ACoP8  October 18, 2017

Neighbor’s Envy Owner’s Pride – Comparator Analysis for Drug Development and Market Access

Post-Approval Decision Making Supported by Modeling and Simulation Based on a Variety of Data Sources

Marc R. Gastonguay, Ph.D. FISoP
Statement of Problem and Context

- Variety of Data Sources

- Post-Approval Decision Making
  - Indirect Comparative Effectiveness
  - Probability of Success in Real World Evidence Trial

- Utility of M&S Given Different Data Sources
  - Limitations
  - Opportunities
A Variety of Data Sources

- Individual-Level Clinical Trial Data
- Individual-Level Patient Registry Data
- Individual-Level Electronic Medical Records
- Summary-Level Literature Meta-Data
A novel model-based meta-analysis to indirectly estimate the comparative efficacy of two medications: an example using DPP-4 inhibitors, sitagliptin and linagliptin, in treatment of type 2 diabetes mellitus

Jorge Luiz Gross,¹ James Rogers,² Daniel Polhamus,² William Gillespie,² Christian Friedrich,³ Yan Gong,⁴ Brigitta Ursula Monz,⁴ Sanjay Patel,⁵ Alexander Staab,³ Silke Retlich³

Trial Summary Data: HbA1c Change from Baseline

A. Linagliptin studies

B. Sitagliptin studies

Probability Distribution for Expected Response

Figure 4  (A) Estimated drug effects on glycaated haemoglobin (HbA1c) for reference population, with no pretreatment washout, over 24 weeks (difference from placebo). (B) Estimated drug effects on HbA1c for reference population, with 4-week washout plus 2-week placebo run-in period, over 24 weeks (difference from placebo). Reference population of 1000 participants, baseline HbA1c 8%, racial composition: 61.5% White, 1.5% Black, 37% Asian.

Figure 5  Posterior distribution for the difference in effect estimates between linagliptin (5 mg) and sitagliptin (100 mg) at 24 weeks. Reference population of 1000 participants (therefore involving $10^5$ simulated patients), baseline glycaated haemoglobin (HbA1c): 8%, racial composition: 61.5% White, 1.5% Black, 37% Asian.

Toujeo Real World Evidence Trial Simulation: Data Sources

**Toujeo Phase III trial data**
(4 studies, ~26 Gb of SAS data)
- 4,681 patients over 4 trials
- 126,548 relevant lab records \( \rightarrow \) 29,012 HgbA1c labs

**MGH EMR Data**
- 314,292 patients \( \rightarrow \) ~65,000 T2DM patients
- 4,917 patients with ≥2 long acting insulin outpatient prescriptions that are at least 30D apart (life)

**GE Health EMR from Quintiles**
(> 300 GB, ~2 billion records of structured data & metadata)
- ~3,000,000 patients
- ~400,000 HgbA1c labs

Slide courtesy of Jeffrey Barrett, ACoP 2017
Gastonguay ACoP2017 Covariate Effects for Labeling
The Big Picture = Systems and Data

Big Data

- MGH Virtual patients from EMR Data – 60,000 Diabetics patients
- Toujeo and Lantus Clinical Trial Data
- Quintiles Electronic Medical Record data
- 3.6 million patients

Sanofi Bridgewater Lab

Sanofi

- Data as input to models
- Analysis-ready datasets

SABRE HPC platform

Clinical Trial Simulation

Data transfer
Activity / Treatment

Metrum – creation of models
Execution of simulators
Support and Training
Sanofi Internal users

Slide courtesy of Jeffrey Barrett, ACoP 2017
Gastonguay ACoP2017 Covariate Effects for Labeling
Pros/cons of aggregate data (AD) MA

- **Pros**
  - Relatively easy access to data from public sources

- **Cons**
  - Not well-suited for inferences about patient-level covariates.
    - Ecological bias/fallacy
    - Aggregate covariate data describes a narrower range of values than individual covariate data
  - For nonlinear models the relationship between the dependent variable and the covariates, e.g., dose or time, is not described by the same function for AD and IPD.
  - Usually no info about correlations among multiple outcomes
  - Model usually not suitable for prediction/simulation of individual outcomes
Alzheimer's Disease Progression Model

Sub-populations
- Normal (N=200)
- MCI (N=400)
- Mild AD (N=200)

Literature Meta-Data
- 73 Trials (1990 to Present)
- Interstudy variability
- Estimate of drug treatment effects (magnitude, onset, offset)

http://www.adni-info.org/

http://www.c-path.org/CAMD.cfm
Tradeoffs Between Summary-level Analysis and Patient-level Analysis
Combining patient-level and summary-level data for Alzheimer’s disease modeling and simulation: a beta regression meta-analysis

James A. Rogers · Daniel Polhamus · William R. Gillespie · Kaori Ito · Klaus Romero · Ruolun Qiu · Diane Stephenson · Marc R. Gastonguay · Brian Corrigan

Objective:
Develop a model to describe the longitudinal progression of ADAS-cog in Alzheimer’s disease patients in both natural history and randomized clinical trial settings, utilizing both IPD and AD.
Hierarchical expectation propagation for Bayesian aggregation of average data

Sebastian Weber†  Andrew Gelman‡  Bob Carpenter †  Daniel Lee†
Michael Betancourt§  Aki Vehtari¶  Amy Racine†
26 Oct 2015

https://arxiv.org/abs/1602.02055

- Details methodology for joint analysis of IPD and AD from one study each [17, 18].
- Readily generalized to multiple IPD and AD studies.
- The AD data likelihood is imputed by simulation.
- That is embedded within an overall Bayesian analysis method involving:
  - Analysis of IPD by HMC (Stan),
  - Analysis of AD data by importance sampling, and
  - Iterative updating of both IPD and AD analyses by expectation propagation.
Why Use Meta-Data at All?

- Comprehensive view of current state of knowledge

- May be only source of estimates for competitor drug effects

- Inferences may be limited to simple treatment mean or SD comparisons

- Combine with individual-level data from other sources
Another Strategy

Build Models Sequentially by Data Source

- Model-Based Meta Analysis for Comparator Mean Effect
- Bayesian Data Analysis
  - Informative Prior Distributions for Comparator Mean Effects based on MBMA
  - Individual-Level Data for Disease Progression and Population Variability
  - Individual-Level Data for Proprietary Asset
- Perform Simulation from Bayesian Posterior Distributions
## Utility of Different Data Types

<table>
<thead>
<tr>
<th>Model-Based Inference or Application</th>
<th>A. Individual-Level Clinical Trial Data</th>
<th>B. Individual-Level Patient Registry Data</th>
<th>C. Individual-Level Electronic Medical Records</th>
<th>D. Summary-Level Clinical Trial Meta Data (e.g. Mean &amp; SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Mean Comparison</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Treatment SD Comparison</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Individual Covariate Effect Estimation</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Covariate Distributions</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Sample Size Calculation</td>
<td>X</td>
<td>in combination with A or D</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Trial Simulation with Individual Inferences or multiple endpoints</td>
<td>X</td>
<td>in combination with A</td>
<td>in combination with A</td>
<td>in combination with A</td>
</tr>
</tbody>
</table>