Model-Based Meta-Analysis of The Effects of Biologics on Induction of Clinical Remission in Crohn’s Disease

Leon Bax, PhD (Certara)
Steven Martin, PhD (Pfizer)
Walter Reinisch (McMaster University)
The Presentation Content

• The idea in the next half hour is to familiarize you what can be done with model-based meta-analysis (MBMA) to inform decision-making in drug development for Crohn’s Disease

• I’ll focus on clinical remission in select biologics, but any selection of drugs or endpoints could be made

• I’ll cover some methodology, some strategy, and show a specific application with detailed numerical and graphical assessments

• I’ll end with a discussion of alternative and additional steps and how this compares to MBMA in other disease areas
Meta-Analysis in Drug Development

• MBMA is only one possible approach to combine and assess aggregate level data from clinical trials

• It is often a good idea to be flexible and use the approach that is most efficient for answering the relevant questions
The differences between PMA, NMA, and MBMA

- **Pairwise meta-analysis**
  - Assesses singular interventions by pairs – i.e. direct comparisons
  - Integration of meta-regression allows for covariate assessments
- **Network meta-analysis**
  - Assesses singular interventions in networks – i.e. direct and indirect comparisons
  - Often uses Bayesian approaches to accommodate integration in health policy decisions
- **Model-based meta-analysis**
  - Assesses variable dose interventions in networks – i.e. direct and indirect comparisons
  - Integrates pharmacological models for dose and time to accommodate integration in drug development decisions
Challenges and Opportunities in Crohn’s Disease Meta-Data

• Challenges
  • Induction and maintenance studies with different methods and time ranges
  • Relatively small datasets with variable results
  • Limited dose ranges and dose response data within drug classes
  • Multiple definitions of CDAI-based binary endpoints

• Opportunities
  • Common reference treatments (placebo)
  • Common and consistent reporting of CDAI endpoints
Use of MBMA to Answer Strategic Questions

• MBMA is typically applied to phase 2 and phase 3 clinical trial data, using biomarker, clinical outcomes, and patient-reported outcomes

• Meta-analysis is the only method that can answer a number of strategic development questions
  o What are the (dose-)response characteristics of competitor drugs for a new compound?
  o How do baseline characteristics or background treatments impact drug response?
  o What is the impact of trial design features (e.g. time, endpoints) on treatment effects?
  o What is the impact of region?
  o How do biomarker and clinical endpoint results compare in existing trials?
  o How do different doses of a new drug compare to existing and developing competitors?
  o Can we predict clinical trial results for different drug, endpoint, population, and design scenarios?
Pragmatic Modeling in MBMA

• MBMA modeling is different from primary data modeling
  o Typically, the number of data points is much smaller
    o E.g. a single trial contributes only one data point (regardless of the sample size)
    o It’s easy to wrongfully attribute between-study differences in outcomes to certain covariates
    o Automated covariate selection and elimination techniques and focusing on P values are a bad idea

• Since there are different questions to be answered, multiple models may be taken forward for simulations
  o Use the model that is most suitable for answering a certain question
    o E.g. If an ED50 of a drug is estimated with substantial uncertainty…
      o … and the goal is to build a reliable landscape of competitor benchmarks with approved doses, it may be better to use a model that simply uses a dose step
      o … and the goal is to estimate a dose response of a number of drugs within a drug class that are in development, it may be better to use the ED50 estimation and borrow information across drugs to be able to estimate effects for different doses
Illustration with Actual Data

• The illustration is based on an actual project but…
  o Proprietary data have been excluded
  o The original dataset has been replaced with a subset from an updated clinical trial database
  o Specific set of biologics have been included

• The search and selection of studies is always summarized in our protocols (including inclusion/exclusion criteria and up-to-date search filters for PubMed)
  o I will not go into detail of the search here, but we looked for clinical trials in PubMed and, in addition, we searched FDA publications, conference websites, and the NCT entries (clinicaltrials.gov)

• To illustrate results, I will use PowerPoint and an interactive report with Shiny apps
  o For completeness and redundancy, I have some screenshots in the slides that follow
  o The exploratory results are shown in PowerPoint and the MBMA results are created with R, Rmarkdown, and Shiny; this will show you why we favor the latter approach over PowerPoint
The Clinical Trial Outcomes Database

Choose Outcomes Database:
- Alzheimer’s Disease
- Chronic Kidney Disease
- Crohn’s Disease
- Diabetes Mellitus - Type 2
- Diabetes Mellitus - Type 2 (Evaluation)
- Dyslipidemia
- Dyslipidemia (Evaluation)
- Hepatitis B Virus
- Hepatitis C Virus
- Migraine

Choose Database Version:
- Version 1, Released on: 2013-11-24
- Version 2, Released on: 2014-08-25
- Version 1, Released on: 2013-07-20

Database Info:
The current version of the database contains information from randomized controlled trials on all current biologic and small molecule drugs that are evaluated for moderate-to-severe Crohn’s disease in patients that failed conventional treatment.

Crohn’s Disease Database Development Protocol
## Trial Flow for Selection of First Analysis Set

The analysis set used in explorations has clinical remission as well as clinical response. The subset with remission data only is created later.

<table>
<thead>
<tr>
<th>excluded</th>
<th>reason</th>
<th>remaining</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Start</td>
<td>74</td>
<td>trials in cd database</td>
</tr>
<tr>
<td>12</td>
<td>endpoint of interest</td>
<td>62</td>
<td>remission</td>
</tr>
<tr>
<td>4</td>
<td>indication of interest</td>
<td>58</td>
<td>cd in adults</td>
</tr>
<tr>
<td>14</td>
<td>study types of interest</td>
<td>44</td>
<td>induction</td>
</tr>
<tr>
<td>0</td>
<td>duration in weeks</td>
<td>44</td>
<td>time &lt;= 54</td>
</tr>
<tr>
<td>5</td>
<td>single (approved/common) routes</td>
<td>39</td>
<td>no ada iv - alica sc - certo iv - uste sc</td>
</tr>
<tr>
<td>3</td>
<td>randomized combination treatments</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>data considerations</td>
<td>35</td>
<td>unusual data patterns in sandborn 2001a</td>
</tr>
</tbody>
</table>
## Characteristics of Included Studies

<table>
<thead>
<tr>
<th>Summary column(s):</th>
<th>trials</th>
<th>arms</th>
<th>patients</th>
<th>publication,year</th>
<th>age</th>
<th>males</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>randomized,drug.class</td>
<td>5</td>
<td>7</td>
<td>1238</td>
<td>2004 (2001 to 2007)</td>
<td>37 (35 to 40)</td>
<td>43 (39 to 49)</td>
<td>68 (64 to 72)</td>
</tr>
<tr>
<td>alpha4-integrin</td>
<td>4</td>
<td>5</td>
<td>714</td>
<td>2012 (2008 to 2014)</td>
<td>37 (36 to 38)</td>
<td>45 (40 to 48)</td>
<td>68 (67 to 70)</td>
</tr>
<tr>
<td>cd22</td>
<td>1</td>
<td>3</td>
<td>323</td>
<td>2012 (2012 to 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corticosteroid</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>2009 (2009 to 2009)</td>
<td>35 (35 to 35)</td>
<td>52 (52 to 52)</td>
<td></td>
</tr>
<tr>
<td>gli2</td>
<td>1</td>
<td>3</td>
<td>75</td>
<td>2010 (2010 to 2010)</td>
<td>41 (39 to 42)</td>
<td>45 (36 to 54)</td>
<td>69 (68 to 69)</td>
</tr>
<tr>
<td>gm-csf</td>
<td>2</td>
<td>2</td>
<td>168</td>
<td>2007 (2005 to 2009)</td>
<td>37 (36 to 38)</td>
<td>49 (43 to 55)</td>
<td>74 (74 to 74)</td>
</tr>
<tr>
<td>icam-1</td>
<td>3</td>
<td>5</td>
<td>258</td>
<td>2002 (1998 to 2007)</td>
<td>39 (36 to 40)</td>
<td>42 (0 to 60)</td>
<td>68 (47 to 82)</td>
</tr>
<tr>
<td>il12/23</td>
<td>3</td>
<td>6</td>
<td>567</td>
<td>2011 (2008 to 2012)</td>
<td>40 (38 to 43)</td>
<td>41 (37 to 54)</td>
<td>75 (72 to 78)</td>
</tr>
<tr>
<td>immunosuppressants</td>
<td>3</td>
<td>6</td>
<td>291</td>
<td>2012 (2007 to 2014)</td>
<td>38 (37 to 42)</td>
<td>52 (41 to 69)</td>
<td>72 (70 to 74)</td>
</tr>
<tr>
<td>intestinal anti-inflammatory agent</td>
<td>1</td>
<td>3</td>
<td>252</td>
<td>2013 (2013 to 2013)</td>
<td>36 (35 to 37)</td>
<td>49 (45 to 51)</td>
<td></td>
</tr>
<tr>
<td>p38 map kinase</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>2002 (2002 to 2002)</td>
<td>36 (28 to 45)</td>
<td>17 (0 to 53)</td>
<td></td>
</tr>
<tr>
<td>placebo</td>
<td>32</td>
<td>32</td>
<td>3135</td>
<td>2008 (1997 to 2014)</td>
<td>37 (26 to 44)</td>
<td>46 (33 to 77)</td>
<td>73 (58 to 83)</td>
</tr>
<tr>
<td>tnf</td>
<td>11</td>
<td>17</td>
<td>1571</td>
<td>2004 (1997 to 2012)</td>
<td>37 (26 to 43)</td>
<td>46 (31 to 70)</td>
<td>72 (66 to 78)</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>35</td>
<td>92</td>
<td>8667</td>
<td>2008 (1997 to 2014)</td>
<td>37 (26 to 45)</td>
<td>46 (0 to 77)</td>
<td>72 (47 to 83)</td>
</tr>
</tbody>
</table>
Our selection resulted in only placebo-controlled trials. Natalizumab and certolizumab have most studies, and abatacept, etanercept, and oncept have the least amount of studies in the network.
There is some time course data so it seems feasible to create a longitudinal model.

Some studies are classified as induction studies but have an induction as well as maintenance phase; we will limit studied time further for formal MBMA.
High correlation between raw endpoint values (left plot) and even higher correlation between endpoint effect sizes (right plot) – this is without model-based corrections.
For most drugs, the raw data exploration does not indicate much of a dose response; this does not necessarily limit the utility of MBMA, i.e. many questions can still be answered.

Dose is already standardized by unit and in some cases also by frequency of administration; before formal MBMA this is checked and further standardized.
Some Special Data Preparations

• Presentation of vedolizumab data from the GEMINI-I study included a plot of effect over binned concentrations
  o Even though only one dose (300 mg) was studied, this allows us to create multiple dose bins
    o The calculation anchors the 300mg dose at the median exposure (24 ug/mL) and replaces the 300 mg arm with 4 arms of proportional doses of 95, 245, 361, and 1098 mg via midpoints of the concentration bins:
      o midpoints = \frac{15.2}{2}, \frac{(15.2+24)}{2}, \frac{(24.0+33.8)}{2}, \frac{(33.8+142.0)}{2}
      o proportional doses = \left(\frac{\text{midpoints}}{24}\right) \times 300 \text{mg}

<table>
<thead>
<tr>
<th>Bins</th>
<th>Dose</th>
<th>Response</th>
<th>Remission</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to &lt;15.2</td>
<td>95mg</td>
<td>20.4</td>
<td>6.1</td>
</tr>
<tr>
<td>15.2 to &lt;24</td>
<td>245mg</td>
<td>32.0</td>
<td>18.0</td>
</tr>
<tr>
<td>24 to &lt;33.8</td>
<td>361mg</td>
<td>36.0</td>
<td>14.0</td>
</tr>
<tr>
<td>33.8 to 142</td>
<td>1098mg</td>
<td>48</td>
<td>22.0</td>
</tr>
</tbody>
</table>
**MBMA Modeling of Clinical Remission**

\[ N_{\text{event,ij}} \sim \text{binomial}(P(\text{event})_{ij}, N_{ij}) \]

\[ f(\theta_{ki}) = \frac{E_{\text{max, class}} \cdot Dose_{ij}}{Dose_{ij} + ED_{50, \text{drug}}} \]

- Number of patients with a remission event at time (t) in treatment arm (j) of trial (i) follows a binomial distribution with probability \( P(\text{event}) \) and sample size \( N_{ij} \)
- The probability of event is a function of a placebo response at time t in that trial \((E_0, t_i)\) and a function of models for the treatment effect \( f(\text{drug, dose, time, } \theta, X) \) dependent on dose, time, fixed effect model parameters \( \theta \) and trial covariates \( X \)
- \( f(\theta_{ki}) \) is typically a general effect or **dose-response model** for specific drugs k
- The purpose of the modeling was to assess relative efficacy of various biologics
  - Potentially predict relative efficacy of compounds beyond their studied time course and doses
A Note on Non-Parametric Placebo Effects

- Absolute treatment effects (placebo + drug) are typically very variable
- This variability is often driven by the placebo response
  - Placebo effects can be very unpredictable
- In our experience, to minimize bias, an effect estimate for each trial based on a non-parametric placebo is preferred over a parametric structure; this is due to the unstructured nature of the variability in placebo
  - Different intercept (placebo or control effect) for each trial; we are often not interested in this part of the model
  - Using this non-parametric placebo we avoid misspecification of placebo effects due to parametric constraints that could result in biased estimations of drug effects
- Relative drug effects are commonly more stable than the placebo effects and can be modeled based on drug, dose, and covariates
  - Difference (ratio or absolute) from placebo or control; this is typically our primary interest
Exploratory Forest Plots – Placebo Only

Placebo responses are extremely variable

It would be challenging to model this if one would be inclined to do so for certain simulations (not needed to model it parametrically for MBMA)
Model Parameter Estimates Including Covariate Impacts

- Covariates and time course models were tested but not included in the final model
  - Covariates were not significant or did not improve fit (including prior TNF experience)

<table>
<thead>
<tr>
<th>Effect parameter</th>
<th>Estimate [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>emaba</td>
<td>-0.244 [-1.13 to 0.641]</td>
</tr>
<tr>
<td>emada.20</td>
<td>0.503 [-0.273 to 1.28]</td>
</tr>
<tr>
<td>emada.40</td>
<td>0.881 [0.152 to 1.61]</td>
</tr>
<tr>
<td>emada.80</td>
<td>1.39 [0.675 to 1.91]</td>
</tr>
<tr>
<td>emcer</td>
<td>0.349 [0.154 to 0.545]</td>
</tr>
<tr>
<td>emeta</td>
<td>-0.815 [-2.36 to 0.725]</td>
</tr>
<tr>
<td>eminf</td>
<td>1.73 [1.01 to 2.46]</td>
</tr>
<tr>
<td>emnat</td>
<td>0.487 [0.273 to 0.701]</td>
</tr>
<tr>
<td>ebone</td>
<td>-0.0246 [-0.868 to 0.819]</td>
</tr>
<tr>
<td>emust.1</td>
<td>0.213 [-0.584 to 1.01]</td>
</tr>
<tr>
<td>emust.3.6</td>
<td>0.653 [0.0333 to 1.27]</td>
</tr>
<tr>
<td>emvde</td>
<td>1.82 [0.112 to 3.52]</td>
</tr>
<tr>
<td>edved</td>
<td>341 [48.5 to 2390]</td>
</tr>
</tbody>
</table>

- Some estimates are negative, none of these are Emax parameters and the general effect estimates are consistent with observed data.
- Due to variability in observed values, many of these estimates are relatively imprecise.
- Vedolizumab could arguably be modeled with a dose step as well, but dose response was kept to illustrate the impact of doses higher than those studied.
Model Diagnostics – Time Course Plots with Fits

Reasonably good fits, without dose response and time course models
Model Diagnostics – Forest Plots with Fits

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>CONTROL TIME</th>
<th>TRAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>19</td>
<td>1.4 (9.5.7)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>2</td>
<td>7.1 (4.8.4)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>4</td>
<td>2.0 (4.3)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>6</td>
<td>2.6 (1.6.2)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>8</td>
<td>2.5 (1.5)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>10</td>
<td>1.0 (2.2)</td>
</tr>
<tr>
<td>raloxifene 5 mg (paw)</td>
<td>12</td>
<td>1.0 (2.1)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>6</td>
<td>1.7 (7.0.3)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>8</td>
<td>1.7 (7.4.9)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>10</td>
<td>3.5 (5.5.7)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>12</td>
<td>2.3 (9.3)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>14</td>
<td>3.2 (7.7)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>16</td>
<td>3.9 (5.3)</td>
</tr>
<tr>
<td>veledolozem 100 mg (paw)</td>
<td>18</td>
<td>4.6 (1.7)</td>
</tr>
</tbody>
</table>

This plot only shows the primary time points, so there are more fits than the ones shown here. This illustrates the disadvantage of forest plots: they become too long when assessing longitudinal data fit.
Model Diagnostics – Dose Response Plots with Fits

Dose response models or dose step models can be applied and result in similar fits; dose response parameters are more imprecise.
The odds ratios are not sensitive to the choice of the placebo effect; the risk differences and risk ratios, are so be careful with those.

We can easily develop a placebo model and use that for our simulations; here it is set at a constant of 17%.
Conclusions

- Clinical outcome databases can be used to provide up-to-date data sources for MBMA of interventions to induce remission in Crohn’s disease.
- The Crohn’s disease data are very variable, in particular placebo effects; a non-parametric placebo estimates can accommodate this variability and facilitate the estimation of drug effects without bias from misspecification of placebo models.
- The most parsimonious MBMA model fits the observed drug effects reasonably well, but there is a lot of uncertainty in parameter estimates:
  - Drug or drug class specific dose response models are not feasible in most cases.
  - Covariates or a time course within the induction period are not integrated.
  - It seems infliximab, natalizumab, ustekinumab, and vedolizumab do best.
Discussion
Discussion

• Simulations can integrate a placebo model so absolute treatment effects can be simulated rather than drug-to-drug odds ratios
  o The placebo effects do have a time course

• A joint endpoint model can increase data mass, but will not integrate extra studies
  o It may (but probably will not) decrease the width of drug effect confidence intervals
    o It can have a negative impact on estimate precision if the endpoint scaling factor is variable across trials

• The data and the modeling for ulcerative colitis is very similar, but there is less data and more variability in endpoint thresholds / definitions
Additional Slides
Joint Endpoint MBMA

• In a joint endpoint meta-analysis, multiple endpoints are modeled simultaneously:

\[ Y_{ijt.ep} = g\left( E_{o_it.ep} + f(Drug_{ij}, Dose_{ij}, X_{ij}, \theta_{(i)})_{ep_0} \cdot (1 + S_{ep>0}) \right) \]

• For each endpoint (ep), a separate treatment effect \( Y_{ijt.ep} \) is based on a non-parametric placebo effect for that endpoint \( E_{o_it.ep} \) and a drug effect function \( f \) for a reference endpoint \( ep_0 \) that is scaled by multiplicative scaling factors \( S_{ep>0} \) for endpoints other than the reference endpoint (if the reference endpoint \( EP_0 \) is modeled, the scaling parameter defaults to 0).

  • For binary endpoints \( Y_{ijt.ep} \) is the probability of an event and is modeled with a logit transform (g); for continuous endpoints g is 1 and there is no transformation

• If the S parameter estimate is close to zero or not significant, this indicates that the drug effects for the modeled endpoints are very similar

• For example, response 100 can be the reference endpoint \( Ep_0 \) and a scaling parameter \( S_{remission} \) can be estimated to associate response with remission
Placebo Model and Joint Endpoint Example

- A joint model can potentially model remission and response together and use the correlation between these endpoints
- A placebo model will have a time course and may depend on prior treatment with TNFs
  - Multiple placebo models and drug effect models can be used to answer different questions or assess sensitivity of the results to model assumptions