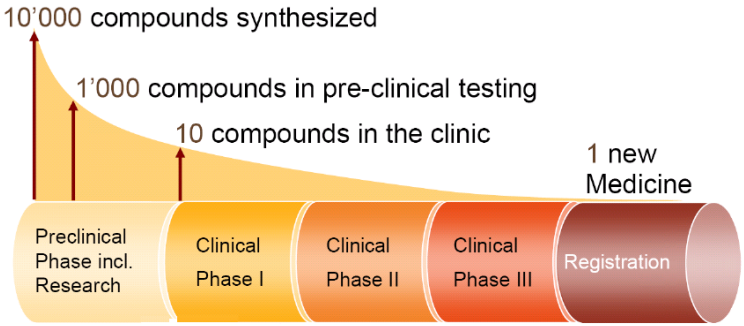
Role of Bayesian methods in evaluating and communicating risk

Lisa Hampson
BAYES 2023, Utrecht
26th October, 2023

NOVARTIS | Reimagining Medicine

Motivation



- Drug development is inherently risky, with on average over 86% of programs starting clinical development failing to secure regulatory approval*.
- At each decision milestone, there are several sources of uncertainty:
  - Is the drug efficacious and safe?
  - Will the planned clinical trials demonstrate that the drug is efficacious and safe?
- Our goal is to evaluate the risk of failure (of the drug, of a trial, of the program) and communicate these risks to stakeholders to support decision making

*Wong et al. *Biostatistics* 2019; 20:273

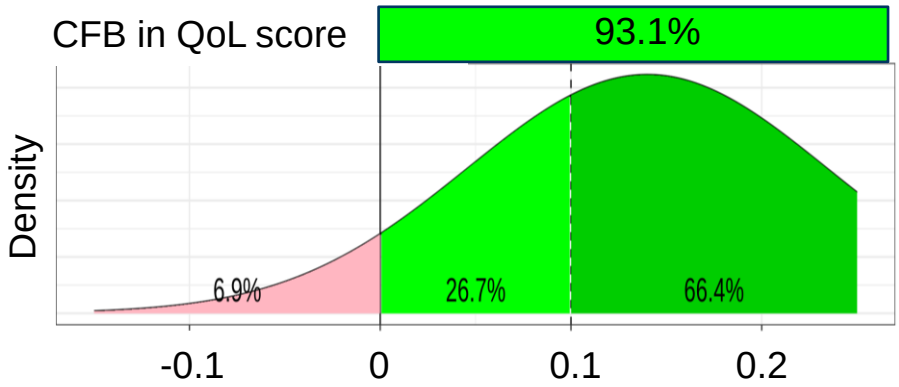
A question-based approach to evaluating risk

Suppose we are at the end of Ph2 and deciding whether to launch Ph3 ...

NOVARTIS | Reimagining Medicine

What is the risk the drug is ineffective?

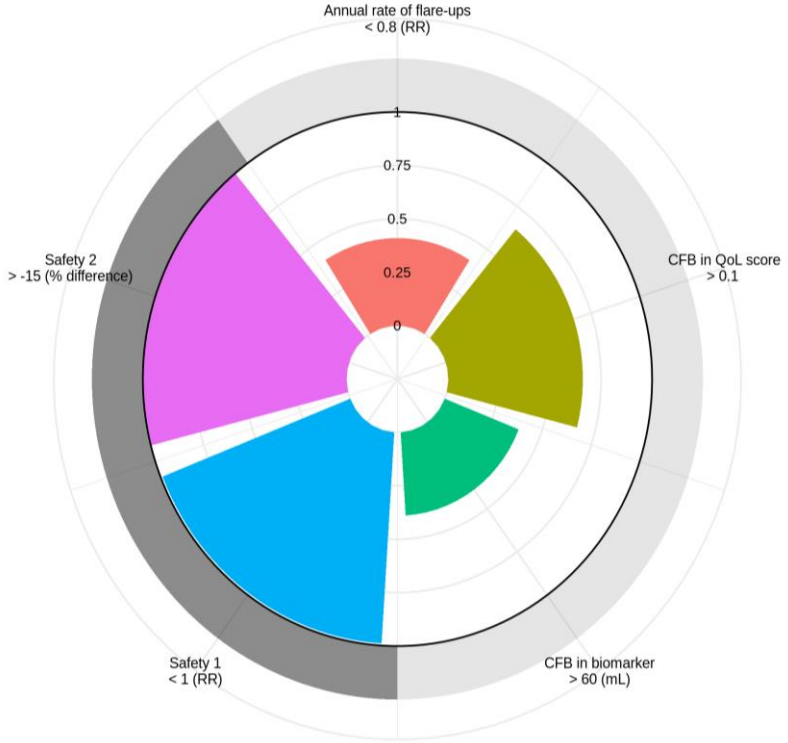
- Posterior distribution quantifies our uncertainty abut the true causal effect of the drug (θ) in light of current data
- Use probabilities to quantify current evidence in relation to target effects



What is the risk the drug is ineffective?

- Posterior distribution quantifies our uncertainty abut the true causal effect of the drug (θ) in light of current data
- Use probabilities to quantify current evidence in relation to target effects
- Evaluating evidence on several endpoints can be useful to understand drug benefit-risk profile

Prob. of clinically relevant treatment effect



NOVARTIS | Reimagining Medicine

What is the probability a future Ph3 clinical trial will meet its success criteria?

- Assurance is the expected probability of a 'successful' trial, averaging across a prior for θ

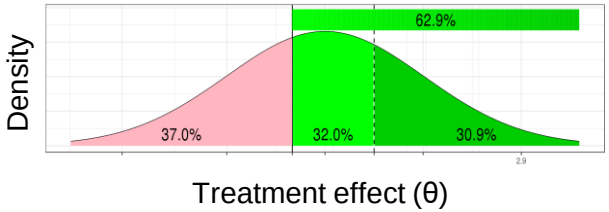
$$\int \Pr(S_3|\theta)\,\pi_0(\theta)\,d\theta$$

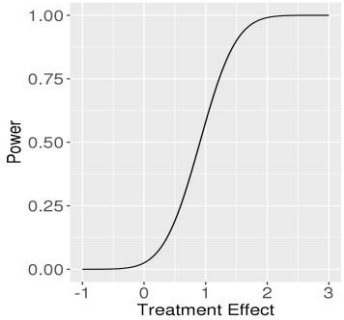
S_3 = success in Ph3 study

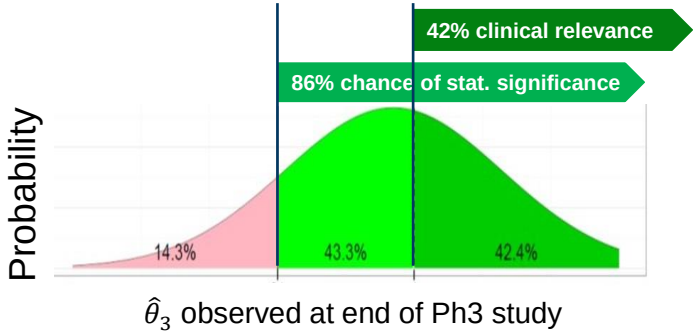
- Prior for the treatment effect θ could be based on data from earlier phase clinical trials. Or it could be the result of an expert elicitation workshop.

O'Hagan et al. *Pharmaceutical Statistics* 2005; 4:187

Calculating assurance







Evidence synthesis

Quantify current evidence. Take *m* samples from distribution of θ

Future trial design

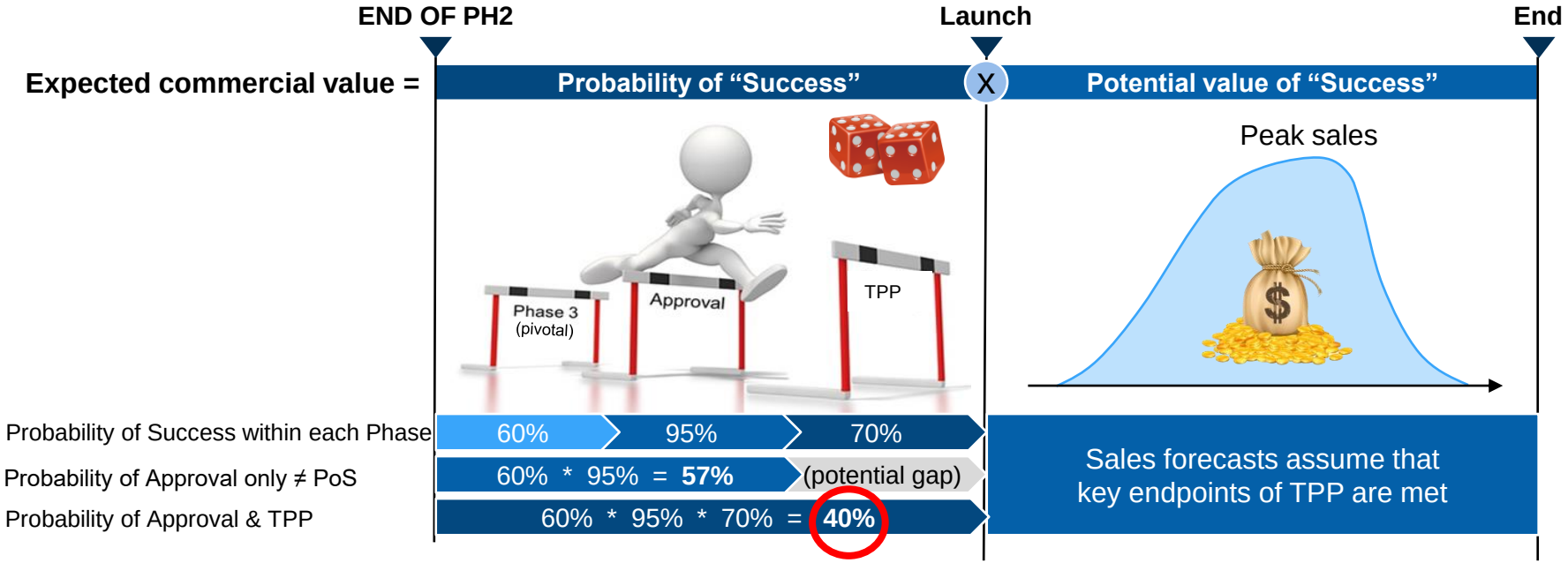
For each prior sample, simulate the future trial

Predictive probability

Calculate proportion of *m* trials in which we meet our success criteria

What is the probability of success for the program?

Success is regulatory approval with key endpoints meeting their TPP targets



TPP = Target Product Profile

Hampson et al. *Pharmaceutical Statistics* 2022; 21:439

NOVARTIS | Reimagining Medicine

A question-based approach to evaluating risk

Suppose we are at the end of Ph1 and now designing Ph2

NOVARTIS | Reimagining Medicine

How effective is this Ph2 trial design at 'derisking' the future Ph3 trial?

- Conditional assurance is assurance of a future study conditional on the success of initial study. For example,

$$\Pr(S_3 \mid S_2) = \int \Pr(S_3 \mid \theta)\, \pi(\theta \mid S_2)\, d\theta$$

S_3 = success in Ph3 trial
S_2 = success in Ph2 trial

- Compare $\Pr(S_3 \mid S_2)$ vs $\Pr(S_3)$ to reveal how effective Ph2 is at derisking Ph3

	Phase II	Phase III
Current assurance of individual phase	25%	22%
Conditional assurance given phase II success		57%

Temple & Robertson. *Pharmaceutical Statistics* 2021; 20:1102

Will running a Ph2 trial at all help us make a better Ph3 go/no-go decision?

- Suppose one design option is to skip Ph2 and make an immediate decision about whether or not to launch Ph3
- Expected Value of Sample Information (EVSI) is often used in HTA settings*:
  - Average gain in net benefit from running a study to learn about a parameter
  - Can we tailor this to obtain the value of Ph2 data for supporting Ph3 go/no-go decision?
- If we only launch Ph3 if expected net rewards exceed c, EVSI of Ph2 is given by:

$$\text{EVSI} = \underbrace{\mathbb{E}_{\hat{\theta}_2}\left[\mathbb{1}\{\mathbb{E}_{\theta|\hat{\theta}_2}[\text{NPV}(\theta)] \geq c\}\mathbb{E}_{\theta|\hat{\theta}_2}[\text{NPV}(\theta)]\right]}_{\text{Average eNPV of best decision based on current evidence } \textit{and} \text{ future Ph2 data}} - \underbrace{\mathbb{1}\{\mathbb{E}_{\theta}[\text{NPV}(\theta)] \geq c\}\mathbb{E}_{\theta}[\text{NPV}(\theta)]}_{\text{eNPV of best decision based on current evidence}}$$

$\text{NPV}(\theta) \approx \Pr_{\theta}(\text{Success in Ph3 and approval}) \times \text{sales} - \text{costs}$

* Welton et al. *Evidence Synthesis for Decision Making in Healthcare*, 2012

Expert elicitation to inform risk evaluations

Defining the prior distribution upon which we base assurance calculations

NOVARTIS | Reimagining Medicine

What is elicitation?

- Elicitation: Process of capturing experts' knowledge on uncertain quantities of interest (QoI) as probability distributions*

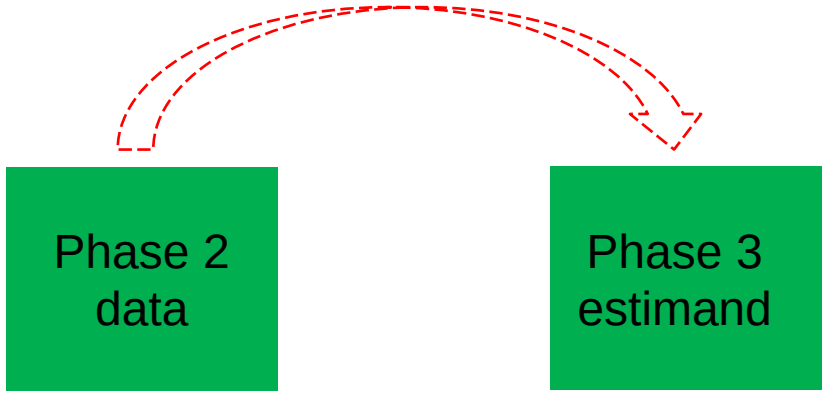
- When can elicitation add most value?
  1. If data are drawn from several sources which are of different relevance to QoI
  2. If we cannot use modelling to translate data to QoI without strong assumptions which are difficult to stress-test and / or communicate to decision makers

- In such settings, it may be more transparent to present the evidence to a group of experts and elicit their judgements

*O'Hagan et al. *Uncertain Judgements*. John Wiley & Sons, 2006

Using elicitation to bridge across differences between Ph2 and Ph3

- Different phases can use different:
  - Endpoints
  - Patient populations
  - Comparator arms
  - Dose regimens



- Relate Ph2 data to QoI in Ph3 by eliciting expert opinion.

SHELF Protocol

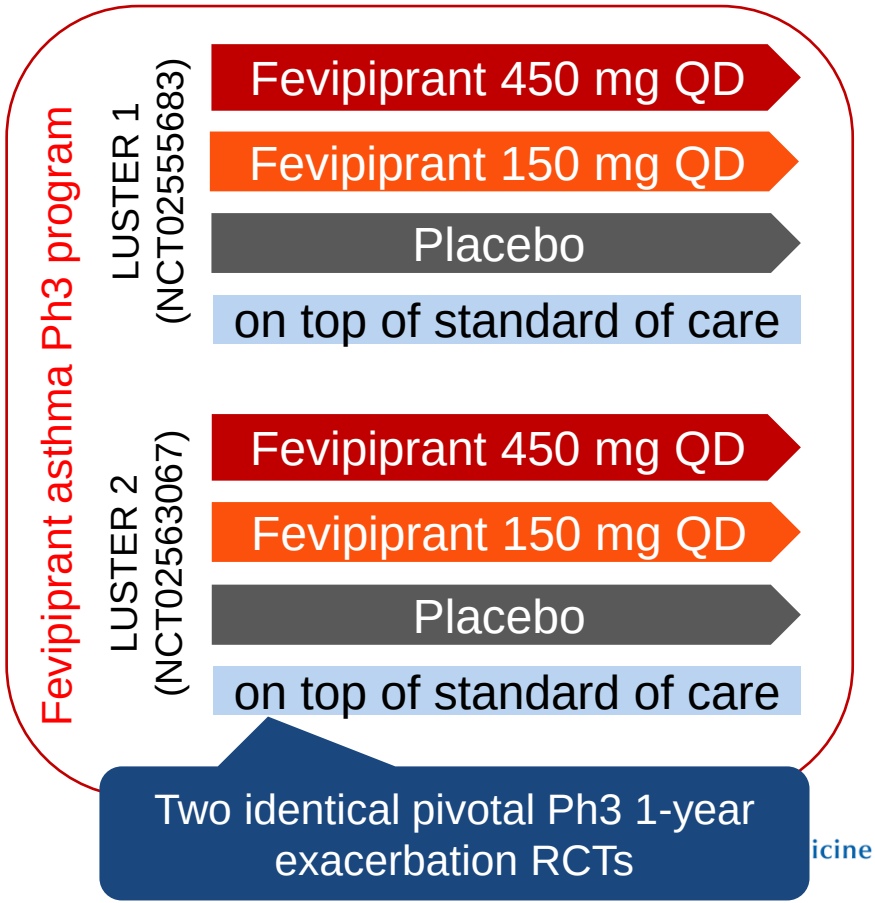
SHeffield ELicitation Framework

- Careful preparation
  - Collate evidence dossier summarizing key information relevant to QoI
- Elicitation workshop
  - Elicit experts' individual judgements
  - Discuss individual beliefs. Align on common understanding of evidence
  - Elicit distribution representing beliefs of **R**ational **I**mpartial **O**bserver (**RIO**)
- Document the elicitation workshop

- Oakley, O'Hagan. *SHELF: The Sheffield Elicitation Framework.* Link
- R SHELF package.

Example of an asthma development program

- Fevipiprant is a treatment for asthma
- Pilot for PoS framework at Novartis
- We calculated the probability of success while the Ph3 program was underway but before DBL.
- Differences between Ph2 vs Ph3:
  - Primary endpoint: Annual rate of asthma exacerbations in Ph3
  - Surrogate endpoint: 1 Ph2 study had measured the surrogate of reduction in sputum eosinophil counts.



Fevipiprant: Linking Ph2 to Ph3

- We want to use the Ph2 results on reduction (%) in sputum eosinophil counts (Y) to predict relative exacerbation rate reduction (%) in Ph3 (X)
- SHELF extension method fits this situation nicely ...

Questions to experts

Several questions were asked. But we focus here on the question:

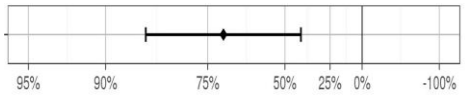
Given that an anti-inflammatory drug reduces sputum eosinophil counts by Y, what do you judge to be the likely value for the relative exacerbation rate reduction X in eligible patients?

Elicitation to map Ph2 data on sputum eosinophil reduction (Y) to exacerbation rate reduction (X)

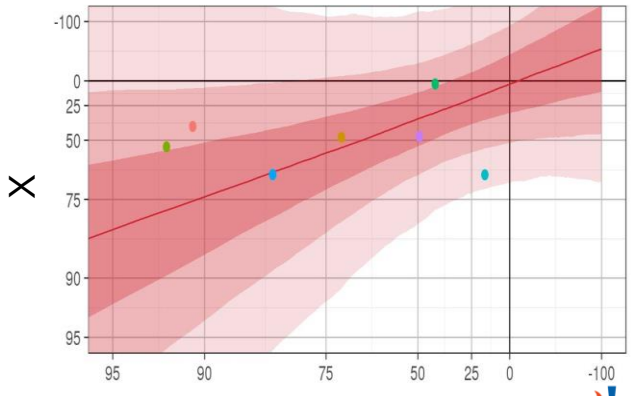
Analyze – Use Ph2 data to create a meta-analytic-predictive (MAP) prior for Y in new Ph3 study

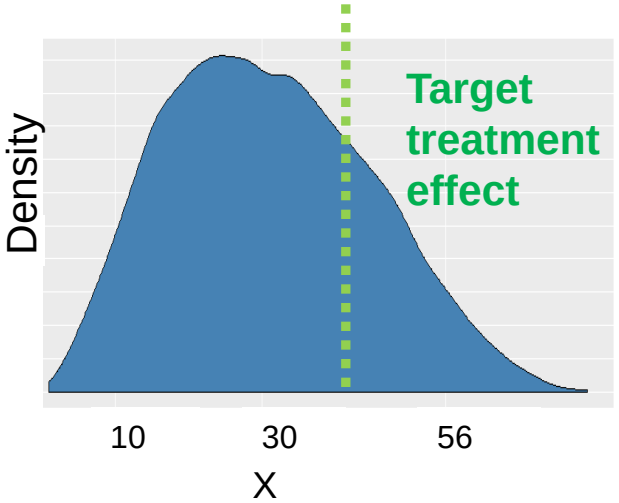
Elicit – Elicit conditional opinions on X under different scenarios for Y

Synthesize – Use expert opinion to translate Ph2 evidence on Y to derive marginal prior for X in Ph3

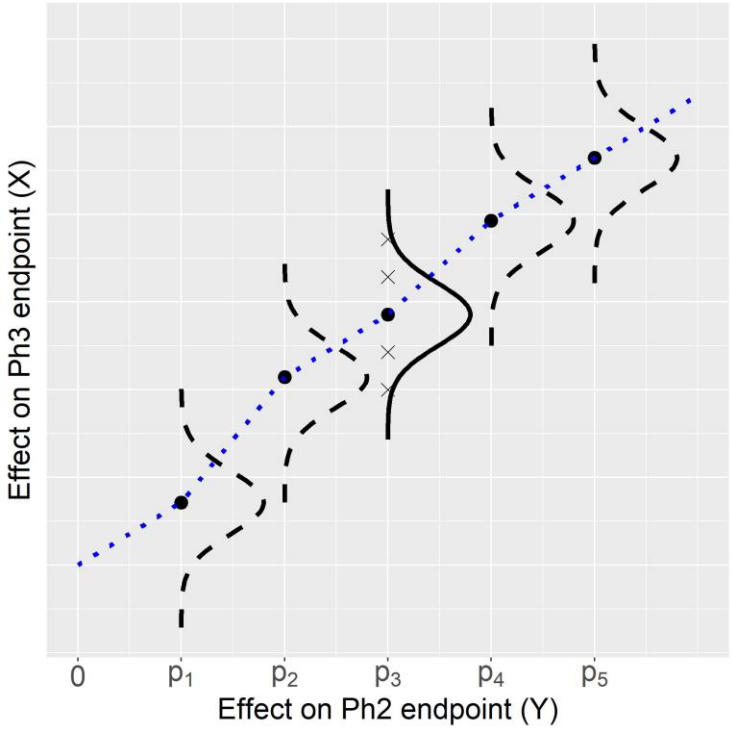


MAP prior for Y in Ph3





SHELF extension method



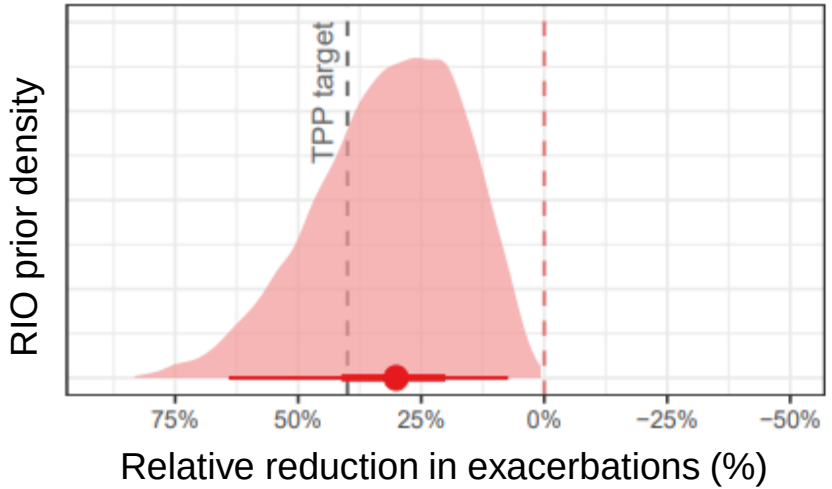
p_1 ... p_5 are percentiles of MAP prior for Y.

– E.g 10th, 25th, 50th, 75th and 90th percentiles

- Step 1: Elicit full prior distribution for X given Y = p_3
- Step 2: Elicit medians of X conditional on Y=p_1 & Y=p_5.
- Step 3: Elicit medians of X conditional on Y = p_2 & Y = p_4.
- Step 4: Identify a suitable model for median(X|Y) and Y
- Step 5: Identify a suitable model for X|Y and Y
- Step 6: Repeatedly sample y* from the MAP prior for Y and then sample x* from X|Y=y*. This generates a set of samples from the marginal prior for X.

SHELF extension method

RIO prior for the fevipiprant example



Prior median: 30.2%
95% Credible Interval: 7.0% to 60.2%

- RIO prior was consistent with the outcome of the LUSTER 1 & 2 Ph3 trials
- Observed reduction in the exacerbation rate was 23% (95% CI: 3 – 39%) based on a pooled analysis of LUSTER 1 & 2 for fevipiprant dose 450mg

Holzhauer B et al. *Pharmaceutical Statistics,* 2022;21:1005

Alternative elicitation strategies

… In case of short timelines or if there are many QoIs

In such cases, we have used abbreviated elicitation strategies which require a shorter elicitation workshop. Below is one example:

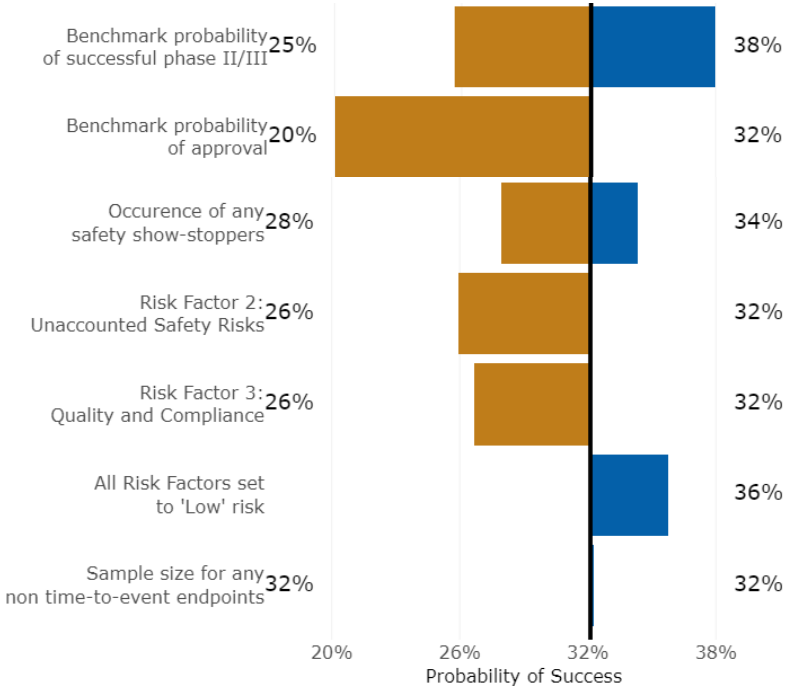
Step 1	Elicit individual judgements on scenarios linking Ph2 QoI to Ph3 QoI. Eg: • Fully translatable: PFS HR of 0.6 implies OS HR of 0.6 • Half translatable: PFS HR of 0.6 implies OS HR of 0.8 • …
Step 2	Combine Ph2 data and Step 1 to derive prior for Ph3 QoI for each expert
Step 3	Linear opinion pooling across experts to obtain ‘consensus’ prior for QoI

QoI = quantity of interest

Communicating risks

Uncertainty about risk evaluations (1)

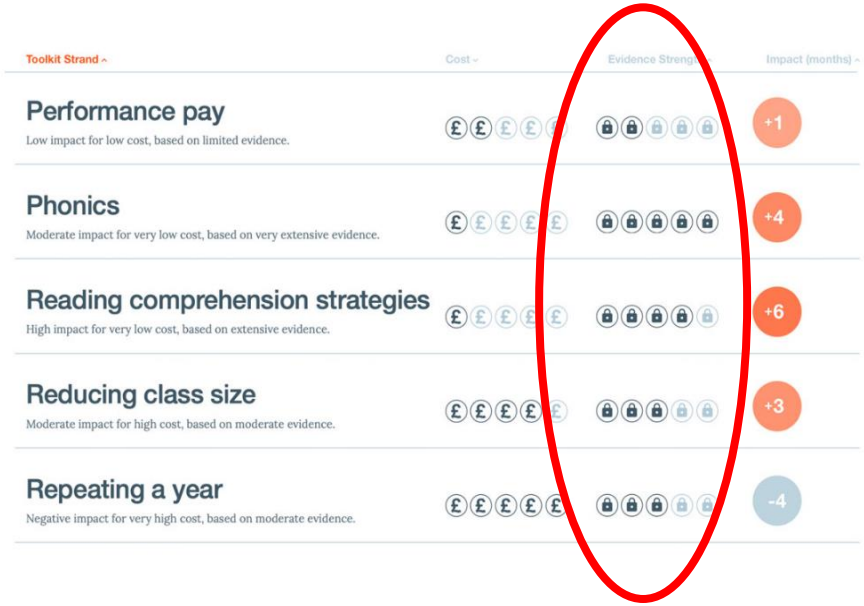
- May be uncertain about the assumptions or quality and relevance of the evidence underpinning a risk evaluation
  - This is referred to as indirect uncertainty*
- Graphics can be helpful to communicate the sensitivity of the risk evaluation to settings or assumptions



*van der Bles et al. *Royal Society Open Science* 2019; 6:181870

Uncertainty about risk evaluations (2)

- Quality and relevance of underlying evidence is usually evaluated qualitatively
  - Several fields have proposed categorical scales for communicating this uncertainty



Source: van der Bles et al. *Royal Society Open Science* 2019; 6:181870; Education Endowment Foundation (Link)

Conclusions

- Different Bayesian metrics may quantify different sources of uncertainty:
  - ✓ Which metric to use depends on the question being asked
- Quantitative summaries need to be transparently communicated, including the limitations of the evidence and robustness to assumptions
- If direct data are unavailable for a QoI, expert elicitation is an attractive solution, but requires a structured process and thorough preparation
- Feedback from the experts: they find the evidence dossier a helpful resource in itself and appreciated the rigorous process and quality of the discussions

Acknowledgements

- Novartis PoS team: Steffen Ballerstedt, Björn Bornkamp, Giovanni Della Cioppa, Björn Holzhauer, Joseph Kahn, Markus Lange, Wen-Lin Luo, Pritibha Singh

Thank you

NOVARTIS | Reimagining Medicine