Model-Based Meta-Analysis for Integration of Data from Multiple Sources

Bill Gillespie, Ph.D.
Lead Scientist and VP
Strategic Consulting Services
Modeling and simulation of clinical outcomes via integration of heterogeneous data

• Objective
  – Provide an introduction to modeling and simulation of clinical outcomes via meta-analysis of data from heterogeneous sources, e.g., individual patient data, summary data on clinical outcomes from public sources and pre-clinical data.

• Outline
  – Rationale
  – Data collection and modeling process
  – 2 case studies illustrating the use and value of model-based meta-analysis in clinical drug development
  – Brief discussion of issues relevant to model-based meta-analysis
Rationale for model-based meta-analysis

• **Why build models via integration of heterogeneous data?**
  – Leverage prior knowledge from all available sources
    • Data on the NCE of interest
      – Preclinical, Phase I safety & biomarkers, clinical safety & efficacy …
    • Knowledge about the target disease & affected physiologic systems
    • Knowledge/data on related compounds
      – From proprietary or public sources
      – Analogs or competitors
  – **Comparison of competing drugs/treatments**
    • Borrow information
    • Shared model parameters are more precisely estimated
    • Comparative inferences are also more precise
  – **Prediction of unobserved clinical outcomes**
    • Build models for preclinical-to-clinical or biomarker-to-clinical relationships
    • Use those models to predict clinical outcomes for NCE’s for which no clinical data is available
Taxonomy of data and data sources

- **Data information level**
  - **Summary data**
    - Summary statistics such as treatment mean, sample size, fraction of patients that experience an event, etc.
  - **Individual patient data**
  - **Endpoint data**
    - Usually LOCF or OC
  - **Longitudinal data**
    - Data observed at multiple times within each patient
    - May be available as either summary or individual patient data
    - Data at early time points may be viewed as biomarkers for later outcomes
    - May allow modeling strategies to adjust for biases in LOCF or OC data
Taxonomy of data and data sources

- **Source of data**
  - Proprietary
    - Individual patient data from clinical trials
    - Preclinical data
  - Public domain
    - Journals
    - Regulatory documents, e.g., FDA “SBA’s”
    - Published abstracts
    - Posters, slides or other documents presented to the public at meetings
    - Web documents, e.g., press releases about new clinical trial results
    - Clinical trial registries
      - [http://www.clinicalstudyresults.org/](http://www.clinicalstudyresults.org/)
      - [http://ctr.gsk.co.uk/welcome.asp](http://ctr.gsk.co.uk/welcome.asp)
Taxonomy of data and data sources

- **Type of data**
  - Preclinical data
    - In vitro
    - In vivo
    - Pharmacokinetics
  - Clinical data
    - Pharmacokinetics
    - Biomarkers
    - Clinical efficacy/safety data in target patient population
The process

- **Establish a collaborative relationship with the development team**
  - What decisions are they facing, e.g.,
    - Design of next trial
    - What treatment regimen(s) (dose, frequency, route, etc.)
    - Go/no-go
    - What indication(s)
  - Identify their key issues, dilemmas and uncertainties
  - Identify where modeling and simulation can enhance their decision-making
    - By more effectively leveraging prior information
    - By making assumptions explicit and quantitative

- **Plan your modeling strategy**
  - Develop a modeling work plan, e.g., objectives, data required, methodology,…
  - Tailor the plan to the specific objectives, e.g., don’t try to develop models to address all potential future objectives
  - Prioritize objectives to maximize value to the development program
The process

- **Identify key information you want to collect**
  - Clinical endpoints
    - Type of data needed: summary data, patient data, …? 
    - For summary data, what summary stats to collect?
  - Potential preclinical or clinical predictors (biomarkers)
  - Drugs
  - Covariates of potential interest or explanatory value
  - Relevant mechanistic knowledge
  - Prior modeling work relevant to the project

- **Construct a database**
  - Establish a consistent process including quality control measures
  - Finding the sources
    - Public sources
      - Literature search
      - Web search
      - FDA web site
      - Clinical trial registries
    - Proprietary sources
      - Preclinical and clinical data on NCE of interest
      - Preclinical and clinical data on previous NCE’s studied for same indication
The process

• **Construct a database (cont.)**
  – **Compiling the public source database**
    • Review a subset of the data sources to assess
      – Range of data presentation, e.g., mean, mean change from baseline, percent change from baseline, etc.
      – Availability of data and covariates
    • Adjust modeling and data gathering plan as necessary
      – Is there enough data to support modeling?
      – Are other related quantities reported?
  • Plan the structure of the database
    – Platform: ASCII, spreadsheet, relational database, etc.
    – Format: nomenclature, multiple data columns vs. one data column with column containing data label, missing data handling, etc.
  • Enter the data
  • QC

• **Execute the modeling work plan**
Meta-analysis of mean LDL-C % change from baseline

- 25 clinical trials
  - 13 trials of statin monotherapy
  - 4 trials of ezetimibe monotherapy
  - 4 trials of statin/ezetimibe combinations
  - 3 trials of gemcabene monotherapy
  - 1 trial of statin/gemcabene combinations

The original work also included modeling of CRP, headache, ALT elevation, myalgia, HDL-C and CHD relative risk

Key question: Can gemcabene/statin compete favorably with ezetimibe/statin?
Statin dose-response: monotherapy and combination with ezetimibe 10 mg/d
Incremental effect of gemcabene decreases with increasing statin doses
Statin/gemcabene is inferior to statin/ezetimibe at higher statin doses

Table 3. Predicted Additional Mean Change in LDL-C Reduction Between Combination Therapy of Atrovastatin With 900 mg of Gemcabene or 10 mg of Ezetimibe and Monotherapy With Atorvastatin

<table>
<thead>
<tr>
<th>Atorvastatin (mg)</th>
<th>900 mg Gemcabene</th>
<th>10 mg Ezetimibe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>5%</td>
</tr>
<tr>
<td>0 mg</td>
<td>-31.9</td>
<td>-35.4</td>
</tr>
<tr>
<td>10 mg</td>
<td>-12.0</td>
<td>-13.9</td>
</tr>
<tr>
<td>20 mg</td>
<td>-8.7</td>
<td>-10.7</td>
</tr>
<tr>
<td>40 mg</td>
<td>-5.5</td>
<td>-7.8</td>
</tr>
<tr>
<td>80 mg</td>
<td>-2.5</td>
<td>-5.3</td>
</tr>
</tbody>
</table>

*The expectation and 90% probability interval are shown.*
Conclusion: Terminate the development program because gemcabene cannot compete favorably with ezetimibe as combination therapy with statins.
Trial simulation to design a Phase II dose finding strategy

• Phase II objective: Efficiently find a dose of drug X that:
  – Is at least non-inferior to the standard of care (drug R 10 mg/d) with respect to both efficacy and safety.
  – With sufficient certainty that we can risk a Phase III program with only one dose level.

• Efficacy
  – Decrease in fraction of patients with a disease-related event

• Probable dose-limiting AE
  – Same biological mechanism as efficacy
Trial simulation to design a Phase II dose finding strategy

- **Available information:**
  - **Drug X:**
    - Clinical pharmacokinetics from Phase I.
    - Pre-clinical response thought to be predictive of clinical outcomes related to mechanism of action (both efficacy and dose-limiting AE)
      - Data for drug X and competitors
  - **Public-source data:**
    - 5 marketed drugs, 27 clinical trials, 77 treatment arms
    - Disease-related events
      - Number of patients with events & total number of patients for each treatment arm
    - Dose-limiting AE’s
      - Number of patients with events & total number of patients for each treatment arm
    - Pharmacokinetics
      - Mean clearance
    - Two categories of patients known to have different risks for both disease events and AE’s (designated group 1 and group 2)
Modeling strategy: Simultaneously model pre-clinical response, and frequency of clinical efficacy and AE events

Model concentration-response for pre-clinical response

Model relationships between pre-clinical & clinical potencies

Optimize trial design and analysis via clinical trial simulations

Predict clinical dose-response of drug X

15 February 2006
Inhibition of preclinical response

![Graph showing inhibition of preclinical response for different drugs.](image-url)
Prevention of disease event: Dose-response by drug

![Graph showing dose-response by drug for different groups.](image-url)
Major AE: Dose-response by drug

![Graph showing dose-response by drug for different groups.](image-url)
Preclinical-to-clinical outcome: Model relationships between pre-clinical & clinical potencies

Drug A
Drug B
Drug C
Drug D
Drug X (median & 90% prediction interval)

log(EC₅₀) for preclinical response

log(EC) for disease event prevention

Drug X (median & 90% prediction interval)
Modeling of disease events and AE’s lead to consistent but highly uncertain predictions of the drug X dose equivalent to drug R 10 mg/d
Clinical design and analysis options considered

• Number of patients per treatment arm (210, 630 or 1050)
• Number (5 or 7) and spacing (linear- or log-spaced) of doses
• Trial analysis
  – Dose-response modeling using conventional logistic regression
  – Bayesian modeling using prior information, e.g., the dose-response model described in the previous slides
  – In either case the dose-response model is used to select a dose with efficacy equivalent to drug R 10 mg/d
• Adaptive pruning of treatment arms for lack of efficacy or excess AE’s
  – Based on frequentist confidence intervals & observed fraction of events for drug R 10 mg/d
  – Based on posterior probabilities from Bayesian modeling
• Trial performance is assessed by whether the selected dose is non-inferior to drug R 10 mg/d (under the simulated “truth”)
  – Fraction of simulated trials where Pr(disease event) and Pr(AE) at the selected dose are less than 1.25 times that for drug R 10 mg/d
Trial simulation results: Effects of number and spacing of doses and sample size

Results using logistic regression without prior information

- Best performance is seen with 7 log-spaced doses
- Only trials with \( \geq 1050 \) patients offer sufficient certainty to consider risking a Phase III program with only one dose level.
Trial simulation results: Impact of using prior information

Bayesian modeling using prior information:

- Increases the probability of selecting a non-inferior dose, particularly in trials with smaller sample sizes
- Also improves the probability of making correct pruning decisions

Bottom line: Simulations led to more efficient designs that can find a non-inferior dose with high probability
Summary of key points from the case study

• Integration of pre-clinical & public-source clinical data permits construction of a model for predicting clinical outcomes for the NCE.

• Leveraging prior information permits more efficient design and analysis of a Phase II trial to select a dose for Phase III:
  – Optimizes range, number and spacing of doses
  – Adaptive pruning assigns patients to most relevant doses
  – Enhances characterization of dose-response & therefore dose selection

  ⇒ Shorter, more informative Phase II program* (reduction by ~400-600 patients or ~4-6 months)

  ⇒ Shorter time to market (at least 4 months and potentially much greater by avoiding incorrect dose selection and a failed Phase III program)

  ⇒ More $$ and better patient care

* Compared to 1050 patient trial using 5 drug X doses and analysis without prior information
Some additional issues relevant to model-based meta-analysis

• When are models based on summary data sufficient?
• Applying individual models to summary data
• Publication bias

• See “additional slides” section for further discussion of:
  – Summary statistics when there are dropouts: LOCF, OC, …
  – Covariates: Interpretation with models based on summary data
  – Combining summary statistics and individual patient data
  – Integrating pre-clinical or biomarker data with clinical data
When are models based on summary data sufficient?

- **What type of model & data is required depends on the planned use of the model:**
  - Population simulations for dose selection or for product profiling (comparing NCE to competitors)
    - Generally a model describing a population statistic (e.g., population mean) as a function of dose is used
    - Then modeling based on summary statistics is sufficient
    - However this limits the ability to describe the influence of patient characteristics because model covariates are estimated based on summary statistics of those covariates
  - Clinical trial simulations to optimize trial design
    - Generally requires a model capable of simulating individual patient responses
    - Usually need to model individual patient data (possibly combined with summary data)
Applying individual models to summary data

- Our usual PK & PD models are strictly relevant only for describing responses in individual organisms---not for summary stats for groups.

- For example, consider a drug with PK following an iv bolus best described by a one compartment model
  - Each patient’s data is described by monoexponential function
  - But the mean concentration time-course of 2 patients’ data is described by a biexponential (unless their elimination rate constants are equal)

- Mechanistic interpretations of model structure and parameter values should be approached cautiously
Publication bias

• Negative results, e.g., results of trials that fail to show a drug is superior to placebo, are often not published or otherwise made available to the public.

• Similarly the results of trials for drugs that failed for some reason (not progressed in development or not approved) are often not available.

• As a result model predictions based on publicly available data may be over-optimistic.

• Methods for detecting and adjusting for publication bias have been described in the statistical literature about meta-analysis.
  – To my knowledge such methods have not been applied to PK/PD model-based meta-analysis
Conclusion: Value of model-based meta-analysis

- Greater precision of model-based predictions and inferences
- Quantitative comparison of competing treatments without direct within-study comparisons
- Prediction of clinical outcomes for NCE’s for which no clinical data is available
- Leverage the above capabilities to enhance strategic decision-making in clinical drug development
  - Optimize clinical trial designs via clinical trial simulations
  - Optimize treatment regimens
  - Go/no-go
  - …
Recommended reading

Submit Questions by clicking on the “Q&A” feature to the left of your screen
Summary statistics when there are dropouts

- Summary statistics such as treatment means are typically summaries of:
  - LOCF (last observation carried forward) data for all “intent to treat” patients or
  - OC (observed cases) data

- Both types of statistics are potentially biased estimates of “as treated” outcomes
  - If assigned treatment affects dropout behavior
  - If the measured quantity systematically changes with time

- LOCF is most commonly reported and may be the only value available
Covariates: Interpretation with models based on summary data

• With summary data most covariates are also in the form of summary statistics, e.g., means of efficacy metrics at baseline, means of demographic quantities, numbers of patients in different categories

• Model predictions of a population statistic for a given covariate value should be interpreted as the predicted value for a population where the specified covariate value is a mean for a population whose individual patient values are distributed similarly to those in the trials used for model development.

• They do not represent predictions for
  – An individual patient with that covariate value
  – A population that is substantially more or less homogeneous than those in the previous trials
Combining summary statistics and individual patient data

• **Method 1:** Trial arms are treated like “super-patients” for trials providing only summary data
  - Same structural model as individuals
  - Residual variances are divided by sample size
  - Summary and individual data may be analyzed simultaneously or sequentially
    - Sequential analysis should use Bayesian approach where the results of the summary data analysis are used as a prior distribution for the individual data analysis

• **Method 2:** Individual patient results are treated as missing data for trials providing only summary data
  - Implemented with simulation-based methods, e.g., MCMC
  - Sampling distribution of summary statistic is estimated via simulation of missing individual data
  - Very computation-intensive
  - But more theoretically rigorous/consistent approach

• For either method both inter-trial and inter-patient variation may be modeled (but implementing both levels of random effects is difficult and klugy with NONMEM)
Integrating pre-clinical or biomarker data with clinical data

- **Objective:** Develop a model describing the relationship between a pre-clinical measurement or a clinical biomarker (let's call them “predictors”) and a clinical outcome.
- When feasible this provides a quantitative strategy for selecting doses and designing early clinical trials before the drug has been administered to patients.
- The specific form of such a model depends on the context so it is difficult to generalize about the specific functional form. Specific considerations include:
  - Mechanistic knowledge about the relationship between:
    - Drug exposure and the predictor
    - Predictor and clinical response
  - Is the predictor response a measure of the drugs’ MOA or a more generalizable predictor of the drugs’ effect on the disease?
  - Are predictor measurements available for comparators for which you have clinical data, and is it plausible that they are predictors of response to those drugs?
- **Such issues influence the feasibility of constructing a suitable model and the extent to which the form of the model is mechanistic or empirical.**
Integrating pre-clinical or biomarker data with clinical data

- **Uncertainty in the predictor-to-clinical model:**
  - Is often large and should be quantified to the extent possible
  - Is critical for making valid model-based inferences
    - Predicting clinical response
    - Selecting doses
    - Designing clinical trials
  - Is usually too large to permit skipping a dose finding trial
  - But it provides clear guidance on the range of doses to explore
  - And provides an excellent tool for simulating proof of concept and dose finding trials to assess how well different trial designs and analysis methods perform over the entire range of uncertainty
Implementing the modeling strategy

- **Hierarchical model**
  - Binomial models for numbers of disease and adverse events
  - Normal model for preclinical responses
  - Inter-trial variation in clinical response
  - Inter-drug variation in the preclinical-to-clinical model to account for model misspecification

- **Bayesian data analysis**
  - Implemented with WinBUGS
    - Easy to implement complicated probabilistic structure
    - Rigorous approach to quantifying uncertainty in model parameters and predictions
  - Relatively non-informative priors
Integrated preclinical-clinical model: disease event, AE and preclinical response

• Preclinical response sub-model

\[ PCR = E_0^{PCR} (1 - E_D^{PCR}) \]

\[ E_D^{PCR} = \frac{c^{\gamma_{PCR}}}{(EC_{50}^{PCR})^{\gamma_{PCR}} + c^{\gamma_{PCR}}} \]

\[ n_{DE} \sim \text{binomial}(p_{DE}, n) \]

\[ \text{logit}(p_{DE}) = E_P^{DE} + I_{\text{group2}} \theta_{\text{group2}}^{DE} - E_D^{DE} \]

\[ E_D^{DE} = \frac{c^{\gamma_{DE}}}{(EC^{DE})^{\gamma_{DE}}} \]

\[ \mu_{EC^{DE}} = \theta_{PCR \rightarrow DE} EC_{50}^{PCR} \]

• Disease event sub-model

\[ \log(EC^{DE}) \sim N(\log(\mu_{EC^{DE}}), \sigma_{EC^{DE}}^2) \]

• AE sub-model

\[ n_{AE} \sim \text{binomial}(p_{AE}, n) \]

\[ \text{logit}(p_{AE}) = E_P^{AE} + I_{\text{group2}} \theta_{\text{group2}}^{AE} + E_D^{AE} \]

\[ E_D^{AE} = \frac{c^{\gamma_{AE}}}{(EC^{AE})^{\gamma_{AE}}} \]

\[ \mu_{EC^{AE}} = \theta_{PCR \rightarrow AE} EC_{50}^{PCR} \]
Integrated preclinical-clinical model:
The gory details

- Preclinical response sub-model
  - $i^{th}$ preclinical response measurement resulting from exposure to drug$ _j$ in the $j^{th}$ experiment

$$PCR_{ij} \sim N\left(\mu_{PCR,j} \left(c_{ij}, drug_j\right), \sigma^2_{PCR}\right)$$

$$\mu_{PCR,j} \left(c_{ij}, drug_j\right) = E_{0,j}^{PCR} \left(1 - E_{D,j}^{PCR} \left(c_{ij}, drug_j\right)\right)$$

$$E_{D,j}^{PCR} \left(c_{ij}, drug_j\right) = \frac{c_{ij}^{\gamma_{PCR}}}{\left(EC_{50}^{PCR} \left(drug_j\right)\right)^{\gamma_{PCR}} + c_{ij}^{\gamma_{PCR}}}$$

$$E_{0,j}^{PCR} \sim N\left(\mu_E^{PCR}, \sigma^2_{E_{0,PCR}}\right)$$

Priors

$$\mu_{E_0^{PCR}} \sim N\left(5000, 10^{10}\right) \quad \frac{1}{\sigma^2_{E_{0,PCR}}} \sim \text{gamma}(0.00001, 0.001)$$

$$\log\left(EC_{50}^{PCR} \left(drug\right)\right) \sim N\left(0, 10^5\right) \quad \gamma_{PCR} \sim \text{uniform}(0.1, 10)$$

$$\frac{1}{\sigma^2_{PCR}} \sim \text{gamma}(0.0001, 0.001)$$
**Integrated preclinical-clinical model:**

The gory details

- **Disease event/AE sub-model**
  
  Number of disease events and number of AE’s resulting from exposure to drug in the kth treatment arm in the mth trial

\[
\begin{align*}
    n_{DE,km} & \sim \text{binomial} \left( p_{DE,km} (\bar{c}_{km}, drug_{km}), n_{km} \right) \\
    n_{AE,km} & \sim \text{binomial} \left( p_{AE,km} (\bar{c}_{km}, drug_{km}), n_{km} \right)
\end{align*}
\]

\[
\begin{align*}
    \text{logit} \left( p_{DE,km} (\bar{c}_{km}, drug_{km}) \right) & = E_{P,m}^{DE} + I_{\text{group2},km} \theta_{\text{group2}}^{DE} - E_{D,km}^{DE} (\bar{c}_{km}, drug_{km}) \\
    \text{logit} \left( p_{AE,km} (\bar{c}_{km}, drug_{km}) \right) & = E_{P,m}^{AE} + I_{\text{group2},km} \theta_{\text{group2}}^{AE} + E_{D,km}^{AE} (\bar{c}_{km}, drug_{km})
\end{align*}
\]

\[
\begin{align*}
    E_{D,km}^{DE} (\bar{c}_{km}, drug) & = \frac{\bar{c}_{km}^{\gamma_{DE}}}{\left( EC_{DE}^{DE} (drug_{km}) \right)^{\gamma_{DE}}} \\
    E_{D,km}^{AE} (\bar{c}_{km}, drug) & = \frac{\bar{c}_{km}^{\gamma_{AE}}}{\left( EC_{AE}^{AE} (drug_{km}) \right)^{\gamma_{AE}}}
\end{align*}
\]

\[
\begin{align*}
    (E_{P,m}^{DE}, E_{P,m}^{AE}) & \sim N \left( \left( \mu_{E_{P,m}^{DE}}, \mu_{E_{P,m}^{AE}} \right), \Omega_{E_P} \right) \\
    (EC_{m}^{DE}, EC_{m}^{AE}) & \sim N \left( \left( \mu_{EC_{m}^{DE}}, \mu_{EC_{m}^{AE}} \right), \Omega_{EC} \right)
\end{align*}
\]

\[
\begin{align*}
    \log \left( EC_{DE}^{DE} (drug_{km}), EC_{AE}^{AE} (drug_{km}) \right) & \sim N \left( \log(\mu_{EC_{DE}^{DE} (drug_{km})}, \mu_{EC_{AE}^{AE} (drug_{km})}), \Sigma_{EC} \right) \\
    \mu_{EC_{DE}^{DE} (drug_{km})} & = \theta_{PCR \rightarrow DE} EC_{50}^{PCR}_{DE} (drug_{km}) \\
    \mu_{EC_{AE}^{AE} (drug_{km})} & = \theta_{PCR \rightarrow AE} EC_{50}^{PCR}_{AE} (drug_{km})
\end{align*}
\]
Integrated preclinical-clinical model: The gory details

• Disease event/AE sub-model (cont.)

Priors

\[
\begin{align*}
\mu_{E_p}^{DE} & \sim N\left(0,10^4\right) \quad \mu_{E_p}^{AE} \sim N\left(0,10^4\right) \\
\Omega_{E_p}^{-1} & \sim \text{Wishart}\left(\text{diag}\left(0.0001,0.0001\right),2\right) \\
\mu_{E_C}^{DE} & \sim N\left(0,10^4\right) \quad \mu_{E_C}^{AE} \sim N\left(0,10^4\right) \\
\Omega_{E_C}^{-1} & \sim \text{Wishart}\left(\text{diag}\left(0.0001,0.0001\right),2\right) \\
\log(\theta_{PCR\rightarrow DE}) & \sim N\left(0,10^6\right) \quad \log(\theta_{PCR\rightarrow AE}) \sim N\left(0,10^6\right) \\
\Sigma_{E_C} & \sim \text{Wishart}\left(\text{diag}\left(0.0001,0.0001\right),2\right) \\
\theta_{group2}^{DE} & \sim N\left(0,10^4\right) \quad \gamma_{DE} \sim \text{uniform}(0.1,10) \\
\theta_{group2}^{AE} & \sim N\left(0,10^4\right) \quad \gamma_{AE} \sim \text{uniform}(0.1,10)
\end{align*}
\]