A prediction model of heterogeneous treatment effects using randomized and observational data

Konstantina Chalkou, Salanti Georgia – University of Bern, Fabio Pellegrini – Biogen International GmbH, Suvitha Subramaniam, Benkert Pascal - University of Basel
Motivation - Effectiveness of drugs in Relapsing-Remitting Multiple Sclerosis (MS)

- Several drugs, compared in Network Meta-Analyses (NMA) #not personalized predictions - Tramacere I. et al., 2015
- We focus on Dimethyl Fumarate, Glatiramer Acetate, and Natalizumab
- Outcome: Relapse MS in 2 years (Yes/No)
- We want to find the drug that minimizes the risk of relapse, subject to patient characteristics: Heterogeneous Treatment Effects
Aim

To develop a three-stage evidence synthesis prediction model to predict the most likely outcome under several possible treatment options while accounting for patients’ characteristics using randomized clinical trials and observational data.
Data

RCTs

- 3 randomized clinical trials (phase III), 2990 observations in total
- **Disease:** Relapsing-remitting Multiple Sclerosis (MS)
- **Outcome:** Relapse MS in 2 years
**Data**

Observational data – Swiss MS Cohort (SMSC)

- **Inclusion criteria:** Patients with confirmed RRMS and at least two-year follow-up period from the baseline visit date

- **Patients:** 935 patients, each one with 1, 2, or 3 treatment cycles (i.e. repeated measures)

- **Observations:** 1752 follow-up cycles

© The HTx Consortium 2019-2023. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 825162.
**Prognostic model**

\[
h(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j \times PF_{ij}
\]

**Risk score**

**Prediction model using IPD Network meta-regression with PF and EM**

**Prediction model with IPD Network meta-regression using only the risk score**

---

**Treatments**

- **Dimethyl Fumarate**
  - Predicted Outcome A

- **Glatiramer acetate**
  - Predicted Outcome B

- **Natalizumab**
  - Predicted Outcome C

- **Placebo**
  - Predicted Outcome D
Treatments

- Dimethyl Fumarate
- Glatiramer acetate
- Natalizumab
- Placebo

**IPD from Observational Data SMSC**

**IPD from RCTs**

**HTE**

**Predicted Outcome A**

**Predicted Outcome B**

**Predicted Outcome C**

**Predicted Outcome D**

**Prognostic model**

\[ h(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j \times P_{F_{ij}} \]

**Risk score**

*Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset*

*Prediction model with IPD Network meta-regression using only the risk score*
**Treatments**

- **Dimethyl Fumarate**
- **Glatiramer acetate**
- **Natalizumab**
- **Placebo**

**IPD from Observational Data SMSC**

- **Dimethyl Fumarate** → Predicted Outcome A
- **Glatiramer acetate** → Predicted Outcome B
- **Natalizumab** → Predicted Outcome C
- **Placebo** → Predicted Outcome D

**Prognostic model**

\[ h(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j \times PF_{ij} \]

**Risk score**

- **#STAGE1**
  - Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

**IPD from RCTs**

**Prediction model with IPD Network meta-regression using only the risk score**

- **#STAGE2**
  - Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

- **#STAGE3**
  - Prediction model with IPD Network meta-regression using only the risk score
Stage 1: Development of the prognostic model

Selection of prognostic factors

- Age
- Sex
- EDSS
- Prior MS treatment
- Months since last relapse
- Disease Duration
- Nr. of prior relapses
- Number of Gd enhanced lesions

8 previously identified prognostic factors (at least 2 times included in pre-existing prognostic models)
Stage 1: Development of the prognostic model

Generalized linear mixed effects model – Bayesian framework

\[ Y_{ij} \sim \text{Bernoulli}(R_{ij}) \]

\[ \text{logit}(R_{ij}) = \beta_0 + u_{0i} + \sum_{k=1}^{P} (\beta_k + u_{ki}) \times PF_{kj} \]

**Notation**

- \( i \): individuals, where \( i = 1, 2, \ldots, N \)
- \( j \): time point, where \( j = 1, 2, 3 \)
- \( PF_{kj} \): kth prognostic factor at j\(^{th}\) time point, where \( k = 1, 2, \ldots, P \)
- \( \beta_0 \): fixed effect intercept
- \( u_{0i} \): random effect intercept
- \( \beta_k \): fixed effect slopes of k\(^{th}\) prognostic factor
- \( u_{ki} \): the individual-level random slopes of k\(^{th}\) prognostic factor
Stage 1: Development of the prognostic model

Shrinkage

Bayesian shrinkage methods use a prior on the regression coefficients

Laplace prior distributions for regression coefficients

$$\pi(\beta) = \prod_{k=1}^{p} \frac{\lambda}{2} e^{-\lambda|\beta_k|}, \ p: number\ of\ regression\ coefficients$$

- O’Hara et al., 2009

Small coefficients → towards zero faster

Large coefficients → smaller shrinkage

- Genkin et al., 2007
Stage 1: Development of the prognostic model

Missing data - Multilevel Joint Modelling Multiple Imputations

4 prognostic factors with missing data

Steps

1. Check for **auxiliary variables** – 1 variable was identified and used as auxiliary

2. Creation of 10 imputed datasets – Use of the same model (i.e. the substantive one) to impute the datasets – mitml R-package

\[
Y_{1ij} = \beta_0 + u_{oi} + \sum_{k=1}^{P} (\beta_k + u_{ki}) \times X_{k,j}
\]

\[
Y_{2ij} = \beta_0 + u_{oi} + \sum_{k=1}^{P} (\beta_k + u_{ki}) \times X_{k,j}
\]

\(Y_{1ij}\) and \(Y_{2ij}\) factors with missing values, \(X_{k,j}\) complete factors used in the substantive model & auxiliary variables, Use of random intercept \((u_{oi})\) and random slope \((u_{ki})\) as in the substantive model
Stage 1: Development of the prognostic model

Missing data - Multilevel Joint Modelling Multiple Imputations

4 prognostic factors with missing data

Steps

3. Application of the Bayesian model to all 10 imputed datasets

4. Pooled estimates via Rubin’s rules for m imputed datasets
### Stage 1: Development of the prognostic model

#### Estimated coefficients

<table>
<thead>
<tr>
<th>Prognostic factors</th>
<th>Estimations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.25</td>
</tr>
<tr>
<td>Age</td>
<td>-0.04</td>
</tr>
<tr>
<td>Disease Duration</td>
<td>0.36</td>
</tr>
<tr>
<td>Edss</td>
<td>0.12</td>
</tr>
<tr>
<td>Gd enhanced lesions</td>
<td>0.00</td>
</tr>
<tr>
<td>Number of previous Relapses (1 vs 0)</td>
<td>-0.08</td>
</tr>
<tr>
<td>Number of previous Relapses (more than 2 vs 0)</td>
<td>0.15</td>
</tr>
<tr>
<td>MonthsSinceRelapse</td>
<td>-0.45</td>
</tr>
<tr>
<td>Treatment Naive</td>
<td>0.15</td>
</tr>
<tr>
<td>Gender</td>
<td>0.28</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.04</td>
</tr>
</tbody>
</table>
IPD from Observational Data
SMSC

Treatments

- Dimethyl Fumarate
- Glatiramer acetate
- Natalizumab
- Placebo

Predicted Outcome

A

B

C

D

HTE

Risk score

Prognostic model

\[ h(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j \times PF_{ij} \]

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

Prediction model with IPD Network meta-regression using only the risk score

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset
Stage 2: Re-calibration and re-estimation of the risk model to RCTs

**Aim**

To update the model to improve predictions for new patients from the new setting (i.e. RCTs)

<table>
<thead>
<tr>
<th>Methods</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Update</td>
<td>0.57</td>
</tr>
<tr>
<td>Update only the intercept (Re-calibration)</td>
<td>0.50</td>
</tr>
<tr>
<td>Update intercept and coefficients (Re-calibration)</td>
<td>0.57</td>
</tr>
<tr>
<td>Model revision (Re-calibration &amp; selective re-estimation)</td>
<td>0.61</td>
</tr>
</tbody>
</table>
Risk of relapse in two years in RCTs

AUC = 0.61
Dimethyl Fumarate
Glatiramer acetate
Natalizumab
Placebo

**Treatments**

**Predicted Outcome**

- A
- B
- C
- D

**IPD from Observational Data SMSC**

**HTE**

**Risk score**

**Prognostic model**

\[ h(y_i) = \beta_0 + \sum_{j=1}^{n} \beta_j \times PF_{ij} \]

**IPD from RCTs**

**Prediction model with IPD Network meta-regression using only the risk score**

**#STAGE1**

**#STAGE2**

**#STAGE3**

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset
Stage 3: IPD Network Meta-regression

\[ Y_{ijk} \sim Bernoulli(p_{ijk}) \]

\[
\logit(p_{ijk}) = \begin{cases} 
  u_j + B \times (\logit R_{ij} - \logit R_j) & \text{if } k = b_j \\
  u_j + D_{bjk} + B \times (\logit R_{ij} - \logit R_j) + G_{bjk} \times (\logit R_{ij} - \logit R_j) & \text{if } k \neq b_j 
\end{cases}
\]

Notation

- \( i \): Individuals
- \( j \): study
- \( k \): treatment
- \( bj \): baseline treatment in study \( j \)
- \( B \): Individual level covariate regression term for Risk / the impact of Risk as prognostic factor
- \( D_{bjk} \): the treatment effect of treatment \( k \) versus placebo / fixed effect
- \( G_{bjk} \): The interaction of treatment and risk. Different for each treatment vs study’s control / the impact of Risk as effect modifier

Saramago et al., 2012
## Stage 3: IPD Network Meta-regression

### Results: Estimation of model parameters

OR for relapse for one unit increase in logit-risk in untreated patients (placebo) - \((\exp(B)) = 2.8\) (2.1, 3.9)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>OR for relapse versus placebo at the study mean risk ((\exp(D))) &amp; 95% Cr. Intervals</th>
<th>OR versus placebo for one unit of increase in the logit risk ((\exp(G))) &amp; 95% Cr. Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natalizumab</td>
<td>0.28 (0.21, 0.37)</td>
<td>0.62 (0.31, 1.15)</td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>0.52 (0.34, 0.78)</td>
<td>0.83 (0.32, 2.10)</td>
</tr>
<tr>
<td>Dimethyl Fumarate</td>
<td>0.43 (0.3, 0.57)</td>
<td>0.96 (0.50, 1.87)</td>
</tr>
</tbody>
</table>

\[
\text{logit}(p_{ijk}) = \begin{cases} 
  u_j + B \times (\text{logit}R_{ij} - \text{logit}R_j) & \text{if } k = b_j \\
  u_j + D_{bjk} + B \times (\text{logit}R_{ij} - \text{logit}R_j) + G_{bjk} \times (\text{logit}R_{ij} - \text{logit}R_j), & \text{if } k \neq b_j
\end{cases}
\]
Stage 3: IPD Network Meta-regression

Results: Estimation of model parameters

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean</th>
<th>Less than 25% Risk</th>
<th>More than 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natalizumab</td>
<td>51%</td>
<td>26%</td>
<td>76%</td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>65%</td>
<td>30%</td>
<td>92%</td>
</tr>
<tr>
<td>Dimethyl Fumarate</td>
<td>62%</td>
<td>23%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Best treatment

Dimethyl fumarate - 3% Absolute benefit compared to Natalizumab

Best treatment

Natalizumab - 17% Absolute benefit compared to Dimethyl Fumarate
Conclusions & further research

Conclusions

The risk score blinded to treatment modifies the absolute benefit of treatments

Further research

We plan to use measures relevant to clinical usefulness to validate the model
Thank you for your attention!

Questions?