 for NSCLC, and from January 1, 2013 through December 31, 2015 for other tumor types. The December 31, 2015 cutoff was applied to align with NDI data availability.
- Sensitivity analyses included patients with the respective diagnoses after December 31, 2013 and through December 31, 2015, as data for advMelanoma, mCRC and mBC were available from this date (N=1622, N=7325 and N=3792, respectively).

Principal findings

- For advNSCLC, sensitivity of mortality information improved from 66% in EHR structured data alone to 91% in the composite dataset, with high date agreement compared to the NDI.
- For other cancer types, sensitivity of the final variable was 85-89%.
- In addition to these increases in sensitivity, high specificity was maintained (>96%).
- Kaplan-Meier survival analyses showed that improving mortality data completeness minimized overestimation of survival relative to NDI-based estimates.
- The practice-level mortality sensitivity showed stepwise improvement as additional data components were added, both in terms of increased median sensitivity and reduced variability across practices.

Table 2: Validation metrics during each step in the development of the mortality variable for the advNSCLC cohort.

	Sensitivity	Specificity	PPV	NPV	Date agreement (exact date)	Date agreement (±30 days)
Structured EHR only (EHR)	66.0% (64.8%, 67.1%)	97.1% (96.5%, 97.6%)	97.8% (97.4%, 98.2%)	58.8% (57.5%, 60.1%)	88.7% (87.7%, 89.7%)	97.0% (96.5%, 97.5%)
SSDI only	34.7% (33.6%, 35.9%)	99.1% (98.7%, 99.4%)	98.7% (98.2%, 99.1%)	43.2% (42.1%, 44.3%)	97.3% (96.6%, 97.9%)	98.5% (98.0%, 99.0%)
EHR-CDD1	84.1% (83.2%, 84.9%)	96.3% (95.6%, 96.9%)	97.8% (97.5%, 98.2%)	75.2% (73.9%, 76.4%)	92.33% (91.6%, 93.0%)	97.4% (97.0%, 97.8%)
EHR-CDD1 + SSDI	88.8% (88.1%, 89.6%)	96.1% (95.4%, 96.7%)	97.8% (97.5%, 98.2%)	81.1% (79.9%, 82.4%)	93.8% (93.2%, 94.5%)	97.5% (97.1%, 97.9%)
EHR-CDD1-SSDI + ABS (final variable)	90.6% (89.9%, 91.3%)	96.0% (95.3%, 96.7%)	97.8% (97.5%, 98.2%)	83.6% (82.5%, 84.8%)	93.5% (92.9%, 94.1%)	97.5% (97.1%, 97.9%)

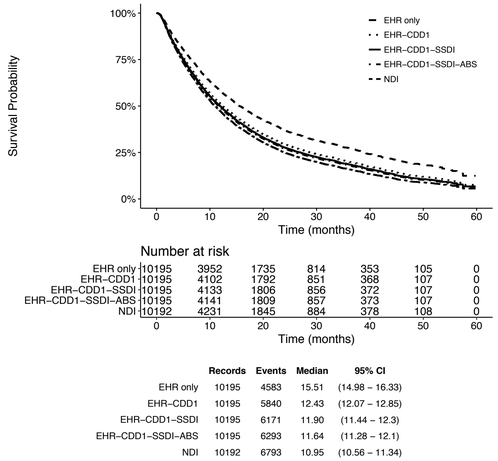
For each step in the variable development process, estimates and 95% CIs are shown. This cohort included patients diagnosed with advNSCLC on or after January 1, 2011 and through December 31, 2015 (N=10195).

Table 3: Validation metrics for different tumor types.

	Sensitivity	Specificity	PPV	NPV	Date agreement (exact)	Date agreement (±30 days)
advNSCLC (N=6840)	89.7% (88.7%, 90.6%)	97.3% (96.7%, 97.9%)	97.9% (97.4%, 98.4%)	87.1% (85.9%, 88.3%)	93.4% (92.6%, 94.2%)	97.5% (97.0%, 98.1%)
advMelanoma (N=1622)	88.4% (86.0%, 90.8%)	98.8% (98.2%, 99.5%)	98.2% (97.1%, 99.3%)	92.3% (90.7%, 94.0%)	95.4% (93.7%, 97.1%)	97.9% (96.7%, 99.0%)
mCRC (N=7325)	85.3% (84.0%, 86.5%)	98.2% (97.8%, 98.6%)	97.0% (96.3%, 97.6%)	90.9% (90.1%, 91.8%)	92.0% (90.9%, 93.1%)	96.7% (96.0%, 97.4%)
mBC (N=3792)	87.0% (85.1%, 88.8%)	98.5% (98.0%, 99.0%)	96.7% (95.7%, 97.7%)	93.7% (92.8%, 94.6%)	91.4% (89.7%, 93.1%)	96.3% (95.2%, 97.4%)

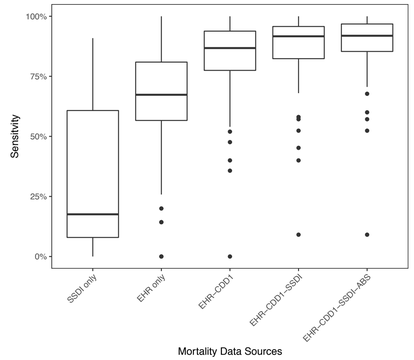
Data are shown for the final mortality variable that comprises structured EHR data, CDD1, SSDI and abstraction of unstructured EHR data. The cohorts here included patients with the respective diagnoses on or after January 1, 2013 and through December 31, 2015, as data for advMelanoma, mCRC and mBC were available from this date. The advNSCLC cohort was restricted to the same date range here to enable comparisons of data across the cohorts. 95% CIs are shown.

Figure 1: Overall survival for advNSCLC determined using indicated mortality data. NDI data were used as the benchmark in this study and were assumed to have 100% completeness. Patients were excluded from this analysis if their death date fell before the advanced diagnosis date.



	Records	Events	Median	95% CI
EHR only	10195	4583	15.51	(14.98 – 16.33)
EHR–CDD1	10195	5840	12.43	(12.07 – 12.85)
EHR–CDD1–SSDI	10195	6171	11.90	(11.44 – 12.3)
EHR–CDD1–SSDI–ABS	10195	6293	11.64	(11.28 – 12.1)
NDI	10192	6793	10.95	(10.56 – 11.34)

Figure 2: Sensitivity of advNSCLC data by practice. Data were restricted to practices with ≥100 patients. Boxplots show the median sensitivity, with lower and upper hinges of the boxes corresponding to the 25-75% interquartile range (IQR); lower and upper whiskers indicate sensitivity within 1.5 IQR of the lower and upper quantiles, respectively; and points outside of the whiskers show the rest of the data.



Limitations

- Since NDI was the gold standard, all validation metrics depend on the quality and recency of NDI data.
- The lack of NDI recency did not allow us to validate data after 2015.
- SSDI quality declines may lead to attenuated contribution to the composite mortality dataset in the future[3–5].
- The EHR-derived composite mortality variable does not include cause of death, which is less important for many cancer-related real-world evidence studies, but may be important for certain research questions and in other therapeutic areas.

Conclusions

- For EHR-derived data to be a reliable foundation of real-world evidence that enables healthcare quality improvement, it needs to be of known and sufficiently high quality.
- Here we improved quality of the mortality variable by linking different death datasets.
  - In addition to enhancing overall quality for the composite variable, adding in complementary data sources leads to more uniform quality across practices.
  - The high quality and recency of this variable makes it suitable for evaluating outcomes in oncology using retrospective and prospective studies that leverage real-world evidence.
- A remaining question is "what is good enough?" with respect to mortality quality. This will likely vary based on the use case. Future work will determine the potential of missing data to impact conclusions for various use cases, including at different sensitivity levels.
- We will continue to monitor data quality over time and evaluate additional sources of death data.

Implications for policy and practice

- Considering the impact of mortality data completeness on endpoints such as survival, we highlight the importance of data quality assessment and advocate for benchmarking to the NDI.
- In parallel, national efforts to define a usable, linkable mortality data source should continue.

References

1. Sherman RE, et al. New Engl J Med 2016; 375:2293–2297.
2. Curtis MD, et al. Health Services Res 2018: published online 14 May.
3. da Graca B, Filardo G, Nicewander D. Circulation 2013: 6:124–128.
4. Blackstone EH, J Thorac Cardio Surgery 2012: 143(1):37–38.
5. Sack K. New York Times, October 8, 2012.

Disclosure statement: M.D.C., S.G., M.T., B.H., A.Z.T., P.Y. and A.P.A. are employees of Flatiron Health, which is an independent subsidiary of the Roche Group. M.D.T., W.B.C., G.C. and B.A. are employees of Genentech, a member of the Roche group. No other disclosures.

Presented at the AcademyHealth Annual Research Meeting (June 24–26, 2018).
Presenting author: Sandra Griffith (sgriffith@flatiron.com)