Predicting loss of Christmas Cheer in Xmas Elves

A Prospective Cohort Study by the North Pole Medical Centre

Nathan Constantine-Cooke (Vallejos Group)

The Problem

Santa, director of the North Pole Medical Centre, noticed a growing problem amongst his beloved elves: many were becoming overwhelmed with a lack of Christmas cheer each year.

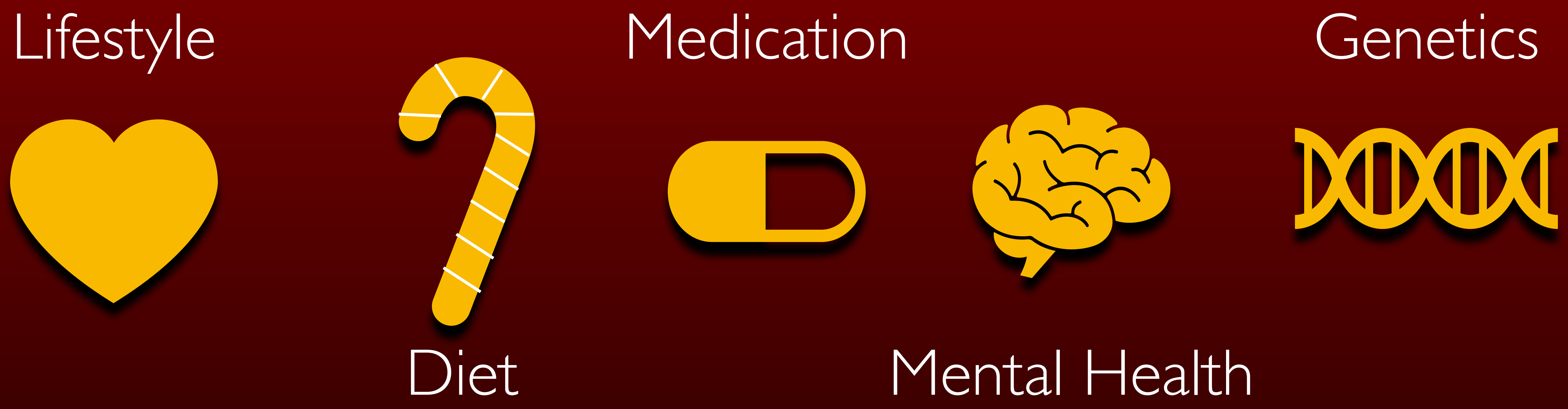
This lack of Christmas cheer may result in Santa's workshop being unable to make enough presents!

The Study

- As such, Santa decide to establish a prospective study to establish associations with loss of Christmas cheer using a time-to-event model.
- All elves who work in the workshop were recruited.
- End point was defined as an elf no longer joining-in with spontaneous sing-a-longs.
- Elves were recruited after the post-Christmas break when all elves have regained their Christmas cheer and were all equally at risk of losing their Christmas cheer.

A Holistic Approach to Xmas Cheer

- Santa and his team suspected that lack of Xmas cheer was driven by complex interacting factors including:



- And therefore asked participants to answer questionnaires based around these factors and to provide genetic samples*

* which happens to be a remarkably similar experimental design to the PREdiCCt study.

Time-to-Event Analysis

In the absence of longitudinal biomarkers

One of the most popular forms of time-to-event analysis is **Cox proportional hazards models** which assumes that the effect of a covariate is to multiply the hazard by some constant.

$$h(t, X) = \underbrace{h_0(t)}_{\text{baseline hazard}} \cdot e^{\underbrace{\beta_1 x_1 + \beta_2 x_2 + \ldots}_{\text{covariates and coefficients}}}$$

But What if We Want to Use Longitudinal Biomarkers?

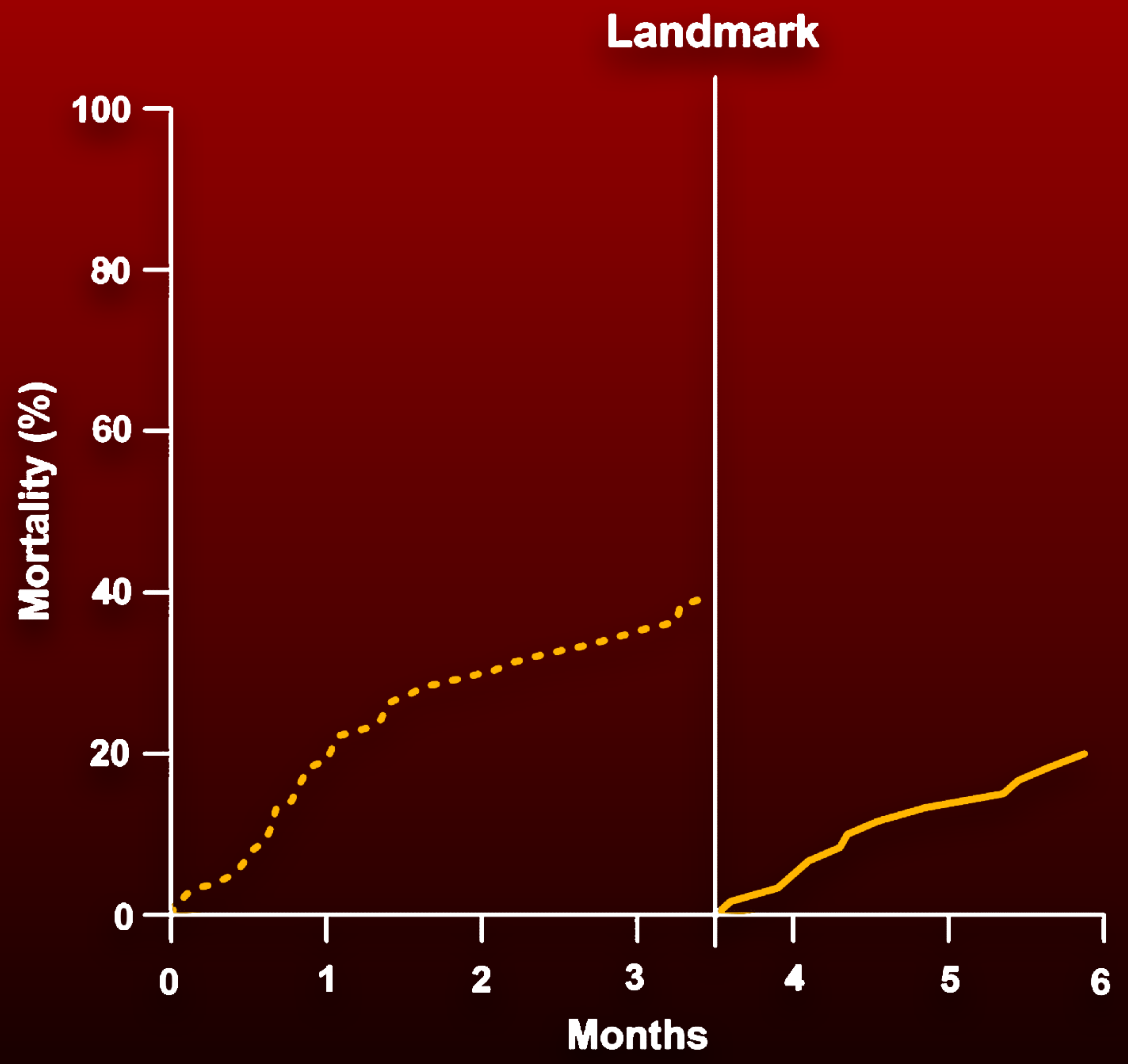
- Suspecting that specific, commonly measured, proteins could act as biomarkers for lack of Xmas cheer, Santa quickly realised that incorporating longitudinal measurements of these proteins could improve accuracy of his models.

- But Cox Proportional hazards models are not suitable for biomarkers!

If we wish to use longitudinally collected biomarkers, then two broad family of approaches are applicable:

Landmarking

Landmarking fits time-to-event models at landmark times using subjects still at risk at the landmark times.



Morgan, C.J. Landmark analysis: A primer. *J. Nucl. Cardiol.* **26,** 391–393 (2019). https://doi.org/10.1007/s12350-019-01624-z
Modified

Landmarking

Advantages

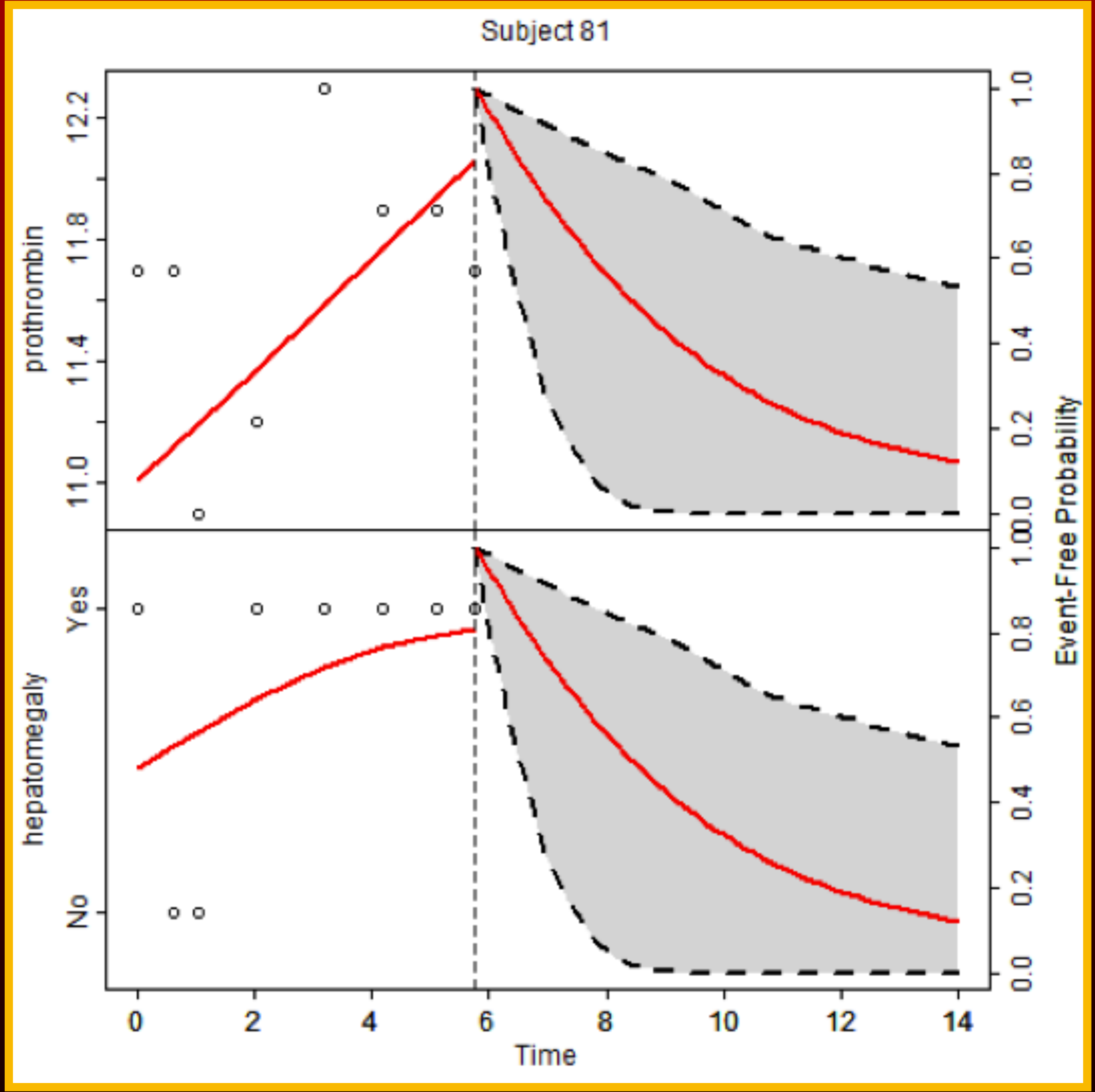
- Computationally fast
- Can be used with any software which supports time-to-event models.
- Requires few assumptions and opportunities for misspecification

Disadvantages

- Does not use all of the longitudinal data available
- As separate models are fitted for each landmark time point, it may be difficult to make inferences for overall disease progression

Joint Models

Joint Models models the survival process and the longitudinal process and then link these models together.



Rizopoulos, D. “The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC”. Journal of Statistical Software, 2016, 72, 1-45

Joint Models

Advantages

- Uses all of the longitudinal data available.
- Allows LMEs to be used which assume measurements are made with error.
- Models the overall progression

Disadvantages

- Computationally slow.
- Generally only available via specialist software (mostly R packages).
- Opportunities for misspecification in the LME and how the models are joined

Shared Random Effects Model

- A characteristic of the longitudinal process, defined as a function of the random effects, is included as a covariate of the survival model.

$$
\begin{cases}
h_i\left(t|\mathcal{M}(t)\right) & = h_0(t)\cdot \exp\{\gamma^T\omega_i + \alpha m_i(t)\}, & \\
y_i(t) & = m_i(t)+\epsilon_i(t), & \epsilon_i(t) \sim \mathcal{N}(0,\sigma^2) \\
m_i(t) & = x_i^T(t)\beta_i + z_i^T(t)b_i + \epsilon_i(t) & b_i(t) \sim \mathcal{N}(0,D)
\end{cases}
$$

Joint Latent Class Model

- Assumes a heterogenous population consists of homogenous latent subgroups which share the same marker trajectory and risk of the event.
- Postulates that the association between the two processes is captured by categorical random effects

$$h_i(t|c_i = g; \zeta_g, \delta_g) = h_{0g}(t; \zeta_g) \cdot \exp\{X_{ei}(t)^T \delta_g\}$$

Comparison of Likelihoods

Does not have a closed-form solution!

Shared Random effects model:

$$\ell(\theta) = \sum_{i=1}^{n} \log \int p(y_i|b_i;\theta) \cdot \left(h(T_i|b_i|\theta)^{\delta_i} \cdot S(T_i|b_i;\theta)\right) \cdot p(b_i;\theta)\, db_i$$

Joint latent class model:

$$\ell(\theta) = \sum_{i=1}^{N} \log \sum_{g=1}^{G} \pi_{ig} \cdot f\left(Y_i|c_i = g;\theta_G\right) \cdot h_i\left(T_i|c_i = g;\theta_G\right)^{E_i} \cdot S_i(T_i|c_i = g;\theta_G)$$

Have to sum across all groups!

But still considerably easier computationally!

The End

- Even with the North Pole’s RUDOLPH HPC available, using shared random effects models would take too much time when applied to all of the elves in his massive cohort.
- With time of the essence if all well-behaved children were to receive their presents this year, Santa decided to instead use joint latent class models to predict which elves needed additional boosts to their Christmas cheer.
- Christmas was saved.
- or something. This xmas cheer metaphor ended up a bit laboured.

Acknowledgements

Vallejos Group

Catalina Vallejos

ChantrioInt-Andreas Kapourani

Yipeng Cheng

Evgenii Lobzaev

Andrew Papanastasiou

Christos Maniatis

Karla Monterrubio

James Liley

Rachel Jackson

Alan O'Callaghan

Lees Group / PREdiCCt

Charlie Lees

Gareth Jones

Spyros Siakavellas

Nik Plevris

Laura Lucaciu

Phil Jenkinson

Lisa Derr

Lauren Murdoch

Santa

Fraser McPhie

PREdiCCt