AN INTEGRATED WEB-BASED APPROACH TO SURVIVAL ANALYSIS AND EXTRAPOLATION FOR ECONOMIC MODELING

Nikolay A Avxentyev[1], Aleksandr S Makarov[2]

[1] *Financial Research Institute, Moscow; Health and Market Access Consulting, Moscow; Pharmaceutical Analytics, New York, United States*

[2] *Health and Market Access Consulting, Moscow; Pharmaceutical Analytics Middle East, Ras al Khaimah, United Arab Emirates*

BACKGROUND

- Survival analysis plays a central role in health technology assessment (HTA) submissions by estimating time-to-event outcomes such as progression or death and supporting long-term projections required for modelling. Analysts typically rely on several separate tools for data extraction, statistical modelling, and exporting results to economic models, which complicates workflows, limits reproducibility, and leads to a steep learning curve
- To overcome these barriers, we developed a **free and non-commercial** integrated online platform designed to perform the entire survival analysis process from data input to economic model export within a single interface

OBJECTIVES

- The primary objective was to develop an integrated web-based platform that unifies all stages of survival analysis and extrapolation from data input and digitization to model fitting and export for economic evaluations
- The goal was to simplify the analytical process by combining validated open-source R packages within a web interface, enabling users to perform robust, reproducible survival analyses without coding or switching between multiple tools

METHODS

- The Surv platform was implemented as a web-based application using Python for the interface and R as the analytical engine:
  - A customized version of WebPlotDigitizer (https://automeris.io/WebPlotDigitizer) was embedded directly into the interface, allowing users to digitize published Kaplan-Meier curves (Kaplan EL et al., 1958) or import coordinate data without switching tools
  - The analytical core employs open-source, validated R packages (“survival”, “flexsurv”, “flexsurvcure”, “survHE” and few others)
- Users can upload individual patient-level data, digitized KM coordinates, or KM images for automatic pseudo-IPD reconstruction using a validated algorithm (Guyot P et al., 2012)
- The platform supports both single- and two-arm analyses with estimation of hazard ratios, 95% confidence intervals, restricted mean survival times (RMST), and performs proportional hazards testing
- For extrapolation, users can fit multiple model classes including parametric distributions (Weibull, Gompertz, Log-logistic, Log-normal, Generalized Gamma, Exponential, as well as their piecewise versions), spline-based models (normal, hazard, and odds scales), and mixture-cure models for flexible extrapolation
- Fitted models are compared by AIC/BIC, and diagnostic plots are automatically produced for each model.
- Final results are exported into Excel templates with built-in survival functions, allowing immediate use in cost-effectiveness or budget-impact analyses.

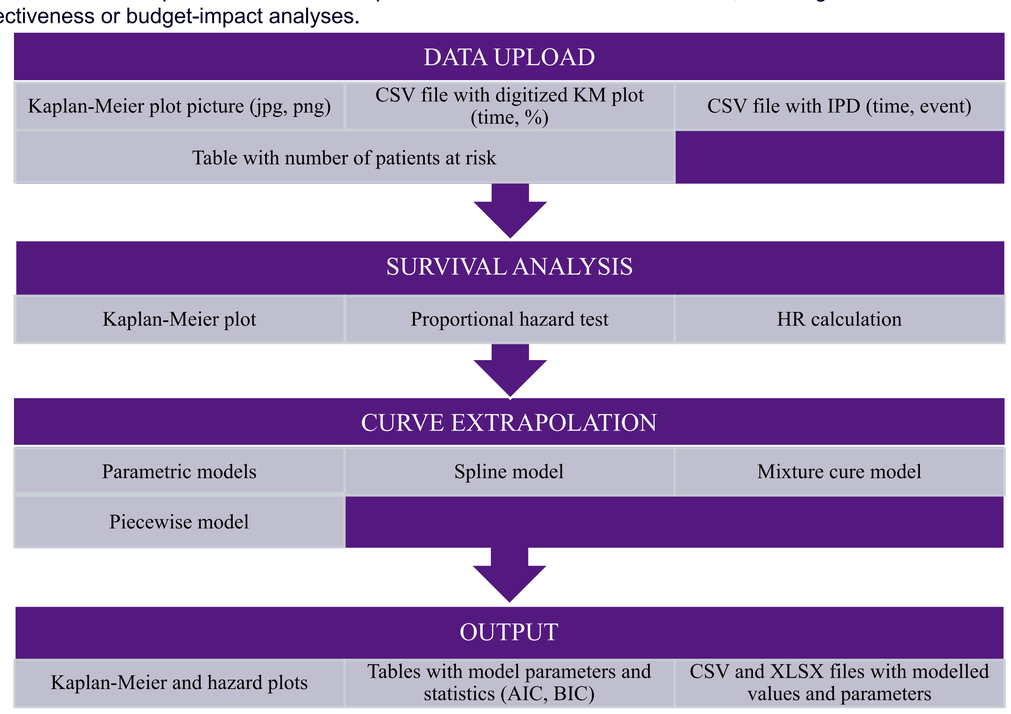

Figure 1. User journey flowchart

RESULTS

- Figure 2 shows the data entry interface of the Surv web platform. The left-hand navigation bar provides direct access to five main modules: Entry Data, Survival Analysis, Standard Parametric Models, Spline, and Mixture Cure Model

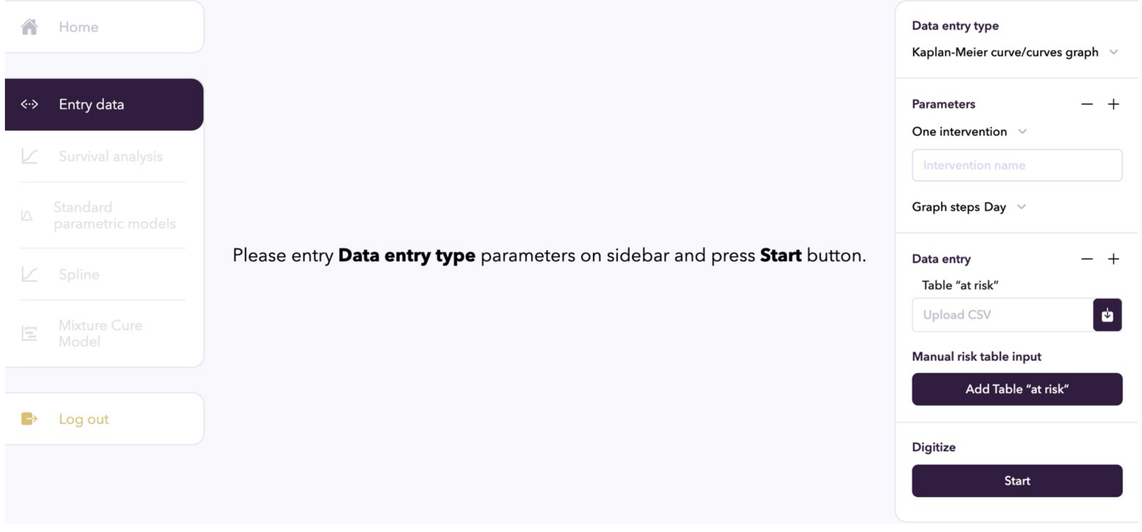

Figure 2. Data Entry module. Left: navigation panel; Right: upload and parameter setup interface.

- Users begin by selecting the data entry type (e.g., Kaplan-Meier curve, individual patient-level data, or digitized points) and defining parameters such as number of interventions, graph time frame and intervention names. CSV files containing risk tables can be uploaded, or data can be entered manually via the built-in “Add Table ‘at risk’” form. Once parameters are set, users initiate the process with the Start button
- After data are imported, the digitization interface enables users to overlay extraction points directly on Kaplan-Meier graphs (Fig. 3). Integrated curve-recognition tools (mask, pen, box, erase) allow fine control over point placement, and all coordinates are displayed dynamically in the panel on the right. The extracted survival data are visualized in real time and stored for subsequent analyses

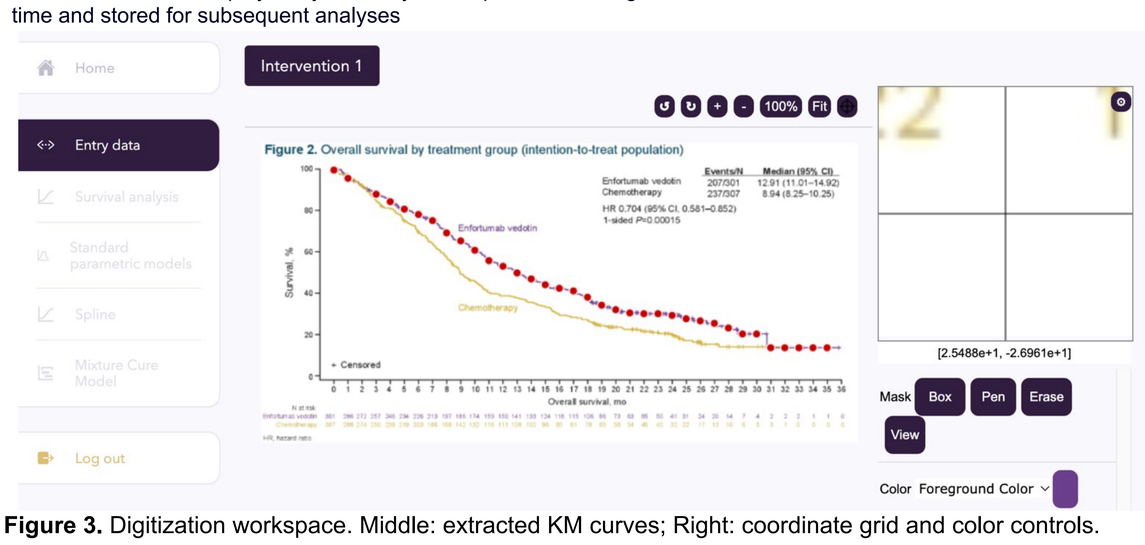

Figure 3. Digitization workspace. Middle: extracted KM curves; Right: coordinate grid and color controls.

- Within the Survival Analysis tab (Fig. 4), users can generate Kaplan-Meier plots instantly access computed statistics:
  - Total number of events, event rate, and log-rank p-value
  - Cox proportional hazards regression (with a single binary covariate – intervention arm) output: hazard ratio and p-value
  - RMST difference and Schoenfeld residual test results

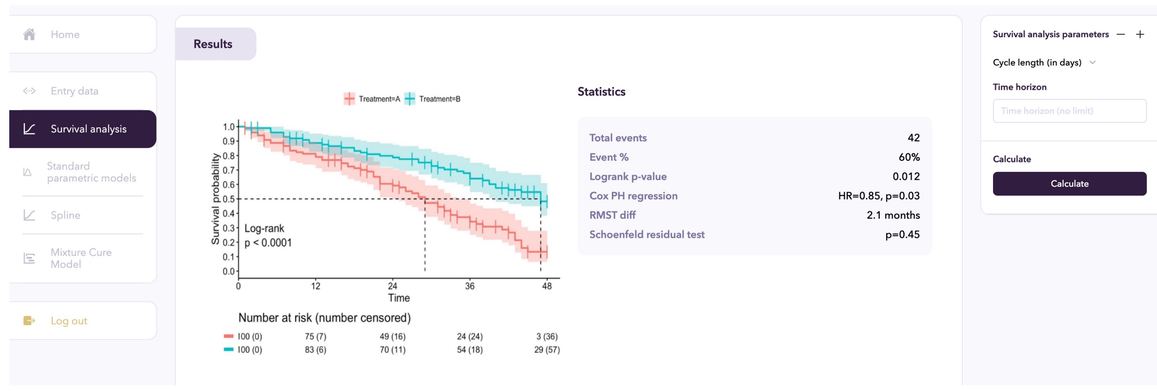

Figure 4. Survival Analysis tab. Middle: KM Plot with extrapolation; Right: statistics panel.

- The Standard Parametric Models section allows direct fitting of multiple distributions (Weibull, Gompertz, log-logistic, log-normal, exponential, and generalized gamma) as well as piecewise models with user-defined time intervals for improved flexibility in long-term extrapolation (Fig. 5):
  - Each fitted curve is overlaid on the observed Kaplan–Meier data for immediate visual comparison. Hazard curves are also presented on separate plot. All graphs update automatically when parameters such as cycle length or time horizon are modified
  - The interface automatically computes AIC and BIC values for each model and displays them alongside survival-under-curve metrics to help users identify the best-fitting function
  - All fitted model outputs (including estimated parameters and information criteria) can be exported as CSV or XLSX files for further processing during economic modelling

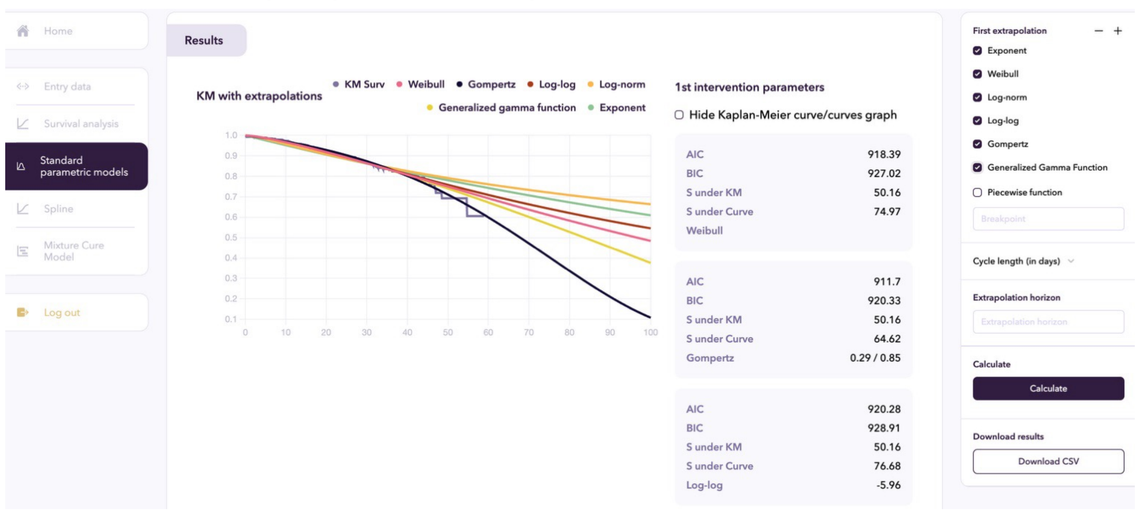

Figure 5. Standard Parametric Models tab. Middle: overlay of fitted curves; Right: model-selection metrics and export panel.

Advanced models

- Beyond the standard parametric module, the Spline and Mixture Cure Model (MCM) tabs provide advanced flexibility for extrapolating survival data
- The Spline module allows users to fit survival curves using natural cubic splines with user defined number of knots and their coordinates, supporting modelling on normal, hazard, or odds scales
- The Mixture Cure Model tab enables users to incorporate the concept of a cured patient fraction, separating long-term survivors from those experiencing events
- Both modules include overlaid Kaplan-Meier visualizations, and export of all fitted parameters for use in downstream cost-effectiveness analyses. The interface layout follows the same structure as in Standard parametric tab

CONCLUSIONS

- The Surv platform consolidates all key steps of survival analysis and extrapolation into a unified, browser-based environment. By integrating curve digitization, statistical modelling, and economic export within one code-free interface, it significantly streamlines analytical workflows and minimizes potential sources of error

References:

1. Rohatgi A. WebPlotDigitizer. [software]. 2011-2024. Available from: https://automeris.io/WebPlotDigitizer.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-481
3. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. BMC Med Res Methodol. 2012;12:9. doi:10.1186/1471-2288-12-9.
4. Therneau T. survival: Survival Analysis. R package version 3.7-0. 2024. Available from: https://CRAN.R-project.org/package=survival
5. Jackson C. flexsurv: Flexible Parametric Survival and Multi-State Models. R package version 2.3.2. 2024. Available from: https://CRAN.R-project.org/package=flexsurv
6. Bender A, Scheipl F, Dimitriadou E. flexsurvcure: Flexible Parametric Cure Models for Survival Analysis. R package version 1.2.0. 2023. Available from: https://CRAN.R-project.org/package=flexsurvcure
7. Baio G. survHE: Survival Analysis for Health Economic Evaluation and Cost-Effectiveness Modeling. R package version 1.1.4. 2022. Available from: https://CRAN.R-project.org/package=survHE