**RESEARCH BRIEF**

# Validation analysis of a composite real-world mortality endpoint for patients with cancer in the United States

Qianyi Zhang MS  |  Anala Gossai PhD, MPH  |  Shirley Monroe MS  |
Nathan C. Nussbaum MD  |  Christina M. Parrinello PhD, MPH

Flatiron Health Inc, New York, New York, USA

**Correspondence**
Qianyi Zhang, MS, Flatiron Health Inc.,
233 Spring St, New York, NY 10013.
Email: qianyi.zhang@flatiron.com

**Abstract**

**Objective:** We expanded the previous assessment of a mortality variable suited for real-world evidence-focused oncology research.

**Data source:** We used a nationwide electronic health record (EHR)-derived de-identified database.

**Data collection:** We included patients with at least 1 of 18 cancer types between January 1, 2011 and December 31, 2017. Patient-level structured data (EHRs, obituaries, and Social Security Death Index) and unstructured EHR data (abstracted) were linked to generate a composite mortality variable.

**Study design:** We benchmarked sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and ±15-day agreement against the National Death Index (NDI). Real-world overall survival (rwOS) was estimated using the Kaplan-Meier method. We performed sensitivity analyses using a smaller patient cohort that underwent next-generation sequencing testing.

**Principal findings:** Compared with the NDI across 18 cancer types (overall N = 160 436): sensitivity, 83.9%-91.5% (17/18 cancer types had sensitivity ≥85.0%); specificity, 93.5%-99.7%; PPV, 96.3%-98.3%; NPV, 75.0%-98.7%; ±15-day agreement, 95.6%-97.6%; and median rwOS estimates ranging from 2.8% to 12.7% greater. Sensitivity analysis results (n = 17 540) were consistent with the main analysis.

**Conclusions:** Across all cancer types analyzed, this composite mortality variable showed high sensitivity, specificity, PPV, NPV, and ±15-day agreement, and yielded median rwOS values modestly overestimated when compared to NDI-based results.

**KEYWORDS**
mortality, observational data, overall survival, real-world data, real-world evidence

**What is known on this topic?**

- The utility of real-world evidence depends on the quality of the underlying data and the integrity of the analytic methods deployed for its generation, therefore demonstrating the validity and accuracy of clinical endpoints is important.

- In oncology research, mortality, as a variable, and overall survival, as an endpoint, are critical, since low sensitivity in mortality surveillance is known to bias overall survival estimates.
- The National Death Index represents the gold standard in mortality data sources, but it is not always sufficiently recent or accessible for contemporary observational research.

**What this study adds?**
- We report refreshed and expanded results obtained with a novel composite real-world mortality variable for oncology studies, generated from multiple structured and unstructured data sources.
- This variable shows high sensitivity and accuracy, and enables the reliable analysis (with negligible bias) of overall survival as an endpoint in large real-world cohorts of patients with cancer.
- The availability of this variable and associated endpoint unlocks the potential application of electronic health record-derived data for multiple research purposes, including comparative effectiveness or generation of external cohorts as contextual references for single-arm clinical trials.

## 1 | INTRODUCTION

As the complexity of clinical research grows, so does the need for additional investigative tools. Real-world data (RWD) refers to the clinical data collected in the course of routine care, via platforms such as electronic health records (EHRs), administrative claims, and/or clinical registries.[1] Recently, real-world evidence (RWE), namely, the clinical insights generated by analyzing those data, has been postulated as a complement or supplement to evidence gathered from clinical trials. Traditionally, RWD have been deployed in areas such as epidemiology or pharmacovigilance. But technologic and methodological capabilities to accrue and analyze data continue to improve, and the potential to use RWD and RWE to support clinical development programs, validate clinical trial findings at a large scale, or to support regulatory or reimbursement decisions is increasing.[2,3] Ultimately, the utility of RWE depends on the quality of the underlying RWD and the integrity of the analytic methods deployed for its generation.[4] Therefore, demonstrating the validity and accuracy of clinical endpoints becomes important.

In oncology (and other potentially fatal diseases), mortality surveillance and associated endpoint analyses (overall survival [OS]) are key clinical research components. In the United States, the National Death Index (NDI) has been the traditional gold-standard mortality data source.[5,6] However, full NDI updates are released only yearly with data delays of up to 2 years, which limits the use of this source as a reference for analyses with high recency. Additionally, substantial NDI use restrictions may limit its accessibility. Historically, the also-public Social Security Death Index (SSDI) served as an alternative, but the 2011 reporting modifications removed some state-sourced data from the SSDI and reduced its overall completeness.[7]

To address the gap in suitable mortality RWD sources, researchers have turned to commercial obituary repositories or EHR data,[8] however, these individual sources have their own shortcomings and their completeness may not be sufficient to support rigorous analyses. As a solution, the combination of multiple mortality data sources may improve the performance of single-source-derived data. Prior work from our team characterized a novel real-world mortality variable for oncology studies, generated as a composite of structured and unstructured EHR-derived data, obituary data (OD), and the SSDI.[9] That report presented validity metrics benchmarking this mortality variable against the NDI in patients with at least one of four cancer types (advanced non-small cell lung cancer [aNSCLC], metastatic colorectal cancer [mCRC], metastatic breast cancer [mBC], and advanced melanoma [aMel]). This present report expands on that prior work by refreshing the results with more recent data for cancer types previously reported, and evaluating this variable across 14 additional cancer types (18 cancer types in total).

## 2 | METHODS

### 2.1 | Data source

This study used the nationwide longitudinal Flatiron Health EHR-derived de-identified database. During the study period, the de-identified data originated from approximately 265 US cancer clinics (∼800 sites of care).[10] The main analysis included patients with at least one of the following 18 cancer types: early breast cancer, mBC, chronic lymphocytic leukemia (CLL), mCRC, diffuse large B-cell lymphoma (DLBCL), advanced gastro-esophageal cancer, hepatocellular carcinoma, advanced head and neck cancer, aMel, multiple myeloma,

malignant pleural mesothelioma (MPM), aNSCLC, ovarian cancer, metastatic pancreatic cancer, metastatic prostate cancer, metastatic renal-cell carcinoma (mRCC), small cell lung cancer (SCLC), and advanced urothelial cancer (additional selection criteria in Suppl. Table 1), with diagnosis documented between January 1, 2011 (January 1, 2013 for mCRC, mProstate, or SCLC, and January 1, 2014 for metastatic pancreatic cancer; documentation of diagnosis *or treatment* was acceptable for CLL) and December 31, 2017 (inclusive). In addition to the main analysis, a sensitivity analysis of the validity metrics was conducted in a cohort of patients sourced from a database of patients who underwent FoundationOne next-generation sequencing tests for their tumors (as part of routine clinical care).[10] This cohort included patients with the 18 cancer types in the main analysis as well as a pooled group of patients with other cancer types, considered as a pan-tumor category.

The study was IRB-approved with a waiver of informed consent.

## 2.2 | Variable

We used multiple RWD sources to generate a composite mortality variable defining vital status (dead/alive) and date of death. The sources were de-identified patient-level structured and unstructured data from the EHR, curated via technology-enabled abstraction, OD, and the SSDI. Manual abstraction of unstructured information was used for cases where death date was not available in the structured sources and there was no recent EHR activity (eg, in the past 60 days).[9]

For subsequent validation analyses, Flatiron Health and NDI records were matched using the NDI-developed probabilistic approach[11] including social security number, first and last name, middle initial, father's surname, sex, race, marital status, state (birth and residence), and date of birth.

## 2.3 | Analyses

Analyses were conducted in each of the 18 cancer types separately and overall, and stratified by the following sociodemographic and clinical characteristics: practice type (academic, community), practice site (for those with ≥100 patients), age group at cohort entry (<35, 35-49, 50-64, 65-74, and ≥75 years), race/ethnicity (White, Black or African American, Hispanic or Latino, Asian, and other/missing), region (Midwest, Northeast, South, West, and other/missing), number of lines of therapy received (0 and among treated patients, the following three separate binary groupings: <3 vs ≥3, <4 vs ≥4, and <5 vs ≥5), timing of NDI-recorded death or last confirmed activity by 6-month interval (2017 H2, 2017 H1, 2016 H2, etc).

Using the NDI as the gold standard, we calculated validity metrics for a series of comparators: the composite mortality variable (comprised of SSDI, OD, structured EHR data, and unstructured EHR data), as well as all single-source and combination components (structured EHR only, OD only, SSDI only, structured EHR + OD, structured EHR + SSDI, OD + SSDI, and structured EHR + OD + SSDI). The metrics calculated were sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and date agreement (exact, ±15-days, and ± 30-days).

Sensitivity was calculated as the percentage of deaths in the NDI that were correctly identified as such by the comparator. Specificity, as the percentage of patients alive (without a death date) in the NDI that were correctly identified as non-deceased by the comparator. PPV, as the percentage of deaths in the comparator that were truly deaths in NDI data. NPV, as the percentage of patients alive (without a death date) in the comparator that were alive (without a death date) in NDI data. Date agreement analyses were restricted to include only patients who had a death date in the comparator. The absence of an NDI death date was considered a disagreement; 15-day date agreement was calculated as the percentage of death dates in the comparator that matched a record in NDI data within a ±15-day window.

For all comparators, including the composite mortality variable, and NDI data, we generated Kaplan-Meier curves and median real-world (rw)OS estimates, using the relevant cohort entry date as the index date (depending on the cancer type: initial diagnosis date, advanced diagnosis date, or metastatic diagnosis date [Suppl. Table 1]), and using the death date as the event date. We used the most recent structured data entry documenting a visit, or the last abstracted end date for oral medications (if available) as the censor date.

**TABLE 1** Ranges of validity metrics for the composite mortality variable across the 18 cancer type-specific cohorts

| | | Composite Mortality Variable (%)[a] | Structured EHR Only (%) | OD Only (%) | SSDI Only (%) |
|---|---|---|---|---|---|
| Sensitivity | | 83.9-91.5 | 54.0-70.7 | 53.8-67.2 | 17.7-32.3 |
| Specificity | | 93.5-99.7 | 95.7-99.9 | 96.9-99.8 | 98.5-99.9 |
| PPV | | 96.3-98.3 | 97.3-98.7 | 96.2-98.9 | 96.1-99.2 |
| NPV | | 75.0-98.7 | 46.4-96.4 | 43.7-97.0 | 28.9-94.1 |
| Date agreement | Exact | 90.7-95.6 | 86.8-91.4 | 93.7-97.0 | 93.8-98.3 |
| | ± 15 days | 95.6-97.6 | 95.8-98.5 | 96.2-98.4 | 95.2-99.1 |
| | ± 30 days | 96.3-97.9 | 96.8-98.6 | 96.2-98.7 | 95.7-99.1 |

Abbreviations: EHR, electronic health record; NPV, negative predictive value; OD, obituary data; PPV, positive predictive value; SSDI, social security death index.
[a]Components of the composite mortality variable: SSDI, OD, structured EHR data, and unstructured EHR data.

**TABLE 2**  Validity metrics for the composite mortality variable across the 18 cancer types combined, overall, and stratified by sociodemographic and clinical characteristics

| Strata | | N (%) | Sensitivity, % (95% CI) | Specificity, % (95% CI) | PPV, % (95% CI) | NPV, % (95% CI) | 15-day date Agreement, % (95% CI) |
|---|---|---|---|---|---|---|---|
| Overall | Overall | 160 436 (100.0) | 89.2 (89.0, 89.4) | 97.4 (97.3, 97.5) | 97.8 (97.7, 97.9) | 87.5 (87.3, 87.7) | 97.0 (96.9, 97.2) |
| Practice type | Community | 145 212 (90.5) | 89.1 (88.9, 89.3) | 97.6 (97.5, 97.7) | 98.0 (97.9, 98.1) | 87.0 (86.8, 87.3) | 97.3 (97.2, 97.4) |
| | Academic | 15 224 (9.5) | 91.0 (90.4, 91.6) | 95.3 (94.8, 95.8) | 95.0 (94.5, 95.5) | 91.5 (90.9, 92.1) | 94.6 (94.0, 95.1) |
| Age group[a] (at cohort entry) | <35 | 1514 (0.9) | 86.9 (84.0, 89.7) | 98.5 (97.7, 99.2) | 96.9 (95.4, 98.4) | 93.1 (91.6, 94.6) | 96.3 (94.6, 98.0) |
| | 35-49 | 10 814 (6.7) | 88.3 (87.4, 89.2) | 97.9 (97.6, 98.3) | 97.1 (96.6, 97.6) | 91.5 (90.8, 92.1) | 96.5 (96.0, 97.1) |
| | 50-64 | 50 435 (31.4) | 89.2 (88.8, 89.6) | 97.7 (97.5, 97.9) | 97.7 (97.5, 97.9) | 89.2 (88.9, 89.6) | 97.0 (96.8, 97.3) |
| | 65-74 | 51 134 (31.9) | 89.3 (88.9, 89.6) | 97.4 (97.2, 97.7) | 97.9 (97.7, 98.0) | 87.4 (87.0, 87.8) | 97.1 (96.9, 97.3) |
| | 75+ | 46 538 (29.0) | 89.4 (89.1, 89.8) | 96.5 (96.2, 96.8) | 97.8 (97.7, 98.0) | 83.6 (83.1, 84.1) | 97.1 (96.9, 97.3) |
| Race/ethnicity | White | 112 116 (69.9) | 91.4 (91.1, 91.6) | 98.0 (97.9, 98.1) | 98.4 (98.3, 98.5) | 89.6 (89.4, 89.9) | 97.8 (97.6, 97.9) |
| | Afr.American | 13 111 (8.2) | 88.0 (87.2, 88.7) | 98.0 (97.6, 98.3) | 98.2 (97.9, 98.5) | 86.5 (85.7, 87.3) | 97.3 (96.9, 97.7) |
| | Hisp/Latino | 433 (0.3) | 84.4 (79.8, 88.9) | 91.6 (87.6, 95.5) | 92.8 (89.3, 96.2) | 82.1 (76.9, 87.2) | 92.3 (88.8, 95.8) |
| | Asian | 3270 (2.0) | 76.3 (74.2, 78.4) | 95.1 (94.0, 96.1) | 93.6 (92.2, 94.9) | 81.0 (79.3, 82.7) | 92.7 (91.3, 94.1) |
| | Other/missing | 31 506 (19.6) | 83.4 (82.8, 83.9) | 95.2 (94.9, 95.6) | 95.7 (95.4, 96.0) | 81.7 (81.1, 82.3) | 94.6 (94.3, 95.0) |
| Region | Midwest | 22 339 (13.9) | 94.1 (93.7, 94.5) | 98.0 (97.7, 98.3) | 98.5 (98.3, 98.7) | 92.3 (91.8, 92.8) | 97.7 (97.4, 98.0) |
| | Northeast | 40 799 (25.4) | 91.5 (91.1, 91.8) | 97.2 (96.9, 97.4) | 97.6 (97.4, 97.8) | 89.9 (89.4, 90.3) | 97.1 (96.8, 97.3) |
| | South | 63 896 (39.8) | 90.9 (90.6, 91.2) | 97.7 (97.6, 97.9) | 98.2 (98.0, 98.3) | 88.8 (88.5, 89.2) | 97.5 (97.3, 97.7) |
| | West | 30 599 (19.1) | 82.4 (81.8, 83.0) | 96.6 (96.3, 96.9) | 96.6 (96.3, 96.9) | 82.6 (82.0, 83.1) | 95.6 (95.3, 96.0) |
| | Other/missing | 2803 (1.7) | 46.8 (44.3, 49.4) | 95.8 (94.8, 96.9) | 92.4 (90.5, 94.3) | 62.4 (60.3, 64.5) | 90.5 (88.4, 92.6) |
| Lines of therapy | Not documented | 44 911 (28.0) | 85.6 (85.1, 86.0) | 97.1 (96.8, 97.3) | 97.4 (97.1, 97.6) | 84.3 (83.8, 84.7) | 96.6 (96.3, 96.8) |
| | <3 L | 92 531 (57.7) | 90.0 (89.7, 90.3) | 97.5 (97.4, 97.7) | 97.8 (97.7, 97.9) | 88.8 (88.6, 89.1) | 97.0 (96.9, 97.2) |
| | 3 L+ | 22 994 (14.3) | 92.9 (92.5, 93.3) | 97.3 (96.9, 97.6) | 98.3 (98.1, 98.5) | 88.9 (88.2, 89.5) | 97.8 (97.5, 98.0) |
| | <4 L | 104 947 (65.4) | 90.2 (90.0, 90.5) | 97.5 (97.4, 97.6) | 97.9 (97.7, 98.0) | 88.7 (88.5, 89.0) | 97.1 (97.0, 97.3) |
| | 4 L+ | 10 578 (6.6) | 94.1 (93.6, 94.7) | 97.3 (96.8, 97.8) | 98.4 (98.1, 98.7) | 90.3 (89.4, 91.2) | 97.9 (97.5, 98.2) |
| | <5 L | 110 593 (68.9) | 90.4 (90.2, 90.7) | 97.5 (97.3, 97.6) | 97.9 (97.8, 98.0) | 88.8 (88.5, 89.0) | 97.2 (97.0, 97.3) |
| | 5 L+ | 4932 (3.1) | 94.5 (93.7, 95.3) | 97.9 (97.2, 98.6) | 98.7 (98.3, 99.1) | 91.1 (89.8, 92.4) | 98.2 (97.8, 98.7) |

Abbreviations: L, line of therapy; NPV, negative predictive value; PPV, positive predictive value.
[a]One patient had unknown age and was not analyzed for stratification by age group.

We calculated absolute and relative comparisons between the median rwOS values using the composite mortality variable and NDI data.

The analysis was conducted using R statistical computing software version 3.3.2.[12]

## 3 | RESULTS

### 3.1 | Validity metrics

In the main study cohort spanning 18 cancer types (N = 160 436 unique patients), the validation analysis comparing the composite mortality variable (SSDI, OD, structured EHR data, and unstructured EHR data) to the NDI showed high sensitivity (ranging from 83.9% to 91.5%), specificity (93.5%-99.7%), PPV (96.3%-98.3%), NPV (75.0%-98.7%), and ±15-day agreement (95.6%-97.6%) (Table 1). Validity

metrics showed high results across all cancer type-specific results, with only slight variability in the sensitivity of the composite mortality variable (Suppl. Table 2, Suppl. Figure 1).

We conducted analyses overall and separately for each cancer type stratified by certain sociodemographic and clinical factors (Table 2). In analyses across the 18 cancer types, there were noticeable differences in sensitivity for the following stratifications (with some strata dropping below sensitivity of 85.0%): US region (94.1% for Midwest, 91.5% for Northeast, 90.9% for South, 82.4% for West, and 46.8% for missing/other region), race/ethnicity (91.4% for White, 88.0% for African American, 84.4% for Hispanic/Latino, 83.4% for other/missing race/ethnicity, and 76.3% for Asian), and practice site for those with at least 100 patients (sensitivity ranged from 41.1% to 100.0%, Median, IQR: 89.9% [83.8%-95.5%]) (Suppl. Figure 3). Only slight sensitivity variations were seen across the rest of the stratifications, all remaining above 85%. In analyses that stratified by 6-month time period of death/last confirmed activity from 2011 to 2017, the sensitivity of the composite

**TABLE 3**  Comparison of median rwOS estimates obtained with the composite mortality variable vs the NDI across 18 cancer types

| Cancer Type | n | Median rwOS, mos (95% CI) Composite Mortality Variable | NDI | Difference Absolute, mos | Relative, % |
|---|---|---|---|---|---|
| eBC | 1669 | NR (NR–NR) | NR (NR–NR) | — | — |
| mBC | 16 473 | 32.4 (31.6-33.3) | 29.9 (29.3-30.6) | 2.5 | 8.4 |
| CLL | 9035 | 203.8 (198.6-211.7) | 197.6 (190.8-203.2) | 6.2 | 3.1 |
| mCRC | 17 232 | 23.2 (22.8-23.7) | 21.6 (21.2-22.1) | 1.6 | 7.4 |
| DLBCL | 4344 | 77.4 (71.4 - NR) | 71.3 (68.8-77.8) | 6.1 | 8.6 |
| aGE | 7169 | 12.4 (12.0-12.8) | 11.6 (11.3-12.0) | 0.8 | 6.9 |
| HCC | 2784 | 19.4 (18.0-21.3) | 17.4 (16.3-19.0) | 2.0 | 11.5 |
| aHNC | 5271 | 15.0 (14.5-15.5) | 14.3 (13.9-14.8) | 0.7 | 4.9 |
| aMel | 7031 | 40.6 (38.4-42.9) | 36.2 (34.4-39.0) | 4.4 | 12.2 |
| MM | 7803 | 61.9 (59.5-64.4) | 57.1 (55.7-59.8) | 4.8 | 8.4 |
| MPM | 1700 | 14.8 (13.7-15.6) | 14.4 (13.3-15.3) | 0.4 | 2.8 |
| aNSCLC | 45 070 | 11.8 (11.5-12.0) | 11.0 (10.8-11.2) | 0.8 | 7.3 |
| Ovarian | 4964 | 53.2 (50.7-57.9) | 48.3 (46.2-50.7) | 4.9 | 10.1 |
| Pancreatic (metastatic) | 5458 | 6.9 (6.6-7.2) | 6.5 (6.3-6.8) | 0.4 | 6.2 |
| Prostate (metastatic) | 8495 | 33.9 (32.9-35.1) | 32.4 (31.7-33.1) | 1.5 | 4.6 |
| mRCC | 5770 | 25.7 (24.5-27.1) | 22.8 (21.3-24.4) | 2.9 | 12.7 |
| SCLC | 4724 | 10.9 (10.5-11.2) | 10.5 (10.2-10.8) | 0.4 | 3.8 |
| Urothelial (advanced) | 6293 | 12.6 (12.1-13.1) | 11.9 (11.4-12.3) | 0.7 | 5.9 |

*Note:* Index dates are either initial diagnosis or advanced/metastatic diagnosis date, variable by cancer type.

Abbreviations: aGE, advanced gastroesophageal; aHNC, advanced head and neck cancer; aNSCLC, advanced non-small cell lung cancer; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; e(m)BC, early (metastatic) breast cancer; HCC, hepatocellular carcinoma; mCRC, metastatic colorectal cancer; MM, multiple myeloma; MPM, malignant pleural mesothelioma; mRCC, metastatic renal cell carcinoma; NDI, National Death Index; NR, not reported; rwOS, real-world overall survival; SCLC, small-cell lung cancer.

mortality variable was largely constant, yet slightly lower for patients with more recent deaths or last activity records. The sensitivity of SSDI-only data were substantially lower for patients with more recent deaths/last activity (Suppl. Figure 2). These trends were largely consistent across cancer type-specific analyses (Suppl. Table 3, A-Q).

## 3.2 | rwOS analysis and estimates

Median rwOS estimates based on the composite mortality variable were longer than NDI-based estimates (differences ranged from 0.4 months longer for MPM, metastatic pancreatic cancer, and SCLC to 6.2 months longer for CLL). Relative differences in median rwOS ranged from 2.8% (MPM) to 12.7% longer (mRCC) (Table 3).

In cancer type-specific analyses, sequentially adding OD, SSDI, and abstracted (from unstructured data) death dates onto structured EHR mortality data resulted in median rwOS estimates progressively closer to those using NDI data (Suppl. Figure 4).

## 3.3 | Sensitivity analysis

To assess the validity of the composite mortality variable in datasets of smaller size and with different selection criteria, we conducted a

sensitivity analysis in a separate cohort (n = 17 540, described in the Methods section). Validity metrics across cancer types (and in a pan-tumor cohort, described in the Methods section) were consistent with the main analyses: sensitivity, >85.0%; specificity, >95.0%; PPV, >96.0%; NPV, >84.0%; and ±15 day agreement, >94.0% (Suppl. Table 2).

## 4 | DISCUSSION

This article expands the results from the prior publication reporting the initial characterization of a composite mortality variable.[9] Consistent with those seminal results, this update showed high sensitivity, specificity, PPV, NPV, and date agreement for the variable across 18 cancer types (the initial four plus additional 14); of note, refreshed results for the four cancer types previously reported were remarkably similar to the prior report.[9] Sensitivity was high overall and did not fall below 84% in any cancer type. Further strengthening the robustness of these findings, a sensitivity analysis produced similar results in a smaller cohort of patients generated using different eligibility criteria (ie, requiring specific genetic testing).

We observed differences in sensitivity across several sociodemographic and clinical characteristics, particularly region, race/ethnicity, and practice site. Examining individual data source components for

each practice site showed that some differences could be due to practice behaviors and documentation patterns, but the range of sensitivities was actually largest for SSDI data. Among patients with the documented region of residence, sensitivity was lower in the Western US as compared to other US regions, possibly driven by the low sensitivity of SSDI-only data. In analyses stratified by race/ethnicity, the lowest sensitivity was for Asian patients across tumor types, although it was unclear what factors were driving that finding. While there were sensitivity variations across tumor types, we could not pinpoint consistent links to disease-specific clinical features, such as indolent diseases with lower sensitivity, due to potentially greater follow-up losses.

Our work shows that quality varies across mortality surveillance tools, and understanding the sensitivity, specificity, and accuracy of a given source is critical. For instance (and similar to the prior report by Curtis et al[9]), this study showed gaps in EHR-derived data that could be addressed by aggregating multiple sources of structured and unstructured data into a composite variable that performs above each one of its single original sources, and, importantly, above structured source pairings.

In the evolving field of RWE, reaching a consensus regarding acceptable quality thresholds for the underlying RWD (for parameters such as completeness or concordance with pre-existing standards) remains an important open issue. Low sensitivity in mortality surveillance is known to bias rwOS estimates,[13-16] and determining the sensitivity threshold at which those biases may have an excessive analytic impact is key. Our benchmarking exercise showed that the biases introduced in rwOS estimates using the composite mortality variable across 18 cancer types were modest in most cases (less than 13% higher in relative comparisons to NDI-based median rwOS). Prior work by Carrigan et al[13] indicated that, within the sensitivity levels achieved by the composite mortality variable, there would be the limited impact of any potential rwOS bias for descriptive research (ie, absolute survival estimates) or comparative effectiveness research comparing two groups analyzed from the same source. However, the impact could be greater on analyses comparing survival across different sources (eg, external control arms).[13] Additionally, the effects of varying sensitivity levels in mortality detection on survival analyses may be contingent on the age of the cohort under study,[15] a point that may warrant further examination in studies of aging populations. Considering all these factors, understanding these different scenarios, and their risk for biased rwOS analyses is important. Future standardization work will be required to define which boundaries for the quality of a data element, mortality in this case, are considered acceptable. This could be solved by setting fixed sensitivity thresholds, or by taking use-case specific approaches (namely, for rwOS comparisons, acceptability thresholds dependent on the magnitude of the expected effect, or on the cohort age). Throughout this line of work, and as it relates to longitudinal data, sustaining benchmarking and validating efforts over time will be important to understand whether and how quality may fluctuate.

This study has limitations inherent to the data sources used. First, the probabilistic process used for NDI record matching may be subject to its own intrinsic limitations (based on the availability of all required elements), which in turn may affect its quality as a reference[5]; in addition, the yearly lag in NDI releases limits the feasibility of any benchmarking exercise for highly recent data. Second, this mortality variable has been developed based on 18 cancer type-specific EHR-derived cohorts, therefore, the performance of the variable depends on the optimization of the underlying rules for data abstraction, such as index date definitions, or hierarchical criteria for adjudication of death dates (when conflicting).

In conclusion, we have developed a composite mortality variable for oncology research that shows high sensitivity, specificity, and accuracy across a wide range of cancer types when compared with the NDI as the gold standard reference. As the components of this variable are aggregated into partial combinations, the resulting interim variables show increasing sensitivity; the full composite variable (a combination of SSDI, OD, structured EHR data, and unstructured EHR data) is the one that consistently reaches the greatest sensitivity and the one we have implemented in our databases. rwOS estimates obtained with this variable showed modest overestimations when compared against NDI-based estimates. This mortality variable represents an important tool for RWE oncology research. Further efforts are needed to improve public sources of mortality data and to establish data quality standards in RWE.

## CONFLICT OF INTEREST

All authors report employment at Flatiron Health, Inc., which is an independent subsidiary of the Roche group, equity ownership in Flatiron Health, Inc. and stock ownership in Roche.

## ORCID

*Qianyi Zhang* https://orcid.org/0000-0002-7850-6701
*Anala Gossai* https://orcid.org/0000-0003-0892-979X
*Nathan C. Nussbaum* https://orcid.org/0000-0002-5054-5186
*Christina M. Parrinello* https://orcid.org/0000-0003-2578-4588

## REFERENCES

1. US Food and Drug Administration. Framework for FDA''S real-world evidence program. US Food and Drug Administration. 2018. https://www.fda.gov/media/120060/download. Accessed September 3, 2020.
2. Raphael MJ, Gyawali B, Booth CM. Real-world evidence and regulatory drug approval. *Nat Rev Clin Oncol*. 2020;17(5):271-272.
3. Bolislis WR, Fay M, Kühler TC. Use of real-world data for new drug applications and line extensions. *Clin Ther*. 2020;42(5):926-938.
4. Gliklich RE, Leavy MB. Assessing real-world data quality: the application of patient registry quality criteria to real-world data and real-world evidence. *Ther Innov Regul Sci*. 2020;54(2):303-307.
5. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major U.S. mortality databases. *Ann Epidemiol*. 2002; 12(7):462-468.
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.

7. da Graca B, Giovanni F, David N. 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.

8. Maynard C. Changes in the completeness of the social security death master file: a case study. *Internet J Epidemiol*. 2013;11(2):1-3.

9. Curtis MD, Griffith SD, Tucker M, et al. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv Res*. 2018;53(6):4460-4476.

10. Singal G, Miller PG, Agarwala V, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non–small cell lung cancer using a clinicogenomic database. *Jama*. 2019;321(14):1391-1399.

11. National Center for Health Statistics. National death index users guide, appendix A: a probabilistic scoring approach for assessing NDI match results. https://www.cdc.gov/nchs/data/ndi/NDI_Users_Guide.pdf. Accessed September 3, 2020.

12. Team RC. *R: A Language and Environment for Statistical Computing (Version R Version 3.3.2)*. R Foundation for Statistical Computing: Vienna, Austria; 2016.

13. Carrigan G, Whipple S, Taylor MD, et al. An evaluation of the impact of missing deaths on overall survival analyses of advanced non–small cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiol Drug Saf*. 2019;28(5):572-581.

14. Ibrahim JG, Chu H, Chen M. Missing data in clinical studies: issues and methods. *J Clin Oncol*. 2012;30(26):3297-3303.

15. Jacobs EJ, Newton CC, Wang Y, Campbell PT, Flanders WD, Gapstur SM. Ghost-time bias from imperfect mortality ascertainment in aging cohorts. *Ann Epidemiol*. 2018;28(10):691-696.

16. Siannis F. Sensitivity analysis for multiple right censoring processes: investigating mortality in psoriatic arthritis. *Stat Med*. 2011;30(4): 356-367.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.