Quantifying the impact of mortality underreporting on analyses of overall survival

Michael T Bretscher, Thibaut Sanglier

F. Hoffmann-La Roche Ltd, Basel, Switzerland

Background

- Clinical outcomes may occasionally be compared across different datasets, e.g. when a historical control group serves as comparator for a single arm trial.
- The two datasets may differ in completeness and accuracy of outcome reporting (e.g. death)
- Mortality underreporting leads to overestimation of overall survival (OS), and differential underreporting between compared datasets leads to biased estimates of mortality hazard ratios.
- Endpoint validation through collection of additional data, or linking to a "gold standard" source of mortality information is often not practical.
- A theoretical framework could help quantify and better understand the impact of mortality underreporting on median OS and mortality hazard ratios.
- Flatiron Health (FIH) developed an electronic health records (EHR)-derived database[1]. The process to capture mortality improved over time, from 76.9% of deaths reported in version 1.0 ($OS_{1.0}$) to 87% in version 2.0 ($OS_{2.0}$)[2]
- **Aim: to validate theoretical predictions of bias by comparison with outputs from the FIH data with known levels of mortality underreporting.**

Methods

Mathematical models (described below) predict OS biases.

Validation of theoretical predictions using real world data:

- A cohort of incident metastatic breast cancer (mBC) patients diagnosed between January 1st 2011 and July 31st 2016 was identified in the FIH EHR-derived database[1]. This allowed for both $OS_{1.0}$ and $OS_{2.0}$ (from diagnosis) to be available for all patients. Data beyond July 31st 2016 was censored.
- Patients were stratified by biomarker status, using well-known prognostic biomarkers.
- Within the five biomarker-based strata, each patient was duplicated to be present with mortality information versions $OS_{1.0}$ (arm 1) as well as $OS_{2.0}$ (arm 2), permitting comparison of mortality endpoints.
- Within biomarker subgroups, hazard ratios of $OS_{1.0}$ (arm 1) vs $OS_{2.0}$ (arm 2) were compared against theoretical predictions (equation 1 below)
- Similarly, a ratio of Kaplan-Meier median $OS_{1.0}$ (arm 1) vs median $OS_{2.0}$ (arm 2) was compared against theoretical predictions (equation 2 below)

Mathematical Models

Parameters

s_1, s_2	proportion of reported deaths ("sensitivity") in arms 1 and 2, respectively
$R_{1,2}$	Observed hazard ratio comparing arms 1 and 2 (biased by mortality underreporting)
$\rho_{1,2}$	True hazard ratio comparing arms 1 and 2 (unbiased, with perfect mortality data)
m_1, m_2	Observed median OS in arms 1 and 2 (biased by mortality underreporting)
μ_1, μ_2	True median OS in arms 1 and 2 (unbiased, with perfect mortality data)

Impact of mortality underreporting on OS hazard ratios

1) $$\frac{R_{1,2}}{\rho_{1,2}} = \frac{s_1}{s_2}$$

"The observed hazard ratio between arms 1 and 2 compares to the true hazard ratio as the sensitivity in arm 1 compares to the sensitivity in arm 2."

Derivation: Let τ_i be the true hazard function in arm i, and let s_i denote the proportion of death events (assumed randomly) reported in the data of arm i. The observed hazard function o_i can then be written as $o_i = s_i \tau_i$. Division of o_1 by o_2 yields the above equation, using $R_{1,2} = {}^{o_1}/_{o_2}$ and $\varrho_{1,2} = {}^{\tau_1}/_{\tau_2}$ for observed and true hazard ratios, respectively.

Impact of mortality underreporting on median OS (assuming exponential survival)

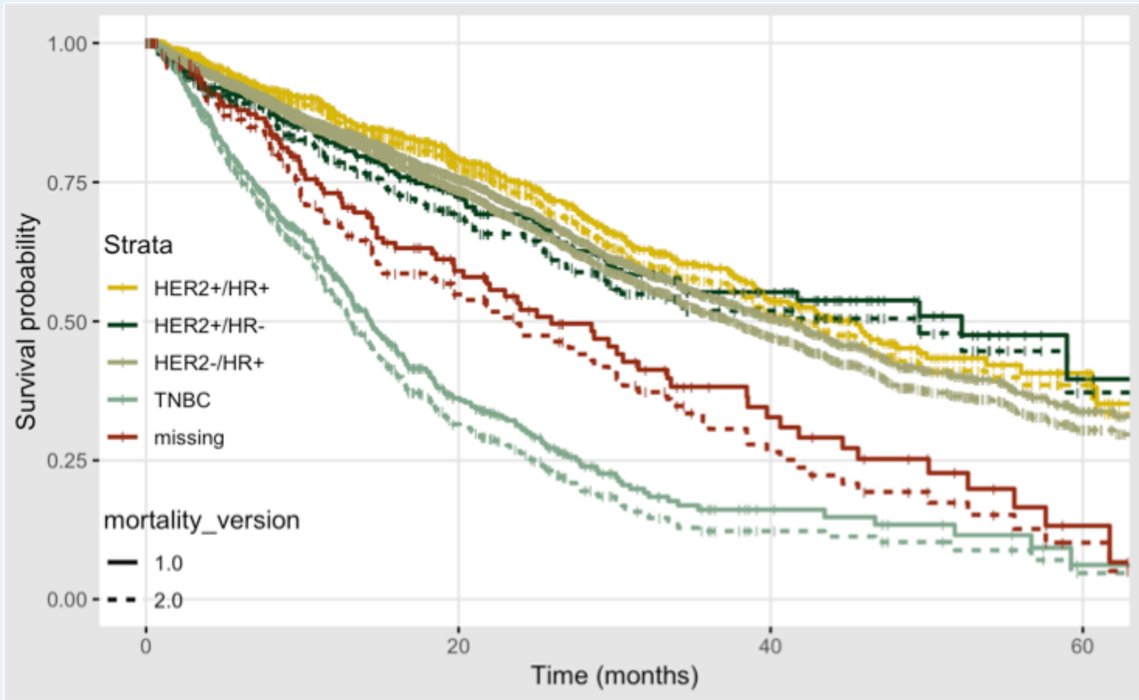
2) $$\frac{m_1}{m_2} = \frac{\mu_1 s_2}{\mu_2 s_1}$$

"The apparent ratio of medians between arms 1 and 2 equals the true ratio of medians divided by the ratio of corresponding sensitivities"

Derivation: Let τ_i be the true hazard function in arm i, and let s_i denote the proportion of death events (assumed randomly) reported in the data of arm i. The observed hazard function o_i can then be written as $o_i = s_i \tau_i$. Both hazards are time-constant in this case, consistent with assuming exponentially distributed survival times. Using the formula for the median of an exponentially distributed variable, observed and true medians in arm i can be written as $m_i = {}^{\ln(2)}/_{o_i}$ and $\mu_i = {}^{\ln(2)}/_{\tau_i}$. The above equation is obtained through division of m_1 by m_2 and simplification using the relationship between true hazard and true median.

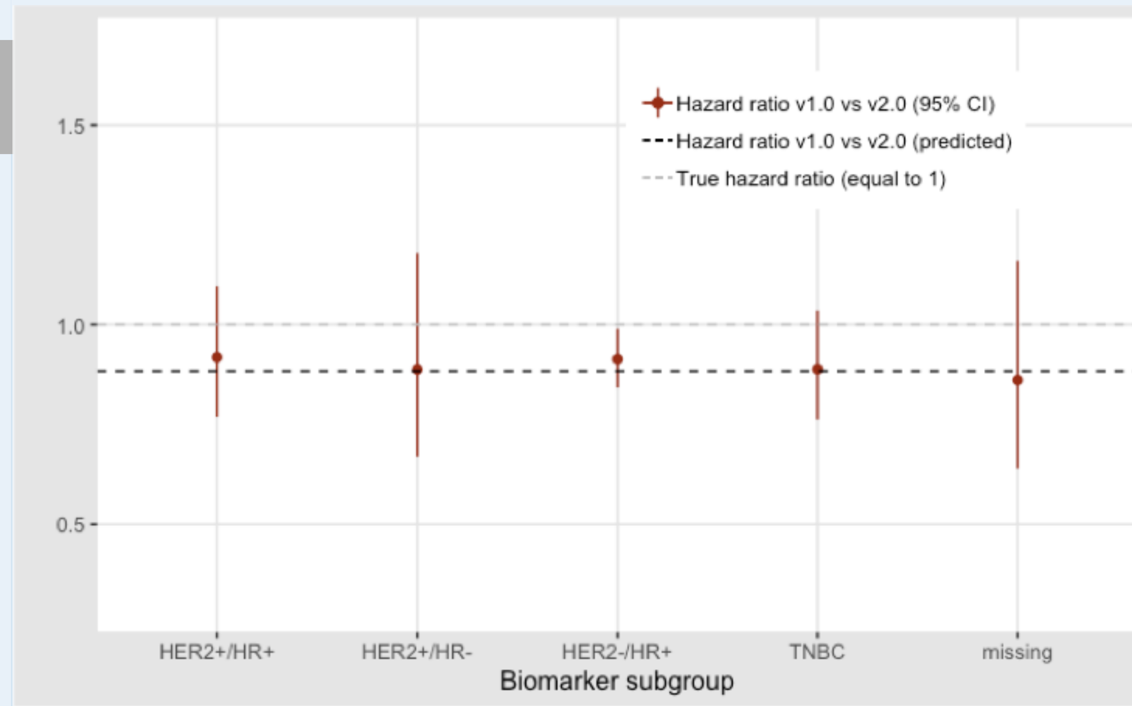
Results

Fig 1: Kaplan-Meier plots of OS, by subgroup and mortality version



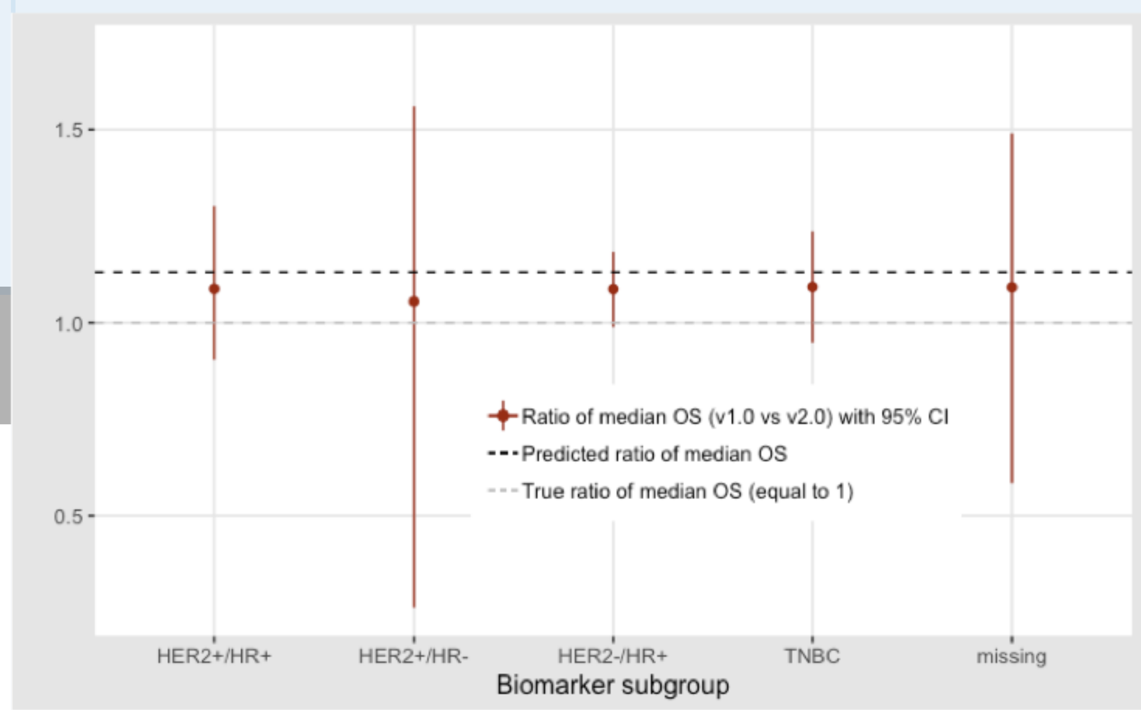
- N=5,483 mBC patients were identified in the FIH data, with subgroups HER2+/HR+ (N=842), HER2+/HR- (N=305), HER2-/HR+ (N=3608), triple negative (N=562), missing biomarker information (N=166).
- Median follow-up: 22.6 months
- OS differed by biomarker subgroup
- Mortality data version 1.0 ($OS_{1.0}$) leads to numerically longer survival due to a higher % of missed deaths

Fig 2: Impact of mortality underreporting on OS hazard ratios



- Fig 2: within-subgroup comparison of the hazard of death (Hazard ratio $OS_{1.0}$ vs $OS_{2.0}$)
- True hazard ratio should be 1.0 with perfect mortality data (same patients in each group)
- Bias due to differential % deaths missed is reasonably well predicted by the mathematical model (eq. 1)

Fig 3: Impact of mortality underreporting on Kaplan-Meier median OS



- Fig 3: within-subgroup comparison of median OS: ratio of K-M median $OS_{1.0}$ vs K-M median $OS_{2.0}$
- True ratio of medians should be 1.0 with perfect mortality data (same patients in each group)
- The model-based prediction (eq. 2) of the "ratio of medians" consistently overestimates the effect of differential mortality underreporting.

Conclusion

- The bias in OS hazard ratios due to differential mortality underreporting is well predicted by the theoretical framework (Fig 2).
- Model-based predictions of bias in the ratio of K-M medians (Fig 3) are consistently high, albeit still within confidence limits.
- This approach could potentially be used to assess the impact of differences in mortality reporting between compared datasets when the sensitivity parameters are approximately known.
- Conversely, the mathematical models could help determine acceptable levels of mortality underreporting that would not alter the conclusions of a particular analysis.

References

1. Flatiron Health database (https://flatiron.com/real-world-evidence/), May 2018, mortality v2.0
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.

34th International Conference for Pharmacoepidemiology and Risk Management, Prague, August 2018