Using prior elicitation and Bayesian thinking to help shape decision making in the pharmaceutical industry

Nicky Best
Statistical Innovation Group, GSK
# The Drug Development Process

![Image of the drug development process]

## Average number of years taken to develop successful medicine

<table>
<thead>
<tr>
<th>Stage</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-clinical testing</td>
<td>4.5 years</td>
</tr>
<tr>
<td>Early phase</td>
<td>5.5 years</td>
</tr>
<tr>
<td>Middle phase</td>
<td>7.0 years</td>
</tr>
<tr>
<td>Late phase</td>
<td>8.5 years</td>
</tr>
<tr>
<td>Approval</td>
<td>11.0 years</td>
</tr>
<tr>
<td>Total</td>
<td>12.5 years</td>
</tr>
</tbody>
</table>

## Average cost to research and develop successful medicine

<table>
<thead>
<tr>
<th>Stage</th>
<th>Average Cost (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-clinical testing</td>
<td>£436 million</td>
</tr>
<tr>
<td>Early phase</td>
<td>£533 million</td>
</tr>
<tr>
<td>Middle phase</td>
<td>£710 million</td>
</tr>
<tr>
<td>Late phase</td>
<td>£916 million</td>
</tr>
<tr>
<td>Approval</td>
<td>£1.1 billion</td>
</tr>
<tr>
<td>Total</td>
<td>£1.15 billion</td>
</tr>
</tbody>
</table>

## Number of medicinal candidates tested to achieve one approved medicine

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number of Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-discovery</td>
<td>5,000 - 10,000</td>
</tr>
<tr>
<td>Drug discovery</td>
<td>10-20</td>
</tr>
<tr>
<td>Pre-clinical testing</td>
<td>5-10</td>
</tr>
<tr>
<td>Phase 1 clinical trial</td>
<td>2-5</td>
</tr>
<tr>
<td>Phase 2 clinical trial</td>
<td>1-2</td>
</tr>
<tr>
<td>Phase 3 clinical trial</td>
<td>1</td>
</tr>
</tbody>
</table>

## Notes:
4. IFPMA analysis, updated for data per Tufts Center for the Study of Drug Development (CSD) database (1998)
Key Milestone Decisions Gates Through Drug Development

- Discovery & Preclinical
  - Commit to Target (C2T)
  - Commit to Candidate Selection (CS)

- Phase I
  - Commit to Phase I (C2FTI)
  - Commit to Medicines Development (C2MD)

- Phase II
  - Commit to Phase II (C2PhII)

- Phase III
  - Commit to Phase III (C2PhIII)
  - Commit to File (C2F)

- Regulatory Agency Review

- Phase IV
  - Commit to Launch (C2L)
  - 1, 2 5 Year Launch Reviews

- Commercialisation
Trends in Pharmaceutical Industry Success Rates

Based on data from a consistent cohort of 20 companies participating each year between 2008 and 2015.

© CMR International, a Thomson Reuters business
Most late phase clinical trials are conducted with 90% power, but the success rate is much less than 90%.

Why is this?
Case study: Cancer trial

Phase 2 study results

HR = 0.75
95% CI (0.46, 1.23)
Case study: Cancer trial

Phase 2 study results

HR = 0.75
95% CI (0.46, 1.23)

Phase 3 study results

HR = 1.02
95% CI (0.89, 1.18)
Case study: Cancer trial

**Phase 2 study results**

- **HR = 0.75**
- **95% CI (0.46, 1.23)**

**Phase 3 study results**

- **HR = 1.02**
- **95% CI (0.89, 1.18)**

How can we better discharge risk?

Should we be surprised?
A protocol might say something like this ...

Assuming a clinically relevant difference of 2 points on the primary endpoint scale, with a standard deviation of 6.2, 200 subjects per arm are required to provide 90% power at the 5% alpha level (two-sided).
A protocol might say something like this...

Assuming a clinically relevant difference of 2 points on the primary endpoint scale, with a standard deviation of 6.2, 200 subjects per arm are required to provide 90% power at the 5% alpha level (two-sided).

We are assuming with 100% certainty that the true effect of the drug is 2 points.
Power is not knowledge

Expert belief about true effect
Power is not knowledge

Expert belief about true effect

Power calculation assumption
## Power and Assurance

<table>
<thead>
<tr>
<th>True effect size</th>
<th>Power</th>
<th>Expert Belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.5%</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>36%</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>90%</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>99.8%</td>
<td>10%</td>
</tr>
</tbody>
</table>
### Power and Assurance

<table>
<thead>
<tr>
<th>True effect size</th>
<th>Power</th>
<th>Expert Belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.5%</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>36%</td>
<td>30%</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td><strong>90%</strong></td>
<td><strong>40%</strong></td>
</tr>
<tr>
<td>3</td>
<td>99.8%</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Power = 90%**

*the probability of success assuming the true (unknown and never known) effect of the drug is 2 points*
## Power and Assurance

<table>
<thead>
<tr>
<th>True effect size</th>
<th>Power</th>
<th>Expert Belief</th>
<th>Power x Belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.5%</td>
<td>20%</td>
<td>0.5%</td>
</tr>
<tr>
<td>1</td>
<td>36%</td>
<td>30%</td>
<td>10.8%</td>
</tr>
<tr>
<td>2</td>
<td>90%</td>
<td>40%</td>
<td>36%</td>
</tr>
<tr>
<td>3</td>
<td>99.8%</td>
<td>10%</td>
<td>9.9%</td>
</tr>
</tbody>
</table>

Power = 90%

*the probability of success assuming the true (unknown and never known) effect of the drug is 2 points*
Power and Assurance

<table>
<thead>
<tr>
<th>True effect size</th>
<th>Power</th>
<th>Expert Belief</th>
<th>Power x Belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.5%</td>
<td>20%</td>
<td>0.5%</td>
</tr>
<tr>
<td>1</td>
<td>36%</td>
<td>30%</td>
<td>10.8%</td>
</tr>
<tr>
<td>2</td>
<td>90%</td>
<td>40%</td>
<td>36%</td>
</tr>
<tr>
<td>3</td>
<td>99.8%</td>
<td>10%</td>
<td>9.9%</td>
</tr>
</tbody>
</table>

**Power = 90%**  
*the probability of success assuming the true (unknown and never known) effect of the drug is 2 points*

**BUT**  
*Assurance (prob of success) = 57%*

*the average of the power calculations, weighted by the belief about how big the true effect size is*
Back to the cancer trial....Let's travel back in time!

– What would you like to know before doing the study that would help you make an investment decision?

– Rewind 10 years
– Designed to have 90% power to detect clinically relevant HR of 0.78
– What do the Phase II data tell us about the treatment effect?
  – Conventional frequentist analysis gives HR = 0.75; 95% CI (0.46, 1.23)
  – Bayesian analysis with ‘ignorance’ prior:
Back to the cancer trial....

- Designed to have 90% power to detect clinically relevant HR of 0.78
- What do the Phase II data tell us about the treatment effect?
  - Conventional frequentist analysis gives HR = 0.75; 95% CI (0.46, 1.23)
  - Bayesian analysis with ‘ignorance’ prior:

\[
\begin{array}{cccccc}
0.25 & 0.5 & 0.75 & 1.0 & 1.5 & 2.0 \\
\text{Pr(HR<0.78)} & = & 0.56 & & & \\
\text{Pr(HR>1)} & = & 0.12 & & & \\
\end{array}
\]
Back to the cancer trial....

- Designed to have 90% power to detect clinically relevant HR of 0.78
- What do the Phase II data tell us about the treatment effect?
  - Conventional frequentist analysis gives HR = 0.75; 95% CI (0.46, 1.23)
  - Bayesian analysis with ‘ignorance’ prior:

\[
\begin{array}{c|c|c}
\text{HR} & \text{Pr(HR<0.78)} & \text{Pr(HR>1)} \\
0.25 & 0.56 & \\
0.5 & 0.56 & 0.12 \\
0.75 & 0.56 & 0.12 \\
1.0 & 0.56 & 0.12 \\
1.5 & 0.56 & 0.12 \\
2.0 & 0.56 & 0.12 \\
\end{array}
\]

![Graph showing posterior distribution and power curve](image-url)
Back to the cancer trial....

- Designed to have 90% power to detect clinically relevant HR of 0.78
- What do the Phase II data tell us about the treatment effect?
  - Conventional frequentist analysis gives HR = 0.75; 95% CI (0.46, 1.23)
  - Bayesian analysis with ‘ignorance’ prior:

![Graph showing Bayesian analysis results]

\[
\text{Pr}(\text{HR}<0.78) = 0.56 \quad \text{Pr}(\text{HR}>1) = 0.12
\]

\[
\text{Assurance} = 68\%
\]
Back to the cancer trial....

Assurance = 68%

- Probability that the trial will meet its primary endpoint based on current (....we are still back in time...) evidence about the treatment effect
  
- Is this probability high or low?

- Phase 2 trial does not exist in a vacuum – what other evidence should we take into account to produce our prior?

- Phase 3 setting ≠ Phase 2 setting
  - Different treatments
  - Different populations
Uncertainty is not Ignorance

- Even if we have only imperfect knowledge about an asset
  - How it performed in a related population
  - What our competitors have found with the same mechanism
  - What I know about the disease (which you might not know)

→ this can be used to help interrogate potential future clinical designs
Uncertainty is not Ignorance

- Even if we have only imperfect knowledge about an asset
  - How it performed in a related population
  - What our competitors have found with the same mechanism
  - What I know about the disease (which you might not know)

→ this can be used to help interrogate potential future clinical designs

- We do this by formally combining knowledge and data, into a “prior distribution”, that represents our best expression of what is known, “just now”, about the true drug effect of our asset
Uncertainty is not Ignorance

– Even if we have only imperfect knowledge about an asset
  – How it performed in a related population
  – What our competitors have found with the same mechanism
  – What I know about the disease (which you might not know)

→ this can be used to help interrogate potential future clinical designs

– We do this by formally combining knowledge and data, into a “prior distribution”, that represents our best expression of what is known, “just now”, about the true drug effect of our asset

– The prior 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?
GSK Prior Elicitation Initiative

– Prior knowledge exists on every project in some form
  – Different levels of uncertainty in predictability or relevance of the information
  – Often a translational gap between historical and current settings
GSK Prior Elicitation Initiative

- Prior knowledge exists on every project in some form
  - Different levels of uncertainty in predictability or relevance of the information
  - Often a translational gap between historical and current settings
- GSK have implemented formal expert elicitation to translate this information into quantitative prior distributions
GSK Prior Elicitation Initiative

- Prior knowledge exists on every project in some form
  - Different levels of uncertainty in predictability or relevance of the information
  - Often a translational gap between historical and current settings
- GSK have implemented formal expert elicitation to translate this information into quantitative prior distributions
- Elicited priors used to:
  - determine assurance (prior predictive probability of success for a future study)
  - design clinical trials (e.g. plan interims, compare development strategies, stagger investment)
  - draw statistical inference (i.e. analysis of study data)
GSK Prior Elicitation Initiative

- Prior knowledge exists on every project in some form
  - Different levels of uncertainty in predictability or relevance of the information
  - Often a translational gap between historical and current settings
- GSK have implemented formal expert elicitation to translate this information into quantitative prior distributions
- Elicited priors used to:
  - determine assurance (prior predictive probability of success for a future study)
  - design clinical trials (e.g. plan interims, compare development strategies, stagger investment)
  - draw statistical inference (i.e. analysis of study data)
- 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 be captured and communicated
GSK Prior Elicitation process

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)
- Frame problem
- Select experts
- Select method
- Prepare evidence dossier

Elicitation phase (experts + facilitator)
- Carry out elicitation
- Training

Post-elicitation phase (facilitator)
- Documentation
Example of Prior Elicitation at GSK

**Decision problem:** Phase 3 planning for fixed dose combination (FDC) of two approved products.

**Relevant Data:** A positive Phase 2 study and a wealth of data and knowledge on individual components and other FDCs.

**Unknown:** How results from the phase 2 study (challenge model) translate to Phase III clinical study (real world situation).
Example of Prior Elicitation at GSK

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

- **Pre-elicitation phase (project statistician & physician + facilitator)**
  - Frame problem
  - Select experts
  - Select method
  - Prepare evidence dossier

- **Post-elicitation phase (facilitator)**
  - Elicitation

**Problem definition (project team)**
- Limited/conflicting evidence; high uncertainty
- Decision problem or statistical model

**Data summaries**
- from GSK reports and published competitor studies
- GSK Historical Data Sets
- Evidence dossier
- Regulatory Reviews
- Journal Articles

**Evidence dossier**

**Decision to conduct elicitation**
Example of Prior Elicitation at GSK

Problem definition (project team)
- Limited/conflicting evidence; high uncertainty
- Decision problem or statistical model

Pre-elicitation phase (project statistician & physician + facilitator)
- Frame problem
- Select experts
- Select method
- Prepare evidence dossier

Elicitation phase (experts + facilitator)
- Carry out elicitation

Post-elicitation phase (facilitator)
- Documentation

Decision to conduct elicitation

Supplementary data:
- Expert 1
- Expert 2
- Expert 3
- Expert 4
- Expert 5
- Expert 6
- Consensus
Example of Prior Elicitation at GSK

**Problem definition (project team)**
- Limited/conflicting evidence; high uncertainty
- Decision problem or statistical model

**Pre-elicitation phase (project statistician & physician + facilitator)**
- Frame problem
- Select experts
- Select method
- Prepare evidence dossier

**Elicitation phase (experts + facilitator)**
- Carry out elicitation
- Training

**Post-elicitation phase (facilitator)**
- Documentation

**Decision to conduct elicitation**

---

**ELICITATION RECORD – Part 1 – Context**

<table>
<thead>
<tr>
<th>Elicitation title</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Session</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Part 1 start time</td>
<td></td>
</tr>
</tbody>
</table>

**ELICITATION RECORD – Part 2 – Distribution**

**Roulette Method**

<table>
<thead>
<tr>
<th>Definition</th>
<th>Define quantity to be elicited (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence</td>
<td>Review of evidence relating to X</td>
</tr>
<tr>
<td>Plausible range</td>
<td>Record the range of plausible values for X elicited from each expert</td>
</tr>
<tr>
<td>Chips in bins</td>
<td>Each expert asked to create histogram representing their beliefs about X. Record histograms in bins here.</td>
</tr>
<tr>
<td>Fitting</td>
<td>Record distributions fitted to each of the experts’ histograms</td>
</tr>
<tr>
<td>Group elicitation</td>
<td>Experts invited to discuss their different distributions and share knowledge and reasoning about differences. Record key points of this discussion, together with the consensus histogram</td>
</tr>
<tr>
<td>Fitting and feedback</td>
<td>Record process of fitting, feedback and revision of the group consensus judgments</td>
</tr>
<tr>
<td>Chosen distribution</td>
<td>Record and show the final fitted distribution</td>
</tr>
<tr>
<td>Discussion</td>
<td>Record experts’ reactions to the process and to the final fitted distribution, plus any difficulties that arose during the elicitation</td>
</tr>
</tbody>
</table>
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

Success = $p < 0.05$ and observed effect $> 0.4$ in both P3 trials

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

Plot shows assurance for 3:3:1:1 randomisation ratio; alternative designs with different randomisation ratios gave almost identical assurance values
1. Elicit a prior for the true treatment effect \textit{conditional} on the drug ‘working’ (e.g. mechanism translating)

Assumed prior distribution for treatment effect \textit{if} drug ‘works’
Managing the tendency for over-optimism in expert opinion

1. Elicit a prior for the true treatment effect \textit{conditional} on the drug ‘working’ (e.g. mechanism translating)
2. Elicit a prior probability that the drug ‘works’
Managing the tendency for over-optimism in expert opinion

1. Elicit a prior for the true treatment effect **conditional** on the drug ‘working’ (e.g. mechanism translating)
2. Elicit a prior probability that the drug ‘works’
3. Combine with ‘placebo-like’ distribution tightly centred around zero

[Diagrams showing assumed prior distributions for treatment effect if drug is placebo-like and if drug works]
Managing the tendency for over-optimism in expert opinion

1. Elicit a prior for the true treatment effect *conditional* on the drug ‘working’ (e.g. mechanism translating)
2. Elicit a prior probability that the drug ‘works’
3. Combine with ‘placebo-like’ distribution tightly centred around zero
   ➔ Mixture prior
Example of Bimodal Prior Elicitation

Problem definition

Decision problem:
- 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 II study
- Combine the prior with the observed phase II data so as to calculate the assurance for potential phase III 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%)
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?
Example of Bimodal Prior Elicitation

Overall mixture prior
- Update this with phase 2 data
- Can make statements about the posterior of the phase 2
- Use in assurance calculations for planning phase 3

- 75% probability it doesn’t work
- 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
– Can make statements about the posterior of the phase 2
– Use 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**?
Challenges and Benefits of Prior Elicitation

- 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
- 13 elicitations conducted at GSK to date
  - positive feedback received from all teams
Challenges and Benefits of Prior Elicitation

- 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
- 13 elicitations conducted at GSK to date
  - positive feedback received from all teams

Practical challenges:
- Experienced, skilled facilitators are essential
- Need at least 2 facilitators, one to lead and one to run software and keep written record of elicitation session
- Logistics extremely challenging
  - 3-6 hour time commitment
  - Face-to-face in same room (VTC an option but not ideal)
- Training of experts is essential
- Need experts who are open-minded
Challenges and Benefits of Prior Elicitation

Benefits:

– Assurances of key outcomes are what decision makers need
  – Power is more or less useless for decision making
  – But you have to bite the bullet of characterising knowledge and uncertainty about true effects → prior distributions
Challenges and Benefits of Prior Elicitation

Benefits:

– Assurances of key outcomes are what decision makers need
  – Power is more or less useless for decision making
  – But you have to bite the bullet of characterising knowledge and uncertainty about true effects → prior distributions

– Formal and rigorous elicitation of expert knowledge enables
  – honest reflection of current state of knowledge
  – robust basis for conversations about asset management, trial design and interpretation of trial findings
Challenges and Benefits of Prior Elicitation

Benefits:

– Assurances of key outcomes are what decision makers need
  – Power is more or less useless for decision making
  – But you have to bite the bullet of characterising knowledge and uncertainty about true effects → prior distributions

– Formal and rigorous elicitation of expert knowledge enables
  – honest reflection of current state of knowledge
  – robust basis for conversations about asset management, trial design and interpretation of trial findings

– All teams at GSK are being encouraged to explore the potential of Prior Elicitation for their projects
  – “One pager” summarising Prior distribution + Assurance required for all major governance board milestones
Challenges and Benefits of Prior Elicitation

Benefits:

– Assurances of key outcomes are what decision makers need
  – Power is more or less useless for decision making
  – But you have to bite the bullet of characterising knowledge and uncertainty about true effects → prior distributions

– Formal and rigorous elicitation of expert knowledge enables
  – honest reflection of current state of knowledge
  – robust basis for conversations about asset management, trial design and interpretation of trial findings

– All teams at GSK are being encouraged to explore the potential of Prior Elicitation for their projects
  – “One pager” summarising Prior distribution + Assurance required for all major governance board milestones

– Impact
  – 25% reduction in a P3 study size (saving >£15M and 8 months)
  – Inclusion of interim futility analyses in several studies
Acknowledgements

Nigel Dallow
Tim Montague
Adam Crisp
John Davies
Graeme Archer
Sam Miller
Maria Costa
Will Powley
Jane Temple
Jon Robertson
Kim Hacquoil
Nicola Scott
Frank Fang
Younan Chen
Grace Zhang
Faiz Ahmad

Sara Hughes
All of GSK Clinical Statistics
Thank you for listening

Any Questions?
Backups
Assurance for Phase 3 Design – Possible Scenarios

- **Assurance = 4%**
- **Assurance = 8%**
- **Assurance = 20%**
- **Assurance = 36%**
- **Assurance = 75%**
- **Assurance = 76%**
Assurance for Phase 3 Design

% Reduction observed in Ph II vs. Assurance (%)

- Orange line: Just Ph II data
- Black line: Ph II data + elicited prior