ISSUE PANEL 1: PRAGMATIC CLINICAL TRIALS TO ESTIMATE TREATMENT EFFECTS: ARE THEY WORTH THE EFFORT?

Presented by the ISPOR Statistical Methods in HEOR Special Interest Group

Monday, 6 November 2017

Panelists

- **Rita M. Kristy, MS,** Senior Director, Medical Affairs Statistics, Astellas Pharma Global Development, Northbrook, IL, USA
- **Helene Karcher, PhD,** Managing Vice-President, Analytica Laser, Basel, Switzerland
- **Christoph Gerlinger, PD, Dr,** Senior Director, Pharmaceuticals Statistics, Bayer AG, Berlin, and Gynecology and Obstetrics, University of Saarland, Homburg, Germany
- **Keith R. Abrams, PhD, CStat,** Professor of Medical Statistics, NIHR Senior Investigator Emeritus & Head, Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK

What is a pragmatic clinical trial?

- Can this intervention work under ideal conditions (explanatory)

vs.

- Does the intervention work under usual conditions (pragmatic)

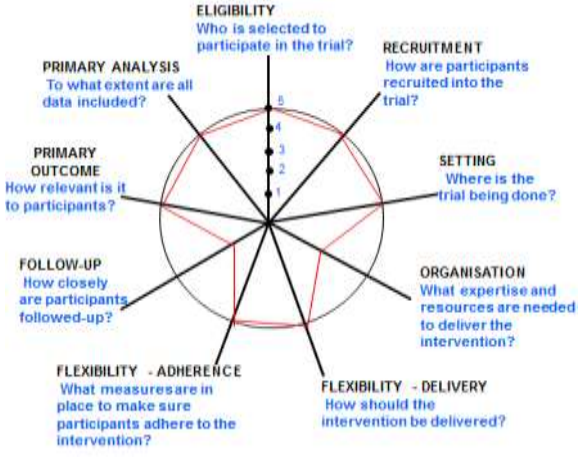
PRECIS-2 Criteria

- Pragmatic-Explanatory Continuum Indicator Summary 2
- Developed and validated to improve issues with the original PRECIS
- 9 domains scored from very explanatory to very pragmatic

PRECIS-2 Wheel

The PRagmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2) wheel.
Adapted from BMJ 2015;350:h2147

Examples of PRECIS-2 wheel



Little P, Moore M, Kelly J, Williamson I, Leydon G, McDermmott L, Mullee M, Stuart B: Ibuprofen, paracetamol, and steam for patients with respiratory tract infections in primary care: pragmatic randomised factorial trial. BMJ 2013, 347:f6041.

Overview of the issue panel

- How to design pragmatic trials
- Using cross-design analysis to overcome limitations of both pragmatic and explanatory studies
- Using of both pragmatic trials and evidence synthesis to overcome limitations of both randomized controlled trials (RCTS)

7

Why all the buzz about pragmatic trials?

Can't we just do RCTs, and complement with observations in usual care practice?

Not any more!

Why design pragmatic trials?

- To prove effectiveness of interventions in the real world (RW)
  - During drug development
  - Around drug launch

  } **Justify future value in the RW → increasingly important!**
  - After launch: comparative effectiveness of already-established products

- To generalize effectiveness measured in pragmatic trials to other RW settings
  - Using predictive modeling

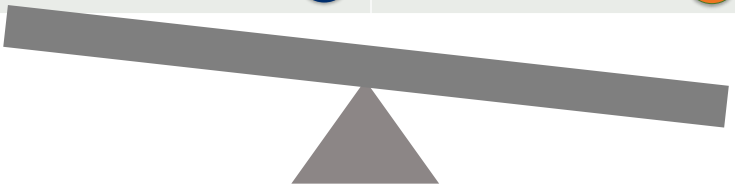
Hurdles to incorporating pragmatism into drug development trials* (review of 39 articles)

1. Known and unknown confounders in real-world trials may lead to inconclusive effect sizes
2. Extensive cost of running such trials due to larger sample size required
3. Operational difficulties in recruiting certain populations, and in minimising measurements/study visits
4. Uncertainty in reactions from regulatory bodies

* Karcher, Nordon, Neumann, Nikodem, Zyla, Chevrou-Séverac, Jimenez, Bala, Abenhaim. Methods to Evaluate Real-World Effectiveness in Pre-Authorization Trials SLR. HTAi 2015.

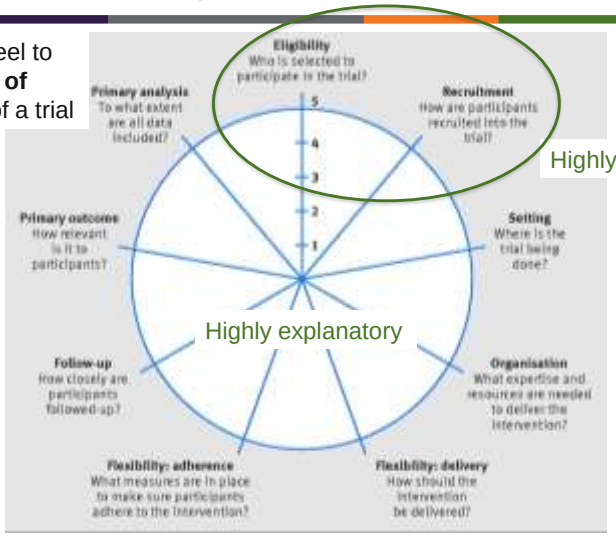
A trade-off between different trial goals

Explanatory	Pragmatic
High internal validity → Difficult to extrapolate effect to other populations / other conditions (–)	High external validity → Generalizable trial results (via predictive modeling) (+)
Homogeneous population and controlled conditions → Little variability in endpoint → Detect effect sizes of investigated drug with small sample sizes (+)	Heterogeneous population and less-controlled conditions → Larger variability in endpoint → Requires larger sample sizes to detect the same effect size (–)



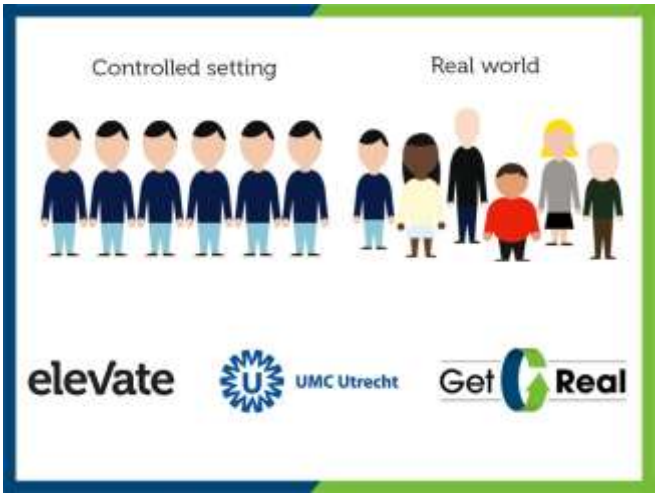
How to pick dimensions/degree of pragmatism for your trial?

PRECIS-2 wheel to appraise **level of pragmatism** of a trial



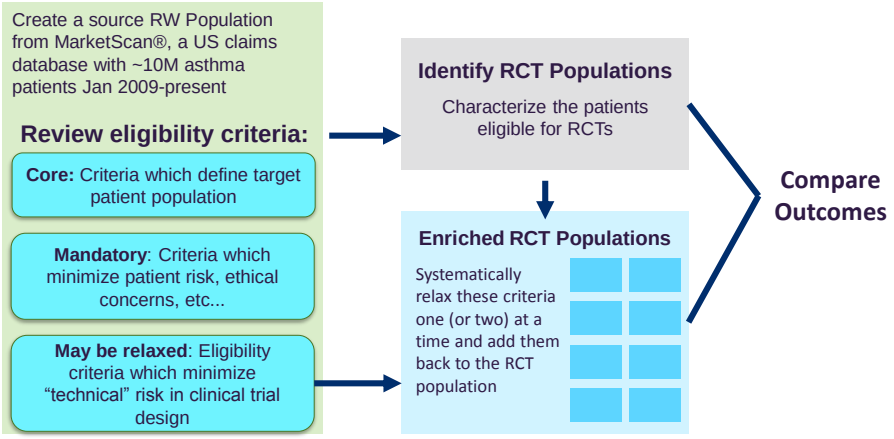
Loudon et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147

Example: quantifying this trade-off when including a more heterogeneous population



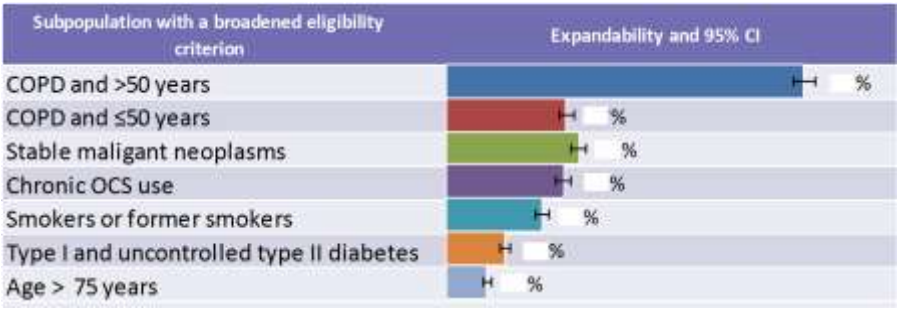
Optimizing trial populations in clinical development
"RCT enrichment" approach – case study in asthma[1]

- Which patients are typically excluded from clinical trials?
- Impact of re-inclusion of these patients on trial recruitment and outcomes?



1. Karcher, Meng, Fu, Loefroth, Cao, Peress. Optimal design of pre-authorization trials for effectiveness evaluation in severe asthma. *Value in Health* 19 (7), A360-A361. 20160

Expandability of the population pool eligible for Phase 3 trials per exclusion criterion (prevalence)



Expandability:

$$\frac{Number\ of\ patients\ re\text{-}included\ into\ Phase\ 3\ eligible\ population}{Number\ of\ Phase\ 3\ eligible\ population}$$

Efficacy and Safety differences in Phase 3 vs re-included real-world populations?

Exacerbations Cardiovascular risk

Subpopulation with a broadened eligibility criterion	Reduction of exacerbation # (mean, 95% CI) in subpopulation	Relative Risk (mean, 95% CI) in # patients with MACE Subpopulation/population
COPD and >50 years		
COPD and ≤50 years		
Stable maligant neoplasms		
Chronic OCS use		
Smokers or former smokers		
Type I and uncontrolled type II diabetes		
Age > 75 years		

Blinded forest plot

Blinded forest plot

Less exacerbations than RCT population — More exacerbations

Less MACE than RCT population — More MACE

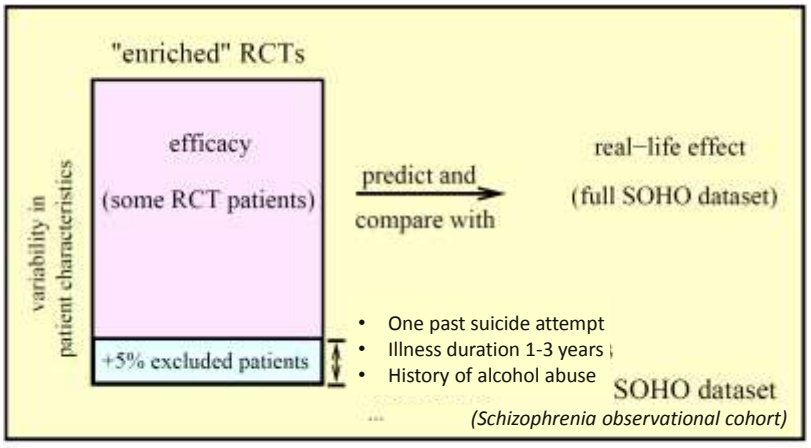
Rationale for decision to relax exclusion criteria based on

1. **Expandability of eligible population** (prevalence) – linked to recruitment speed
2. **Efficacy and safety** in the re-included populations

17

MACE: Major Adverse Cardiac Event

The "RCT enrichment" approach in schizophrenia[1,2]



1. Karcher, [..] Nordon. The "RCT enrichment": a novel simulation method to add patient heterogeneity into Phase III trials. Under review at *BMC Med Res Methodol.* 2017
2. Nordon, [..], Karcher. Trial exclusion criteria and their impact on the estimation of antipsychotic drugs effect: a case study using the SOHO database. Schizophr Res. 2017

18

Schizophrenia case: results to choose the degree/type of pragmatism for a new trial

- The best choice of population enrichment factor (=pragmatic dimension) to predict real-life effects was found to be driven by:
  - Size of the excluded real-life population
    - Re-including “illness duration 1-3 years” and “number of past suicide attempts > 1” increased the most the pool of schizophrenia patients eligible for Phase 3 trials.
  - Change in outcome in patients with this factor
    - Patients with a practice type “private” and illness duration between 1-3 years had the most different outcome from typical Phase 3 patients.

- The trial statistical power is calculable for each set of eligibility criteria via simulations of virtual RCTs with the more heterogeneous population.

Conclusion: how to design pragmatic trial design?

- Early demonstration of value in the RW is essential
  - Pragmatic trial are one important part of this demonstration

- Need to reach a compromise between demonstrating drug effect & learning about effectiveness
  → Carefully select dimensions/degree of pragmatism in a trial

- Methods exist to quantify how much adding each pragmatic feature to the trial:
  - Will benefit in terms of generalizability of its results
  - May compromise (but also sometime improve!) detection of effect sizes

RCTs and pragmatic trials – why not take the best of both worlds?

PD Dr. Christoph Gerlinger
Dr. Tatsiana Vaitsiakhovich
Dr. Anna Filonenko

Problem statement: Another view on RCTs

*"Drugs are tested by the people who manufacture them, in poorly designed trials, on **hopelessly small numbers** of **weird, unrepresentative patients**, and analysed using techniques which are flawed by design, in such a way that they **exaggerate the benefits** of treatments."*

Ben Goldacre, Bad Pharma
www.badscience.net

Problem statement: Evidence sources with different strengths and limitations

- Randomized clinical trials
  - Highly selected subset of the total patient population
  - Protocol-driven procedures and treatments
  - High internal validity (indispensable for drug licensing)
  - But, low external validity

- Pragmatic trials
  - More representative of clinical practice
  - But, internal validity is limited due to confounding, selection bias, channeling, …

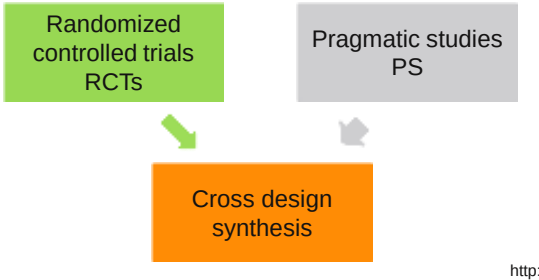
Idea: Combine the strengths of pragmatic and randomized trials

Several methods proposed in the literature

- Confidence profile method
- Network meta-analysis and indirect treatment comparison
- Cross-design synthesis
- Direct modeling of bias
- Bayesian hierarchical methods

Cross design synthesis

- **Cross design synthesis** is a novel strategy for medical effectiveness research, advancing knowledge on medical treatments based on the results of randomized clinical trials **and** real life evidence
- Cross design synthesis combines the results from studies that have different **complementary** designs



http://www.gao.gov/products/PEMD-92-18 2017-10-19

Kaizar 2011 paper

- Framework for cross design synthesis
- based on
  - Rubin‘s causal model
  - Stratification (within and between study designs)
  - Linear model for the relationship of errors between strata

- CAVE! Several typos in the formula in the appendix!

Kaizar paper - stratification

- Study type stratification
  - Randomized vs. Observational
  - Reflects differential treatment selection error
- Population stratification
  - RCT inclusion criteria met or not
  - Reflects sample selection error

RCT inclusion criteria	RCT	OS
met	subjects in RCT	subjects in OS_{in}
not met	n.a.	subjects in OS_{ex}

Kaizar paper – estimators

Strata specific estimators

RCT inclusion criteria	RCT	OS
met	$d_{RCT} = \bar{x}^{RCT}_{active} - \bar{x}^{RCT}_{control}$	$d_{OS_{in}} = \bar{x}^{OS_{in}}_{active} - \bar{x}^{OS_{in}}_{control}$
not met	n.a.	$d_{OS_{ex}} = \bar{x}^{OS_{ex}}_{active} - \bar{x}^{OS_{ex}}_{control}$

Cross design estimator unbiased if treatment selection error for the patients in the PS, fulfilling the inclusion criteria of the RCT, and the patients, who do not, is constant

$$d_{CDS} = d_{RCT} + \frac{n_{OS_{ex}}}{n_{OS}} \left(d_{OS_{ex}} - d_{OS_{in}} \right)$$

$$S^2_{CDS} = \frac{S^2_{RCT_{active}}}{n_{RCT_{active}}} + \frac{S^2_{RCT_{control}}}{n_{RCT_{control}}} + \frac{n^2_{OS_{ex}}}{n^2_{OS}} \times \left(\frac{S^2_{OS_{in_{active}}}}{n_{OS_{in_{active}}}} + \frac{S^2_{OS_{in_{control}}}}{n_{OS_{in_{control}}}} + \frac{S^2_{OS_{ex_{active}}}}{n_{OS_{ex_{active}}}} + \frac{S^2_{OS_{ex_{control}}}}{n_{OS_{ex_{control}}}} \right)$$

Example - introduction

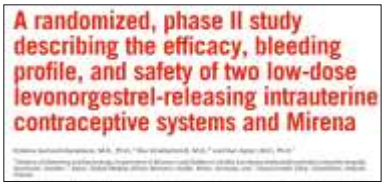
- Data from literature
- Indication: Long acting reversible contraceptives (LARC)
- Research question: How long do women adhere to the method
- Assumption for example: similar adherence for all different LARCs (as the data per product were not in the PS publication)
  - RCT data only from adults
  - PS data from all ages

Photos: www.your-life.com/en/contraception-methods/long-acting-contraception

Example – Data

Randomized trial

- 738 women
- Age 20-41 (mean 32.1)


A randomized, phase II study describing the efficacy, bleeding profile, and safety of two low-dose levonorgestrel-releasing intrauterine contraceptive systems and Mirena

- Data taken from supplemental figure 1 of online publication.
- Kaplan-Meier estimates were re-calculated considering dropout for „lost to follow-up“ and „other“ as censored (to mimic OS publication as far as possible)

Pragmatic trial

- 3203 women
- Age 14-45 (mean 25.7)


RESEARCH

GYNECOLOGY

Three-year continuation of reversible contraception

- Lost to follow-up and dropout „wish to get pregnant“ considered as censored

Example – Data and Results

Continuation rates by study type and age group

	Year 1	Year 2	Year 3
PS age 14-19	**82. 1** (78.0-85.6)	**68.0** (63.0-72.5)	**52.6** (47.2-57.7)
PS age 20-45	**86.3** (85.0-87.6)	**76.2** (74.5-77.8)	**69.2** (67.4-71.0)
RCT age 20-41	**90.5** (88.4-92.7)	**82.4** (79.6-85.2)	**79.9** (77.0-82.9)

$$CDS_{Year\ 3} = 79.9 + \frac{405}{3203}(52.6 - 69.2) = \mathbf{77.8}$$

95% Confidence Interval: 74.8 - 80.8

Example – Strengths and Limitations

- CDS estimator adjusted RCT result for excluded adolescents
  - No huge impact: -2.1 %-point difference in 3-year continuation rate
  - but only 12.6% adolescents in PS

CDS estimator based on publications could not adjust for other possible patient selection biases in RCT

- E.g., 99.3% caucasian in RCT vs. 45.0% in PS
  - Would need analyses on matched individual patient data
- Even with individual patient data one could not adjust for geographic location
  - RCT from northern and central Europe
  - PS from St. Louis, Missouri, USA

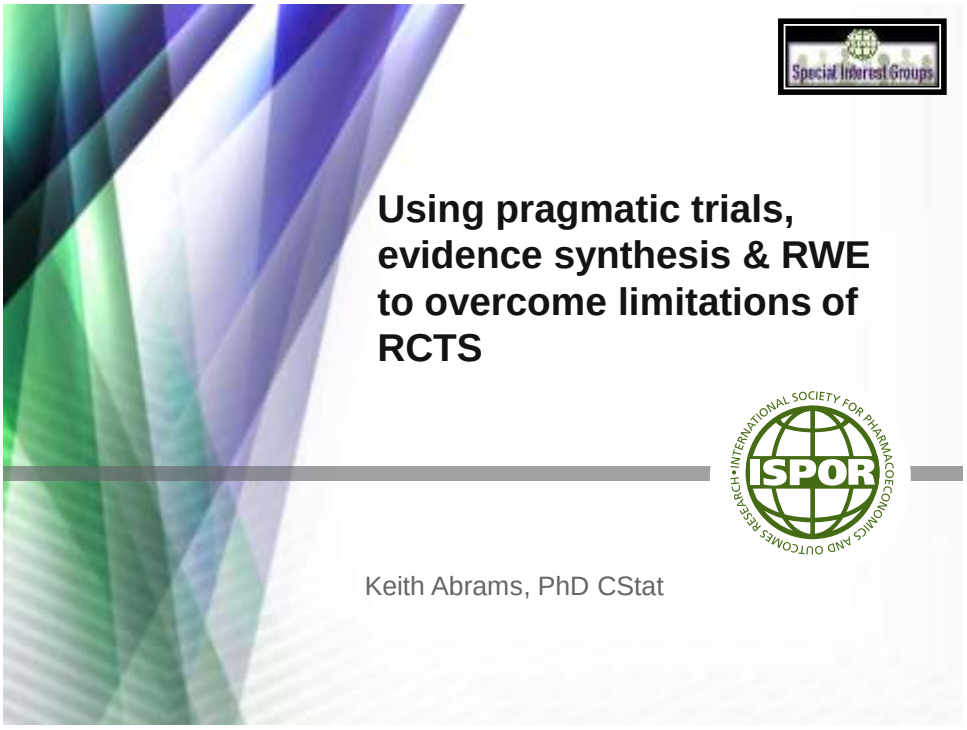
Topics for discussion

- Should we combine results from studies with complementary designs?
  - RCTs, Pragmatic Studies, Real World Evidence, where is the limit?
- Are more methods and evaluation of treatment effect heterogeneity and effect modifiers needed?

Literature

- **Recommended reading**
- Kaizar, E. E. (2011), Estimating treatment effect via simple cross design synthesis. Statist. Med., 30: 2986–3009. dx.doi.org/10.1002/sim.4339
- Verde P.E., Ohmann C. (2015), Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. Res Synth Methods. Mar;6(1):45-62. doi: 10.1002/jrsm.1122
- clinicalstudydatarequest.com
- **Other**

- GAO Report available from www.gao.gov/products/PEMD-92-18
- Kristina Gemzell-Danielsson, Ilka Schellschmidt, Dan Apter, A randomized, phase II study describing the efficacy, bleeding profile, and safety of two low-dose levonorgestrel-releasing intrauterine contraceptive systems and Mirena, Fertility and Sterility, Volume 97, Issue 3, March 2012, Pages 616-622.e3
- Justin T. Diedrich, Qiuhong Zhao, Tessa Madden, Gina M. Secura, Jeffrey F. Peipert, Three-year continuation of reversible contraception, American Journal of Obstetrics and Gynecology, Volume 213, Issue 5, November 2015, Pages 662.e1-662.e8
- Abraham M, Zhao Q, Peipert JF. Young Age, Nulliparity, and Continuation of Long-Acting Reversible Contraceptive Methods. Obstet Gynecol. 2015 Oct;126(4):823-9.



Problems with regulatory Phase 3 RCTs

- Population – often restricted, and not (totally) representative of broader target population to be treated
- Length of follow-up – often restricted to shorter term surrogate outcomes
- Other concomitant medication may be limited (and not appropriate for all jurisdictions) or excluded

- All these problems mean that decision makers (especially HTA) are faced with considerable uncertainty.

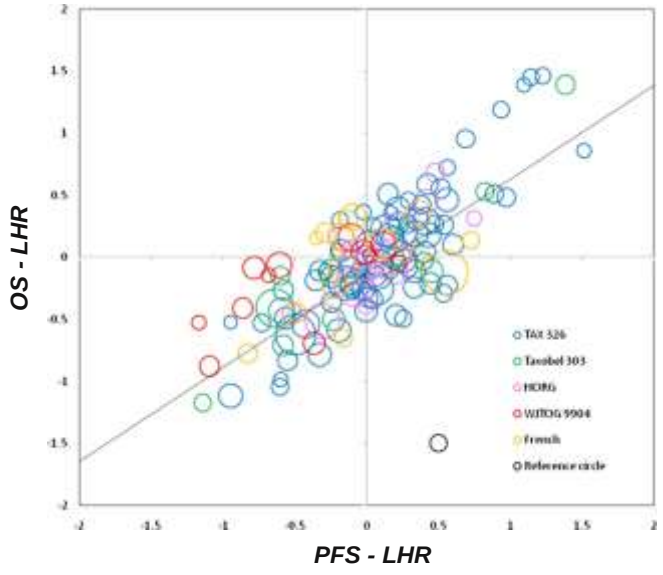
Solutions to these problems …

- Undertake modelling (extrapolation) of RCTs to target population using longer term (patient/decision maker) relevant outcomes …
  - How to generalise to broader target population? Eg IMI GetReal case study in NSCLC using propensity score-re-weighting
  - How to map from shorter term to longer term outcomes? Eg PFS & OS in NSCLC using meta-regression
- Undertake a pragmatic RCT to address these problems
- Or do both … as Decision Makers will require evidence quickly(!) after regulatory approval – the 'best' option will very often depend on context & disease/outcomes

http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.5%20and%201.6%20Combined%20Report%20-%20NSCLC_webversion.pdf

Laporte *et al*. *BMJ Open* 2013;3:e001802. doi:10.1136/bmjopen-2012- 001802

OS & PFS in NSCLC (Laporte *et al*, 2013)



Solutions to these problems …

- Undertake modelling (extrapolation) of RCTs to target population using longer term (patient/decision maker) relevant outcomes …
  - How to generalise to broader target population? Eg IMI GetReal case study in NSCLC using propensity score-re-weighting
  - How to map from shorter term to longer term outcomes? Eg PFS & OS in NSCLC using meta-regression
- Undertake a pragmatic RCT to address these problems
- Or do both … as Decision Makers will require evidence quickly(!) after regulatory approval – the 'best' option will very often depend on context & disease/outcomes

http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.5%20and%201.6%20Combined%20Report%20-%20NSCLC_webversion.pdf

Laporte et al. BMJ Open 2013;3:e001802. doi:10.1136/bmjopen-2012- 001802

39

pRCT as a solution …

- Population – broader than regulatory RCT, but how broad is broad?
- Length of follow-up & outcomes – longer term patient and DM relevant outcomes, but how can these be captured and in timely manner?
- Standard practice allowed along side experimental treatments, but how do we capture what other treatments patient receive?

- Potential solution to these problems -> **nested pRCT** on a patient platform (based on EHRs) together with a cohort of non-randomised patients – **Trial within Cohorts (TWICs)** or **Comprehensive Cohort Design** approaches.

https://www.twics.global/

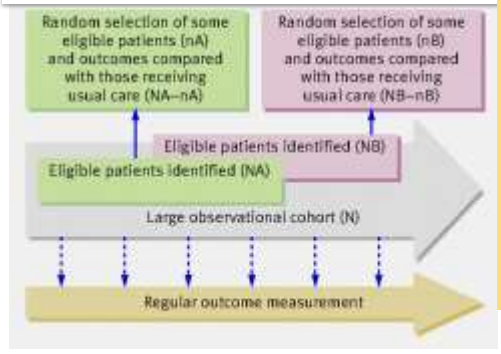
Schmoor *et al*. Stat Med.1996 Feb 15;15(3):263-71.

40

TWICs Approach

Example - Campbel *et al*. *J Health Serv Res Policy*. 2005 Oct;10(4):220-5.

Modified Zelen-design in arthritis evaluating intensive physiotherapy.



Potential Benefits:

- Facility for multiple RCTs
- Long term outcomes as standard
- Ongoing information as to the natural history of condition with SC
- Increased comparability between each RCT within cohort
- Increased efficiency
- More consistent indirect comparisons

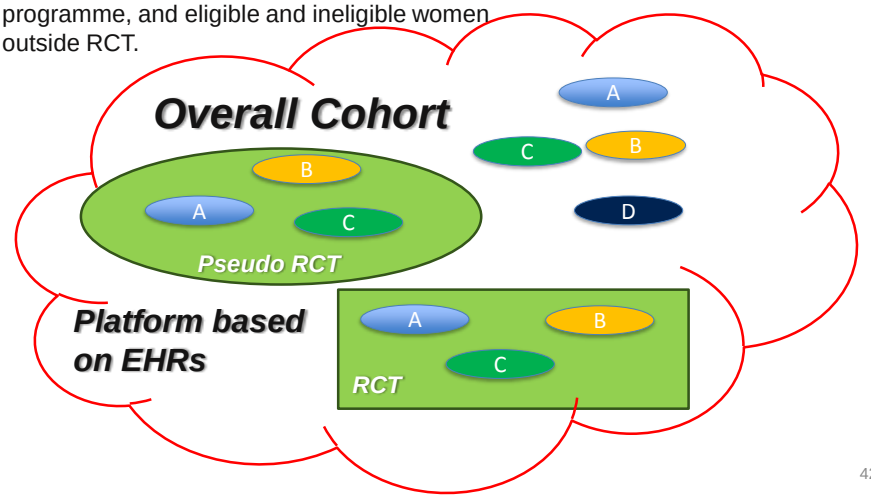
41

Relton *et al. BMJ* 2010;340:bmj.c1066

Comprehensive Cohort Approach

Example – Porthouse *et al*. *QJM* 2004;97:569.

Fracture rates in elderly women within RCT of fracture prevention programme, and eligible and ineligible women outside RCT.



42

Discussion

The Evolution of Value in Health Care

What role for Pragmatic trials?

Topics for Discussion

Helene

- Why all the buzz about pragmatic trials? Can't we just do RCTs, and complement with observations in usual care practice?
- Does it depend on the indication (or other factors?) if it is worth conducting pragmatic trials?
- Pragmatic trials help uncover (relative) effectiveness of interventions in usual care settings. Aren't there alternatives to conducting pragmatic trials to answer this question (e.g., using observational /registry data)?

Christoph:

- Should we combine results from studies with complementary designs?
  - RCTs, Pragmatic Studies, Real World Evidence, where is the limit?
- Are more methods and evaluation of treatment effect heterogeneity and effect modifiers needed?

Keith:

- Do TWICs or CCSs (using patient platforms) allow us to design more efficient RCTs and indirect comparisons?
- Does the use of patient platforms allow longer (and more efficient) follow-up that would otherwise be considered in RCTs?
- Are they more suited to non-pharmacological interventions?

Sign up as a Review Group Member

- Join ISPOR Special Interest Groups
- Need ISPOR membership number
- Give business card to Clarissa Cooblall or e-mail ccooblall@ispor.org
- Submit your evaluation using the ISPOR app