AI-Assisted Time-to-Event Projection: A Case Study and Broader Potential

MSR22

K. Jack Ishak,[1] J. Jaime Caro[1]

[1]Thermo Fisher Scientific, Montreal, Canada

Background

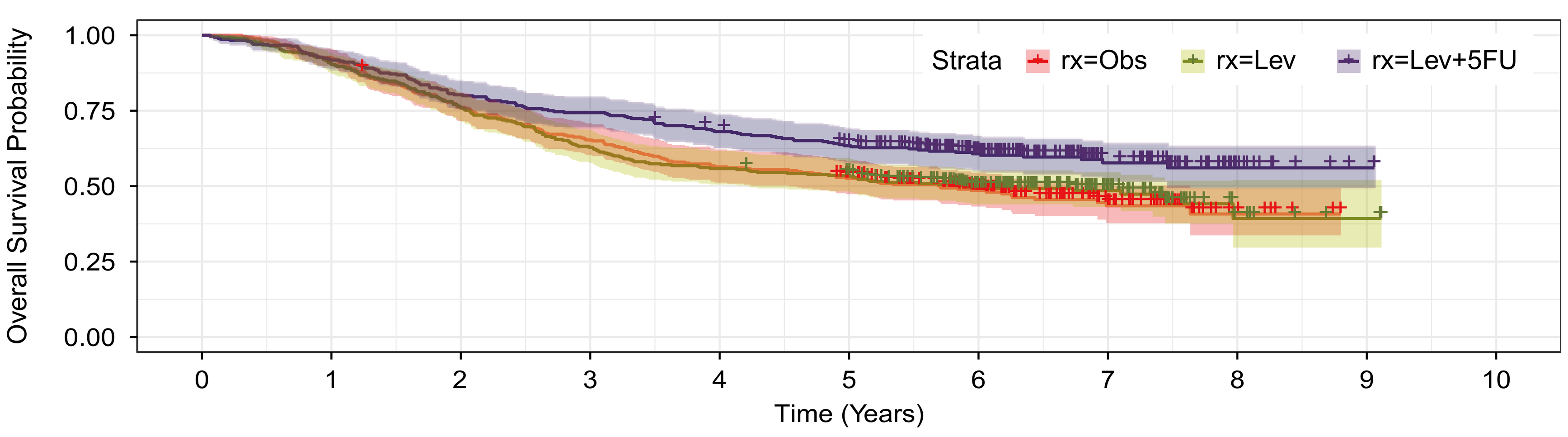
- Time-to-event projection involves balancing clinical information about the disease or treatments being studied and statistical fit considerations to ensure the selected model is appropriate.
- We explored the use of large language models (LLMs) (e.g., Chat-GPT, Gemini) to assist with understanding the prognosis of the study population to establish plausible ranges of life expectancy or median survival times.

Case Study: Overall Survival in Stage B/C Colon Cancer

Data Source and Analysis

- Randomized trial demonstrating benefit of adjuvant chemotherapy for colon cancer[1]
  - Compared levamisole (Lev) and levamisole with 5-FU (Lev+5FU) to observation alone (Obs)
- Outcome of interest: overall survival (OS) (**Figure 1**)

Figure 1. Observed time-to-death distribution with Lev+5FU, Lev. and Obs

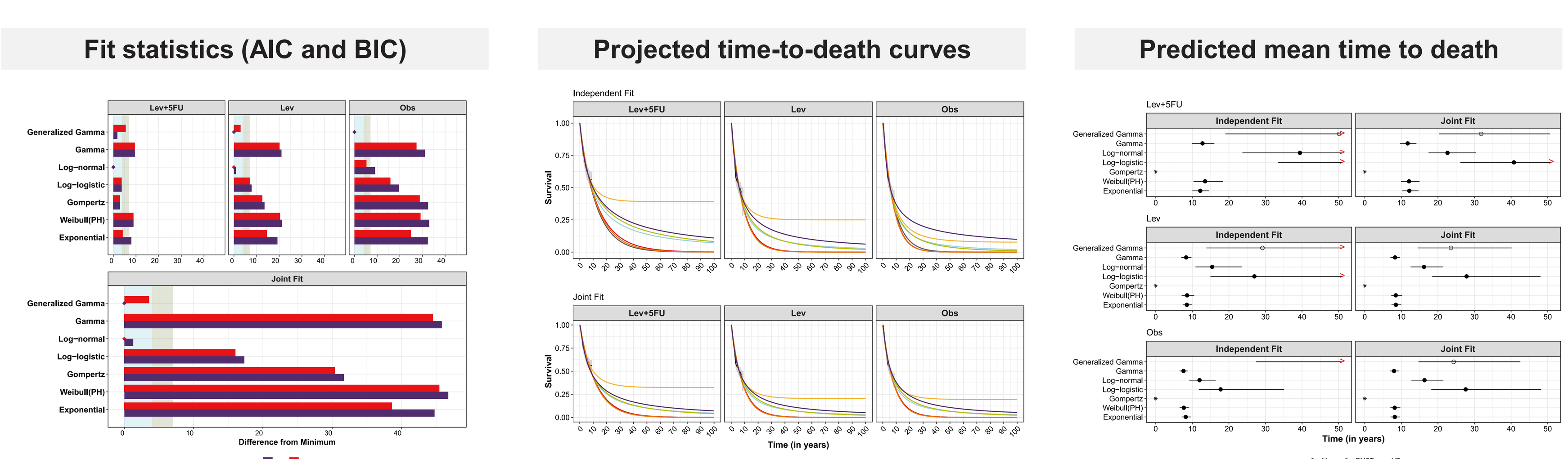


Abbreviations: Lev = levamisole; Lev+5FU = levamisole with 5FU; Obs = observation alone

Parametric Fitting Analyses

- Standard parametric analyses with supporting diagnostic assessments were carried out to select an optimal fit.
  - Analyses considered joint fitting (i.e., assuming proportional hazards) and separate fitting with exponential, Weibull, Gompertz, log-logistic, log-normal, gamma, and generalized gamma distributions.
- Results from analyses suggest hazards are not proportional over time, and log-normal and generalized gamma provide best statistical fit and produce projections that are middle of the road (i.e., between more aggressive estimates from Weibull, for example, and implausibly optimistic fits from Gompertz).
- Selected results from fitting analyses are shown in **Figure 2**.

Figure 2. Fit statistics, projections and predicted meant event times from projection analyses



*Point estimate, lower bound or upper bound not estimated.
RMST calculated up to 99th percentile of the fitted distribution.
Abbreviations: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Lev = levamisole; Lev+5FU = levamisole with 5FU; NE = not estimated; Obs = observation alone; PH = proportional hazard; RMST = restricted mean survival time

AI as a Clinical Assistant

- We iteratively prompted the LLM to provide and request information as described below (**Figure 3**).
- The LLM provided detailed responses at each step, including relevant cautionary/advisory points, such as:
  - The current applicability of the observed results (since the data are from a study conducted pre-1990s)
  - Looking for convergence to general mortality after 10–15 years "as eventually, the excess cancer mortality risk should diminish."
  - Possibility of functional cure supporting 25+ years of maximum survival
- The LLM selected the jointly fitted log-normal distribution based on fit, plausibility, and stability of predictions (e.g., mean survival) and suggested sensitivity analyses to deal with limitations of selected fits (i.e., optimistic tails).

Figure 3. Overview of information exchange with LLM to obtain insights to guide model selection

Context & Background Information
- Assign persona (clinical expert advising analysis)
- Ask to review the study being analyzed
- Ask to summarize the study population and clinical profile

A Priori Expected Survival
- Ask for estimates of the expected survival of participants of the study, considering the period in which the study is conducted
- Provide characteristics of population (% male, age, etc.; LLM had difficulty extracting this information)

Interpretation and Recommendations
- Provide a structured summary, including predicted mean, median and maximum survival times, best statistical fit, and those with comparable or similar fit (AIC/BIC within 4-7 points)
- Ask for recommendations on which fit to select for use in an economic model

Dealing with Implausible Tails
- Ask for suggestions for how to deal with implausible predictions or tails from the selected distribution

Summary
- Ask for a summary of the background, recommendations and suggestions for handling limitations of fits

Abbreviations: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LLM = large language models

Insights about Population Profile

- The LLM characterized the population as follows:

> In short, the cohort you are modeling is a **high-risk but relatively fit** population that had undergone complete removal of their primary tumor, with nodal involvement being the dominant driver of their poor long-term prognosis.

- It cautioned about bias and applicability of results from the study, and considerations for projection analyses:

> When extrapolating the survival tail of your model, be mindful that **non-cancer mortality** related to toxicity/poor compliance might have been higher in the older cohort than it would be today.

A Priori Expected Survival

- The LLM estimated the median survival to be around 3.5–4.5 years and noted that long-term survivors would pull the mean (expected) survival to 5–7 years.
- It suggested survival could extend to 25 years or more in some patients since "a small but significant fraction of patients are cured by surgery alone" and a potentially young age at diagnosis in some patients.

> The combination of these high-risk features suggests the true mean OS for your specific cohort would lean toward the **lower end** of the typical range for all Stage C patients, making **5 to 7 years** a clinically sound estimate for the mean survival of the observation group.

Interpretation and Recommendations

- It provided a recommended fit with rationale explaining the selection (as well as rationale for excluded models – not shown) and suggested sensitivity analyses with alternative distributions.

Metric	Log-normal (Joint Fit)	Rationale
Goodness-of-Fit	Best fit according to the **BIC** and competitive with the **AIC**'s best fit (Generalized Gamma). BIC penalizes models with more parameters, favoring parsimony—a desirable trait for extrapolation.	
Predicted Median OS	**8.18** years (6.68−9.93)	This is clinically plausible for an effective adjuvant regimen in this high-risk population.
Predicted Mean OS	**22.62** years (17.42−30.39)	While high, this is a **finite, numerically stable estimate** and is far more reasonable than the highly inflated means from the Generalized Gamma and Log-logistic models.
Long-Term Plausibility	10% event-free at **50.87** years.	This extrapolation, though optimistic, is contained within a clinically plausible timeframe for a proportion of younger patients to survive.

Dealing with Implausible Fits

- The LLM's assessment that the tail of jointly fitted log-normal is "optimistic" was questionable, particularly the 99th percentile stretched to hundreds of years.
- When prompted for recommendations on how to best deal with the implausible tails, the LLM proposed three strategies with supporting explanations to implement these:
  - Blending with General Population Mortality
  - Imposing a "Cure Point" (Curtailment)
  - Restricting the Mean Survival Time (RMST)

Summary of Insights

- The LLM produced a report of the insights requested during the exchange and a summary of key points.

> The study population consists of high-risk **Stage C resected colon cancer patients** who were medically fit for chemotherapy, characterized by poor prognostic indicators such as a high proportion of patients having multiple positive lymph nodes. The *a priori* expected mean overall survival for the untreated group is realistically conservative, estimated to be between **5 and 7 years**, with maximum plausible survival limited by natural causes to roughly 25 to 35 years. For the economic model base case, the **Log-normal distribution (Joint Fit)** is recommended as it achieves the **best balance** of statistical stability and goodness-of-fit (BIC), ensuring reliable modeling of relative treatment effects. The critical limitation of this fit, however, is its tendency to produce **implausible long-term survival tails** (centuries long); consequently, its use mandates a **blending strategy** with **General Population Mortality (GPM)** rates to ensure the final survival predictions are clinically realistic.

Conclusions

- This case study shows that LLMs can be helpful in providing clinical insights to aid analysts to interpret results from time-to-event fitting.
- The LLM demonstrated an understanding of considerations for projecting time-to-event curves (e.g., behavior of long tails and impact on mean vs. median OS) and flagged relevant clinical issues on use of the data.
- Responses from the LLM on a priori expected survival were particularly useful as this type of information is not always readily available to analysts.
  - The LLM's responses were supported with explanations of its reasoning (e.g., risk factors driving estimates).
  - The LLM produced a clear summary of its recommendations, capturing the salient points from the exchange.
- Responses from the LLM may be influenced by the extent of information provided and how this is structured.
  - We summarized results to provide to the LLM for interpretation in which we identified best-fitting models (based on AIC/BIC) and labeled others as comparable or similar, as appropriate.
  - Changing this format and the type of information provided could affect responses; further exploration of this would be beneficial.
- While this case study focused on clinical insights and interpretation, LLMs can also help with suggestions of analyses to perform, coding support, and communication of results.

References

1. CG Moertel, et al. *N Engl J Med*, 332:352-358, 1990.

Disclosures

All authors are employees of PPD™ Evidera™ Health Economics & Market Access, Thermo Fisher Scientific. The development of this poster was funded by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Sean Smith and Shani Berger of Thermo Fisher Scientific.

Presented at ISPOR —The Professional Society for Health Economics and Outcomes Research—Europe | 9–12 November 2025 | Glasgow, Scotland, UK