Prior elicitation to support quantitative decision making in drug development

Juan J Abellan
Advanced Biostatistics and Data Analytics, GSK

Acknowledgements:
Nicky Best
Nigel Dallow
Tim Montague

EFSPi Decision Making in Drug Development,
Suresnes (Paris), 12 Dec 2018

Background

Back in 2012 at GSK
– Awareness of high rate failure in late development due to poor quantification of uncertainties on key aspects
  – Dose selection for Ph III
  – Efficacy
  – …

It was in fact an industry-wide problem…

JAMA (2014) 311:378-384
**Background**

Quantitative science party promoted a “Re-engineering Phase 2” initiative based on 4 main pillars:

- Emphasis on dose-response studies
  - To ensure at the end of Ph2 there is strong evidence of the doses for Ph3
- Futility analysis and predictive inference
  - To include early stops for futility to mitigate risks due to uncertainties
  - To predict Ph3 results based on evidence accumulated after Ph2 (assurance)
- Use of historical data
- Quantitative Sciences Peer Review Forum
  - Space for Stats and clinical pharmacology M&S to discuss quantitative aspects of study designs

Evidence accumulated included:

- Actual data (e.g. PK data, results from previous clinical studies, publications, ...)
- Scientific knowledge of the molecule/mechanism
- Clinical experience of treating patients
- …

Different levels of uncertainty in predictability or relevance of the information
- Often a translational gap between previous and current studies

In 2014, GSK implemented a formal expert elicitation process to translate prior data and expert knowledge into quantitative prior distributions
Prior elicitation

Quantification of risks

- Represents our best expression of what is known, “just now”, about the true drug effect of our asset
- Elicited priors can be used to interrogate potential clinical trial designs and development plans, in order to assess their utility
  - Which of three trial designs has the highest probability of success?
  - Should we incorporate an interim futility test, because our current state of knowledge is too diffuse?
  - Should we go straight to Phase 3? Do we believe enough in our drug now to make that commitment?
- Additional by-products of the elicitation process include:
  - Dedicated time for team to discuss all relevant data
  - Transparency of beliefs and rationale for those beliefs
- Enables uncertainty to be appropriately captured and communicated to decision makers
  - Elicited priors now provided as supporting information at most major governance board milestones at GSK

Prior elicitation

Key elements

Experts
- Typically ~6 people (clinicians, scientists and 1 statistician)
Facilitator
- Chairs the session
Technical facilitator
- Documents the session and runs software to fit distributions
Evidence dossier
- Experts’ judgements should differ only because of their expertise and interpretation of the evidence not from having different data!
- The dossier should summarise the main relevant evidence
Prior elicitation

The process

Decision to conduct elicitation

Pre-elicitation phase (project statistician & physician + facilitator)

Problem definition (project team)

Limited/conflicting evidence; high uncertainty

Decision problem or statistical model

Elicitation phase (experts + facilitator)

Frame problem

Select experts

Select method

Prepare evidence dossier

Post-elicitation phase (facilitator)

Carry out elicitation

Training

Documentation

Based on SHELF:
SHeffield ELicitation Framework
(O’Hagan and Oakley)
http://www.tonyohagan.co.uk/shelf/
Includes documents and software to aid elicitation

Decision problem: Phase III planning for fixed dose combination (FDC) of two approved products.

Relevant Data: A positive Phase II study and a wealth of data and knowledge on individual components and other FDCs.

Unknown: How results from the phase II study (challenge model) translate to Phase III clinical study (real world situation).
Prior elicitation

The process

Elicitation aim: to elicit true mean treatment difference between FDC and monotherapy

Decision to conduct elicitation

Problem definition (project team)

Limited/conflicting evidence; high uncertainty

Decision problem or statistical model

Pre-elicitation phase (project statistician & physician + facilitator)

Select experts

Select method

Prepare evidence dossier

Elicitation phase (experts + facilitator)

Data summaries from GSK reports and published competitor studies

Evidence dossier

Post-elicitation phase (facilitator)

Consensus

-0.5 0.0 0.5 1.0 1.5 2.0

True treatment difference

Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6

Training

Carry out elicitation

Documentation
Prior elicitation

The process

Decision to conduct eliciton

Pre-elicitation phase (project statistician & physician + facilitator)
- Problem definition (project team)
- Select experts
- Select method

Elicitation phase (experts + facilitator)
- Frame problem
- Decision to conduct elicitation
- Prepare evidence dossier
- Training
- Carry out elicitation

Post-elicitation phase (facilitator)
- Documentation

Communicating priors to decision makers

Belief distribution about true size of treatment effect

- Model-based predictions
  - Multiple uncertainties in statistical model
  - Available data insufficient to estimate parameters well
  - Low precision for predicting phase 3 treatment effect
- Consensus belief distribution
  - More informative than model based prior: based on experts' knowledge in addition to available data
  - Strong conviction that FDC could not lead to true outcome being worse than monotherapy
  - Treatment effects > 1 would be exceptional

- Sample sizes above ~1500 per arm yield negligible gains in assurance

SUCCESS = p < 0.05 AND observed effect > 0.4 IN BOTH P3 TRIALS

0 20 40 60 80 100
0 0.5 1.0 1.5 2.0

-0.5 0.0 0.5 1.0 1.5 2.0

True treatment difference (FDC - monotherapy)

Proposed sample size (N=1575)

Blockbuster
Clinically beneficial
Not viable

Assurance (Probability of Phase 3 Success)

1000 1500 2000 2500 3000

Total sample size, N

0 20 40 60 80 100

Proposed model predictions
Consensus
Individual experts
Model 1 predictions
Model 2 predictions
Structuring the elicitation - what to elicit?

- The quantities to be elicited should be defined in such a way that the expert is able to apply his/her knowledge without necessitating “mental gymnastics”
- Elicitation questions should be framed in a language that is familiar to the experts, e.g. observable quantities rather than model parameters
- Depends on the context and on the experts

Structuring the elicitation
Example – eliciting a prior for a hazard ratio

A. Elicitation to inform a cardiovascular outcomes trial
- Main source of evidence was several previous trials reporting the hazard ratio for the same endpoint for competitor molecules
- Experts felt comfortable directly eliciting their beliefs about the HR
B. Elicitation for a rare disease with a novel endpoint

- No previous comparative studies
- Only relevant evidence was on disease progression rates from a natural history study, plus some limited PK-PD data on the molecule of interest.
- Elicited experts' beliefs about
  1) Proportion of patients who would progress by 18 months on placebo
  2) Relative difference in this proportion between active and placebo.

Assuming exponential event times, (1) & (2) \( \Rightarrow \) prior for the HR

Consensus prior

- Judgements elicited from several experts to cover range of scientific opinion and expertise
- But (ideally) a single prior is needed for decision-making
  - SHELF protocol uses behavioural aggregation for consensus prior
  - Alternative is mathematical aggregation (weighted average)

<table>
<thead>
<tr>
<th>Benefits of Behavioural Aggregation</th>
<th>Risks of Behavioural Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encourages sharing knowledge</td>
<td>Difficulty of managing the experts</td>
</tr>
<tr>
<td>Avoids using an arbitrary</td>
<td>Difficulty of ensuring all opinions are</td>
</tr>
<tr>
<td>mathematical rule</td>
<td>treated on their merits</td>
</tr>
<tr>
<td>Consensus prior intended to</td>
<td>Experts required to ‘put themselves in</td>
</tr>
<tr>
<td>represent view of a Rational</td>
<td>someone else’s shoes’</td>
</tr>
<tr>
<td>Impartial Observer</td>
<td></td>
</tr>
</tbody>
</table>
Achieving an aggregate prior may be challenging

An example

- Setting:
  - Planning for PIII trial of existing drug in new indication with unmet medical need
  - Heterogeneous patient population
  - No well-established disease-severity index
  - Elicitation of response rate on Standard of Care (SOC)

Individual expert priors
- Expert 1 & 2 based on clinical experience (primary care)
- Expert 3 & 5 based on literature and tertiary care experience
- Expert 4 based on literature allowing for heterogeneity

Mathematical average

EFSPI Decision Making in Drug Development
Achieving an aggregate prior may be challenging

**An example**

- Setting:
  - Planning for PIII trial of existing drug in new indication with unmet medical need
  - Heterogeneous patient population
  - No well-established disease-severity index
  - Elicitation of response rate on Standard of Care (SOC)

![Consensus](chart)

- **Consensus**
  - Reflects discussion around patient heterogeneity and expectation that patient population for trial likely to be more severe

---

Another example

- Setting:
  - Planning for PIII trial in rare disease with high unmet medical need
  - Novel clinical endpoint
  - No historical data
  - Elicitation of flare rate on placebo

![Individual expert priors](chart)

- **Individual expert priors**
  - Some experts believed incl/excl criteria could lead investigators to artificially include stable patients by adapting background therapy
  - Expert 5 (red) assumed that stable patients wouldn’t be enrolled
Achieving an aggregate prior may be challenging

Another example

- Setting:
  - Planning for PIII trial in rare disease with high unmet medical need
  - Novel clinical endpoint
  - No historical data
  - Elicitation of flare rate on placebo

Consensus priors

2 consensus priors considered to highlight risks to decision makers:
- Study enrolls subjects prone to flare (expert 5 prior, red)
- Study enrolls more stable patients who are less likely to flare (mathematical average of other experts, blue)

Achieving an aggregate prior may be challenging

Another example

Expert beliefs about placebo flare rate

Expert beliefs about relative risk of flare on active vs placebo

Assurance (Predicted probability of success) for planned PIII trial

<table>
<thead>
<tr>
<th></th>
<th>N=40</th>
<th>N=60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 5</td>
<td>78%</td>
<td>89%</td>
</tr>
<tr>
<td>Placebo prior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled</td>
<td>43%</td>
<td>57%</td>
</tr>
<tr>
<td>Placebo prior</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Placebo flare rate is a key determinant of assurance
- Expert 5 had a higher expectation than other experts due to different beliefs about stability of patients recruited
- Inc/Excl criteria modified and blinded sample size readjustment planned to mitigate risk of low placebo flare rate
Managing the tendency for over-optimism in expert opinion may also be a challenge

1. Elicit a prior probability that the drug ‘works’
2. Elicit a prior for the true treatment effect conditional on the drug ‘working’ (e.g. mechanism translating)
3. Marginal prior for drug effect is then a mixture distribution:

Example of Bimodal Prior Elicitation

Setting:
- Rare disease with history of studies failing in this disease area
- Ongoing Phase 2 study
- Early stages of planning Phase 3

Elicitation Aim:
- Elicit experts beliefs without the ‘bias’ of observing the phase 2 study
- Combine the prior with the observed phase 2 data so as to calculate the assurance for potential phase 3 designs
Example of Bimodal Prior Elicitation

Elicitation

1. Prior belief that drug works ('causes some relevant biological activity')
   - Consensus was 25% (range: 10 to 40%)

2. Conditional on drug working, how efficacious is it?

   ![Best Fitting Distributions](image)

   Overall mixture prior
   - Update this with phase 2 data
   - Use this phase 2 posterior in assurance calculations for planning phase 3

   ![75% probability it doesn't work](image)
   If it does work, then centred around a 30% reduction in slope
Example of Bimodal Prior Elicitation

Overall mixture prior
- Update this with phase 2 data
- Use this phase 2 posterior in assurance calculations for planning phase 3

What actually happened....
- Phase 2 results were negative
  - Planning for Phase 3 did not go ahead
- Retrospective assurance calculation for Phase 2 study: assurance=21%
  - Should we have planned interim futility analysis?
  - Retrospective mid-trial futility analysis showed trial could have stopped early, saving ~5 months and >200 doses across remaining subjects

Feedback from experts

“It is the process itself which is most valuable for the team, uncovering heterogeneity among expert views in a totally transparent way”

“Allowed internal team to have a clear and honest discussion with external experts without either side trying to say what other side wants to hear”

“The negotiation among experts and the exchange of rationale for probabilities was probably the most valuable part”

“It challenges your views - often entrenched and biased.”
Discussion

- Prior elicitation enables project teams to utilize historical data, prior knowledge from experts, and collective thought for a more robust output on study design and/or analysis
- Over 40 elicitations conducted at GSK to date
  - Majority for Phase 2, 3 and 4 projects
  - Mostly run as face to face or VTC sessions lasting 2-4 hours
  - Number of experts typically 5-7
  - 25% include external experts
  - positive feedback received from all teams
- All teams at GSK expected to explore the potential of Prior Elicitation for their projects
  - White paper and GSK technical paper provide guidance on prior elicitation and assurance
  - Details of Prior distribution + Assurance + MDE required for all major governance board milestones

Thank you

References:
