Joint or simultaneous modelling of related longitudinal and time-to-event data for a network of treatments across multiple data sources

Maria Sudell[1]*, Catrin Tudur-Smith[1] and Ruwanthi Kolamunnage-Dona[1]
[1]Department of Biostatistics, University of Liverpool mesudell@liverpool.ac.uk

1 Introduction

Joint models for longitudinal and time-to-event data simultaneously model an outcome measured repeatedly over time (longitudinal) e.g. blood pressure, and the time until a specified event (time-to-event) e.g. death. For data from a single study, joint models are preferable to separate longitudinal or time-to-event models when a relationship exists between the longitudinal and time-to-event outcomes. Joint models have been expanded to allow for clustered data from multiple sources (e.g. studies in a meta-analyses) [1,2]. However, these methods are limited to cases where the same set of treatments are examined in every data source. This research extends joint modelling methodology to allow analysis of multi-data source data, where sources may examine different subsets of the possible treatments for a given clinical condition.

2 Standard Multi-Data Source Joint Model

Longitudinal sub-model

$$Y_i(t_i) = X_{1i}(t_i)\beta_1 + Z_i^{(2)}(t_i)b_i^{(2)} + Z_i^{(3)}(t_i)b_i^{(3)} + \varepsilon_i$$
$$= \eta_{Li}(t_i) + \varepsilon_i$$

Time-to-event sub-model

$$h_i(t) = h_0(t)\exp\left(X_{2i}\beta_2 + f(\eta_{Li}(t))\right)$$
$$h_i(t) = \lambda\gamma t^{\gamma-1}\exp\left(X_{2i}\beta_2 + (\eta_{Li}(t))\right)$$

Association structure

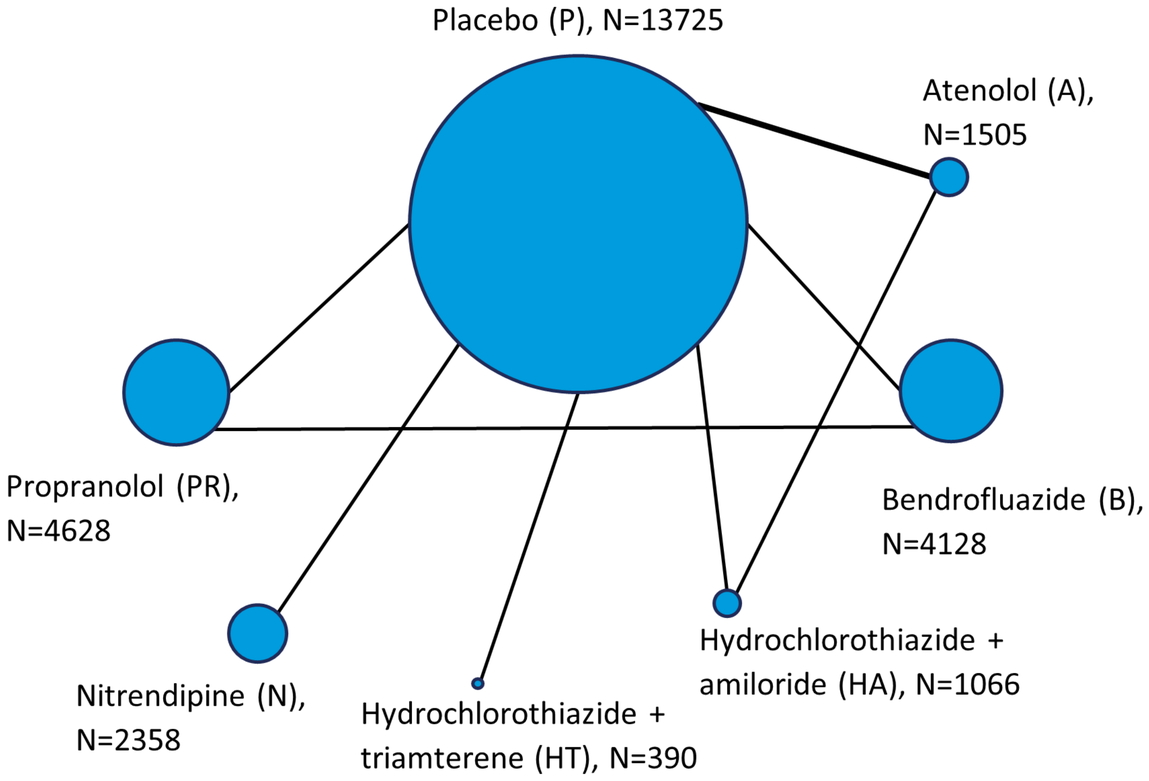
$$f(\eta_{Li}(t)) = \eta_{Li}(t)$$

The standard multi-data source joint model involves a longitudinal linear mixed effects sub-model, with fixed or population level effects, and random effects at the individual (level 2) and data source (level 3) level. The time-to-event sub-model is a proportional hazards model with Weibull baseline hazard

Many association structures have been proposed in the literature [3], here we consider the current value association structure which shares the unobserved true value of the longitudinal outcome ($\eta_{Li}(t)$) between the sub-models

3 Extension to NMA

The standard joint model above assumes all included data sources compare the same set of treatments. This is often not the case in healthcare, where different studies, hospitals, sites may examine different subsets of the possible treatments for a condition. Given the populations from the different sources are comparable, we can use Network Meta-Analytic (NMA) techniques [4] to analyse all the treatments across a network of data sources.



Above is a network of treatments compared in a subset of 5 studies from the INDANA dataset [5]. The node size represents the number of individuals assigned to a given treatment, and the connector width represents the number of studies with data for each treatment comparison.

This dataset involve longitudinally measured Systolic Blood Pressure (SBP), and time until death for 5 studies, containing 27440 individuals and 336419 longitudinal measurements.

4 Basic Parameters in NMA

Treatment effects in an NMA are represented using basic parameters $\beta_{trt(1,g)}$, which hold the treatment effect of each treatment h compared to a Network Reference Treatment (NRT) – the standard or most commonly used treatment (here using placebo), which is labelled treatment 1. Treatment comparisons between other treatments can be made by combining basic parameters:

Treatment 3 vs NRT treatment 1: $\beta_{trt(1,3)}$

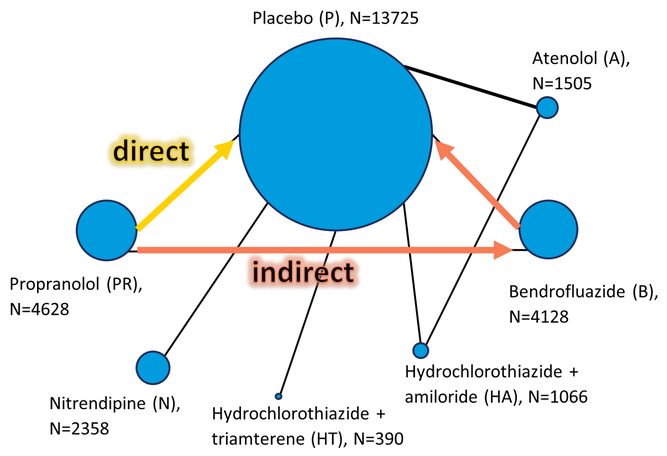
Treatment 2 vs treatment 3: $\beta_{trt(2,3)} = \beta_{trt(1,3)} - \beta_{trt(1,2)}$

5 Consistency across joint models

When comparing treatments in NMA, both direct and indirect information can be available (e.g. on graph, direct and indirect information for Propanolol vs Placebo)

Consistency is the direction of treatment effect agreeing between indirect and direct information

Consistency equations are the combinations $\beta_{trt(2,3)} = \beta_{trt(1,3)} - \beta_{trt(1,2)}$ that combine basic parameters to form treatment comparison



Consistency equations should be specified for each term that involves treatment in the model. In joint NMA consistency holds when the equations are specified on longitudinal treatment effects β_{trt_L}, and time-to-event treatment effects β_{trt_S}, even though longitudinal treatment terms are shared through the association structure

$$\beta_{trt_{S(g,h)}} + \alpha_c\beta_{trt_{L(g,h)}} = \left(\beta_{trt_{S(1,h)}} - \beta_{trt_{S(1,g)}}\right) + \alpha_c\left(\beta_{trt_{L(1,h)}} - \beta_{trt_{L(1,g)}}\right)$$
$$= \left(\beta_{trt_{S(1,h)}} + \alpha_c\beta_{trt_{L(1,h)}}\right) - \left(\beta_{trt_{S(1,g)}} + \alpha_c\beta_{trt_{L(1,g)}}\right)$$

6 Selecting study specific random effects

Another consideration is allowing treatment terms to be assigned random effects. Treatments examined will vary by data source, so random effects needed will vary by study

A method is required to update a master covariance matrix, containing estimates of variability for random effects comparing each treatment to then NRT, that can be subsetted to provide study specific random effects covariance matrices

1. Define master covariance matrix for random effects e.g. $b_k^3 \sim N(0, A)$
2. Partition matrix into blocks related to and not related to treatment (b_k^3 ordered accordingly)
$$A = \begin{pmatrix} A_{(not\_trt)} & A_{(trt,non\_trt)} \\ A_{(trt,non\_trt)} & A_{(trt)} \end{pmatrix}$$
3. Define block diagonal selection matrix B_k for each study to extract study specific covariance matrix A_k from master matrix A_k
4. $$B_k = \begin{pmatrix} B_{k(not\_trt)} & 0 \\ 0 & B_{k(trt)} \end{pmatrix}$$
$B_{k(not\_trt)}$ is an identity matrix
$B_{k(trt)}$ is a matrix of 0, 1 and -1 defining the required combinations of treatments
5. Extract study specific covariance matrix A_k to correctly control each study's random effects
$$A_k = B_k A B_k^T$$

7 Running joint NMA model in Stan

Model implemented in R [6] using Stan [7]. Initial values where possible informed by standalone fits, uninformative priors if not.

Currently time-consuming to fit (several days if not longer for single chain)

Current work is focusing on approaches to allow these methods to run fast enough to be practically useful. Approaches being investigated include:

- Use of condor systems to parallelise (unsuitable for TREs)
- SMC samplers (requires some parallelisation to be effective)
- Variational bayes approximation

8 Discussion

This research has extended joint modelling for longitudinal and time-to-event data in the multi-data source setting to allow for differing subsets of potential treatment options between data sources. The methodology has been implemented in R and Stan, but further work is required to make the modelling approach sufficiently efficient to be practically useful

References

[1] Sudell, M., Tudur Smith, C., Gueyffier, F., & Kolamunnage-Dona, R. (2018). Investigation of 2-stage meta-analysis methods for joint longitudinal and time-to-event data through simulation and real data application. Statistics in medicine, 37(8), 1227–1244. https://doi.org/10.1002/sim.7585, [2] Sudell M, Kolamunnage-Dona R, Gueyffier F, Tudur Smith C. Investigation of one-stage meta-analysis methods for joint longitudinal and time-to-event data through simulation and real data application. Statistics in Medicine. 2019; 38: 247–268. https://doi.org/10.1002/sim.7961, [3] Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., & Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Statistics in medicine, 34(14), 2181–2195. https://doi.org/10.1002/sim.6141, [4] Dias, S. (2018) *Network meta-analysis for decision making / by Sofia Dias, University of Bristol, Bristol, UK [and four others]*. Hoboken, NJ: Wiley.
[5] Gueyffier F, Boutitie F, Boissel JP, et al. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. Therapie. 1995 Jul-Aug;50(4):353-362. PMID: 7482389. [6] R Core Team (2024). _R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ [7] Stan Development Team. 2025. Stan Reference Manual, v2.36.0. https://mc-stan.org