Quantifying Bias in Real-World Studies: A New Hope for RWD Acceptance or Are HTAers Gonna Hate?

Speakers

Sreeram Ramagopalan Global Head Real-World Evidence

F. Hoffmann-La Roche Ltd, Switzerland

Miguel Hernán
Professor of Biostatistics and Epidemiology

Harvard T.H. Chan School of Public Health, USA

Seamus Kent
Senior Adviser, HTA

Flatiron Health, UK

Pilar Pinilla Domínguez
Associate Director, NICE International

National Institute for Health and Care Excellence, UK

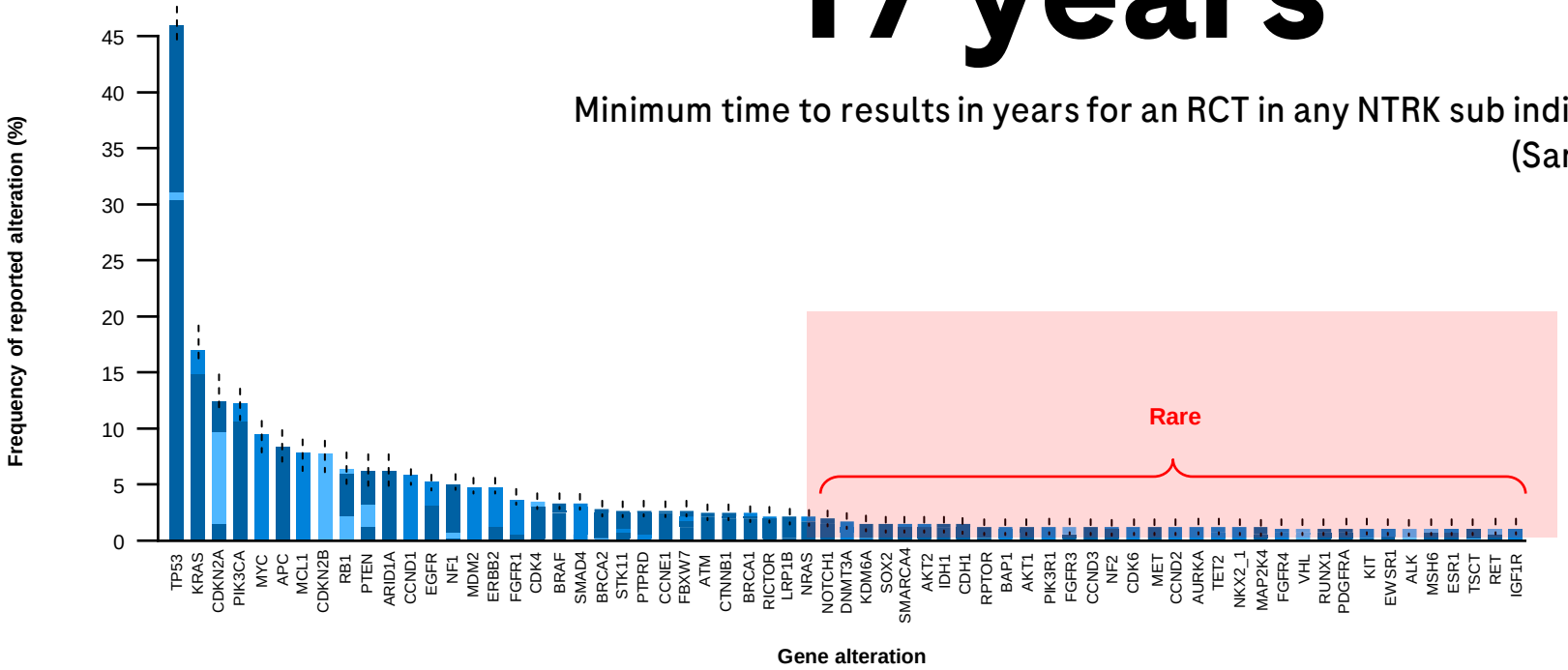
Kristian Thorlund
Senior Principal Scientist

Cytel Inc., Canada

Potentially targetable alterations across all cancers

17 years

Minimum time to results in years for an RCT in any NTRK sub indication (Sarcoma)



Synthetic/External Control Arms in HTA submissions

HTA concerns

- HTA agencies are concerned about bias when combining experimental and observational data

- Some examples of issues related to internal validity that have been cited:
  - Unmeasured confounding in SCAs derived from RWD
  - Residual confounding by variables that are not commonly recorded in RWD, such as performance status, or unadjusted due to large amounts of missingness
  - Insufficient harmonization of covariates and outcomes

- Can quantitative bias analysis (QBA) help?

Addressing issues of internal validity in synthetic control arm (SCA) analyses

External adjustment and quantitative bias analysis (QBA)

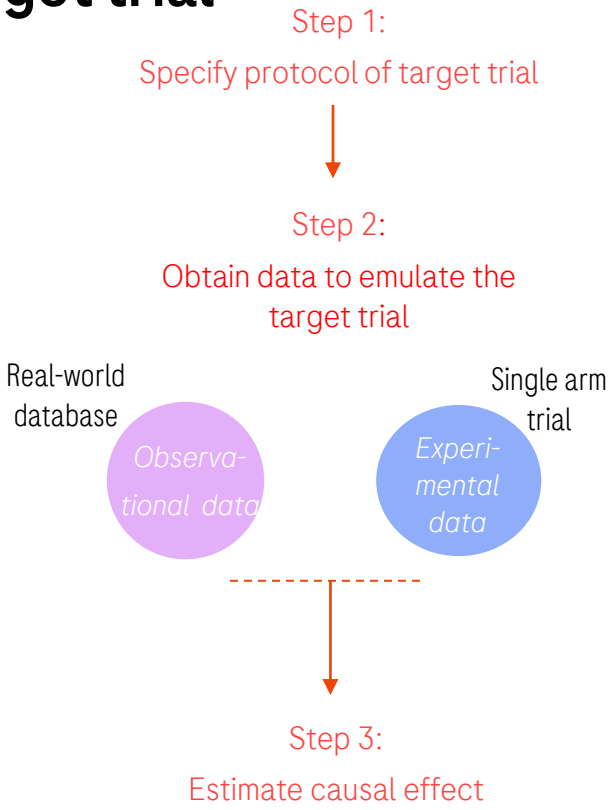
How to estimate the causal effect of treatment vs. standard of care?

- **Using experimental data**: Conduct a target trial
  - Randomly assign eligible individuals to active treatment or standard or care
  - Compare the outcomes between groups
- **Using observational data**: Emulate a target trial
  - Find eligible individuals receiving active treatment or standard of care
  - Compare the outcomes between groups after adjustment for confounders

- Using **both experimental and observational data**
  - Assign eligible individuals to active treatment (experimental data)
  - Find eligible individuals receiving standard of care (observational data)
  - Compare the outcomes between groups after adjustment for confounders
- Also known as "**synthetic control arm**" (**SCA**) analysis

Overview: Emulation of control group of target trial

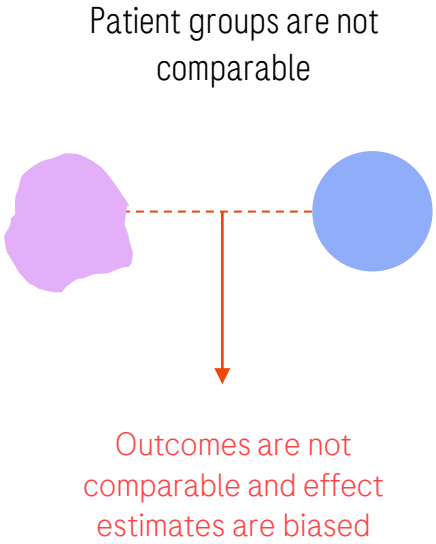
- Step 1: Specify the protocol of the target trial
  - Eligibility criteria
  - Treatment strategy (comparison is standard-of-care)
  - Outcome(s)
  - Start and end of follow-up
  - Statistical analysis
- Step 2: Obtain data to emulate the target trial
  - Recruit and follow eligible participants to treatment group of target trial (i.e., experimental data)
  - Select eligible individuals for standard-of-care group using a healthcare database (i.e., observational data)
- Step 3: Use statistical methods to adjust for differences between arms and estimate the causal effect



What can go wrong?

1. **Treatment and control groups not comparable at baseline**
   - Different distribution of risk factors at baseline
   - Requires measuring the risk factors and adjusting for them

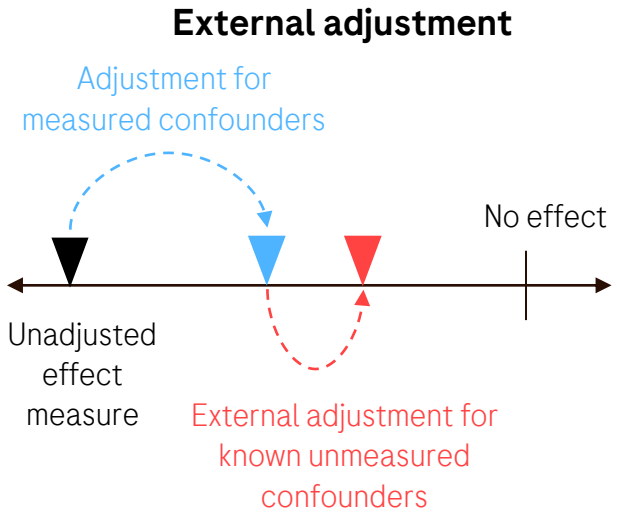
2. **Treatment and control groups comparable at baseline but not later**
   - Different adherence, intensity of monitoring, concomitant treatments, outcome ascertainment, etc. during follow-up
   - Potentially the most serious problem
   - Unclear it can be adjusted away
   - Need to design single-arm trial as pragmatic trial to increase comparability with real world data



Treatment and control groups not comparable at baseline

- **Identify confounders**
  - If confounders are measured, adjust for them
  - If confounders are unmeasured, use external information about them to correct for bias
    - External adjustment
    - Deterministic or probabilistic sensitivity analysis

- **Use negative outcome controls**
  - Outcomes for which the effect estimate should be null
  - If confounding had been adequately adjusted for



Quantitative Bias Analysis for the Assessment of Bias in Comparisons between Synthetic Control Arms from External Data and Lung Cancer trials (Q-BASEL)

Results and discussion

Collaborators

Roche

NICE National Institute for Health and Care Excellence

Cytel

Imperial College London

Putnam
An Ashfield Advisory Company

Universiteit Leiden
The Netherlands

Rationale

- Using RWD to form ECAs in the absence of RCTs remains at risk of residual bias incl. from:
  - Residual confounding – unmeasured, poorly measured, or unknown confounders
  - Missing data or misclassification on key confounders
- In oncology, ECA has worse outcomes (on avg) than trial arm => exaggerated treatment effect

- Quantitative bias analysis can be used to either
  - Directly model bias and adjust for it
  - Explore the sensitivity of results to potential bias

- Quantitative bias analysis is
  - A potentially powerful tool for addressing these residual biases head-on
  - Consistent with approaches for modelling uncertainty in HTA

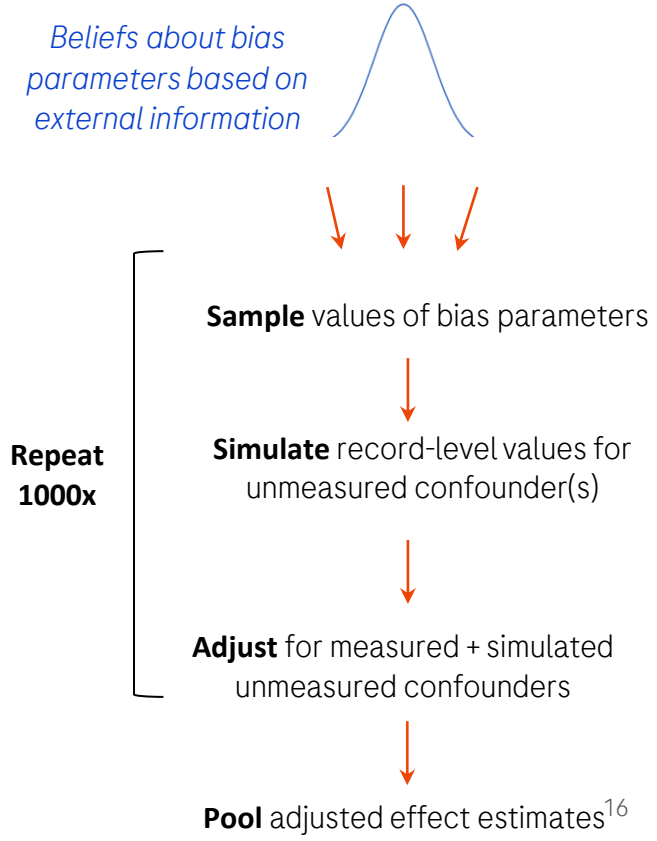
Q-BASEL – study methods

- Key questions/aims
  - Is external bias adjustment feasible and does it improve the validity of ECA studies?
  - Can tipping point analyses add value to decision making?
  - Identify recommendations for the use of QBA methods in HTA
- Overview of process
  - Estimate intention-to-treat treatment effects for 14 RCTs in aNSCLC
  - Select **external control arms** for each study from real-world Flatiron Health database by trial emulation
  - **Adjust** for confounding and estimate hazard ratios **for overall survival (OS)** targeting an observational ITT[2] estimand
  - For each emulation, summarize external information on important sources of bias and **compute bias-corrected hazard ratios**
  - **Compare** randomized and non-randomized bias-corrected estimates
  - Use **tipping point** analyses to explore risk of bias from unknown confounding and missing data on ECOG

Data and methods for external bias adjustment

- Key unmeasured confounders varied by trial emulation but were most commonly:
  - Presence of specific metastases
  - Recorded positive test for variants in EGFR/ALK or PD-L1 expression
- To model confounding bias we need 3 pieces of information:
  1. Prevalence of the confounder
  2. (Conditional) association with treatment assignment
  3. (Conditional) association with outcome (overall survival)
- These data can come from multiple sources
  - Expert opinion
  - Published literature (from systematic review)
  - Internal or external data (e.g., from data recorded in the RCTs)

- Process of bias adjustment



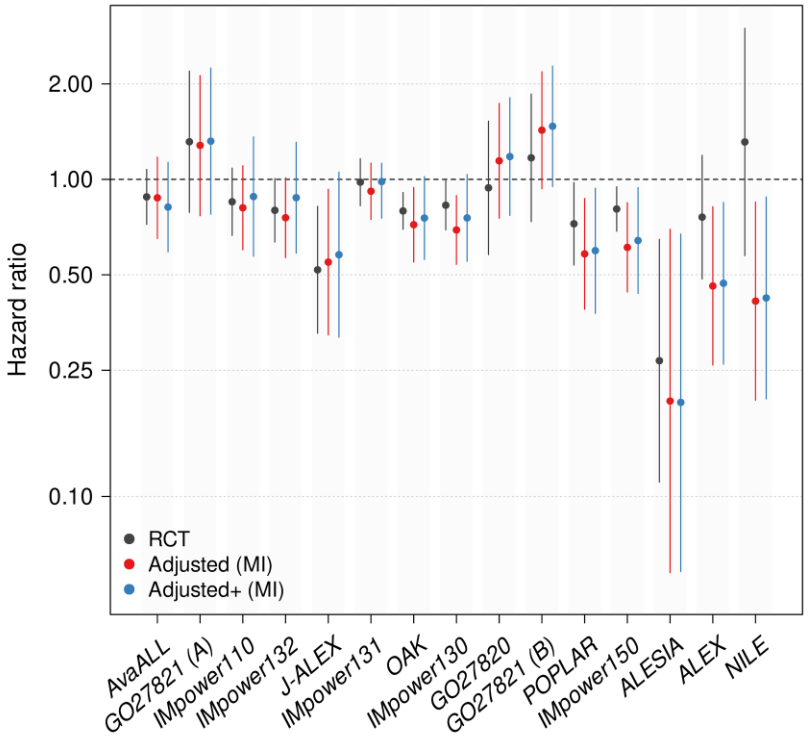
Results – external bias adjustment

- Emulation using ECA was good for 11 of 14 trials
- Bias adjustment tended to improve emulation (and reduced treatment effects)
- But differences in bias-adjusted and unadjusted results were typically small
- Bias adjustment performed best where able to use data on confounders recorded in the trial
- Emulation was most challenging were there was strong residual confounding (e.g., selection into trial based on expected life expectancy)

Adjustment	Mean Standardized mean difference (SMD)	Mean difference from RCT (log HR)
Adjusted	0.75	0.16
Bias-adjusted	0.65	0.12

Estimated hazard ratios and 95% CI across 14 trials

(Ordered by increasing difference between RCT and Adjusted MI point estimates)

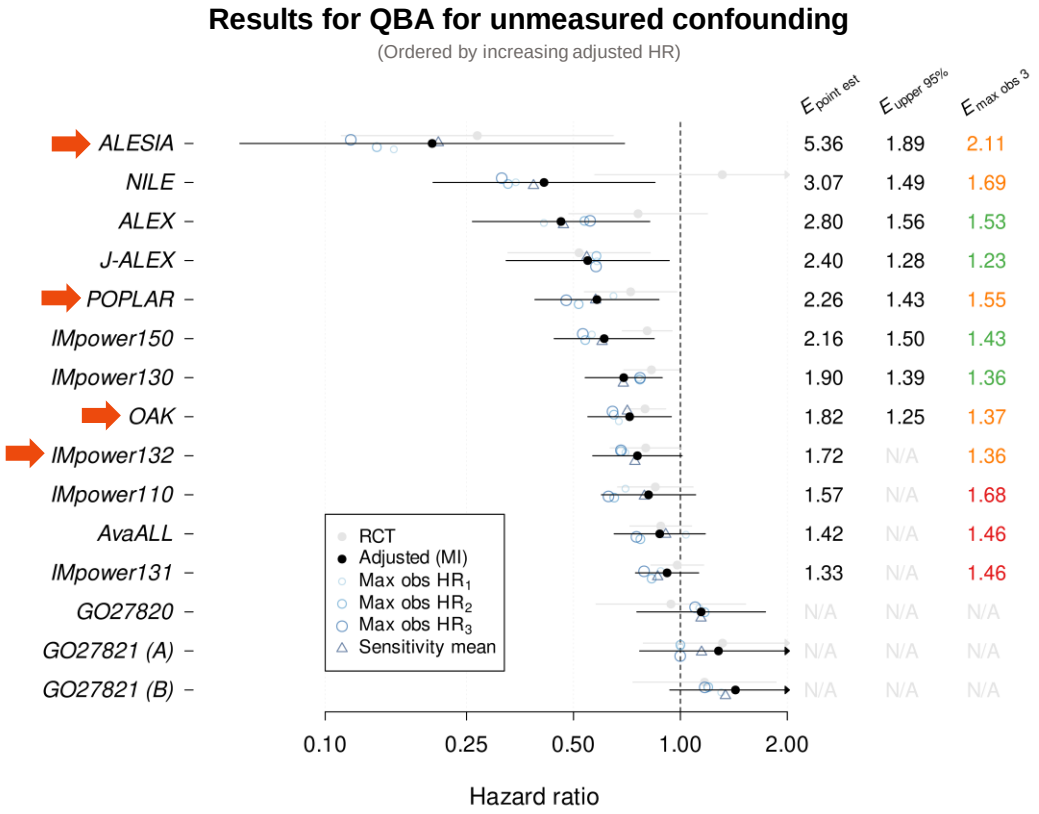


Methods for addressing unknown confounding

- Previous analysis assumed we knew the key unmeasured confounders. There may be **unknown unknowns**.
- HTA reviewers may wish to know: How large would residual confounding have to be to materially change conclusions?
- This threshold could be many things including:
  - Null treatment effect or insignificant effect
  - Cost-effectiveness threshold
- We calculated **E-values** which tells us the minimum strength of confounding required to pass a decision threshold – a form of tipping point or threshold analysis
  - A large E-value implies confounding needs to be large
- E-values can be contextualised by:
  - Some defined threshold of reasonable level of confounding
  - Comparison against observed confounders

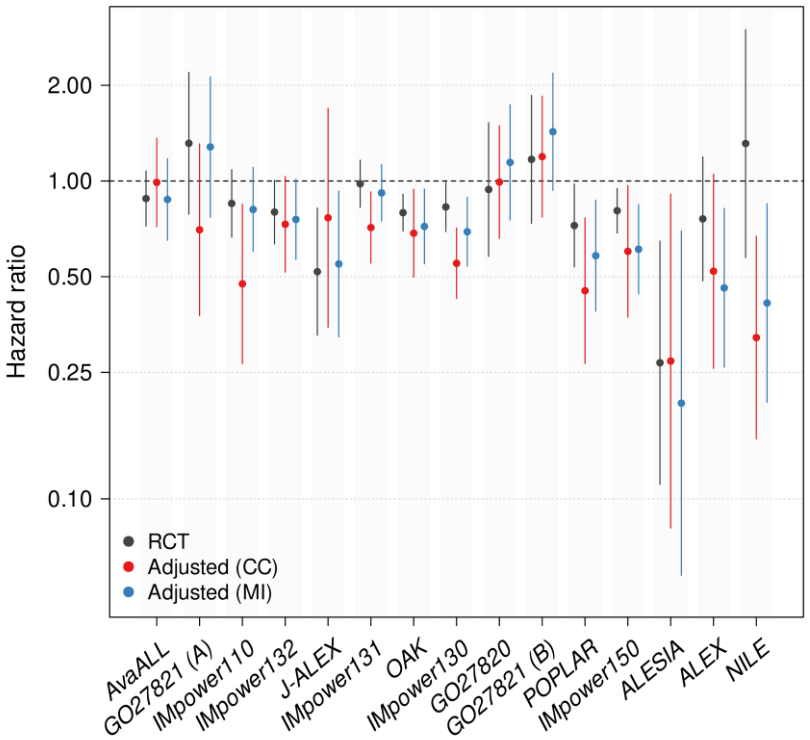
Results from QBA for unmeasured/residual confounding

- E-values were typically higher than strength of confounding for observed variables
- They are a simple method to explore the sensitivity of results to unknown confounding
- High or low E-values were not always associated with poor or better emulation
- Interpretation against a threshold may not be appropriate? How is that threshold decided?



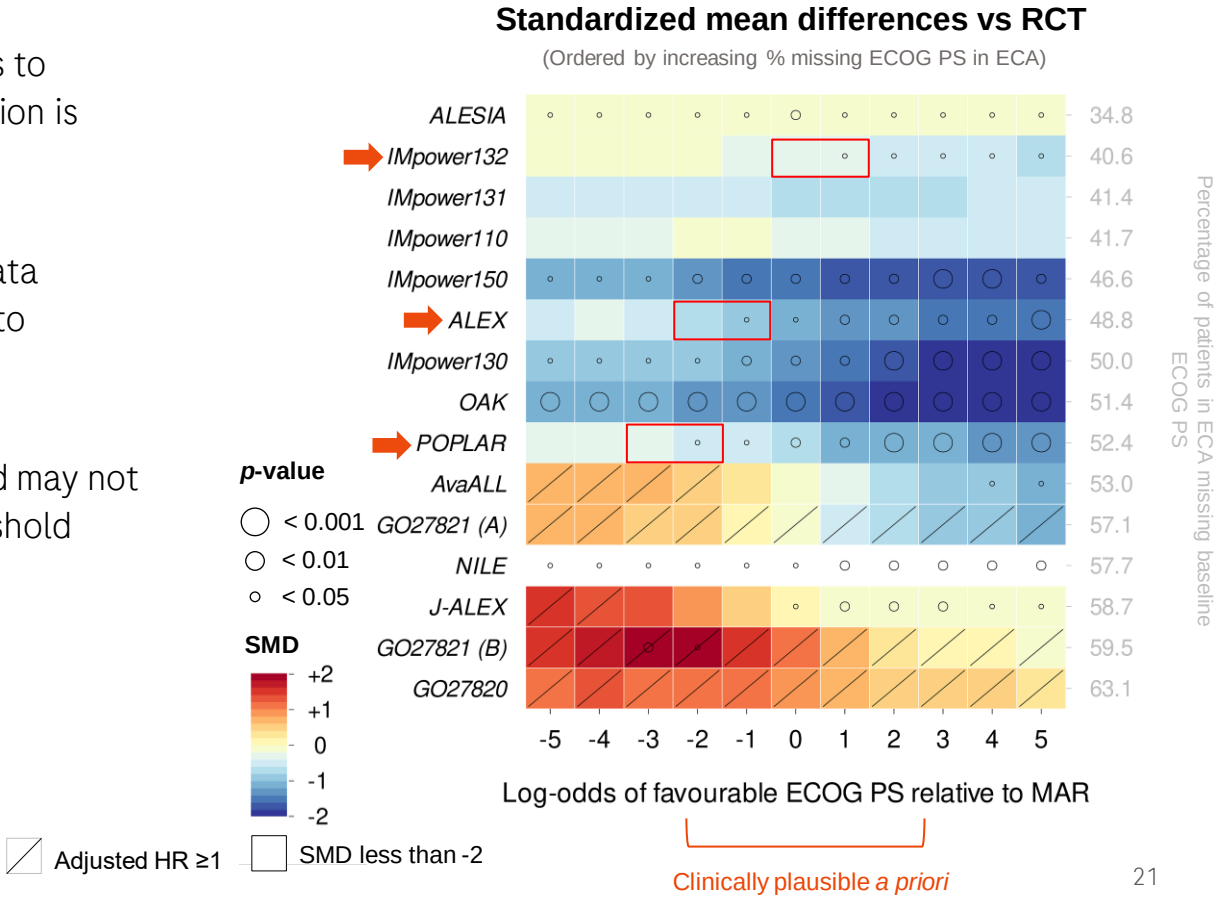
Methods for addressing missing data on ECOG PS

- Data on baseline ECOG status is missing for between **35%-63%** patients across the 15 ECAs
- Base-case analysis used multiple imputation assuming a missing at random mechanism
- External bias adjustment was considered but exert opinion suggested any missing not at random effect was likely to be small – instead we used **tipping point analysis**
- **Question**: How much better or worse than expected would ECOG scores have to be to materially affect conclusions?



Results – missing data tipping point analysis

- Exploring the sensitivity of results to deviations from the MAR assumption is simple and useful
- Sensitivity of results to missing data mechanism is not always related to success of trial emulation
- Interpretation against a threshold may not be appropriate? How is that threshold decided?



Conclusions

- Legitimate concerns about residual bias are likely to remain after controlling for observed confounders using real-world external control data
- Quantitative bias analysis is a potentially useful mechanism to address these concerns through direct bias adjustment or tipping point analyses
- We showed that it was feasible (if resource intensive) to identify external data on key sources of bias and adjust for these in analyses – and provided a template for doing so
- Tipping point approaches may be particularly useful where external data is lacking
- The results also suggest residual bias is not expected to be a major problem in aNSCLC studies (in general) after controlling for key observed confounders
- Can QBA improve confidence in the principled use of RWE for decision making?

Handling uncertainty: QBA and NICE's RWE Framework

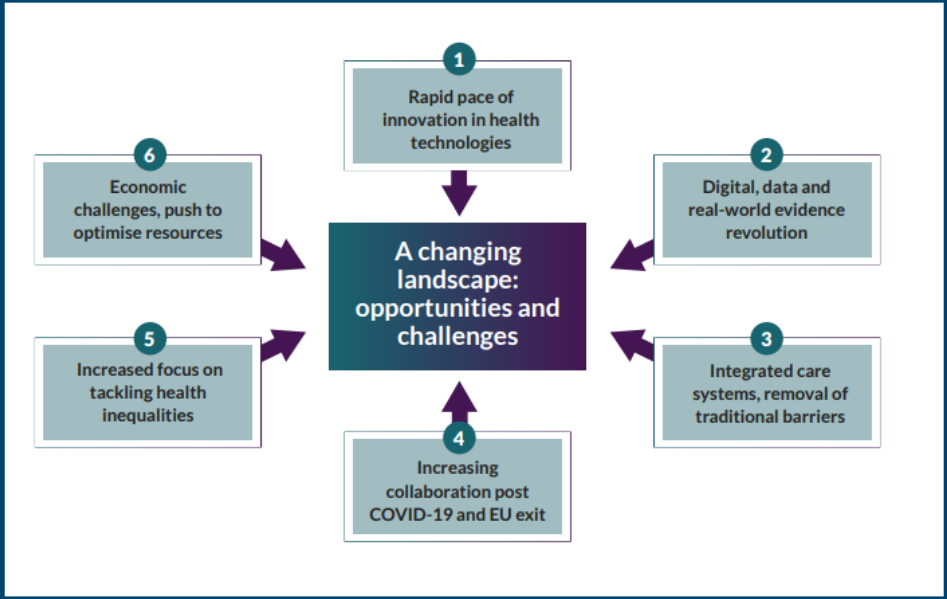
Pilar Pinilla Dominguez

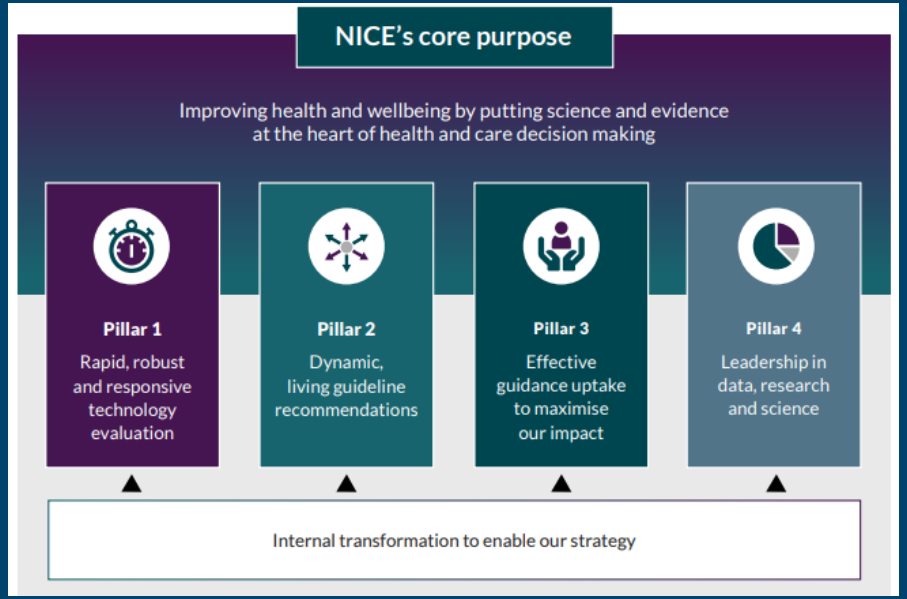
Associate Director – NICE International

ISPOR EU: 07/11/2022

NICE National Institute for Health and Care Excellence

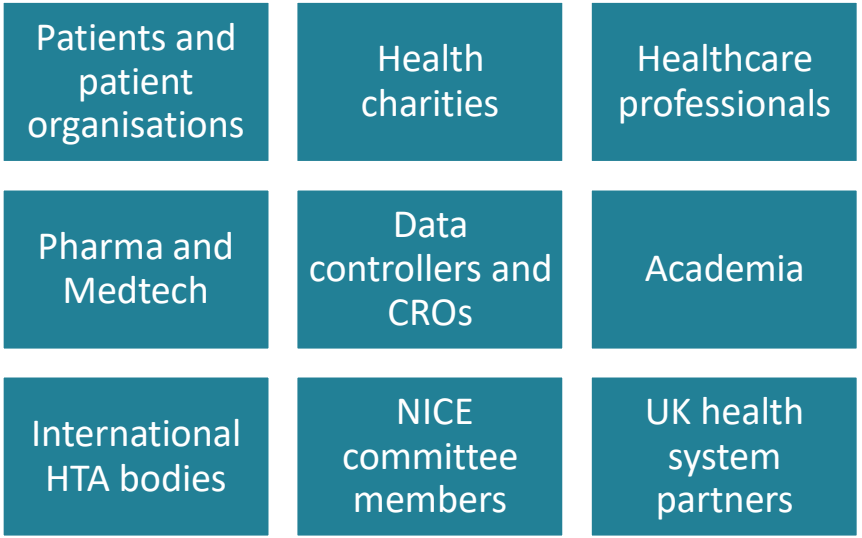
Context


1
Rapid pace of innovation in health technologies
2
Digital, data and real-world evidence revolution
3
Integrated care systems, removal of traditional barriers
4
Increasing collaboration post COVID-19 and EU exit
5
Increased focus on tackling health inequalities
6
Economic challenges, push to optimise resources
A changing landscape: opportunities and challenges


NICE's core purpose
Improving health and wellbeing by putting science and evidence at the heart of health and care decision making
Pillar 1
Rapid, robust and responsive technology evaluation
Pillar 2
Dynamic, living guideline recommendations
Pillar 3
Effective guidance uptake to maximise our impact
Pillar 4
Leadership in data, research and science
Internal transformation to enable our strategy

RWE Framework – purpose and development

- Help to deliver NICE's ambition to use real-world evidence to fill evidence gaps and speed up patient access to innovative interventions
- Framework informed by existing best-practice guidance for using RWD, a series of multistakeholder workshops, internal/external consultation - Published June 2022
- Describes best-practices for planning, conducting, and reporting real-world evidence studies
- Improve the quality and transparency of real-world evidence studies
- Improve committee trust in real-world evidence and enable informed critical appraisal



Principles of evidence generation

1 Transparency

Generate evidence in a transparent way and with integrity from study planning through to study conduct and reporting.

2 Data suitability

Ensure data is trustworthy, relevant and of sufficient quality to answer the research question.

3 Methods

Use analytical methods that minimise the risk of bias and characterise uncertainty.

Tools			
	RWE Registries StartRWE Open code Target trial appr.	HDRUK Innov. gateway DataSAT StartRWE SPIFD	Bias reporting template Target trial approach ROBINS-I StartRWE

Real-world ECA studies

- Increasing subsetting of patients and complexity in treatment pathways
- Single Arm Trials (SATs) and RWD controls increasing as a proportion of submissions to NICE and HTA bodies
    - 13-fold increase in SAT submissions (2011 – 2019)
    - Use of RWD External Controls increased 22% as a proportion per year[1]
- For studies using external control, differences between data sources may limit comparability:
    - availability and operational definitions of key study variables
    - data collection processes
    - patient characteristics
    - treatment settings
    - care pathways
    - time periods

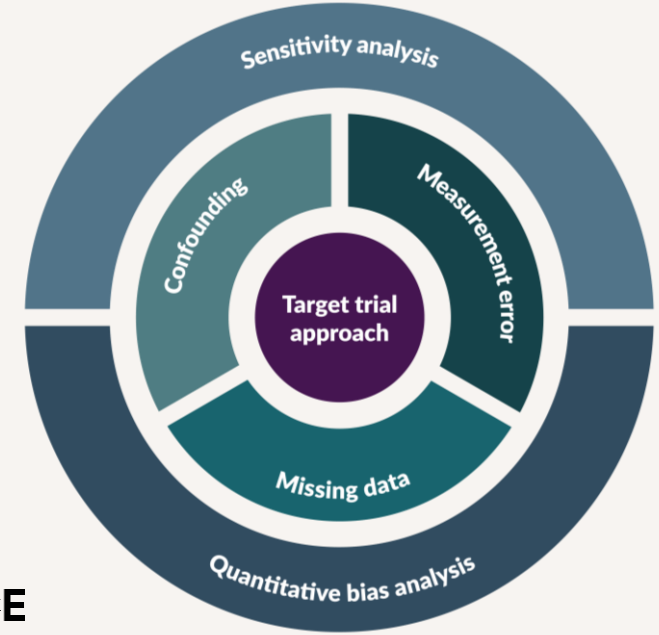
NICE

1. Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, Kim J. Use of external comparators for health technology assessment submissions based on single-arm trials. Value in Health. 2021 Aug 1;24(8):1118-25.

Real-world evidence studies of comparative effects

Real-world evidence can be used **in the absence of trial evidence or to complement it** to answer a broader range of questions about the effects of interventions in routine settings.

Here we present best-practices for cohort studies (including trials using real-world data to form external control). Other study designs including quasi-experimental designs might be most appropriate for some interventions.



Design studies to emulate the preferred randomised controlled trial – use a "target trial approach"

Identify potential confounders and address these considering observed and unobserved confounding

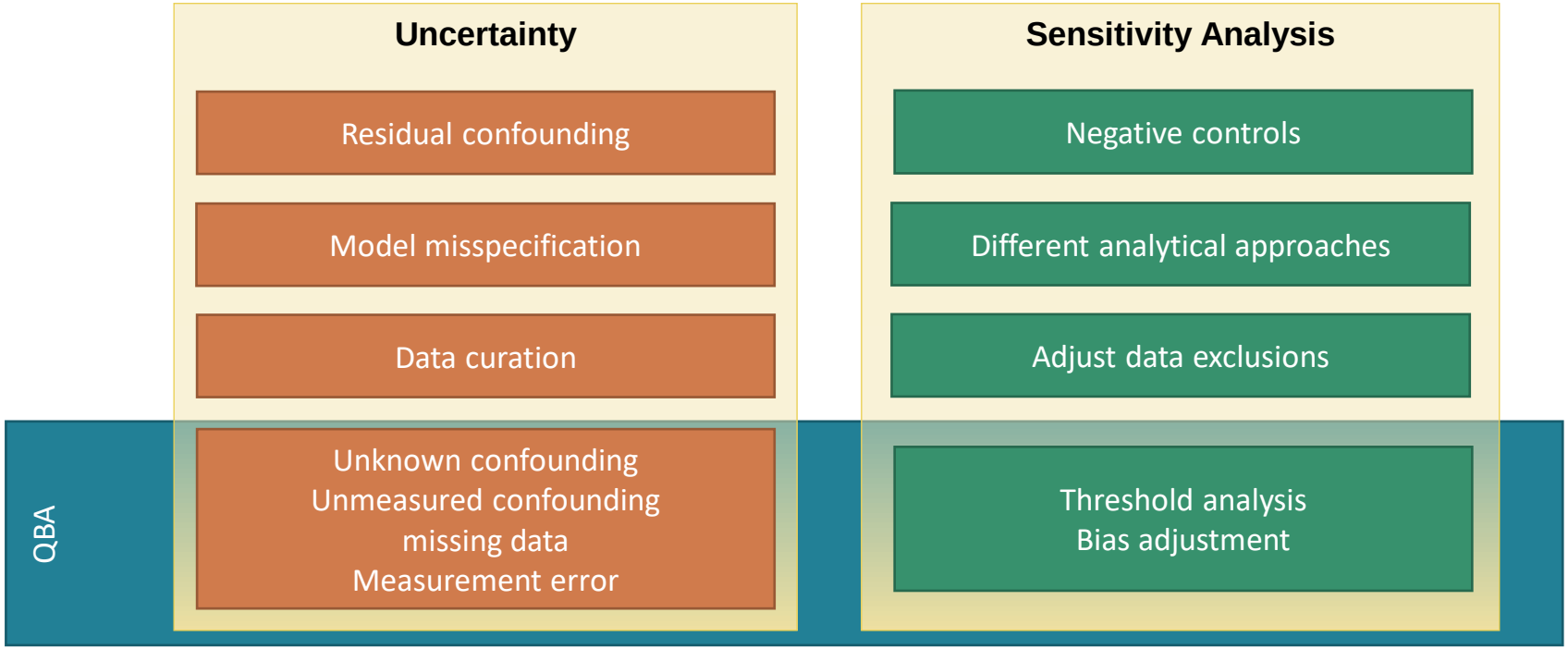
Consider the impact of bias from informative censoring, missing data, and measurement error – address appropriately where required

Use sensitivity and bias analysis to assess the robustness of study findings

NICE

Assessing robustness

Focus on areas where the impact of bias, assumptions, uncertainty are greatest – justify choice, pre-specify where possible



About QBASEL (why we're excited about it)

Trial emulation

- Use best practice External Control Arm methods to reduce bias and compare adjusted effect estimates to RCTs effect estimates

The QBA process

- Identify main sources of concern (unaccounted-for bias) e.g. due to missing data, measurement error, unmeasured confounding, unknown confounding
- Investigate these using QBA methods e.g.
  - **Bias adjustment for unmeasured confounding:** use a transparent and systematic process to identify external information on 1) the prevalence of important confounders 2) their association with the outcome and 3) imbalance across treatment assignment.

    Adjust for the most probable impact of the bias, accounting for already measured confounders.
  - **Tipping point analysis:** investigate plausible scenarios for missing data, confounder mismeasurement, unknown confounding, and whether these would be likely to meaningfully impact results
- Need to validate approach using another data source with different "concerns"

NICE

QBA for HTA

Advantages

- Facilitates open discussion between developers and committees about common uncertainties in Single Arm Trials + External Control Arm studies
- Developers can show that they have identified the key threats to validity in their analyses, and have thoroughly investigated these
- Committees can move from dismissals based on general sense of "uncertainty" to highlighting key biases of concern and what level of uncertainty is acceptable

Agreed approach

- Elements of best practice approach to bias adjustment:
  - Systematic collection of external data
  - Expert elicitation
  - Appropriate weighting
- Operationalising:
  - Framework
  - Flowchart
  - Early HTA engagement
  - Where is QBA not useful
  - When to use tipping point approaches

Future work

- Different disease areas
- Different treatments
- Available data quality
- Likelihood of unknown bias
- Availability of external information

Summary

- NICE's RWE Framework **describes best-practices** for planning, conducting, and reporting real-world evidence studies
- Single Arm Trials and RWD controls are increasing as a proportion of submissions to NICE and HTA
- These studies are affected by bias stemming from differences between data sources
- Quantitative bias analysis describes a group of methods that can investigate and quantify the potential direction and magnitude of bias in analysis, and can adjust for this bias
- Conversations and expectations between developers and HTA organisations regarding key sources of bias and acceptable levels of uncertainty can be supported using QBA, and facilitated by early engagement with HTA organisations.

NICE National Institute for Health and Care Excellence

Thank you

Pilar Pinilla-Dominguez
www.nice.org.uk/NICEInternational
pilar.pinilla-domiguez@nice.org.uk
@pilarpinilladom

Panel discussion