Bayesian inference using Hamiltonian Monte-Carlo algorithm for non-linear joint modelling in the context of cancer immunotherapy

Marion Kerioui\textsuperscript{1,3}

Supervised by Jérémie Guedj\textsuperscript{1} and Solène Desmée\textsuperscript{2}

\textsuperscript{1}INSERM UMR 1137, "Infection, Antimicrobials, Modeling, Evolution", Paris
\textsuperscript{2}INSERM UMR 1246, "methodS in Patients-centered outcomes & HEalth ResEarch", Tours
\textsuperscript{3}Genentech/Roche Clinical Pharmacology, Paris, France

June, 13\textsuperscript{th} 2019
Clinical data IMvigor210 (phase 2) and IMvigor211 (phase 3) trials: patients suffering from advanced or metastatic bladder cancer who did not respond to chemotherapy and treated with Atezolizumab immunotherapy treatment.

Immunotherapy:
- New treatments based on immune system stimulation (Atezolizumab targets PD-L1 to prevent its interaction with its receptor on immune cell),
- Showed impressive results in several cancer, including bladder cancer,
- But also apparition of new types of response (*Hyper-progression, Pseudo-progression*), higher variability in response than with chemotherapy.
Challenges induced by immunotherapy in clinical development:

- Define characteristics of patients to treat and predictive biomarkers of the response to treatment,
- Combinations with other treatments,
- New endpoints to evaluate treatment adapted to the diversity of responses.

⇒ There is a need to develop mathematical models that can characterize the kinetics of response to immunotherapies in order to optimize clinical development and improve patients follow-up and care.
Two main observed responses to treatment:

Longitudinal data

- $y_i$: vector of longitudinal measurements,
- Contains early information in response to treatment,
- Can be modelled in a mixed-effects model framework.

Time-to-event data

- $T_i$: observed event time
- $\delta_i$: event indicator $= \begin{cases} 1 & \text{if event observed} \\ 0 & \text{if event not observed} \end{cases}$
**Joint Models**

The probability to not observe the biomarker depends on current (unobserved) biomarker value

- "Poor responders" are more likely to drop out or to experience the event
- "Good responders" are overrepresented as time goes by

⇒ Sample is not representative (informative censoring), induce bias

⇒ **Joint modelling**

---

2 Rizopoulos et al. (2012) Chapman and Hall/CRC
**Joint Models**

The probability to not observe the biomarker depends on current (unobserved) biomarker value

- "Poor responders" are more likely to drop out or to experience the event
- "Good responders" are overrepresented as time goes by

⇒ Sample is not representative (informative censoring), induce bias

⇒ **Joint modelling**\(^1,2\)

**Longitudinal part** - Mixed-effect models

\[ y_i(t) = X(t, \psi_i) \times (1 + e_i(t)) \]

- **\( X \):** process of interest (Tumor size) **possibly non-linear**
- **\( \psi_i = \tau(\mu, \eta_i) \):** individual longitudinal parameters
- **\( e_i(t) \sim \mathcal{N}(0, \sigma^2) \):** residual error

\(^1\) Tsiasis et al. (1995) Journal of the American Statistical Association
\(^2\) Rizopoulos et al. (2012) Chapman and Hall/CRC
The probability to not observe the biomarker depends on current (unobserved) biomarker value

- "Poor responders" are more likely to drop out or to experience the event
- "Good responders" are overrepresented as time goes by

⇒ Sample is not representative (informative censoring), induce bias

⇒ **Joint modelling**

**Longitudinal part** - Mixed-effect models

\[ y_i(t) = X(t, \psi_i) \times (1 + e_i(t)) \]

- \( X \): process of interest (Tumor size) **possibly non-linear**
- \( \psi_i = \tau(\mu, \eta_i) \): individual longitudinal parameters
- \( e_i(t) \sim N(0, \sigma^2) \) residual error

**Survival part** - Hazard function for patient \( i \)

\[ h_i(t|\psi_i) = h_0(t) \exp(\beta \times f(t, \psi_i)) \quad \text{for} \quad t \geq 0 \]

\[ S_i(t|\psi_i) = P(T_i \geq t) = \exp \left[ - \int_0^t h_i(u|\psi_i) \, du \right] \]

- Link function \( f \) depends on \( \psi_i \)

---

Non-Linear Joint Model

Use of mechanistic models can be suited to characterize biomarker kinetics:

- **Many measurements** of biomarker in the context of clinical trial\(^1\),
- High **biological complexity** of the tumor size kinetics,
- Exacerbated in the context of **immunotherapy** by the complex interaction between the drug, the immune response and the tumor.

⇒ Biomarker kinetics is described by a **non-linear mixed-effects model**,

- Increase of the likelihood expression complexity,
- Requires high performance algorithm.
- Inference in a frequentist framework can be done by maximum likelihood using SAEM (Stochastic Approximation of EM Algorithm)\(^2\),\(^3\).

---

\(^1\) Desmée et al. (2016) Biometrics
\(^2\) Desmée et al. (2015) AAPS
\(^3\) Tardivon et al. (2018) CPT
Bayesian Inference and HMC algorithm

The complex likelihood expression of non-linear joint models already requires high-performance algorithm for inference:

- Bayesian approach offers a natural framework to include prior information to increase identifiability,
- A new inference tool could help to go further in modelisation?

Stan\(^1\) bayesian software:

- Hamiltonian Monte-Carlo algorithm\(^2\) known to have good convergence properties for complex models (Hamiltonian dynamics),
- No-U-Turn Sampler Version\(^3\) optimized version of HMC algorithm.

Until now:

- Joint model inference with Stan limited to linear description of the longitudinal process (R package rstanarm),
- No published work using Stan in nonlinear joint model or nonlinear mixed-effects model inference.

⇒ We aim to assess HMC for non-linear joint model parameters inference

\(^1\)Carpenter et al. (2017) Journal of statistical software
\(^2\)Neal (2011) Handbook of Markov Chain Monte Carlo
\(^3\)Hoffman & Gelman (2014) Journal of Machine Learning Research
Simulation Study

Simulation framework build on real data:

- Pattern of the simulated trial,
- Maximum Likelihood estimates for simulation values.

Evaluation Criteria:

- Relative Estimates Error on Posterior mode, mean and median,
- Coverage Rates.

\[ \text{To assess HMC algorithm for non-linear joint modeling population parameters inference} \]

Clinical Data Analysis

Cross-Validation method for link function selection

Posterior Analysis:

- Estimated posterior density of population parameters,
- Characteristics of the final posterior distribution (mean, median, maximum, standard deviation, credibility interval),
- Individual fits of tumor size and survival probability, with 95% credibility intervals.
Mechanistic model for tumor size kinetics

Longitudinal part

We rely on the **Sum of the Longest Diameters (SLD)** of the target lesions as a marker of the tumor size kinetics.

\[ SLD(t) = \begin{cases} 
BSLD \cdot e^{g \cdot t} & \text{if } t < tx \\
BSLD \cdot e^{g \cdot tx} \times (\phi e^{-d(t-tx)} + (1-\phi) e^{g(t-tx)}) & \text{if } t \geq tx
\end{cases} \]

\[ \Rightarrow TTG = \frac{\log \left( \frac{d\phi}{g(1-\phi)} \right)}{g+d} + tx \]

\( t \) : time since inclusion (days)
\( tx \) : time elapsed between inclusion and treatment onset

BSLD : SLD at inclusion time (mm)
\( d \) : tumor decreasing parameter (day\(^{-1}\))
\( g \) : tumor growth parameter (day\(^{-1}\))
\( \phi \) : proportion of cells that responds to treatment

Stein-Fojo model\(^1\)

\(^1\) Chatterjee et al. (2017) CPT Pharmacomet Syst Pharmacol
Building a simulation framework

Simulation of tumor size and survival data based on IMvigor210 Phase 2 clinical trial:

- \( y_{i,j} = \text{SLD}(t_{i,j}, \psi_i) \times (1 + e_{i,j}), \ e_{i,j} \sim \mathcal{N}(0, \sigma^2), \)
- \( h_i(t|\text{SLD}(t, \psi_i)) = \frac{1}{\lambda} \exp(\beta \times \text{SLD}(t, \psi_i)), \) exponential base hazard function.

<table>
<thead>
<tr>
<th>Fixed effects ( \mu )</th>
<th>Transformation</th>
<th>Standard deviation ( \omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSLD(mm)</td>
<td>60</td>
<td>log-normal</td>
</tr>
<tr>
<td>( d(\text{day}^{-1}) )</td>
<td>0.0055</td>
<td>log-normal</td>
</tr>
<tr>
<td>( g(\text{day}^{-1}) )</td>
<td>0.0015</td>
<td>log-normal</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.2</td>
<td>logit-normal</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.18</td>
<td>-</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>1450</td>
<td>-</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.01</td>
<td>-</td>
</tr>
</tbody>
</table>

100 datasets of 100 patients, measurements every 9 weeks for 2 years
Sensitivity analysis to prior distributions
Evaluation Criteria

Relative Estimates Error of a population parameter $\theta$ estimated on dataset $k$:

$$\text{REE}^k = \frac{\hat{\theta}^k - \theta^*}{\theta^*} \times 100.$$ (1)

Credibility intervals based on ordered posterior sample of size $L$ $\left(\hat{\theta}^k_{(l)}\right)_{l \in \{1, \ldots, L\}}$:

$$\hat{C}I_{\alpha}^k = \left[\hat{\theta}^k_{(L \times \alpha/2)}; \hat{\theta}^k_{(L \times (1 - \alpha/2))}\right]$$ (2)

Coverage rates:

$$\text{Coverage Rate}_\alpha = \frac{1}{K} \sum_{k=1}^{K} 1\{\theta^* \in \hat{C}I_{\alpha}^k\}$$ (3)
Relative Estimate Errors on point estimates
Coverage rates of 95% credibility intervals
**Clinical Data**

*Figure:* Spaghettis-plot of the tumor sizes, estimated overall survival probability by Kaplan-Meier and its 95% confidence interval on clinical data.
Cross-Validation for link function selection

Cross-Validation on patients using the posterior predictive density\(^1\):

\[
p(y_i^{(-m)} | D^m) = \int p(y_i^{(-m)} | \theta) p(\theta | D^m) d\theta
\]

- Monte-Carlo approximation on population parameters
  \[
p(y_i, T_i, \delta_i | D^{(-m)}) = \frac{1}{L} \sum_{l=1}^{L} p(y_i, T_i, \delta_i | \theta_l^{(-m)}),
\]
- Inference on random effects \(p(\eta_i | \theta_l^{(-m)}, y_i, T_i, \delta_i),\)
- Monte-Carlo approximation on random effects
  \[
p(y_i, T_i, \delta_i | \theta_l^{(-m)}) = \frac{1}{S} \sum_{s=1}^{S} \left[ \prod_{j=1}^{n_i} p(y_{ij} | \theta_l^{(-m)}, \eta_i^s) p(T_i, \delta_i | \theta_l^{(-m)}, \eta_i^s) \right].
\]

⇒ Selection of the link function which maximized score.

\(^1\)Vehtari & Lampinen (2002) Neural Computation
**CROSS-VALIDATION PROCEDURE RESULTS**

Joint Model for clinical data analysis:

- \( y_{i,j} = \text{SLD}(t_{i,j}, \psi_i) \times (1 + e_{i,j}), \ e_{i,j} \sim \mathcal{N}(0, \sigma^2), \)

- \( h_i(t|\text{SLD}(t, \psi_i)) = \frac{k}{\lambda} \left( \frac{t}{\lambda} \right)^{k-1} \exp(\beta \times f(\text{SLD}(t, \psi_i))). \)

Selection between the 4 following link functions:

- No link model \( f(\text{SLD}(t, \psi)) = 0, \)

- Current SLD value \( f(\text{SLD}(t, \psi)) = \text{SLD}(t, \psi), \)

- Current Slope of SLD \( f(\text{SLD}(t, \psi)) = \frac{\partial \text{SLD}(t, \psi)}{\partial t}, \)

- Time-to-growth, \( f(\text{SLD}(t, \psi)) = \text{TTG}(\psi) = \frac{\log(\frac{d\phi}{g(1-\phi)} + dx)}{g+d} + tx, \)

<table>
<thead>
<tr>
<th>CV Score</th>
<th>No Link</th>
<th>Current SLD</th>
<th>Models</th>
<th>Current Slope</th>
<th>Time-To-Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-23.44</td>
<td>-22.68</td>
<td></td>
<td>-22.23</td>
<td>-23.11</td>
</tr>
<tr>
<td>Link parameter</td>
<td>0</td>
<td>0.01 (0.001) mm(^{-1})</td>
<td>2.56 (0.70) day.mm(^{-1})</td>
<td>-0.009 (0.001) day(^{-1})</td>
<td></td>
</tr>
</tbody>
</table>
**Posterior density on real data**

**Figure:** Posterior density of current SLD slope model population parameters on clinical data depending on the prior information scenario.
## Posterior density characteristics on real data

<table>
<thead>
<tr>
<th></th>
<th>Maximum</th>
<th>Mean</th>
<th>Median</th>
<th>Sd</th>
<th>RSd(%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSLD (mm)</td>
<td>61.43</td>
<td>61.77</td>
<td>61.63</td>
<td>2.25</td>
<td>3.65</td>
<td>[57.34;66.29]</td>
</tr>
<tr>
<td>$d$ (day$^{-1}$)</td>
<td>0.0059</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0011</td>
<td>18.79</td>
<td>[0.0040;0.0084]</td>
</tr>
<tr>
<td>$g$ (day$^{-1}$)</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.00036</td>
<td>14.01</td>
<td>[0.0010;0.0021]</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.17</td>
<td>0.21</td>
<td>0.21</td>
<td>0.083</td>
<td>38.99</td>
<td>[0.074;0.39]</td>
</tr>
<tr>
<td><strong>Standard deviations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSLD (mm)</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.028</td>
<td>4.22</td>
<td>[0.60;0.72]</td>
</tr>
<tr>
<td>$d$ (day$^{-1}$)</td>
<td>1.09</td>
<td>1.06</td>
<td>1.05</td>
<td>0.15</td>
<td>14.34</td>
<td>[0.80;1.37]</td>
</tr>
<tr>
<td>$g$ (day$^{-1}$)</td>
<td>0.86</td>
<td>0.89</td>
<td>0.89</td>
<td>0.14</td>
<td>16.02</td>
<td>[0.60;1.21]</td>
</tr>
<tr>
<td>$\phi$</td>
<td>4.05</td>
<td>4.23</td>
<td>4.18</td>
<td>0.52</td>
<td>12.2</td>
<td>[3.36;5.35]</td>
</tr>
<tr>
<td><strong>Survival</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.19</td>
<td>1.14</td>
<td>1.14</td>
<td>0.12</td>
<td>10.7</td>
<td>[0.922;1.41]</td>
</tr>
<tr>
<td>$\lambda$ (day)</td>
<td>659</td>
<td>694</td>
<td>679</td>
<td>91</td>
<td>13.1</td>
<td>[549;915]</td>
</tr>
<tr>
<td>$\beta$ (day.mm$^{-1}$)</td>
<td>2.06</td>
<td>2.56</td>
<td>2.45</td>
<td>0.70</td>
<td>27.2</td>
<td>[1.47;4.24]</td>
</tr>
</tbody>
</table>

**Table:** Posterior density characteristics of current SLD slope model parameters with inference under the low prior information scenario
**Individual fits and 95% credibility intervals**

**Figure:** Individual fits and 95% credibility intervals of real data patients under the current SLD slope model with inference under the low prior information scenario on population parameters.

---

1. Kerioui et al. (2019) *preprint version*
A full Bayesian inference for non-linear joint model is now possible.

- Some remaining talking points:
  - Sensitivity to prior information,
  - Integration method for survival probability computation,
  - Further exploration for Bayesian model selection.
⇒ A full Bayesian inference for non-linear joint model is now possible.

- Some remaining talking points:
  - Sensitivity to prior information,
  - Integration method for survival probability computation,
  - Further exploration for Bayesian model selection.

- These results open the way to further work for a better understanding of the large variability between patients in the response to atezolizumab:
  - Impact of new lesions appearance on survival (recurrent events)\(^1\),
  - Modelling individual lesions and intra-patients variability in response to treatment,
  - Comparison with chemotherapy arm,
  - Prediction of the phase 3 outcome.

---
\(^1\) Krol et al (2018) Stat in Med
Acknowledgements

- IAME INSERM UMR 1137, Paris
- SPHERE INSERM UMR 1246, Tours
- René Bruno (gRED), Jin Jin (gRED), François Mercier (pRED), Ben Wu (gRED), Genentech/Roche Clinical Pharmacology Paris