Consilium

Salmonson & Hemmings

BAYESIAN APPROACHES IN DRUG DEVELOPMENT

Rob Hemmings, July 2024

The Cancer Drug Development Forum

DISCLAIMER

Consultant to industry;

C-Path EU Board and Scientific Advisory Committee

I am not a Bayesian. I am not a Frequentist. Hopefully, a balanced introduction to the topic including attractions, challenges and opportunities.

Lots of material borrowed from brilliant people, including Stephen Ruberg, Analytix Thinking.

A starting point for further reading is "**Application of Bayesian approaches in drug development: starting a virtuous cycle**", Ruberg et al, Nature Reviews Drug Discovery volume 22, pages 235–250 (2023)

FLOW

1. What have we been doing?
2. What is Bayesian statistics and how is it different?
3. Available regulatory guidance
4. Uses
5. Summary

WHAT HAVE WE BEEN DOING?

- Frequentist statistics
  - P-values / Confidence intervals etc.

- They make statements about data, e.g., **P (data | belief)**
- These are based on “long-run properties”

- Start with disbelief (null hypothesis), but might change my belief about the ”truth” (alternative hypothesis) if the data are sufficiently extreme

WHAT IS THIS NOVEL APPROACH?

- Not so novel, perhaps…

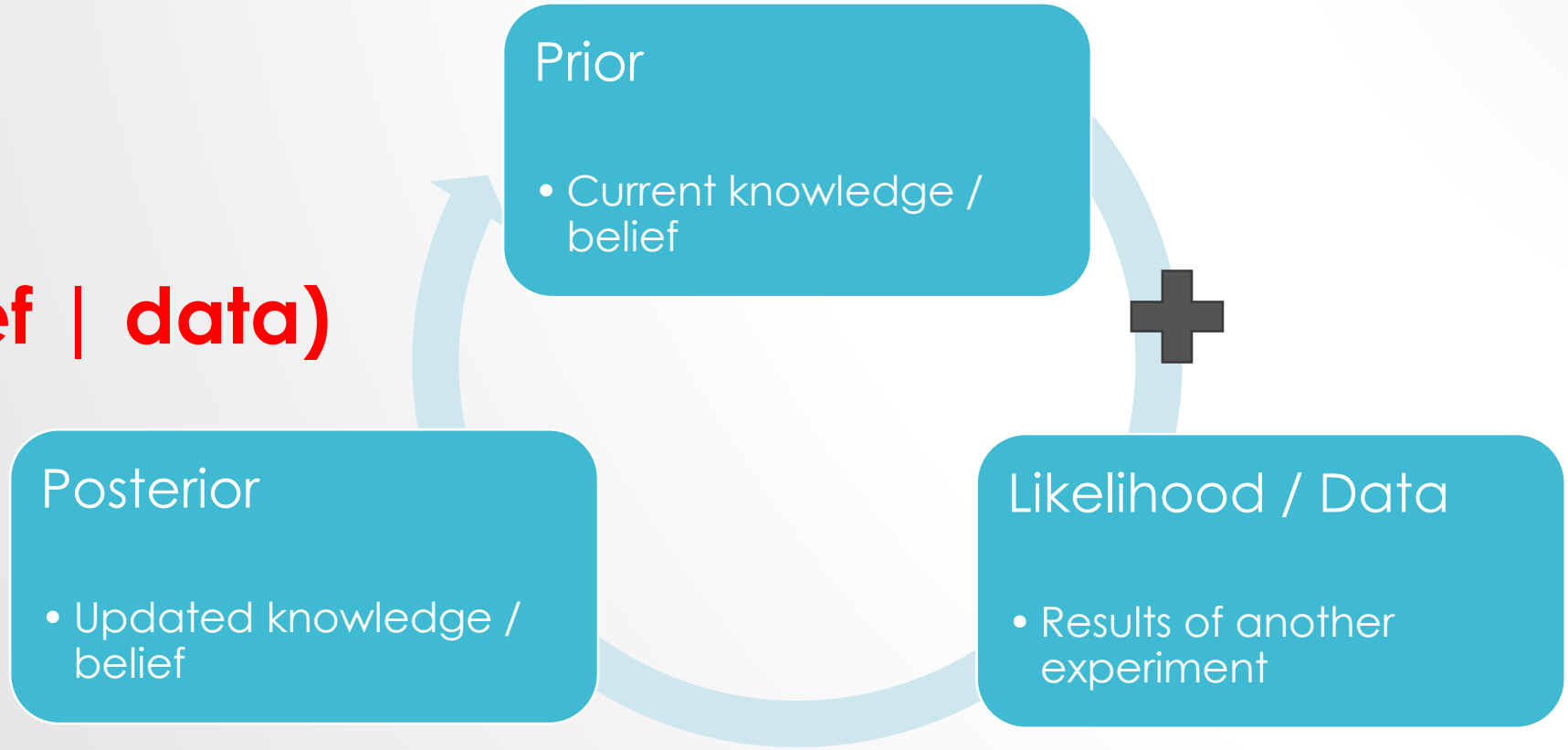
- Rev. Thomas Bayes, in response to an edition of "The Doctrine of Chances", said in 1763:
- *"The only thing I have is what I observe in natural phenomena. I must use the data I observe and infer what State of Nature (i.e. hypothesis) is most likely to be true."*
- **P (belief | data)**

WHAT IS THIS NOVEL APPROACH?

- A P-value does not express the probability that the drug is not effective
- Recall the Prosecutor's Fallacy:
  - Probability of innocence given the evidence is wrongly assumed to equal an infinitesimally small probability that that evidence would occur if the defendant was innocent.
  - P (innocence | evidence) ≠ P (evidence | innocence)
- Diagnostics as a medical example:
  - A false positive rate of 1 in 1000 [P(positive test | no disease)], must not be interpreted as P(no disease | positive test).

WHAT IS THIS NOVEL APPROACH?

P (belief | data)



WHAT IS THIS NOVEL APPROACH? THE PRIOR

Prior

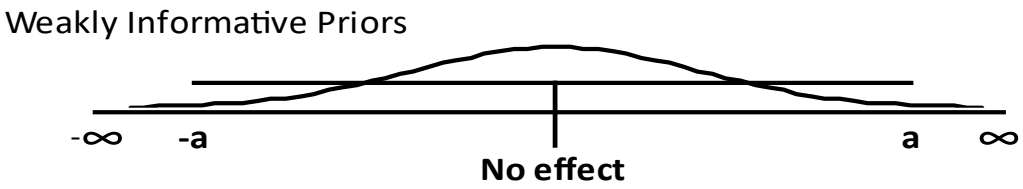
Current knowledge / belief

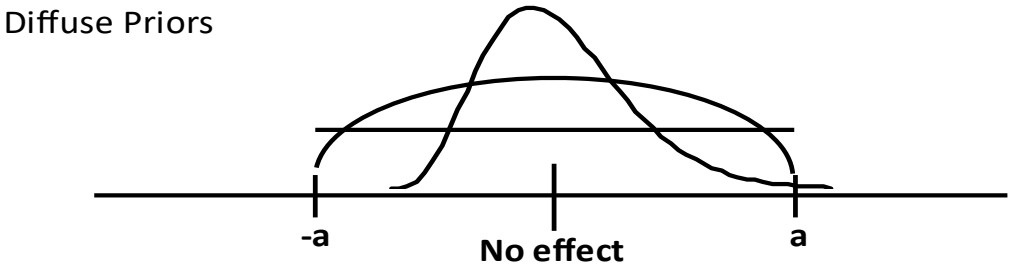
- An expression of what is known and what is not, including expression of uncertainties, based on
  - Any and all relevant and available data
  - Opinion

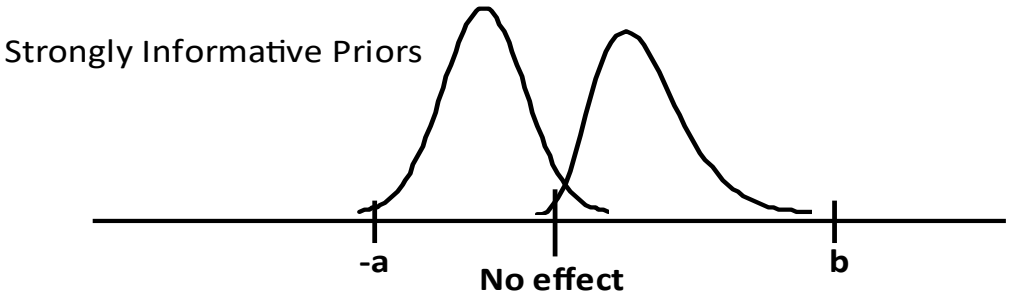
WHAT IS THIS NOVEL APPROACH? THE PRIOR

Prior

Current knowledge / belief



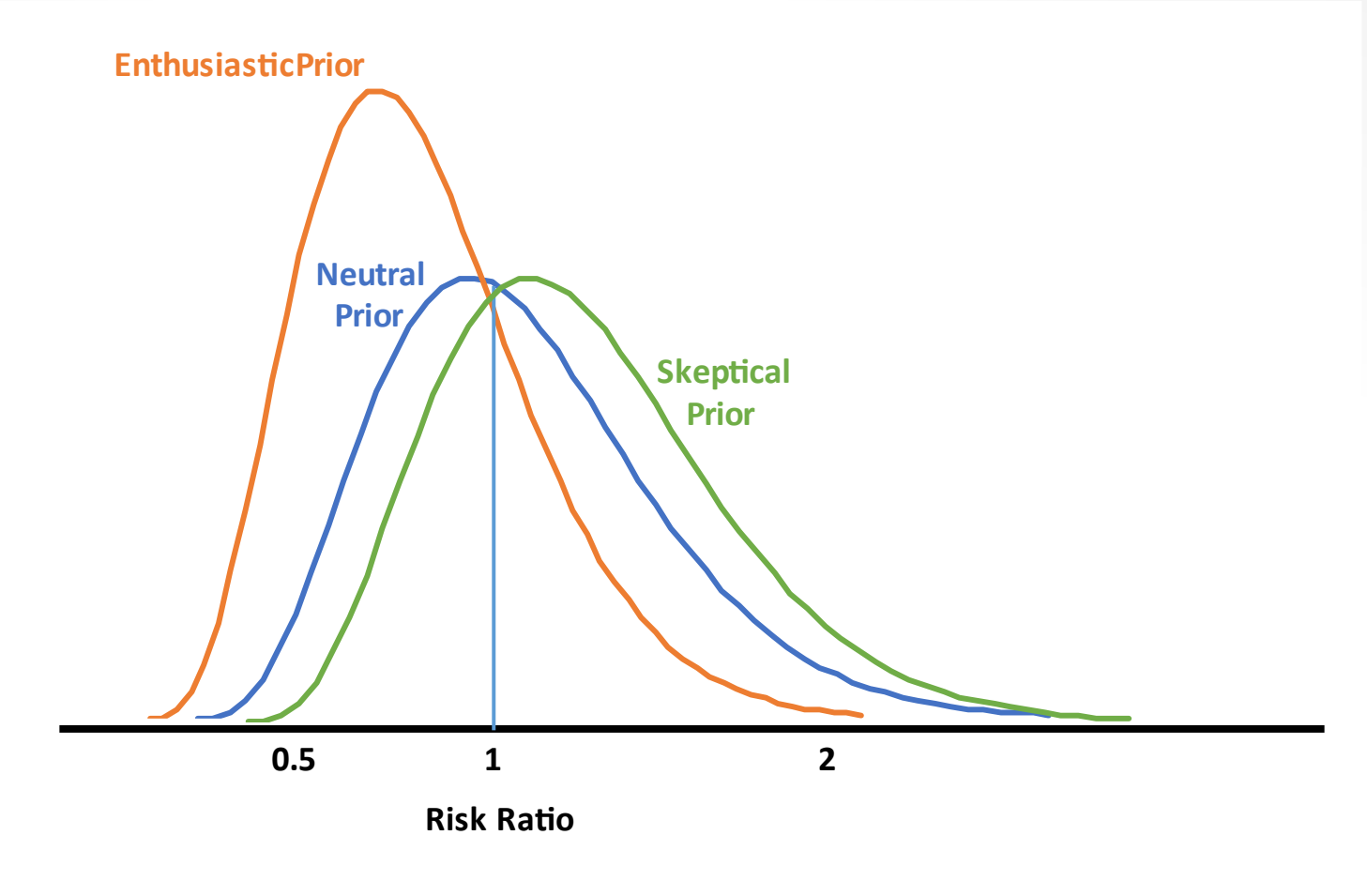




WHAT IS THIS NOVEL APPROACH? THE PRIOR

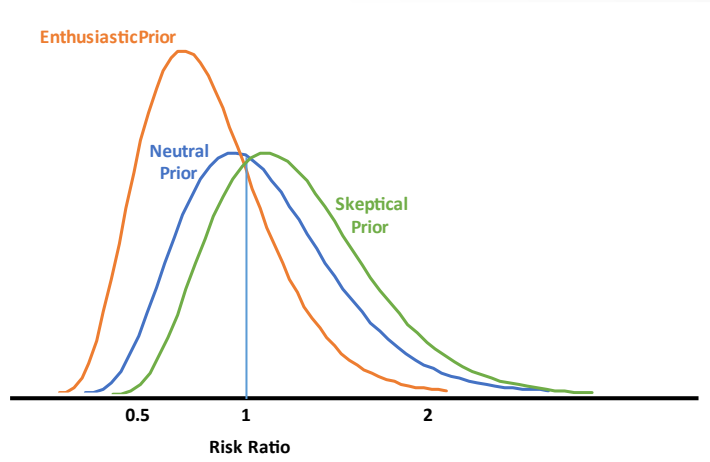
Prior

Current knowledge / belief



FIRST KEY DISCUSSION - WHO PICKS THE PRIOR, AND HOW?

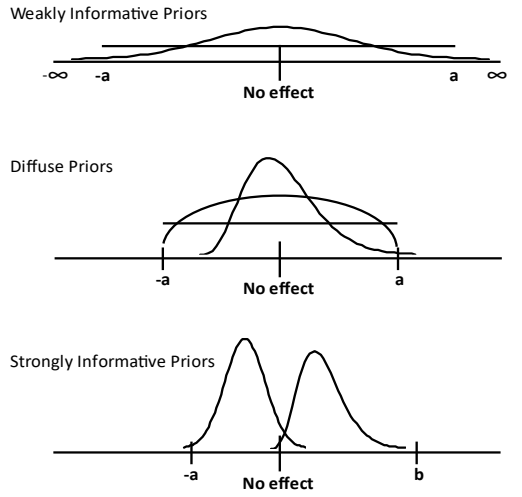
- Selecting a Prior:
  - Who can argue against transparently documenting current knowledge and uncertainties (and isn't it better to write down in advance than using for interpretation after the fact)?
  - But who gets to decide on the Prior, and how?



- Shape / location?
- Sponsor / Regulator / Both?
  - Sponsor too optimistic?
  - Regulator too cautious?
  - Agreement between parties operationally challenging? Agreement between individuals might be impossible!
- Let the data pick the Prior (at least to some extent)?

(OVER-)SIMPLIFIED EXAMPLE

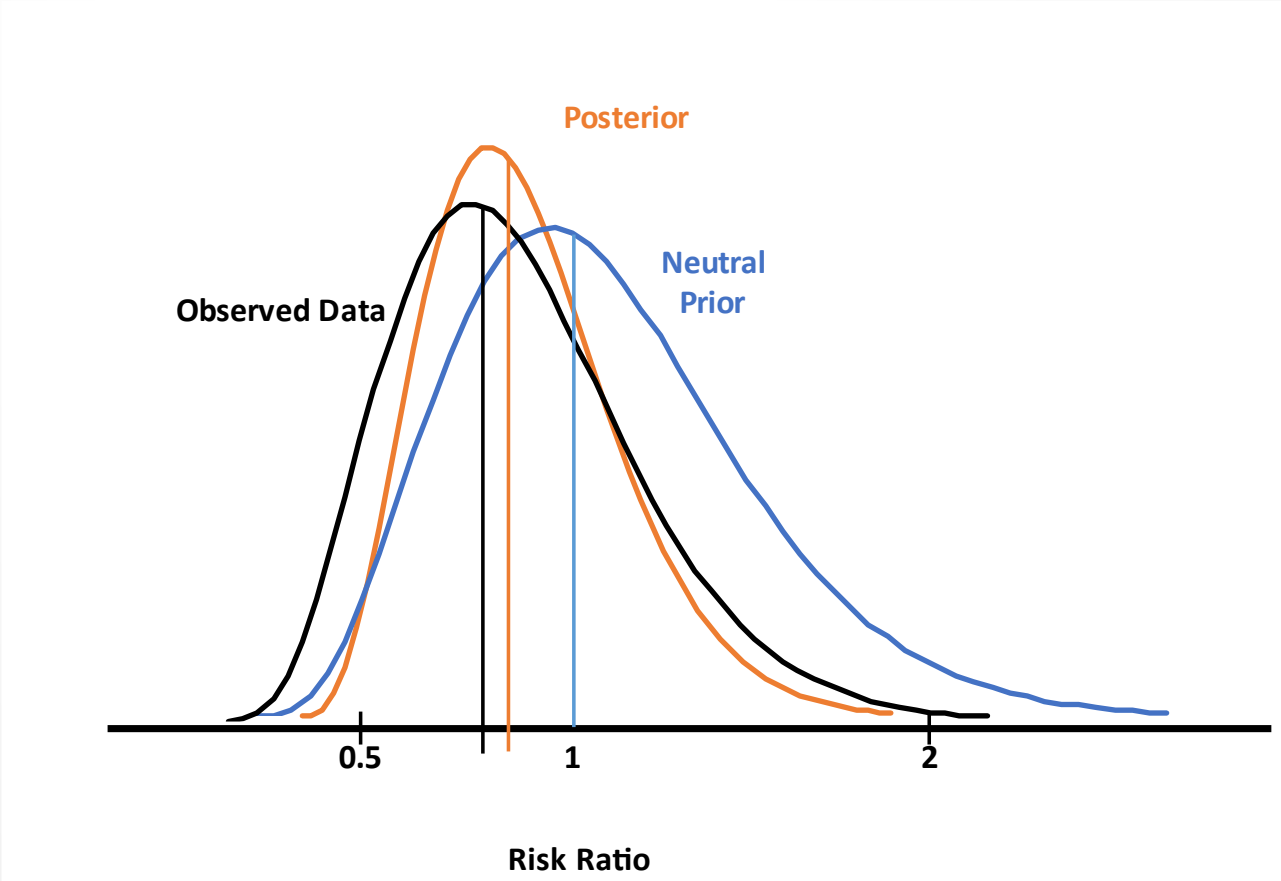
- Drug X has been developed in adults for a condition that also impacts children.
- Efficacy has been "established" in adults, estimated as X with 95% CI (Y, Z).
- Disease in children is very similar, though various factors (e.g., maturation, backbone treatment) might impact Drug X's effect size
- What is known / What can be predicted about the drug effect in children.
  - Is X a good estimate, or too optimistic? Take Y?
  - Reflecting other uncertainties by shape and location of prior



WHAT IS THIS NOVEL APPROACH?

Posterior

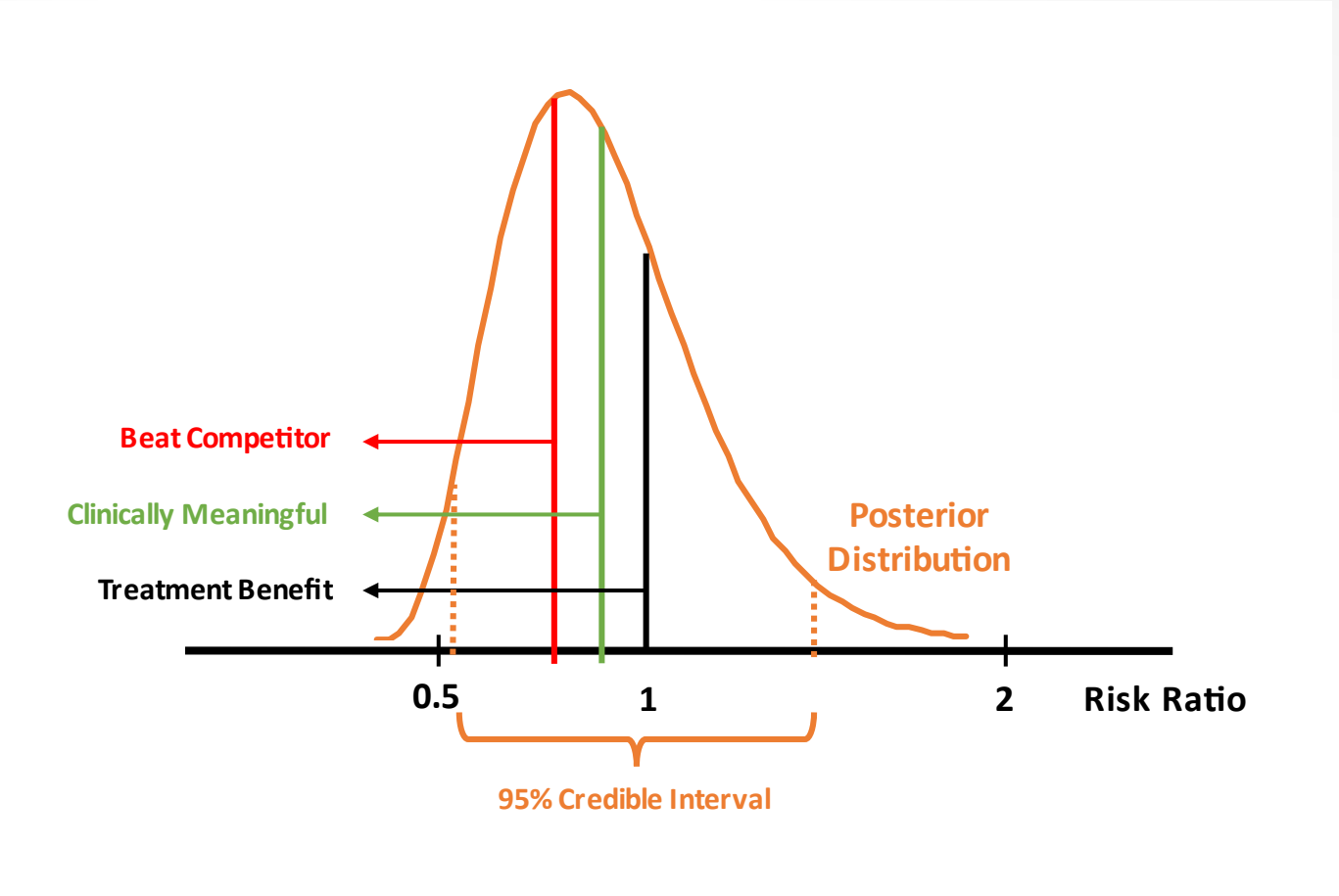
Updated knowledge / belief



WHAT IS THIS NOVEL APPROACH?

Posterior

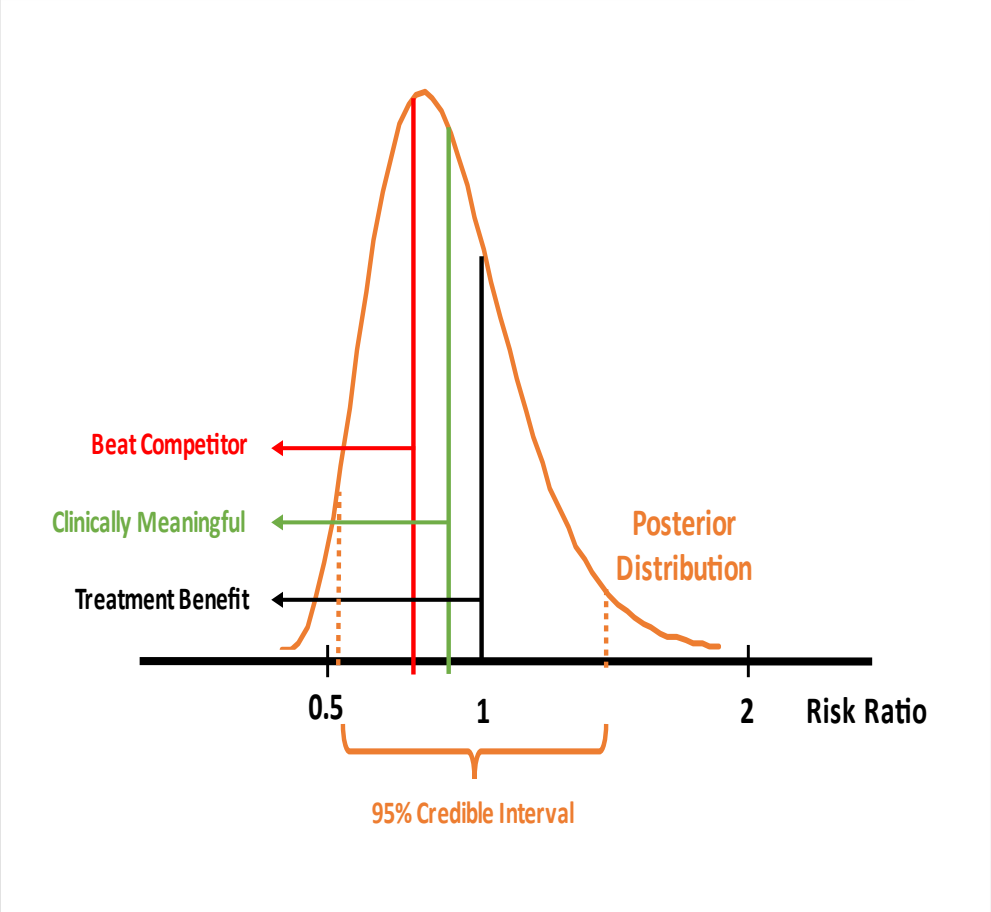
Updated knowledge / belief



WHAT IS THIS NOVEL APPROACH?

- Frequentist reporting gives:
- Effect size, 95% confidence interval
- P-value reflecting a statistical test, expressing how extreme data vs null hypothesis

WHAT IS THIS NOVEL APPROACH?



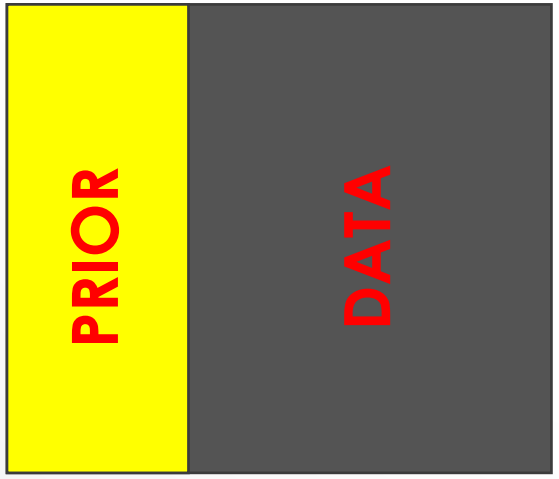
- Bayesian reporting gives
- Effect size, 95% **credible interval**
- No test. Various probability statements, e.g., P (efficacy > 0), P (efficacy > X)

(OVER-)SIMPLIFIED EXAMPLE

- Strongly informative prior that there will be some effect.
  - ORR 40% in adults, science and data might support P (ORR in children > 20%) > 80%; P (…> 40%) is low, P (<10%) is low.
- Data estimates ORR in children 9/30 = 30%, 95% CI (15%, 50%)
- Update Prior with Data to give posterior for inferences:
- P (drug effective in children | data)
- P (drug effect > 20% in children | data) > 95%

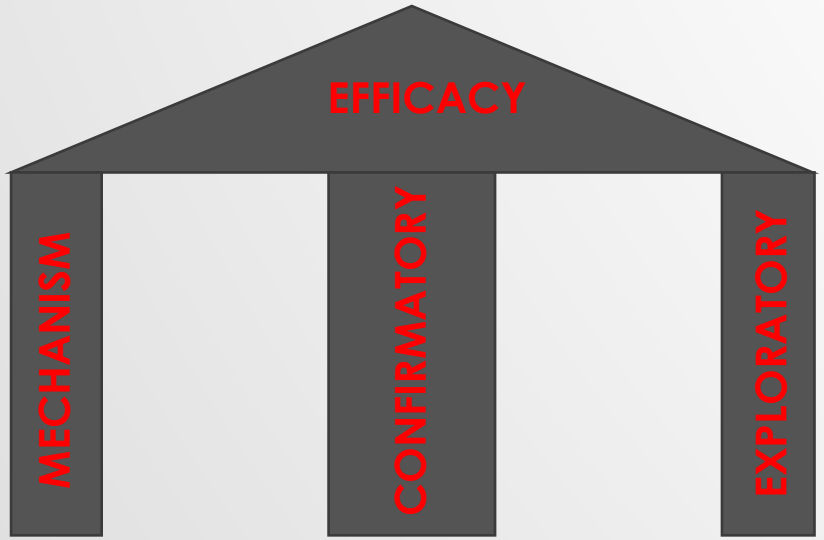
SECOND KEY DISCUSSION – SHOULD WE REDUCE THE AMOUNT OF EVIDENCE?

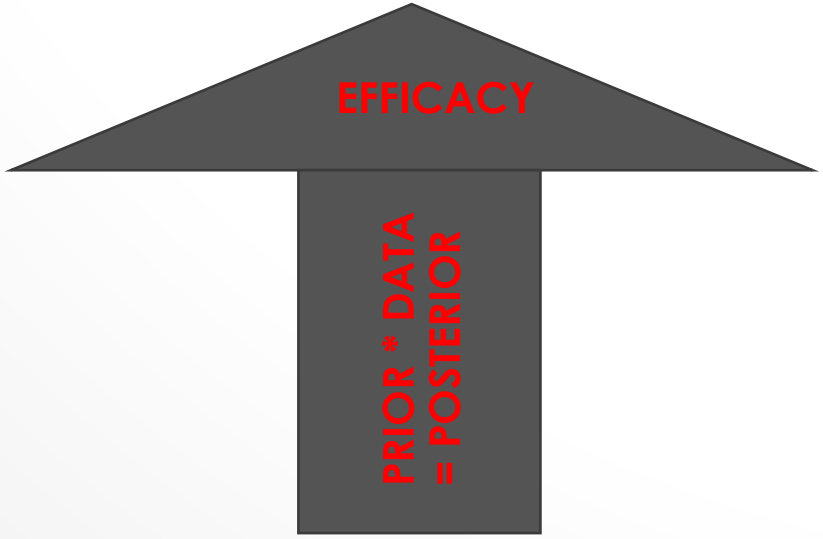
- How many points are in a game of tennis?
- How strong does evidence have to be to establish efficacy?
  - 2 adequate and well-controlled studies, except when…
- If available exploratory data are incorporated informative Priors – can confirmatory trials be smaller to reach the same standards of evidence?



SECOND KEY DISCUSSION – SHOULD WE REDUCE THE AMOUNT OF EVIDENCE?

- How many points are in a game of tennis?
- How strong does evidence have to be to establish efficacy?
  - 2 adequate and well-controlled studies, except when…
- If available exploratory data are incorporated informative priors – should then confirmatory trials be smaller?





SECOND KEY DISCUSSION – REDUCE THE AMOUNT OF CONFIRMATORY EVIDENCE GENERATED?

- Confirmatory studies planned to show $P<0.05$
- **One key strength** of the **Frequentist** framework **as currently applied** is to give clear standards and expectations.
- We run one or more confirmatory experiments that either succeed or fail with well-understood consequences.

SECOND KEY DISCUSSION – REDUCE THE AMOUNT OF CONFIRMATORY EVIDENCE GENERATED?

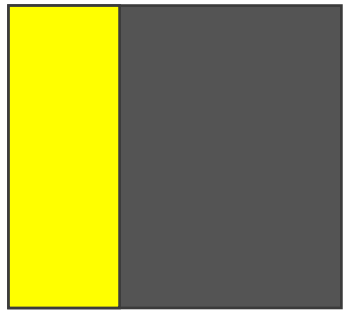
- Confirmatory studies planned to show P<0.05
- **One key strength** of the **Frequentist** framework **as currently applied** is to give clear standards and expectations.
- We run one or more confirmatory experiments that either succeed or fail with well-understood consequences.

- **One key weakness*** is to dichotomise results
- If p=0.055 is there no information?
- Two phase 3 studies, p=0.02 and p=0.06. Sufficient evidence of efficacy?
- Two phase 3 programmes with one completed study (P=0.4 and p=0.06, respectively). Do both require second pivotal study with same design?

Note: the weakness* here is more a problem of application that of Frequentist statistics per se.

SECOND KEY DISCUSSION – REDUCE THE AMOUNT OF CONFIRMATORY EVIDENCE GENERATED?

- Use of informative priors with equivalent success criteria would result in less evidence being generated overall
  - For some, this can be a disadvantage
  - For others, this is precisely the strength.
  - Context is important.
- Is it the same strength of evidence?



- Requires changes in mindset, e.g., must avoid double-counting the evidence

REGULATORY GUIDANCE

- ICH E9
  - *Because the predominant approaches to the design and analysis of clinical trials have been based on frequentist statistical methods, the guidance largely refers to the use of frequentist methods … This should not be taken to imply that other approaches are not appropriate: the use of Bayesian and other approaches may be considered* ***when the reasons for their use are clear*** *and* ***when the resulting conclusions are sufficiently robust.***
- ICH E11A
  - *If data external to the trial are incorporated into the analysis, the reporting should explicitly describe this and discuss how and when these data were originally generated and where they were reported,* ***along with a justification as to why they are considered to be appropriate to include***.
- ICH E20 (under development)

REGULATORY GUIDANCE / INITIATIVES

- FDA
- Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials
- Article / Podcast _ https://www.fda.gov/drugs/cder-small-business-industry-assistance-sbia/using-bayesian-statistical-approaches-advance-our-ability-evaluate-drug-products
  - By the end of the second quarter of FY 2024, the FDA expects to convene a public workshop to discuss aspects of complex adaptive, Bayesian, and other novel clinical trial designs. By the end of FY 2025, FDA also anticipates publishing draft guidance on the use of Bayesian methodology in clinical trials of drugs and biologics.

REGULATORY GUIDANCE / INITIATIVES

- FDA
- CDER Center for Clinical Trial Innovation (C3TI), Bayesian Supplemental Analysis (BSA) Demonstration Project

https://www.fda.gov/about-fda/cder-center-clinical-trial-innovation-c3ti/bayesian-supplemental-analysis-bsa-demonstration-project

- C3TI will partner with sponsors to integrate Bayesian analysis in parallel to frequentist analysis during their trial, providing an opportunity for both CDER and sponsors to learn new methods without impacting review criteria

REGULATORY GUIDANCE / INITIATIVES

- FDA
- Regulatory Science Impact Story

Impact Story: Using innovative statistical approaches to provide the most reliable treatment outcomes information to patients and clinicians | FDA

Using Bayesian hierarchical models, CDER statisticians are improving our understanding of how drugs affect different groups of patients.

REGULATORY GUIDANCE / INITIATIVES

- EMA / Europe
- Complex Clinical Trials: "*With frequentist approaches, adjustment for multiple null-hypothesis significance testing (type I error control) is a central consideration in regulatory submissions… When using a Bayesian methodology, it is of importance that the methodology allows for an evaluation of corresponding issues….*"
- EMA Methodology Working Party – workshop on Bayesian statistics Q3, 2024, Reflection Paper on Bayesian methods.

COMMON USE CASES

- Interim decision making
- Dose-response modelling
- Borrowing external data
- Extrapolation
- Subgroup analysis
- Meta-analysis

- **But can't Frequentist approaches do these too?**

BETTER FOR POST-APPROVAL STUDIES?

- e.g., for Conditional MA, an MAH is required to complete a comprehensive dataset and demonstrate that benefit-risk remains positive
- Some scenarios might be well suited to Bayesian, updating knowledge (estimates, probabilities) of efficacy and safety with post-approval data – and hence updating B-R assessment.
- Framework arguably better suited than whether or not a confirmatory trial is successful, or fails.
- For example:
  - Approval based on single arm trial showing ORR superior to SOC.
  - RCT shows effect vs SOC on OS p=0.06 – which probably strengthens evidence for efficacy
  - Remove from market, Leave on market, Ask for more data?
- Other post-approval work – variations, line extensions?

I DON'T HAVE TO WORRY ABOUT MULTIPLE TESTING?

- If there is no hypothesis testing (no hypothesis, no test), there is:
  - no concept of the test result being an error (type I and type II error)
  - no problem with multiple testing

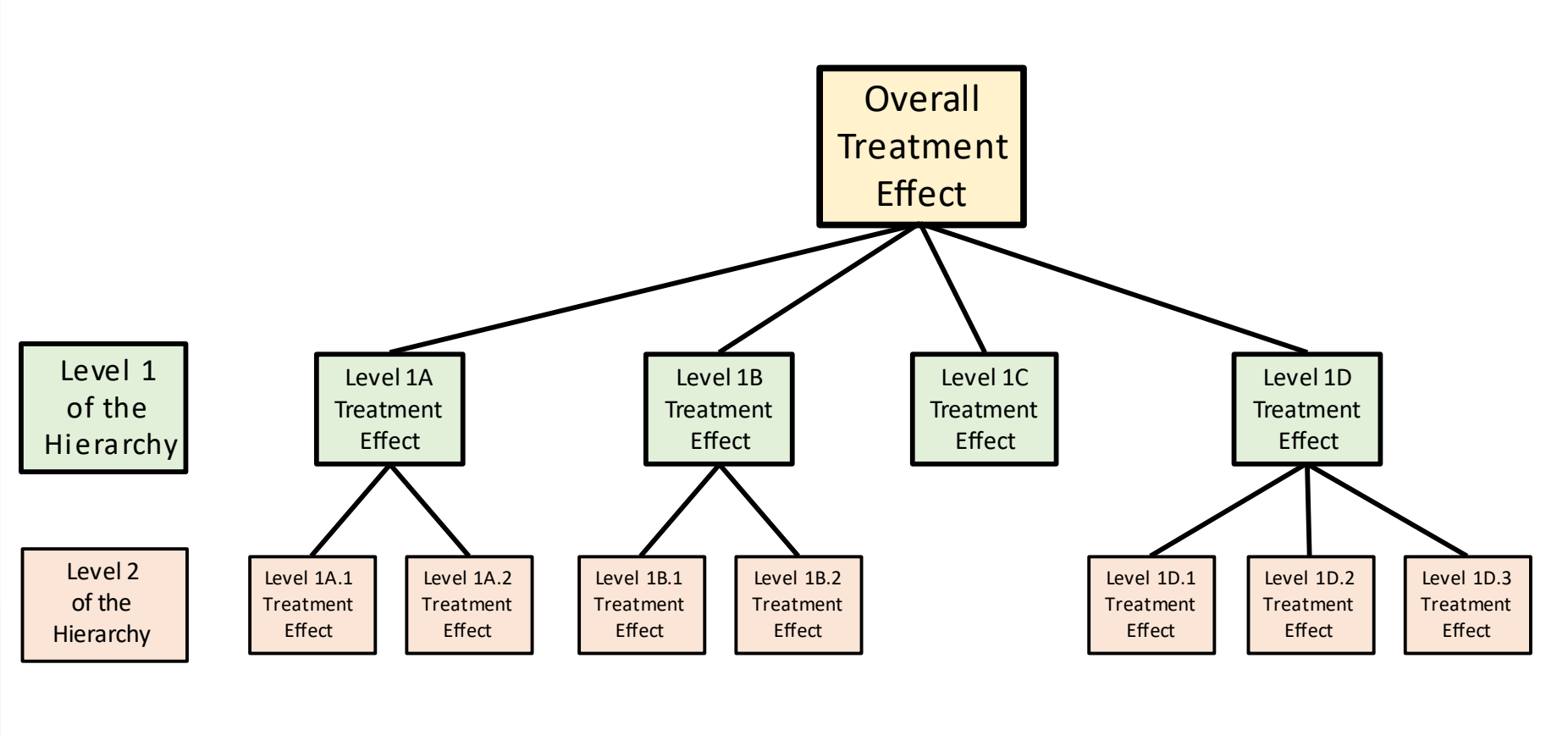
- But plays of chance in data can still be misleading, right?
- True, though data are the Bayesian's truth

- Emerging practice has "a third way" with Bayesian statistics incorporating some Frequentist principles and standards.

INVESTIGATION OF SUBGROUPS

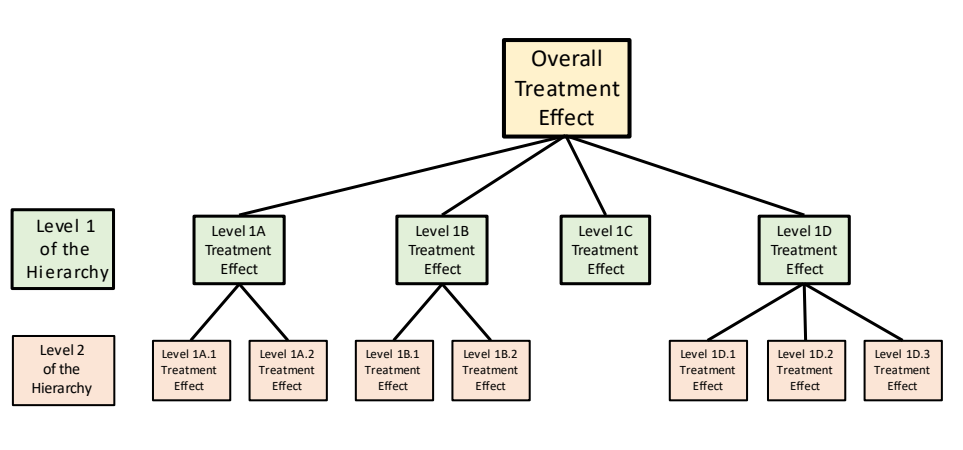
- A notoriously difficult problem
- Does a subgroup result reflect a true effect modification, or a play of chance?
- EMA guidance tells us to look at all available evidence to help determine credibility:
  - Biological plausibility
  - Replication
- Do it in advance and you are Bayesian!
  - Though the "Prior" might change between designing a trial and its analysis.

BAYESIAN HIERARCHICAL MODELS (BHM)



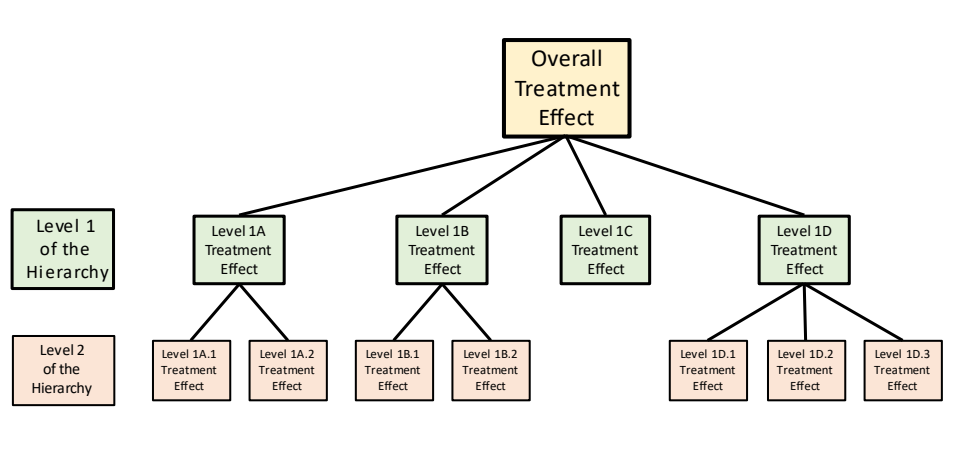
INVESTIGATION OF SUBGROUPS

- Pharmacology directed at particular tumour biology
- Different a priori belief on efficacy by biomarker status
  - + / -
  - Continuum



INVESTIGATION OF SUBGROUPS

- Pharmacology directed at particular tumour biology
- Different a priori belief on efficacy by biomarker status
  - + / -
  - Continuum



- Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE)
- Used a BHM to examine the effectiveness of targeted therapies for NSCLC according to patients' biomarkers status
- Adaptive randomization towards likely most effective treatment

CONCLUSIONS

- Attractive to make direct statements and inferences about drug effects
  - P (drug effect | data)
- Specification of Prior is a strength, but an operational challenge
- Changes to standards for approval should be made consciously and based on policy, not based on choice of statistical approach.
- Better use of data, and patients?
  - Counts failed studies explicitly; intuitive way to accumulate evidence and to bridge to related populations.
- Methods, Mindset or Both?
- Renewed momentum at regulatory agencies, so that Bayesian and Frequentist approaches can exist in harmony.

LIGHT READING

- Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology, Brannath et al
  - STATISTICS IN MEDICINE, Statist. Med. 2009; 28:1445–1463. DOI: 10.1002/sim.3559
- Critical aspects of the Bayesian approach to phase I cancer trials, Neuenschwander et al
  - STATISTICS IN MEDICINE, Statist. Med. 2008; 27:2420–2439. DOI: 10.1002/sim.3230
- A proof of concept phase II non-inferiority criterion, Neuenschwander et al
  - STATISTICS IN MEDICINE, Statist. Med. 2011. DOI: 10.1002/sim.3997
- OPTIM-ARTS—An Adaptive Phase II Open Platform Trial Design With Application to a Metastatic Melanoma Study, Poon et al
  - STATISTICS IN BIOPHARMACEUTICAL RESEARCH 2020, VOL. 00,NO. 0, 1–12
  - https://doi.org/10.1080/19466315.2020.1749722