Bayesian Borrowing Approaches for Rare Disease Settings: *Applications to Basket Trials and Pediatric Studies*

Emma Mackay
Scientific Advisor - Statistics
Inka Health

Presentation to the IMPaCT Lab
June 26, 2024

Disclosures

- Advisor at Inka Health, an HEOR analytics consulting company
- No conflicts of interest related to this presentation
- Opinions expressed are my own

Background (1 / 2)

- Rare disease settings provide a particular challenge for evaluating the efficacy of new drugs:
  - Can be difficult to recruit enough patients to run a conventional well-powered randomized controlled trial (RCT)
- Growing use of unconventional methods, e.g.:
  - Borrowing from historical trials to augment small concurrent control arms
  - Fully-externally controlled trials
  - Newer trial designs like basket trials (recruit patients with multiple disease subtypes who share a common druggable target–e.g. cancer mutation)

Background (2 / 2)

- Small sample size challenges also arise in many pediatric trials–can we borrow information from similar trials in adult populations?
- Growing focus on 'precision medicine' in drug development is resulting in increasingly narrowly defined patient populations (e.g. patients with a specific cancerous genetic mutation)
- Growing receptiveness to the use of Bayesian borrowing methods and synthetic or hybrid control arms where conventional trials are impractical/infeasible[1,2,3]

[1] US FDA. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Guidance for Industry and FDA Staff. 2010.

[2] US FDA. Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products: Guidance for Industry. 2020.

[3] US FDA. Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry. 2019.

Challenge

- Rationale of conducting an RCT:
  - Randomize patients to experimental or control treatment so that only treatment received differs systematically between treatment arms $\rightarrow$ Allows us to infer causal effect of treatment assignment on outcomes
  - Can design trial to achieve type-I and type-II error operating characteristics
- Where it is difficult or infeasible to design a well-powered RCT, we can borrow information from data sources external to the trial, however this:
  - Introduces risk of bias as external data sources are not subject to randomization (differences in patient populations other than treatment received risk confounding treatment effect estimates)
  - Makes it difficult to achieve target type-I and type-II error operating characteristics

How to Approach Borrowing from External Data Sources (1 / 2)

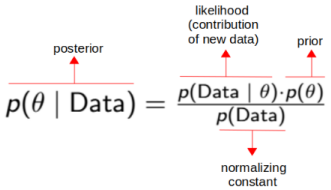
- Care needed in identification of external data source (e.g. historical control arm)
  - Similar patient population (e.g. similar eligibility criteria for historical control, and similar baseline characteristics)
- Appropriate application of methods for adjusting for observed differences in potential confounders
  - Could control for confounders using parametric model (outcome regression)
  - Could construct synthetic control arm with similar baseline characteristics using matching or inverse probability of treatment weighting (IPTW) methods

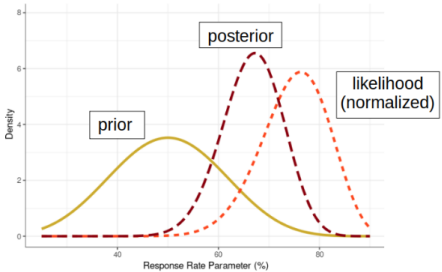
How to Approach Borrowing from External Data Sources (2 / 2)

- But what about residual heterogeneity across populations/data sources?
  - Can down-weight the information contribution of the external data when there is evidence of meaningful differences in (post-adjustment) outcomes between data sources $\rightarrow$ "dynamic borrowing'
    - E.g. borrow less information from an external control to supplement a small concurrent control arm in an RCT when outcomes differ meaningfully
    - Or when estimating, say, the disease control rate for the control arm for a fully-externally controlled trial, heterogeneity in outcomes across external data sources should be reflected in the precision of estimate of the disease control rate
  - Focus of this presentation

Why Use a Bayesian Approach for Rare Diseases?

- Provides a principled framework for incorporating external information:
  - Start with our prior (which can be informed by external data)
  - Update our beliefs after observing new data
- Conducive to sequential 'Bayesian updating'
- Posterior inference allows us to quantify the amount of evidence in favour of a conclusion and allows for more nuanced decision rules
- See Mackay & Springford (2023) for additional discussion[4]

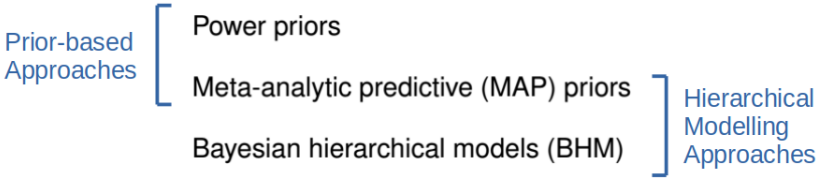




[4] Mackay EK, Springford A. Evaluating treatments in rare indications warrants a Bayesian approach. Frontiers in Pharmacology. 2023.

Methods Roadmap

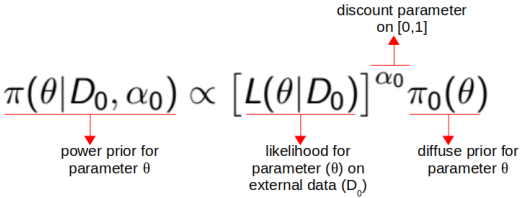
- Will cover three approaches to Bayesian borrowing:



- Typically external data is only available for standard of care / control treatments
  - Could use external data to form a prior for a parameter in a concurrent control arm or as a stand-in for a non-existent control arm
- Without loss of generality, examples will focus on binary response endpoints (will use external data to inform estimates of the control treatment response rate)

Power Priors

- Power priors[5,6]
  - Down-weight the external data by means of a discount parameter, $\alpha_0 \in [0, 1]$



  - As $\alpha_0 \to 0$ we ignore the external data (no pooling)
  - As $\alpha_0 \to 1$ we give it full weight (full pooling)
  - How to choose α_0?: 'tipping point' approach[7], target effective sample size for borrowing[8], dynamic borrowing based on consistency between data sources[5,6]

[5] Ibrahim JG, Chen MH. Power prior distributions for regression models. Stat. Sci. 2000.

[6] Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: Theory and applications. Statistics Med. 2015.

[7] Best N, ... Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing. Pharm. Stat. 2021.

[8] Richeldi L, ..., Maher TM. Trial of a preferential phosphodiesterase 4B inhibitor for idiopathic pulmonary fibrosis. NEJM. 2022.

- Demonstration of dynamic borrowing for binary response endpoint using a beta-binomial model with normalized power prior[9]
  - Observed response rate is $p = y/n$ in current data and $p_0 = y_0/n_0$ in historical data with $n = n_0 = 100$ patients
  - We use a Beta$(1, 1)$ prior for θ and use either a Beta$(1, 1)$ prior or fixed value for α_0 and report posterior means and 95% CrIs for the response rate for two scenarios: $(p, p_0) = (0.4, 0.2)$ and $p = p_0 = 0.4$

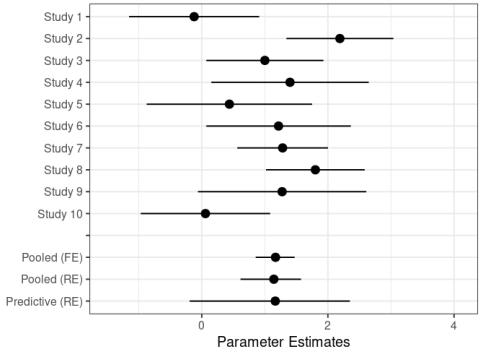
Scenario	**Prior for α_0**	**Response Rate (θ)**
Inconsistent ($p = 0.4, p_0 = 0.2$)	Beta$(1, 1)$	0.368 [0.275, 0.468]
Inconsistent ($p = 0.4, p_0 = 0.2$)	Full pooling ($\alpha_0 = 1$)	0.301 [0.241, 0.367]
Consistent ($p = p_0 = 0.4$)	Beta$(1, 1)$	0.401 [0.325, 0.480]
Consistent ($p = p_0 = 0.4$)	No pooling ($\alpha_0 = 0$)	0.401 [0.309, 0.498]

- The normalized power prior is able to improve precision and also partially mitigates bias when data sources are incompatible
- However, we also show that the amount of borrowing can be sensitive to choice of prior for α_0

[9] Mackay EK, Springford A. Impact of Hyperprior Choice for Bayesian Dynamic Borrowing via a Normalized Power Prior. JSM Proceedings. Alexandria, VA: American Statistical Association. 2023 Oct. https://doi.org/10.5281/zenodo.10001953

Meta-analytic Predictive (MAP) Priors (1 / 2)

- Meta-analysis approach to construct an informative prior (e.g. for the average response under a SoC treatment)[10]
- Since response may vary across trial populations, we want our prior to incorporate both within-trial and between-trial uncertainty
- Idea is to conduct a random-effects meta-analysis and use the posterior predictive distribution (predicted SoC response in a new trial) as our prior



[10] Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. Clinical Trials. 2010.

Meta-analytic Predictive (MAP) Priors (2 / 2)

Basic Setup

- Have control arm response data $D_h = (Y_h, n_h)$ for $h = 1, ..., H$ historical trials
- Assume that $\theta^*, \theta_1, ...\theta_H \sim N(\gamma, \sigma^2)$ where θ^* is a the log-OR in a hypothetical new trial
- Can compute the MAP prior for parameter θ in our concurrent control arm as the posterior predictive distribution $[\theta^*|D_1, ..., D_H]$

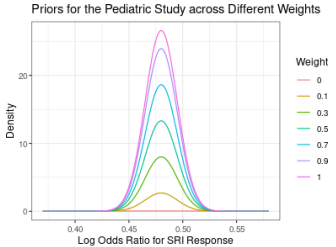
- Low cross-trial heterogeneity $\rightarrow$ greater effective sample size borrowed
- Can be made more robust to prior-data conflict (Robust MAP) by using a weighted mixture between the MAP prior (f_{MAP}) and a vague prior (f_V)[11]
  - $w \cdot f_{MAP}(\theta) + (1 - w) f_V(\theta)$

[11] Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014.

- Post-marketing pediatric study required by FDA to evaluate efficacy of belimumab in pediatric SLE patients for SLE responder index (SRI) endpoint
- Analysis to supplement limited pediatric trial population by borrowing from adult trials via robust MAP prior
- Informative prior for pediatric log odds ratio of SRI response (δ) was constructed using a robust MAP approach following a meta-analysis of two adult studies:

$$\delta \sim w \cdot N(0.48, 0.015^2) + (1 - w) \cdot N(0, 8.27^2)$$

- Assessed weight w on MAP component required to reach efficacy tipping point
- Concluded that amount of borrowing from adult trials to reach tipping point was acceptable, leading to approval



[12] US FDA. BLA 125370/s-064 and BLA 761043/s-007 multi-disciplinary review and evaluation benlysta (belimumab) for intravenous infusion in children 5 to 17 years of age with SLE. 2021. https://www.fda.gov/media/127912/download

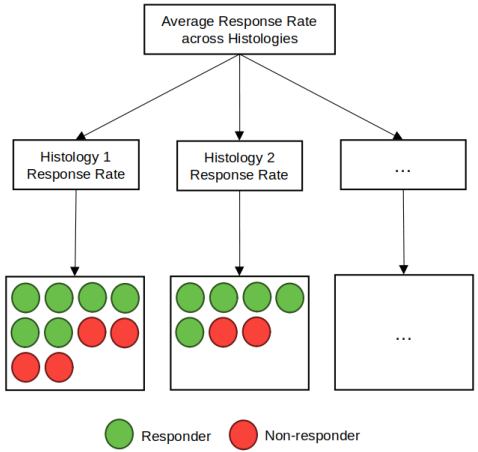
Basket Trials

- Basket trials recruit patients with multiple disease subtypes (e.g. lung cancer, breast cancer) as long as they are positive for the mutation/biomarker that the experimental drug targets
  - These trials are usual single-arm (lack a control arm)
  - Typically done to increase sample sizes where the targeted mutation/biomarker is very rare

- To pool the data or not to pool?
  - Complete pooling ignores potential heterogeneity in response across tumour types $\rightarrow$ results may not generalize
  - No-pooling $\rightarrow$ back to problem of small sample sizes

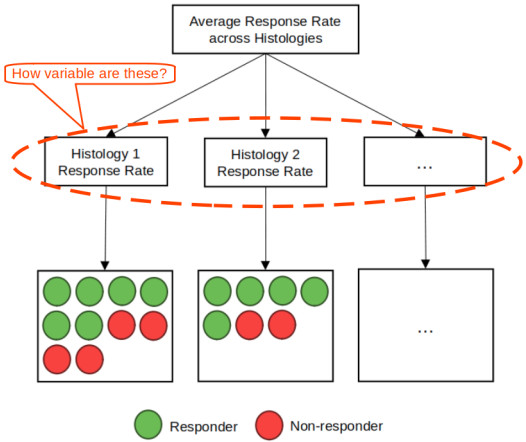
- Bayesian hiearchical models (BHM) allow for partial pooling–a middle-ground between the extremes of complete pooling and no pooling
- Allows response rates to differ across histologies but assumes they are related (“exchangeability assumption”)
- Amount of partial pooling (or “borrowing”) across histologies depends on degree of heterogeneity in responses across histologies

Bayesian Hierarchical Model



- Heterogeneity parameter is estimated based on the trial data
- High heterogeneity $\rightarrow$ little borrowing
- Low heterogeneity $\rightarrow$ more borrowing

Bayesian Hierarchical Model



Example from Murphy et al. (2021)[13]

- Analysis of a basket trial for larotrectinib in NTRK-fusion-positive solid tumours
- Model for histologies $k = 1, ..., K$:

$$r_k \sim \text{Binom}(n_k, p_k)$$
$$\text{logit}(p_k) = \theta_k$$
$$\theta_k \sim N(\mu, \sigma^2)$$

$$\text{priors: } \mu \sim N(-0.8473, 10)$$
$$\sigma \sim \text{Unif}(0, 5)$$

- Partial pooling yields histology-specific response rates that are shrunken towards the average–particularly in the case of tumour types with very few patients

Probabilities of Response for Each Histology

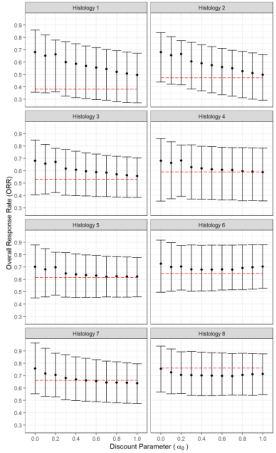
Histology	Observed Response	Estimated Response under BHM Mean (95% CrI)
Soft-tissue sarcoma	10/11 (90.9%)	88.1% (66.0% – 99.1%)
Salivary gland	10/12 (83.3%)	81.8% (58.0% – 96.8%)
IFS	7/7 (100%)	93.3% (70.5% – 100%)
Thyroid	5/5 (100%)	91.6% (63.0% – 100%)
Lung	3/4 (75.0%)	72.6% (30.4% – 97.8%)
Melanoma	2/4 (50.0%)	52.5% (12.4% – 89.4%)
Colon	1/4 (25.0%)	32.0% (2.6% – 75.5%)
GIST	3/3 (100%)	88.3% (49.3% – 100%)
Cholangiocarcinoma	0/2 (0%)	21.0% (0.0% – 75.7%)
Appendix	0/1 (0%)	30.0% (0.1% – 89.7%)
Breast	0/1 (0%)	30.0% (0.1% – 90.1%)
Pancreas	0/1 (0%)	29.8% (0.1% – 89.7%)

[13] Murphy P, Claxton L, Hodgson R, Glynn D, Beresford L, Walton M, Llewellyn A, Palmer S, Dias S. Exploring heterogeneity in histology-independent technologies and the implications for cost-effectiveness. Medical Decision Making. 2021.

Combined Approach Example from Mackay et al. (2023)[14]

- Uses a combined BHM and power prior to facilitate:
  1. Partial pooling of information across histologies under an exchangeability assumption, and
  2. Partial borrowing from adult basket trial data to supplement pediatric trial data using a power prior
- We demonstrate the approach using simulated data under a scenario where borrowing from the adult data is clinically appropriate
- Righthand figure shows how overall response rate (ORR) estimates change with increased borrowing weight on the adult data

Figure 1. Pediatric ORR Estimates (Posterior Median and 95% CrIs) for Various Borrowing Weights (Red Line Indicates True ORRs)



[14] Mackay E, Springford A, Heeg B, Arora P, Thorlund K. Combating Sample Scarcity: A Novel Bayesian Approach to Pediatric Basket Trials [Abstract]. Value in Health. 2023. https://doi.org/10.1016/j.jval.2023.09.2141. Poster available here.

Or... A Fully Hierarchical Approach from Mackay et al. (2023)[15]

- For cohort $j \in \{0, 1\}$ (pediatric or adult) and histology $k = 1, ..., K$ we model the number of responders r_{jk} out of n_{jk} patients at risk as follows:

$$
\begin{aligned}
r_{jk} &\sim \text{Binom}(n_{jk}, p_{jk}) \\
\text{logit}(p_{jk}) &= \mu + \gamma_k + (\eta_k + \delta) \cdot 1\{j = 1\} \\
\gamma_k &\sim \text{N}(0, \sigma_\gamma^2) \\
\eta_k &\sim \text{N}(0, \sigma_\eta^2)
\end{aligned}
$$

- where the γ_k's capture cross-histology heterogeneity in response, the η_k's capture heterogeneity in the relative adult vs. pediatric response across histologies, and δ allows for average response rates to be shifted between adult and pediatric populations
- Approach allows for partial borrowing of information from adult populations to supplement limited pediatric sample sizes

[15] Mackay E, Springford A, Heeg B, Arora P, Thorlund K. A Novel Information Borrowing Approach for Evaluating Response in Pediatric Basket Trials with Limited Sample Sizes [Abstract]. 2023. Presented at Bayes 2023 in Utrecht, NL. Slides available here.

- Need for an approach to perform indirect treatment comparisons between therapies trialled in basket trial settings for health technology assessment (HTA) purposes:
  - Increased uptake of basket trials for drugs targeting NTRK-fusions (larotrectinib, entrectinib, repotrectinib), BRAF V600 mutations (vemurafenib, dabrafenib + trametinib), dMMR/MSI-H tumours (pembrolizumab, other PD-1/PD-L1 immune checkpoint inhibitors?)
  - Potential need to compare outcomes in new basket trials against mutation-positive real-world patients who may receive different histology-specific standard of care therapies (see for example Chen et al. 2024[16])
- Established population-adjusted indirect comparison (PAIC) and external control arm (ECA) methods are not well-suited to basket trials settings due to extremely small sample sizes split across multiple tumour histologies

[16] Chen Y, Martin P, Inoue LY, Basu A, Carlson JJ. Tackling Challenges in Assessing the Economic Value of Tumor-Agnostic Therapies: A Cost-Effectiveness Analysis of Pembrolizumab as a Case Study. Value in Health. 2024.

- Ongoing work[17,18]–pre-print will be forthcoming shortly
- Method allows for prognosis to differ by histology via a histology-specific random effect to mitigate confounding due to imbalances in histology.
- Model assumptions:
  - (i) relative treatment effects are constant across histologies,
  - (ii) histologies are exchangeable (variability in prognosis across histologies can be modelled as random effects),
  - (iii) the distribution of prognostic factors within each histology is similar between basket trials, and,
  - (iv) there is overlap in included histologies between the two trials.

[17] Mackay E, Springford A, Nagamuthu C, Dron L. MSR46 A bayesian hierarchical modelling approach for indirect comparison of response outcomes in histology-independent therapies [Abstract]. Value in Health. 2022.

[18] Mackay E, Springford A, Nagamuthu C, Dron L, Dias S. MSR73 Bayesian hierarchical models for indirect treatment comparisons of histology-independent therapies for survival outcomes [Abstract]. Value in Health. 2023.

- For treatment $j \in \{0, 1\}$ and histologies $k = 1, ..., K$ we model the number of responders r_{jk} out of n_{jk} patients at risk as follows:

$$r_{jk} \sim \text{Binomial}(n_{jk}, p_{jk})$$
$$\text{logit}(p_{jk}) = \mu + d \cdot 1\{j = 1\} + \beta_k$$
$$\beta_k \sim \text{N}(0, \sigma^2)$$

- where μ is an intercept term, d is the relative treatment effect (log odds ratio), and the β_k terms are histology-specific random effects
- We also consider relaxing assumption (i)–that the relative treatment effect, d, is constant across histologies by replacing d above with:

$$\delta \sim N(d, \tau^2)$$

- We use weakly informative priors for parameters μ and d, and Half-Cauchy$(0, 1)$ priors for σ and τ

- We use published data on response by histology for two drugs targeting NTRK fusions–larotrectinib and entrectinib–trialled in separate sets of single-arm basket trials[19,20]
- Tabular data is reported after restricting to adult patients and pre-processing

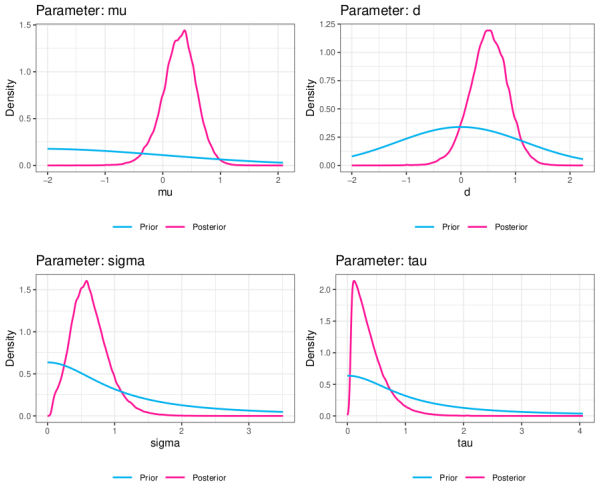
Number of Responders and Observed Overall Response Rate (%)

Tumour Type	Larotrectinib	Entrectinib
Sarcoma	17 / 23 (74%)	15 / 26 (58%)
Thyroid	17 / 22 (77%)	7 / 13 (54%)
Salivary	18 / 20 (90%)	20 / 24 (83%)
Lung	9 / 12 (75%)	14 / 22 (64%)
Colorectal	4 / 8 (50%)	2 / 10 (20%)
Melanoma	3 / 6 (50%)	0 / 0 (–)
Breast	3 / 4 (75%)	5 / 7 (71%)
Pancreatic	1 / 2 (50%)	3 / 4 (75%)
Cholangiocarcinoma	1 / 2 (50%)	1 / 1 (100%)
Unknown Primary	1 / 1 (100%)	1 / 3 (33%)
Appendix	0 / 1 (0%)	0 / 0 (–)
Hepatocellular	0 / 1 (0%)	0 / 0 (–)
Neuroendocrine Tumours	0 / 0 (–)	2 / 5 (40%)
Gynecologic	0 / 0 (–)	1 / 2 (50%)
Head and Neck	0 / 0 (–)	2 / 2 (100%)
Adenocarcinoma of Upper GI Tract	0 / 0 (–)	1 / 1 (100%)
Neuroblastoma	0 / 0 (–)	0 / 1 (0%)
Pooled	**74 / 102 (73%)**	**74 / 121 (61%)**

[20] Hong et al. Larotrectinib in patients with TRK fusion-positive solid tumours: a pooled analysis of three phase 1/2... The Lancet Oncology. 2020.

[20] Demetri et al. Correction: Updated Integrated Analysis of the Efficacy and Safety of Entrectinib in Patients with NTRK Fusion–Positive Solid Tumors. Clinical Cancer Research. 2022.

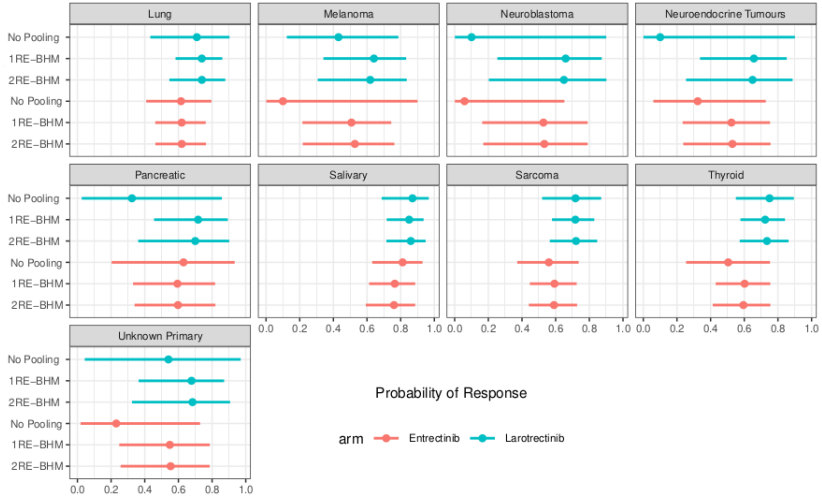
- Posterior for d captures treatment effect estimate (log odds ratio for response)
- Posterior for σ suggests some evidence of heterogeneity in response across histologies
- Posterior for τ indicates that that evidence for heterogeneity in relative treatment effects across histologies is more modest



Towards Indirect Treatment Comparisons for Basket Trials

Estimated Response Rates for a Subset of Tumour Types

- Substantial reduction in width of 95% credible intervals (CrI) from the no-pooling scenario to the 1RE-BHM and 2RE-BHM models
- Differences between 1RE-BHM and 2RE-BHM 95% CrIs are modest– unsurprising given minimal evidence of heterogeneity in relative treatment effects

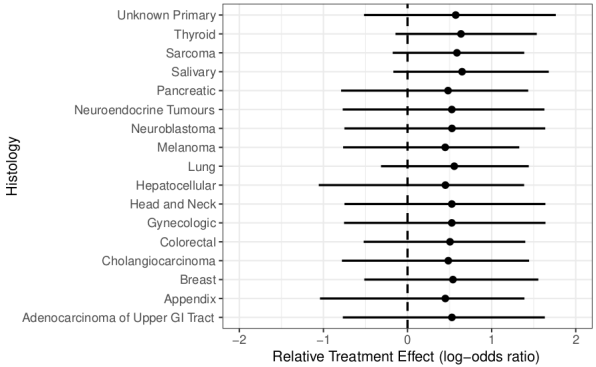


Pooled Relative Treatment Estimates
Effect (log-odds ratio)

Model	Estimate [95% CrI]
Complete-Pooling	0.510 [-0.038, 1.067]
1RE-BHM	0.555 [-0.019, 1.153]
2RE-BHM	0.512 [-0.200, 1.162]

- Difference in treatment estimates between models is modest
- 2RE-BHM has the widest 95% CrI as it incorporates observed heterogeneity cross-histology heterogeneity
- Posterior probability of superiority for larotrectinib exceeds 80% for all histologies (but is below conventional two-sided 97.5% threshold)

Relative Treatment Effect Estimates and 95% CrI by Histology



Concluding Remarks

- Bayesian borrowing approaches present a structured way to leverage all available external data sources when faced with severe data limitations in the evaluation of rare diseases
- Growing receptiveness to their use as part of complex and innovative clinical trial designs (CID)
- These methods can allow for flexible incorporation of disparate data sources under different structural modelling assumptions (e.g. aggregate-level and individual patient data)
- However extreme care needs to be taken to assess (i) suitability of data sources, (ii) structural modelling assumptions, and (iii) sensitivity of results to key inputs (priors, etc.)

Thank You!

Contact: Emma Mackay,
emma@inka.health