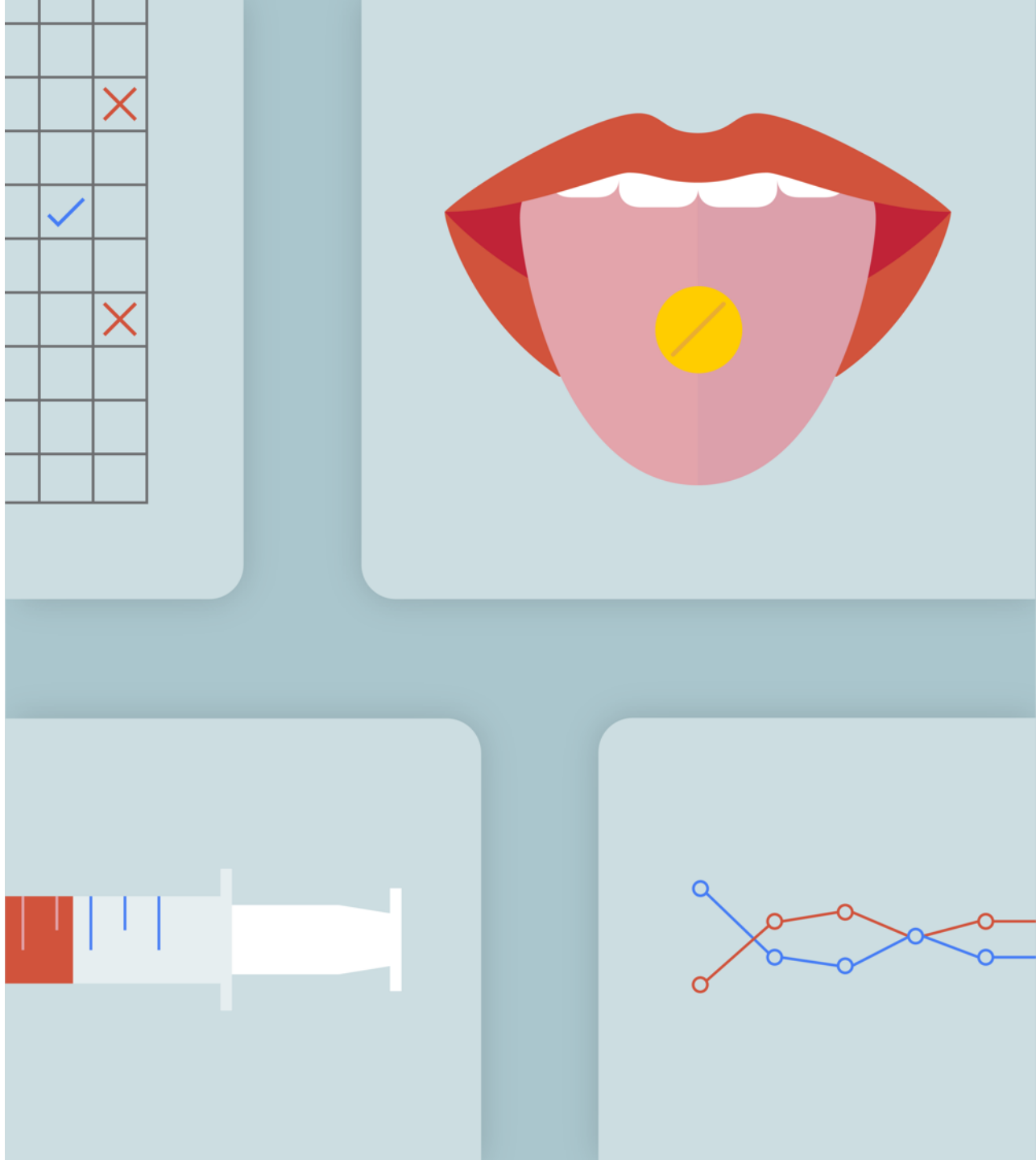
Milou Hogervorst, Lourens Bloem & Wim Goettsch

Update Results GetReal Single-Arm Trials & External Controls



Utrecht University

Pharmacoepidemiology
and Clinical Pharmacology

Content

- Goals of overarching SAT & EXT control study
- Current progress
- Initial results on:
  - Literature review
  - Regulatory review
  - HTA review
- Suggestions for recommendations stage 2 and 3
- Discussion

Three overarching goals SAT project

1. Literature review for background, context, and landscaping of external comparators
   - Including a comparison of regulatory and HTA approaches and recommendations that have been identified

2. Interviews and workshops with experts from different stakeholder groups to inform recommendations and the development of guidance, based on stage 1
   - Deliverable: report/paper to publish in peer-reviewed journal

3. External engagement through conferences/webinars and with professional bodies, stakeholders and groups to facilitate/enable translation of recommendations into practice and, if necessary, the development of guidance/guidelines.
   - e.g. ISPE, ISPOR, EMA, FDA. EUnetHTA, Duke Margolis RWE Collaborative

Specific deliverables for the Institute will need to be better clarified nearer the time.

Three overarching goals SAT project

1. Literature review for background, context, and landscaping of external comparators
   - Including a comparison of regulatory and HTA approaches and recommendations that have been identified → **CURRENT STAGE**

2. Interviews and workshops with experts from different stakeholder groups to inform recommendations and the development of guidance, based on stage 1
   - Deliverable: report/paper to publish in peer-reviewed journal

3. External engagement through conferences/webinars and with professional bodies, stakeholders and groups to facilitate/enable translation of recommendations into practice and, if necessary, the development of guidance/guidelines.
   - e.g. ISPE, ISPOR, EMA, FDA. EUnetHTA, Duke Margolis RWE Collaborative

Specific deliverables for the Institute will need to be better clarified nearer the time.

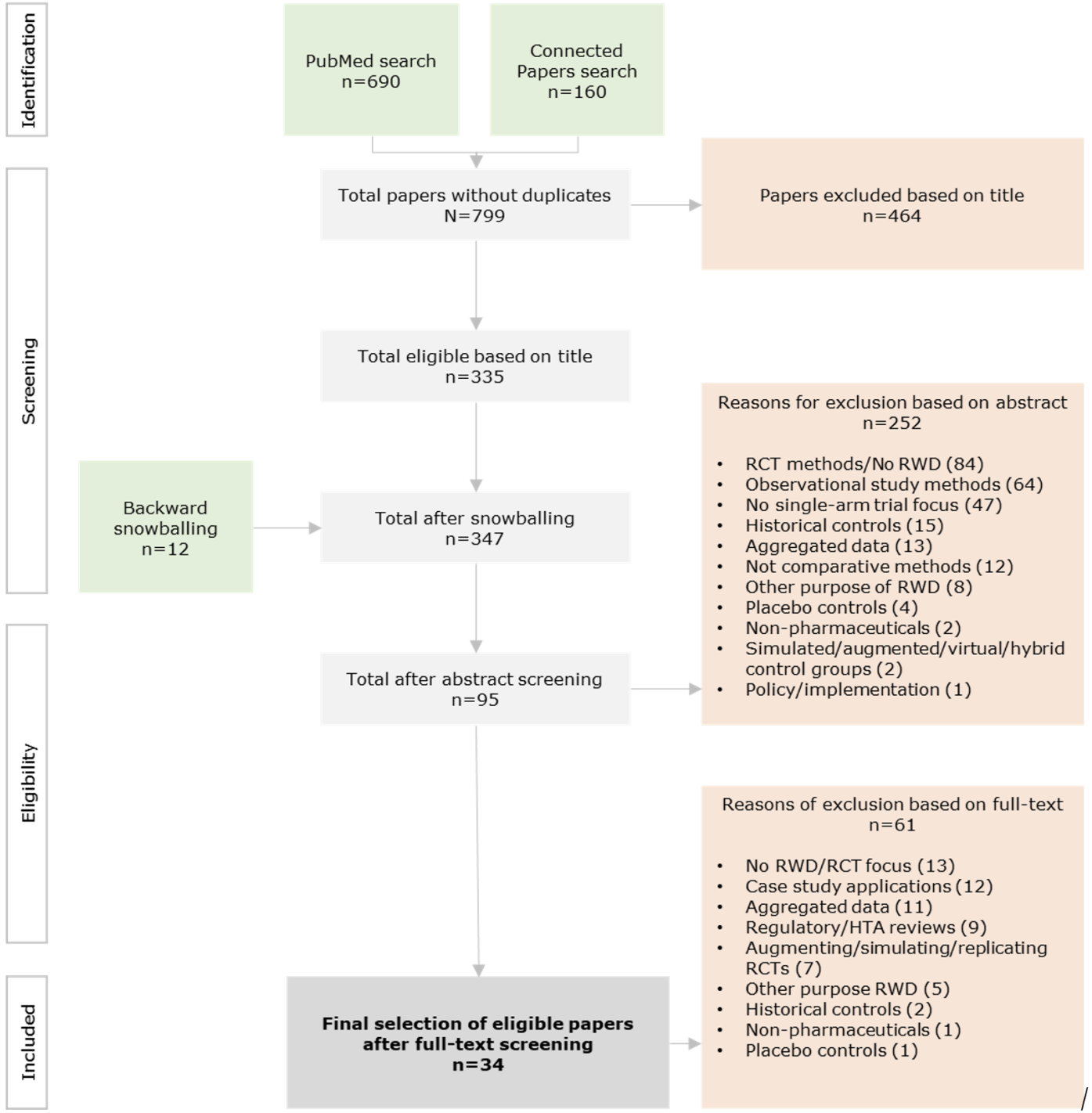
Current progress

- Literature review
- Guidelines regulators & HTA
- Review EPARs 1995-2020
  - Medicines initially authorized based on SATs
- Review HTA reports 2016-2020
  - Medicines authorized by EMA based on SATs, assessed by England, France, Germany, Netherlands, EUnetHTA

→ All have been performed, currently finalising analysis results

Literature Review

Overview of current developments in the field of external comparators, targeting specifically real-world data (RWD) based external comparators and methods for indirect comparisons to uncontrolled (single-arm) trials.



Main results of review (1)

- General
  - Emulate as much as possible single-arm trial (prevent typical observational biases
  - Data sources & quality important (garbage in / garbage out)
  - Transparency is needed on protocols

- Generating control arms
  - Studies a priori designed and preferably collected data
  - Key checks on *in and exclusion criteria*, *identical exposure status*, *outcome measurement identical* and *timing (i.e. OS/PFS time to event)*
  - Propensity score can be used, different methods are suggested
  - Cohort should be prepared before actual comparison
  - Careful check balance baseline characteristics before comparison

Main results of review (2)

Methods for comparing control arm to single arm trial

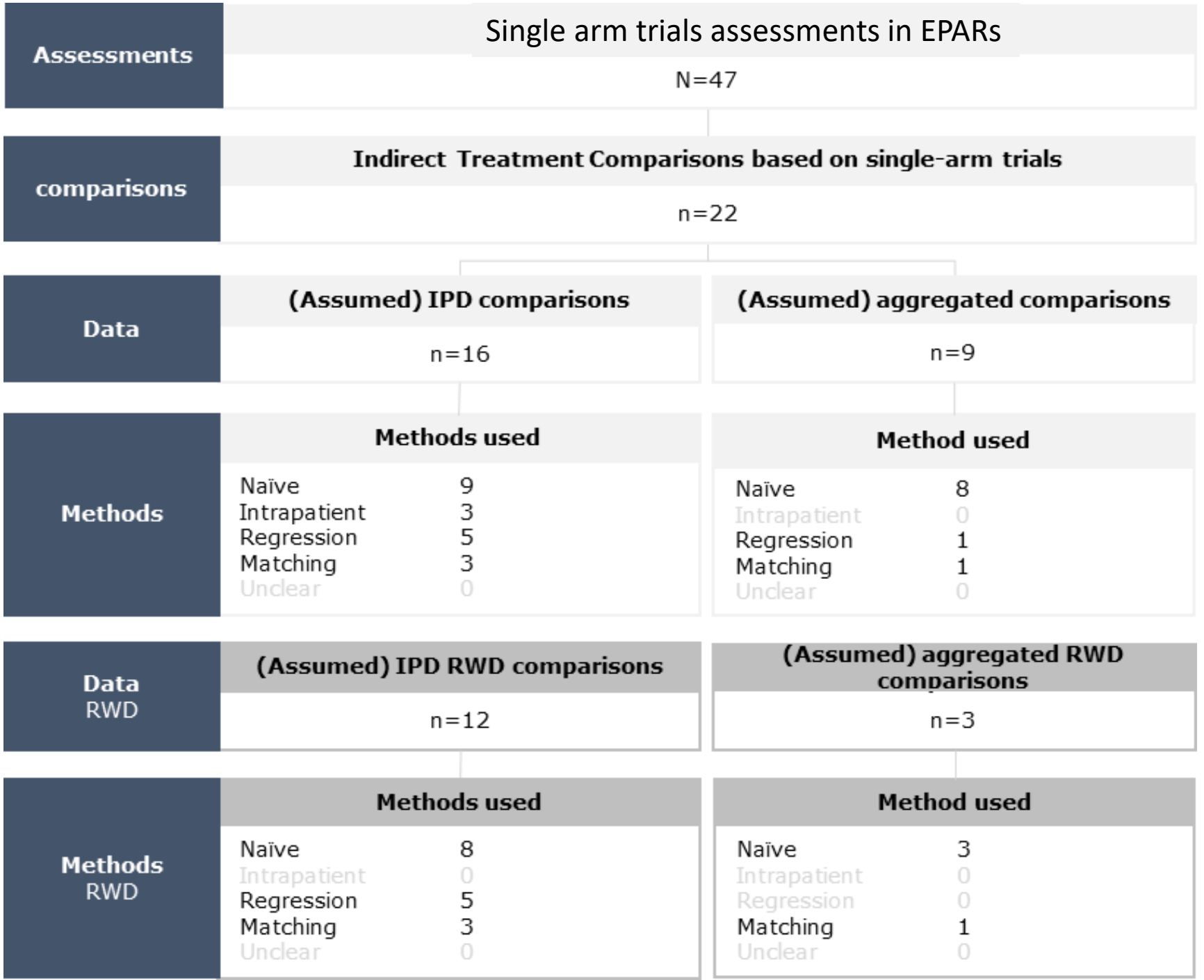
- Methods using propensity scores are most described
  - Weighting, (imbalance) adjustment, (fine) stratification, matching, marginal structural modelling
- Other methods
  - Naïve, (imbalance adjustments), restriction methods, matching methods, stratification methods, instrumental variable methods, ….
- Multiple sensitivity analyse important to visualize residual bias
  - Quantitative Bias analysis for instance Monte Carlo Simulations, ….
  - Formalising the range of bias estimates into Bayesian priors

Overview regulatory and HTA guidelines

- FDA predominantly concerns the generation of the control arm
  - Emulate as much as possible single-arm trial (prevent typical observational biases)
- EMA not yet published

- NICE (DSU)
  - Focus on comparison methods (order of preference)
  - Multi-Level Network Meta Regression (ML-NMR), Simulated Treatment Comparison (STC), Matching Adjusted Indirect Comparison (MAIC)
- HAS
  - Not just SAT with RWD, a priori study design, no specific methods preferred, justification lack of randomization
  - Describe issues that are important for the analysis of results

Assessment of methods used in regulatory market authorization assessments (2005-2020)



Assessment of HTA reports (2016-2020)

HTA Single-Arm Assessments

No. of single-arm trials assessed in HTA (drug-indication combinations N=92)

HAS	23
IQWiG	36
NICE	26
ZIN	7

Single-Arm comparisons

No. of Indirect Treatment Comparisons based on single-arm trials (drug-indication combinations n=66)

HAS	14
IQWiG	19
NICE	26
ZIN	7

Data used Any external source

IPD (any external data source; n=25)

HAS	5
IQWiG	9
NICE	10
ZIN	1

Aggregated (any external data source; n=39)

HAS	8
IQWiG	8
NICE	17
ZIN	6

Methods

Naïve (6)

HAS	0
IQWiG	3
NICE	3
ZIN	0

Intra-patient (9)

HAS	1
IQWiG	3
NICE	4
ZIN	1

Regression (4)

HAS	2
IQWiG	0
NICE	2
ZIN	0

Matching (8)

HAS	3
IQWiG	3
NICE	2
ZIN	0

Naïve (24)

HAS	1
IQWiG	8
NICE	11
ZIN	4

Intra-patient (1)

HAS	0
IQWiG	0
NICE	0
ZIN	1

Regression (8)

HAS	3
IQWiG	1
NICE	3
ZIN	1

Matching (21)

HAS	7
IQWiG	2
NICE	11
ZIN	1

Data used RWD comparison

IP (external RWD source; n=17)

HAS	4
IQWiG	7
NICE	6
ZIN	0

Aggregated (RWD source; n=18)

HAS	4
IQWiG	3
NICE	8
ZIN	3

Methods RWD comparison

Naïve (5)

HAS	0
IQWiG	3
NICE	2
ZIN	0

Intra-patient

HAS	0
IQWiG	0
NICE	0
ZIN	0

Regression (4)

HAS	2
IQWiG	0
NICE	2
ZIN	0

Matching (8)

HAS	3
IQWiG	3
NICE	2
ZIN	0

Naïve (12)

HAS	0
IQWiG	4
NICE	6
ZIN	2

Intra-patient

HAS	0
IQWiG	0
NICE	0
ZIN	0

Regression (4)

HAS	2
IQWiG	1
NICE	1
ZIN	0

Matching (11)

HAS	3
IQWiG	1
NICE	6
ZIN	1

Key results of HTA reports

- Multiple comparisons in one assessment

- In 50% of assessments RWD is used as comparator

- In majority of cases only summary data on external comparator data available
- Large variations per agency on how external controls are considered
- Exact use methods often not described
  - Data source (IPD vs aggregated)
  - Methods (naïve, matching, regression)

- For aggregated data naïve comparison are commonly used (even for IPD!)
- Methodological more sound methods are often not used

Conclusions

- Methods described in literature in more detail than mentioned in guidelines
- Literature focuses on details, and the relevance of these details to bias prevention
  - whereas guidelines are general
  - descriptions in regulatory/HTA reports are brief, difficult to assess potential biases remaining/unaddressed
- Guidelines align with literature preferences
- Literature suggests different preferred methods (ML-NMR) than those that are used in regulatory and HTA assessments.

Recommendations

- There is a need for a set of clear definitions because many very different terms are used. Therefore, It is confusing to create an overview of distinctions between methods
- Need for guidelines on consistent reporting of methods & data used to assess external comparators
- Preferably an analysis plan for external comparisons should be defined a priori
- Transparency is key, devil is in the details!
- Dialogue is needed to decrease discrepancy between the willingness to use more sophisticated methods and the trust in these methods/ability to correctly use and interpret these methods.
  - Literature suggests different preferred methods (ML-NMR) than those that are used in regulatory and HTA assessments.

- Externally controlled trials should as much as possible emulate RCT design
- Use preferably IPD results in less biased comparative methods (but is often not available). There is a need for better access to IPD.
- Assert that efforts have been made to investigate the robustness of sensitivities of the underlying methods for employing synthetic controls: 1) run multiple comparisons using different methods, 2) do these methods show concordant results?