How Accurately do Mixture Cure Models Predict 'Cure' with Increasingly Immature CAR-T Therapy Data?

MSR2

Harper S[1], Hansell N[1], Russell J[2], Butler K[1]

[1] York Health Economics Consortium; [2] Bristol Myers Squibb (BMS)

Background:

- Chimeric antigen receptor T-cell (CAR-T) therapy is a type of personalised treatment that has potential to provide a 'cure'* for a proportion of those with haematological cancers that were previously considered difficult or impossible to treat.
- A mixture cure model (MCM) is a type of survival model that uses time-to-event data from clinical trials to account for cured and non-cured sub-populations within the overall trial population, usually considered appropriate for modelling CAR-T survival.[1-2.]
- Immature, highly censored trial data makes it difficult for health technology assessment (HTA) bodies to be certain about predicted cured and non-cured proportions.
- Recently published data are available from ZUMA-1, a clinical trial observing the clinical and safety of a CAR-T treatment in late-line diffuse large B-cell lymphoma (DLBCL).
- **Objective**: To quantify the impact of increasingly mature clinical trial data on the predictive capabilities of MCM.

* 'cure' defined as having the same risk of death and disease recurrence as the age-matched general population.

Project framing

Identified literature gap

- There is limited evidence available to aid in determining whether mixture cure models can be effectively used with immature trial data to predict long-term survival.

Key objectives

- To analyse published data from a CAR-T clinical trial and establish whether early data cuts with high levels of censoring can reliably predict long-term survival outcomes.
- To provide advice regarding the use of early trial data within HTA submissions for CAR-T therapies.

Dataset generation and truncation

Context

- This project examined ZUMA-1, a clinical trial for a CAR-T therapy in late-line DLBCL.
- The survival analysis from this trial (Kaplan-Meier [KM] plots) was extracted to estimate the original dataset. The 'reconstructed' dataset is termed a 'pseudo' dataset.

Methods

- A full pseudo dataset, in the form of KM and individual participant data were developed for ZUMA-1.
- The pseudo dataset was truncated to generate three pseudo early trial data cuts (representing increasingly mature data). These cuts were validated against an early ZUMA-1 published data cut.

Survival analysis

Methods

- MCMs were used to generate parametric distributions using the pseudo early data cuts and extrapolated to 60 months.
- The core and extrapolated distributions of the early data cuts were compared to the full pseudo dataset.
- Key statistical outputs were reported.

Rationale

- Comparing the extended distributions to the full pseudo dataset allowed for an understanding as to whether MCMs could reliably predict the long-term survival outcomes from immature data when evaluating CAR-T therapies.

Verification

- The outcomes were validated against the published ZUMA-1 dataset and the associated NICE technology appraisal TA559 (axi-cel for DLBCL).
- The findings were compared to previous published literature on the topic. [1-2]

Reporting and Conclusions

Methods

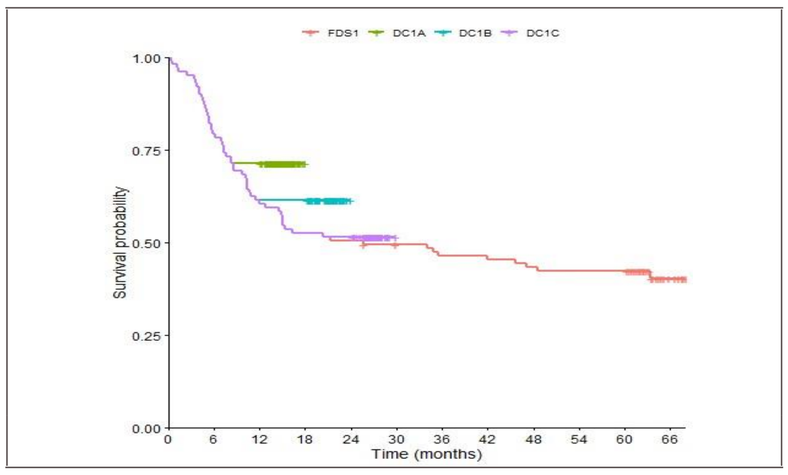
- Published 'ZUMA-1' trial KM data[3] with 63.1 months median follow-up were digitised[4] to generate pseudo individual participant data (IPD) for the complete trial follow-up, named "FDS1" in this analysis.
- FDS1 was truncated to replicate three early trial cuts with increased censoring, representing new data cuts. The shape of the reconstructed Kaplan-Meier curves were validated with trial publications where available (see Table 1).
- MCMs were fit to the full dataset and the early trial cuts, using standard parametric distributions for the 'uncured' population.[5]
- The four analyses were compared using graphed long-term survival predictions, goodness-of-fit statistics, median survivals and cure fractions.

Table 1. Summary of the pseudo early data cuts generated from FDS1

Element of trial	DC1A	DC1B	DC1C
Minimum follow-up	12 months	18 months	24 months
Censoring proportion	70%	60%	50%

The generated early pseudo data cuts and FDS1 were plotted and overlaid, as shown in Figure 1.

Figure 1. Generated pseudo datasets



The generated pseudo datasets were validated externally

- Generated full pseudo-IPD datasets were validated and deemed reflective of the original published evidence.[3]
- The early pseudo data-cut at 24-months follow-up (DC1C) appears to be a good visual match and is validated compared to published data that reflects a similar time period.[6]

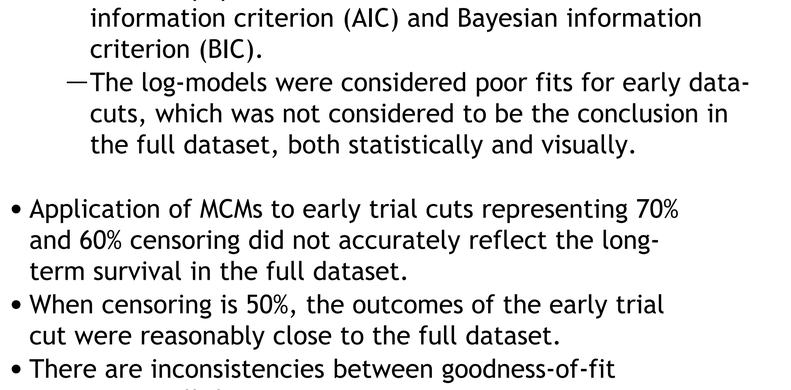
Results

- The key results of the analysis are presented in Table 2 and in Figure 2.
  - —In each of the early data-cuts the cure-fraction is overestimated relative to that of the full dataset (FDS1).
  - —The extent of overestimation lessens with iteratively more mature data.
  - —Gompertz was considered a distribution that fit the uncured population well in each dataset based on Akaike information criterion (AIC) and Bayesian information criterion (BIC).
  - —The log-models were considered poor fits for early data-cuts, which was not considered to be the conclusion in the full dataset, both statistically and visually.
- Application of MCMs to early trial cuts representing 70% and 60% censoring did not accurately reflect the long-term survival in the full dataset.
- When censoring is 50%, the outcomes of the early trial cut were reasonably close to the full dataset.
- There are inconsistencies between goodness-of-fit statistics in all datasets.

Table 2. Key results in each of the analysed datasets

Dataset (% censored, min. follow up)	Median survival range	Cure fraction range	Rank 1 AIC	Rank 1 BIC
DC1A (70%, 12 months)	N/A	67.3% - 71.7%	Gompertz	Gompertz
DC1B (60%, 18 months)	N/A	58.8% - 61.9%	Gen. Gamma	Gompertz
DC1C (50%, 24 months)	35.1 - 44.4 months	46.8% - 51.1%	Gompertz	Gompertz
FDS1 (43%, 60 months)	25.0 - 28.0 months	40.4% - 43.7%	Loglogistic	Gompertz

Figure 2. Five-year KM with predicted survival overlayed for each dataset



DC1A: 70% censoring and 12-months minimum follow up

DC1B: 60% censoring and 18-months minimum follow up

DC1C: 50% censoring and 24-months minimum follow up

FDS1: Full ZUMA-1, five-year follow up

Results Discussion

Predictive capability of MCM was less accurate when IPD censoring was over 60%

- DC1A (70% censoring) and DC1B (60% censoring) did not accurately reflect long-term mature survival shown in FDS1. MCM does not accurately capture the cure fraction.

Predictive capabilities of MCM was sufficiently accurate when 50% of events were observed

DC1C (50% censoring) and FDS1 were closely matched, with only slight differences. The risk that subsequent data, following DC1C, will materially change the cure fraction is estimated to be low.

Overestimation of the cure fraction was observed in early data cuts, which resulted in a predicted survival overestimation

- In DC1A and DC1B the cure fraction is overinflated from high amounts of IPD censoring that generate 'false' plateaus in survival (see figure 1).
- To avoid false plateaus, reasonable timepoint to consider recipients of CAR-T therapies as cured should be clinically validated to ensure robust application of an MCM.
- If this timepoint is not validated it is recommended to explicitly state the limitations and uncertainty associated with using MCMs and immature data, or to fit simpler survival models whilst explicitly discussing the uncertainty around cure.

Exponential distributions in MCMs are unlikely to be appropriate for long-term survival predictions

- All exponential models rank poorly for goodness-of-fit across all degrees of data maturity and, visually, matched less appropriately.

Gompertz and log-model functions have properties that effectively capture the hazard profile of the uncured population

- Log-normal and log-logistic fit closely to the KM for FDS1 visually and in terms of goodness-of-fit statistics.
- The Gompertz distribution fits well to all mature and immature datasets according to the goodness-of-fit statistics.
- Gompertz and log-model distributions allow for the application of a gradually increasing or decreasing hazard over time, which this analysis indicates may be the most appropriate and logical approach for modelling the uncured population.

Goodness-of-fit statistics may not be appropriate for determining the distribution with the best fit in the longer-term survival predictions

- There were inconsistencies between goodness-of-fit statistics in early data-cuts and FDS1. Clinical validation should lead curve selection.

Conclusion

- These results have been generated by reconstructing the IPD of patients treated with CAR-T therapy using published five-year ZUMA-1 overall survival curves and truncating the reconstructed IPD to proxy earlier interim data cuts.
- The results have been externally validated against published interim data cuts of ZUMA-1, other CAR-T treatments, and similar publications.[1-2]

Based on the results and external validation checks, the statistical suitability of extrapolating overall survival of CAR-T treatments may be most appropriate with a maximum of 50% event censoring.

- At 50% censoring, the predicted cured proportion is between 47% to 52%, whereas the full dataset (43% censoring) has a cured proportion of between 40% to 44% (a mid-point divergence of 7%).The cure fraction is still inflated but the estimation of empirical cure via MCM is considered more appropriate than capturing cure in other model structures.
- At 60% censoring, the predicted cured proportion is between 59% to 62%, resulting in a mid-point divergence of 18% compared with the full dataset.
- At 70% censoring, the predicted cured proportion is between 67% to 72%, resulting in a mid-point divergence of 28% compared with the full dataset (42%).
- Mid-point divergences at 60% and 70% OS censoring demonstrate that MCM is likely to be generally inappropriate as a model structure. Clinical validation of the cure fraction is always important, but is particularly important when censored OS events exceed 50% of the trial population.

References

1. Vadgama S, et al. *Value Health 2022*;25(6):1010-17.
2. Peterse EF, et al. *PharmacoEconomics-Open 2023*;7(6):941-50.
3. Neelapu SS, et al. *Blood 2023*;141(19):2307-15.
4. Guyot P, et al. *BMC Med Res Methodol 2012*;12:1-13.
5. Rutherford M, et al. *Flexible methods for survival analysis tsd 2020*. London: National Institute for Health and Care Excellence.
6. Locke FL, et al. *Lancet Oncol 2019*;20(1):31-42.

Presented at ISPOR Europe 2024; November 17-24, 2024; Barcelona, Spain

Email: samuel.harper@york.ac.uk