# The Impact of Missing Deaths on Survival Analyses Conducted in an Oncology EHR Database


**Gillis Carrigan*, Samuel Whipple*, Michael D. Taylor*, Aracelis Z. Torres**, Anala Gossai**, Brandon Arnieri*, Melisa Tucker**, Philip P. Hofmeister**, Peter Lambert*, Sandra D. Griffith**, William B. Capra***

*Genentech/Roche

**Flatiron Health


## Abstract


Real-world evidence (RWE) based on data derived from electronic health records (EHRs) is increasingly used in pharmacoepidemiologic research. In order to fully utilize RWE, high-quality outcomes data are needed, and mortality-based outcomes are particularly important for many disease areas, including oncology. However, in RWE sources, mortality data may not capture all deaths. The aim of this study was to assess the impact of missing death data on survival analyses conducted in an oncology EHR-derived database. Three main analytic frameworks were evaluated in patients with advanced non-small cell lung cancer (aNSCLC): absolute risk represented by median overall survival rates, relative risk (hazard ratios [HRs]) for mortality captured in comparative analyses within the database, and external control arm analyses comparing an experimental group from a gold standard data source with a control group from the EHR-derived database. The study drew on data from the National Death Index (NDI) as a gold standard, which provided a benchmark for assessing the impact of imperfect mortality capture. Overall, small differences were observed in the HRs for the EHR-derived cohort across comparative analyses when compared to HRs obtained from the gold standard data source. When only one treatment arm was subject to estimation bias (i.e., external control analyses), the bias was slightly more pronounced. The magnitude of all observed biases increased when lower sensitivities were generated through simulation. With high mortality sensitivity, the impact on survival analyses is minimal, with only modest evidence of impact associated within external control arm applications.


## Background

- Real-world evidence (RWE) based on data derived from electronic health records (EHRs) is increasingly used in pharmacoepidemiologic research.
- In order to fully utilize RWE, high-quality outcomes data are needed, and mortality-based outcomes are particularly important for many disease areas, including oncology.
- However, in RWE sources, mortality data may not capture all deaths.

## Objectives

- The aim of this study was to assess the impact of missing death data on survival analyses conducted in an oncology EHR-derived database.
- Specifically, we wanted to examine the results of three main analytic applications in patients with advanced non-small cell lung cancer (aNSCLC) at different mortality sensitivity levels with those obtained using a gold standard data source with an assumed mortality sensitivity of 100%.

## Methods

Design:
aNSCLC patients with advanced stage disease diagnosed between 1/1/11 and 12/31/15 in an oncology EHR were included in the study. Three analytic use cases were evaluated: absolute risk estimation, relative risk comparing 2 groups within the EHR, and relative risk comparing simulated single arm trial data to an external control identified within the EHR.

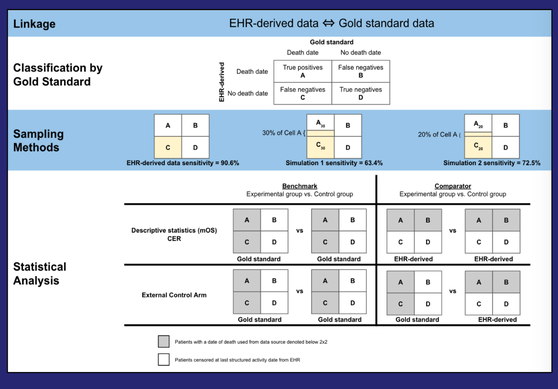
Setting:
The aNSCLC cohort was selected from the Flatiron Health EHR database. National Death Index (NDI) data were obtained on patients in the aNSCLC cohort.

Exposures or Interventions:
The analyses included a number of survival comparisons for groups defined on the basis of drug exposure or clinical characteristics (e.g., biomarker status).

Main Outcome:
The primary outcome was overall survival (OS), defined as time from first line treatment initiation or advanced diagnosis to death. Patients were censored at their last known structured activity date in the EHR prior to the end of the analysis period. The gold standard for the mortality endpoint was death data from the NDI.

Statistical Analysis:
The Kaplan-Meier method was used to estimate median OS (mOS), and Cox proportional hazards models were used to estimate hazard ratios (HRs).

### Study Overview





Disclosure Statement

G. Carrigan, S. Whipple, M. Taylor, B. Arnieri, P. Lambert and W. Capra are employees of Genentech, a member of the Roche Group. A. Torres, A. Gossai, M. Tucker, P. Hofmeister and S. Griffith are employees of Flatiron Health , which is an independent subsidiary of the Roche Group and source of the study data. No other disclosures.


## Results

**Figure 1.** Censoring patterns of those patients with missing death dates

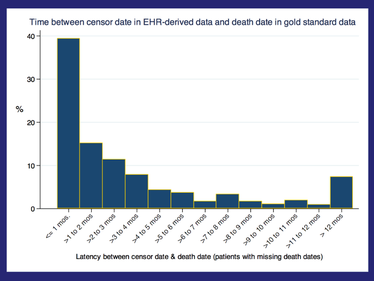



**Table 1.** Impact of missing deaths on measures of absolute risk (mOS)

| Median Overall Survival (months) and 95% CI | | | | | | | |
|---|---|---|---|---|---|---|---|
| Exposure Group | Simulation 1 (sensitivity= 63.4%) | Bias | Simulation 2 (sensitivity= 72.5%) | Bias | EHR-derived (90.6%) | Bias | Gold standard data |
| Platinum treatment | 12.3 (11.7, 13.0) | 36.7% | 11.2 (10.7, 11.7) | 24.4% | 9.5 (9.1, 9.9) | 5.6% | 9.0 (8.7, 9.4) |
| Other Chemo | 11.8 (10.1, 13.6) | 53.2% | 10.1 (8.8, 11.9) | 31.2% | 8.2 (7.3, 9.6) | 6.5% | 7.7 (6.9, 8.7) |
| EGFR+ | 20.7 (17.8, 25.9) | 46.8% | 19.4 (15.9, 22.2) | 37.6% | 15.0 (13.4, 18.9) | 6.4% | 14.1 (12.1, 17.3) |
| EGFR- | 21.2 (18.9, 25.3) | 43.2% | 19.2 (17.1, 21.4) | 29.7% | 16.0 (14.5, 18.3) | 8.1% | 14.8 (14.0, 17.0) |
| KRAS+ | 33.4 (28.5, NA) | 40.9% | 29.9 (24.9, 36.2) | 25.8% | 24.3 (22.3, 29.4) | 2.5% | 23.7 (21.7, 27.3) |
| KRAS- | 16.3 (15.2, 17.2) | 39.3% | 14.6 (13.9, 15.3) | 24.8% | 12.4 (11.8, 13.1) | 6.0% | 11.7 (11.3, 12.4) |

**Figure 2.** Impact on comparative analyses conducted with EHR data

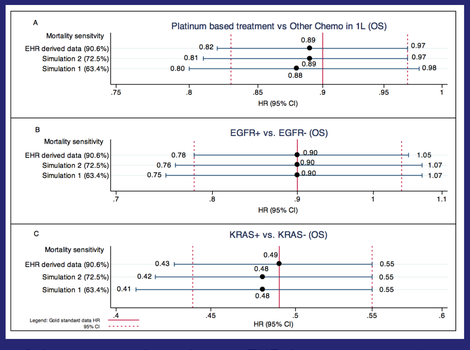
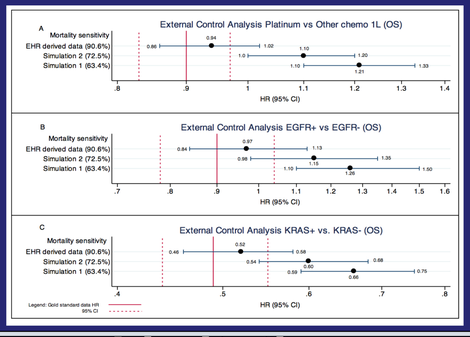



**Figure 3.** Impact on analyses that use EHR data as an external control arm




## Limitations

- This study leveraged data from community-based oncology clinics in the US, and patterns of missing data may be different in academic centers or in other countries.
- Despite the minimal impact on most conclusions observed in aNSCLC, it is unclear how this would expand to other cancer types with longer mOS.
- This study did not consider the mechanism for missing deaths, further work is needed to describe the presence and degree of informative censoring in these data and understand its impact.

## Conclusions

- Overall, small differences were observed in the HRs for the EHR-derived cohort across comparative analyses when compared to HRs obtained from the gold standard data source.
- When only one treatment arm was subject to estimation bias (i.e., external control analyses), the bias was slightly more pronounced.
- The magnitude of all observed biases increased when lower sensitivities were generated through simulation, in particular, in the external control arm application, however, with high sensitivity (e.g., ~90%) only modest evidence of impact was associated within external control arm applications.
- EHR mortality data with high sensitivity limits the potential for missing deaths to bias OS estimates. This enables robust conclusions and decision making based on RWD.

### Bibliography


1. Sean Khozin, Gideon M. Blumenthal, Richard Pazdur. Real-world Data for Clinical Evidence Generation in Oncology. JNCI J Natl Cancer Inst (2017) 109(11).
2. Curtis MD, Griffith S, Tucker M, Taylor MD, Capra WB, Carrigan G, Holzman B, Torres AZ, You P, Arnieri B, Abernethy AP. Development and Validation of a High-Quality Composite Real-World Mortality Endpoint. Health Services Research May, 2018 4-17.
3. Wu YX, Furnary AP, Grunkemeier GL. 2008. Using the National Death Index to Validate the Noninformative Censoring Assumption of Survival Estimation. Annals of Thoracic Surgery 2008;85:1256-60.