Real World Evidence for Coverage Decisions: Opportunities and Challenges

A Report from the 2017 ICER Membership Policy Summit

March 2018

Grace Hampson, MSc
Senior Economist
Office of Health Economics

Adrian Towse, MPhil
Director
Office of Health Economics

Bill Dreitlein, PharmD, BCPS
Director of Pharmaceutical Intelligence
Institute for Clinical and Economic Review

Chris Henshall, PhD.
Independent Consultant

Steven D. Pearson, MD, MSc
President
Institute for Clinical and Economic Review

Table of Contents

ABOUT ICER

The Institute for Clinical and Economic Review (ICER) is an independent non-profit research institute that produces reports analyzing the evidence on the effectiveness and value of drugs and other medical services. ICER's reports include evidence-based calculations of prices for new drugs that reflect the degree of improvement expected in long-term patient outcomes, while also highlighting price levels that might contribute to unaffordable short-term cost growth for the overall health care system.

ICER's reports incorporate extensive input from all stakeholders and are the subject of public hearings through three core programs: the California Technology Assessment Forum (CTAF), the Midwest Comparative Effectiveness Public Advisory Council (Midwest CEPAC) and the New England Comparative Effectiveness Public Advisory Council (New England CEPAC). These independent panels review ICER's reports at public meetings to deliberate on the evidence and develop recommendations for how patients, clinicians, insurers, and policymakers can improve the quality and value of health care.

ABOUT OHE

The Office of Health Economics (OHE) has over 50 years' experience of conducting high quality research on the economics of innovation and the life sciences industry, the organisation and financing of health care, and the role for outcomes research and health technology assessment. The OHE's current work programme is supported by research grants and consultancy revenues from a wide range of UK and international sources.

The OHE is a not-for-profit company limited by guarantee. It is a registered UK Charity (registration number 1170829). Its work is overseen by its Editorial and Research & Policy Committees reporting to its Board of Trustees.

ACKNOWLEDGEMENTS

The authors would like to thank Alexandra Ellis and Dan Ollendorf at ICER for providing wording on the use of network meta-analysis, and, also, the nine interviewees for sharing their expertise and insights into current the opportunities and challenges for real world evidence. All remaining errors are ours.

EXECUTIVE SUMMARY

Real world evidence

The data landscape is changing. Our capacity for rapid data accumulation and data interpretation is advancing exponentially. Computer learning, natural language processing, and the evolution of electronic health records are revolutionising the potential availability and use of real world data sources to improve health.

At the same time, traditional randomised controlled trial (RCT) evidence may be declining as smaller patient populations (related to more personalised medicine) make it harder to design studies, and the per-patient and set up costs of conducting RCTs are rising (related to increases in complexity and external standards). Interest in pragmatic clinical trials (PCTs) that combine randomisation with more real world circumstances has grown with the potential use of routine data sources to record patient events and outcomes, transforming the costs, size and feasibility of such trials. This changing environment is creating new opportunities for the use of real world evidence (RWE).

This paper

This paper sets out the potential opportunities and important challenges and limitations that must be addressed in considering options for using RWE to inform insurer coverage decisions. The primary purpose of developing the paper was to stimulate discussion at the 2017 ICER Policy Summit meeting. A separate paper is available that summarises the authors reflections and proposed ways forwards based on the discussions that were had at the meeting.

Current uses of RWE and related initiatives

RWE is already utilised for multiple purposes in the US and globally. We provide a brief overview of existing use as a platform to discuss gaps and further developments.

Current use in the US:

1. Drug development: RWE is used to identify targets for the development of new therapies and design the drug development pathway.
2. Regulatory approval decisions: Use of RWE in initial FDA regulatory decisions has been limited to date to circumstances where an RCT is not practical, but the FDA has recently (August 2017) released guidance on the use of RWE to support regulatory decision-making for medical devices and is required to issue RWE guidance for drugs under the 21st Century Cures Act.
3. FDA safety monitoring and safety signals: FDA use of RWE to monitor post-approval drug safety is much more established, most recently via the Sentinel Initiative.
4. HTA assessments and payer coverage decisions - initial decisions: Payers use epidemiological data, based in part on claims data, at this stage of decision making to generate estimates of the potential population they cover that could require the treatment, and to estimate potential costs and cost offsets.

5. HTA assessments and payer coverage decisions – reassessments: RWE gives decision makers the opportunity to reconsider coverage, discounts and formulary tiering in light of how the products are performing in their relevant population.
6. Outcomes-based contracting: Patient outcomes are tracked in order to support contract agreements tying level of reimbursement to real-world clinical performance. To date, outcomes-based agreements have not featured prominently in the US health care system because of the difficulty of collecting data to support such agreements, but interest appears to be growing.

The focus of the 2017 ICER Policy Summit was on RWE for coverage decisions, thus the majority of this paper focuses on points 4 and 5 above.

Challenges associated with the use of RWE

Whilst some stakeholders see great value in RWE and are exploring ways to make use of these data sources, our literature review and interviews with experts conducted to inform this paper identified inherent limitations and numerous challenges associated with RWE application to coverage decisions. Acceptance of an expanded future role for RWE is not universal, particularly if it is seen as reducing the amount of RCT evidence available. Among the challenges associated with RWE explored in this paper are: bias and confounding; incomplete data; data mining; access to data; and the lack of universally accepted methodological standards.

Key opportunities for RWE

In this paper we discuss the current uses of RWE that offer important opportunities in the near term for expanded application to coverage decision-making, as well as areas that will require longer-term efforts for future development.

1. Improving current uses of RWE:

- Evaluation of drug effectiveness, safety, and adherence in real-world patients;
    - Using RWE to evaluate the durability of benefits and side effects over a longer period than studied in RCTs
    - Exploring population subgroups in which clinical benefit is (likely to be) greatest
    - Gaining comparable evidence on the new drug and on the comparator ("usual care")
    - Evaluating benefits when used outside of the initial indication
    - Leveraging the advantages of pragmatic clinical trials to inform all aspects of the evaluation of drug effectiveness and safety
- Evaluation of comparative effectiveness through indirect comparisons (network meta-analysis) enriched with outcomes from real-world patients
- Evaluation of outcomes that are not measured during the standard development process, for example any "other benefits" or wider elements of value such as the impact on productivity
- Evaluation of budget impact and cost-effectiveness in a real world setting.

2. Key opportunities for the future:

- Innovative study designs such as nested trials within cohorts can combine benefits of collecting data from real world settings while incorporating best practice methods (i.e. randomisation methods from traditional RCTs).
- Real time monitoring of patients: Wearables, and the Apple Research Kit, for example, will enable some data collection to become routine. These could reduce the cost of evidence generation, expand the available dataset, and allow remote monitoring for better medication management.
- Developing adaptive regulatory pathways linked to coverage with evidence development to monitor the safety and effectiveness of new treatments with highest uncertainty. The ability to collect post-launch RWE is crucial for accelerated access arrangements to be successful from both payer and innovator perspectives.

1. INTRODUCTION

1.1. Real World Data and Real World Evidence

The capacity of the US health care system to generate and interpret large amounts of data is advancing exponentially. Computer learning, natural language processing, and the evolution of electronic health records (EHRs) are revolutionising potential availability and interpretability of real world data sources to improve health. Indeed, recent estimates suggest that around 83% of US office-based physicians and 84% of hospitals now use some form of EHR (BPC, 2016). The ability of patients to use data to manage their own conditions is also increasing, with one survey claiming that 70% of American adults are tracking at least one indicator of their health (BPC, 2016). For some, this will be done via patient wearables, such as Fitbits and the Apple Watch. Social media can also play a role. Twitter data has been used to predict emergency department visits due to influenza outbreaks (Nagar et al., 2014) and Facebook and Twitter data have been examined to explore whether safety signals can be picked up before they are reported to the FDA (Pierce et al., 2017). These technologies are providing new functionality, and can, in principle, contribute to a much richer and larger data set for predictive purposes.

Concurrently, exclusive reliance upon traditional randomised controlled trial (RCT) evidence may be declining for several reasons. Firstly, as medicine becomes more personalised, or targeted, study sizes decrease, making it harder to conduct RCTs. Secondly, the per-patient and set up costs of conducting RCTs have been rising, with increases in complexity and the need to adhere to external standards such as those set by the International Conference on Harmonization (ICH). Pragmatic clinical trials (PCTs) offer the hope that the advantages of randomisation can be combined with the added relevance of results obtained in more real-world clinical settings, but these trials often remain more expensive and more complex to design than traditional RCTs.

From the perspective of health technology assessment (HTA) organizations and payers, as FDA and manufacturers have increased the use of accelerated regulatory pathways, more new drugs are coming into the market with limited RCT evidence, and in some cases no RCT evidence, while many drugs continue to be tested in narrow, refined patient populations that do not represent well the clinical diversity of real-world populations. Assessments of clinical and cost-effectiveness thus must look to RWE where possible to supplement the information generated in pivotal studies.

It is therefore crucial that developers of innovative products and the HTA/payer community understand the potential for RWE to inform coverage decisions and price negotiations. To understand this potential, the opportunities presented by RWE must be viewed in balance with its important limitations, both conceptual and practical. Data from EHRs are often incomplete, incompatible with those from other institutions, or hard to obtain given privacy concerns. RWE arising from observational studies can never escape concerns about biases introduced by unknown confounders, while these datasets are also more prone to data mining than those from RCTs.

In the end, payers need to understand clearly how RWE can be helpful, and how it can mislead. They need to know how to determine the level of trust they can have in any analysis of RWE, whether of their own data or of data from another source. And manufacturers and payers need further guidance on how to work together to communicate effectively and collaborate in the generation, interpretation, and application of RWE to important health policy decisions.

The purpose of this paper is not to recap existing discussions around RWE. We recognize that there is a great diversity of issues that relate to RWE and that other organizations are also debating how best to advance policy and practice in this area. Instead, we summarize the key concerns that have been raised in this area, and set out various opportunities that could in principle be realised through best use of RWE for coverage decisions, providing these concerns can be effectively navigated, mitigated, or resolved. This paper briefly sets out:

- A working definition of RWE for the purpose of this paper and the 2017 ICER Policy Summit;
- Current uses of RWE in US health care;
- Current uses and research initiatives related to RWE outside of the US.

We then explore the following in greater detail:

- Key opportunities for the use of RWE;
- Challenges associated with these key opportunities;
- Ways in which the opportunities can be progressed.

1.2. Approach to preparing the paper

In preparing this Briefing Paper we undertook a literature review, supplemented with nine interviews with experts in this field:

- The literature review was designed to be pragmatic rather than systematic, and was undertaken to identify key opportunities and challenges for the use of RWE, rather than to identify all papers on RWE;
- The interviewees were selected and invited to interview by ICER or the Office of Health Economics (OHE). They included two representatives from US payers, one representative from the National Institute for Health and Care Excellence (NICE) in the UK, three industry representatives, one representative of an organisation working on the collection of RWE, and two academics, all of whom are experienced in the field of RWE.
- We also viewed two September webinars – the FDA-Duke Margolis meeting on the use of RWE in regulatory settings, and the National Academies of Sciences Engineering and Math (NASEM) meeting on incentives for the use of RWE to improve the efficiency of innovation.

2. WHAT IS RWE AND HOW IS IT USED?

2.1. Definition

A full review of the various existing definitions of real world evidence is outside the scope of this paper. Instead, for the sake of clarity throughout this report and the 2017 Policy Summit, we adopt the following definitions that are adapted from the FDA:

> ***Real-world data (RWD)*** *are data relating to patient health status and/or the delivery of health care collected either prospectively or retrospectively from observations of routine clinical practice. Examples of RWD include data derived from EHRs, claims and billing data, data from product and disease registries, patient-generated data including in home-use settings, and data gathered from other sources that can inform on health status, such as mobile devices*.
>
> ***RWE*** *is the clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of RWD"* (FDA, 2016)[1].

Even though evidence must be routinely collected to qualify as RWE according to the FDA definition, authors writing on behalf of the FDA clarify that the distinguishing feature of RWE is the setting in which the evidence is collected (i.e. it is collected in health care settings rather than in a research environment), and *not the presence or absence of a planned intervention or the use of randomization* (Sherman et al. (2016). As such, we believe that it is useful in the US context to include PCTs within this definition. This is not a view shared by all of our interviewees.

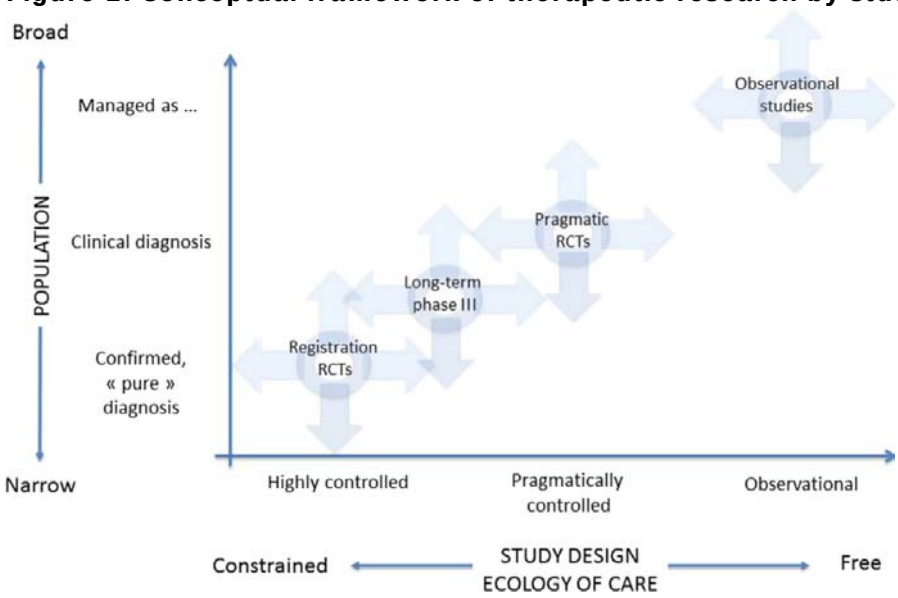
Figure 1, shown on the following page, taken from Roche et al. (2013), provides a framework which shows how observational studies (for example case-control studies; cross-sectional studies; cohort studies) and PCTs compare to RCTs. RWE offers less constrained study designs and broader populations as compared to more traditional RCTs. These attributes offer the possibility of obtaining evidence that more accurately reflects the "true" outcomes of care in real-world care settings, where patients, clinicians, treatment regimens, and all facets of care may be more representative than those involved in the highly controlled RCTs and other trials traditionally used for drug development studies. More traditional RCTs, on the other hand, minimize risks that unidentified biases in patient and clinician selection will provide misleading, invalid results. The complementary strengths and weaknesses of controlled trials and of RWE are often expressed in terms of a trade-off between "internal" and "external" validity, or of a trade-off between internal validity and generalizability.

2.2. What is happening with use of RWE outside of the US?

The expert interviewees indicated that there is a trend towards more reliance on RWE for payer decision making outside the US. Several thought that Europe in particular was making more progress in this respect. We briefly review the examples in the literature and those most often referred to by our interviewees as examples that the US may be able to learn from.

[1] Note that these definitions are from a guidance document for devices (FDA, 2016)

Figure 1: Conceptual framework of therapeutic research by study design



Source: Roche et al., 2013

2.2.1.Examples of the use of RWE in HTA / payer decision making

In Europe, RWE is a recognised tool for accelerated access programs across several countries (Gill et al., 2016). One example of a conditional reimbursement scheme that relies on RWE is provided by NICE in the UK. NICE's observational data unit focuses on commissioning through evaluation (CTE) (this is a type of coverage with evidence development, or CED). CTE enables a number of patients to access treatments that are not widely funded by the UK's National Health Service (NHS), whilst data on their outcomes (clinical and patient experience) are collected. The purpose of CTE is to provide RWE for technologies that look to be beneficial, but have insufficient evidence to support national level implementation at launch. A formal research protocol is developed between NICE and NHS England that is strictly followed to avoid data mining. Participating centers submit data for assessment via NICE. The evidence is reviewed throughout and at the end of the scheme, and a national recommendation made. There are currently six of these schemes in operation - all involving procedures or devices, but a drug (rituximab for membranous nephropathy) has been accepted with the first patients receiving treatment in early 2018. One of our expert interviewees commented that there can be difficulties managing expectations. Once a hospital has committed to providing a technology, or surgeons have learned how to perform a new procedure, it can be difficult to remove it should a positive recommendation not be given.

We can also note that revised arrangements for the Cancer Drugs Fund limit use of a designated budget to coverage with evidence development schemes approved by NICE. To date two drugs and three indications are being reimbursed on this basis. The core evidence generation however – of progression free survival to overall survival – is being collected in continuing RCTs.

2.2.2. IMI GetReal

The Innovative Medicines Initiative (IMI) is a consortium of industry, academia, HTA agencies, regulators and patient organisations across Europe. In 2013, IMI launched GetReal,[2] a project which looks at how methods of RWE generation (defined as observational data) could be adopted earlier in the development and decision making processes to supplement RCT data. The project has several work packages which are looking across:

- the acceptability and usefulness of RWE, and approaches to the analyses of RWE, in estimating the effectiveness of new medicines;
- the scientific validity of RWE study designs and analytical approaches, to better inform pharmaceutical R&D and healthcare decision makers on their potential for use in assessment of effectiveness;
- the operational challenges of performing RWE studies early in the medicine development process and developing practical solutions to better inform their planning and delivery;
- best practices in identifying and sharing evidence syntheses and predictive modelling of different types of data to estimate effectiveness of medicines.

The IMI GetReal project has published its outputs. These are accessible on the website http://www.imi-getreal.eu/. Resources include: i) RWE Navigator, which is an interactive web-based educational resource on the types of RWE that can be used at various stages of drug development, the potential issues that may be incurred, and options for mitigating or overcoming the challenges, and ii) various publications and presentations that provide methodological guidance, recommendations, and other information gathered via the various work packages outlined above. Other IMI initiatives in this area include ADAPT SMART[3] (which we discuss below), DO-IT (Big Data for Better Outcomes), and PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium)[4].

2.2.3.EMA Adaptive Pathways Pilot and IMI ADAPT SMART

For regulatory decision making, Europe's European Medicines Agency (EMA) is seeking to develop what it calls "adaptive pathways" to mirror the accelerated pathways available to manufacturers through the FDA.[5] This exploration of adaptive pathways was part of the multi-stakeholder IMI ADAPT SMART (Accelerated Development of Appropriate Patient Therapies: a Sustainable, Multi-stakeholder Approach from Research to Treatment-outcomes) initiative. Two points are of note. The project struggled to get payer acceptance of reimbursement, even on a conditional basis, with limited evidence, and part of the reason for this was a reluctance on their part to accept that there were "methodologically sound strategies of real-world evidence collection to support the assessment of both efficacy and effectiveness" (EMA, 2016).

[2] http://www.imi-getreal.eu/
[3] http://adaptsmart.eu/
[4] http://www.imi-protect.eu/
[5] http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp

2.2.4. Innovative study designs: pre-launch PCT

The Salford Lung Studies (SLS) are highly cited examples of successful, pre-launch randomized PCTs. They were undertaken by a collaboration of eight UK NHS organisations and were sponsored by GlaxoSmithKline (GSK). The first study (Vestbo et al., 2016) – described below – evaluated the use of an investigational drug for the treatment of chronic obstructive pulmonary disease (COPD), while the second study (Woodcock et al., 2017) evaluated the drug for the treatment of asthma. The studies, which employed similar designs and methods, are considered to be the first of their kind.[6]

The SLS-COPD study was conducted in a clinical setting in the UK and included a larger (N=2,802), broader, more inclusive and representative (i.e. including those with comorbidities) population than is typical of a standard pre-launch RCT. The study was designed to maintain scientific rigour, whilst making the results as applicable to real clinical practice as possible in a pre-launch setting to inform both regulatory and payer decision making.

The study was able to show that COPD patients on a once-daily new inhaler had a significantly lower annual rate of moderate/severe COPD exacerbations (p=0.02), compared with patients receiving usual care. The incidence of serious adverse events was similar between groups. Results were published in September 2016 (Vestbo et al., 2016). The EMA accepted the study for purposes of fulfilling a post-approval safety study requirement, and CHMP opinion was rendered in April 2017, indicated that the post-approval commitment had been fulfilled.

The authors reported that the innovative design of the trial was a strength, and comment that this "*allowed important factors in usual clinical care, such as adherence, frequency of dosing, and persistence of good inhaler technique, to come into play*". However, they note that results need "careful interpretation", as the open label nature of the trial could have introduced bias. They conclude that their findings "*challenge the automatic transfer of findings from efficacy studies to clinical guidelines*", explaining that whilst RCTs are essential for new treatments, their populations are not necessarily representative of the patients who will need the treatment (Vestbo et al., 2016).

2.3. Use of RWE and related initiatives in the US

2.3.1. Use of RWE in the US

Interviewees identified six key areas in which RWE is being used in the US currently:

- **Drug development**: Epidemiological data helps to identify targets for the development of new therapies and can help inform decisions around the most appropriate drug development pathway. For example, evidence on adherence to similar therapies, population size, and clinical disease progression can help in making decisions about whether to move to the next development phase, and if so, what that phase should look like in terms of study design.
- **Regulatory approval decisions**: Use of RWE in initial FDA regulatory decisions to date has been limited (see Jarow et al., 2017 for a summary of use to date). Importantly, the 21st Century Cures Act (2016) specifically directs the FDA to

[6] https://www.nihr.ac.uk/life-sciences-industry/documents/Generic-casestudy-2016_SLS_web.pdf

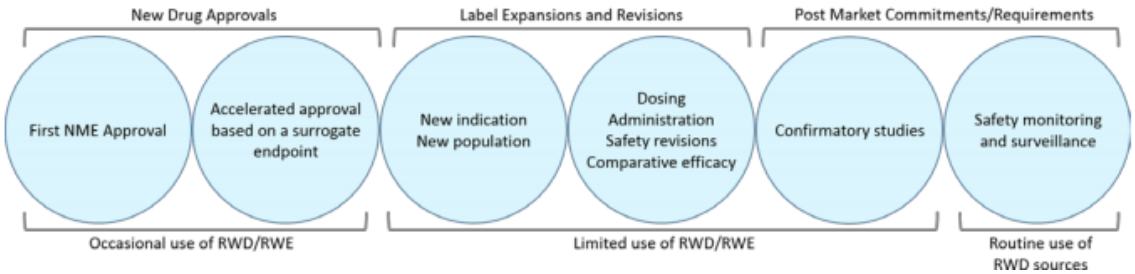
consider the use of RWE for approval of new indications for FDA-approved drugs. The FDA is currently working on how to approach this requirement.

- **Post-approval monitoring of safety signals**: FDA use of RWE to monitor post-approval drug safety is much more established. The most prominent effort is the FDA's Sentinel Initiative[7], a national electronic system that collects evidence on events and other targeted outcomes from a range of data partners who compile their analysis using EHRs, claims data, and other data generated within their system. However, there have been suggestions that Sentinel is not having a measurable impact: to date, despite costing $20 million per year, it has led to no drugs being removed from the market, and only two label changes. The FDA explain that this is partly because for the first eight years Sentinel was a pilot programme (Mini-Sentinel) and only became fully operational in 2016 (Findlay, 2017).

  Another example is the FDA's Adverse Event Reporting System (FAERS)[8], which collects information on adverse event and medication errors from healthcare professionals and product manufacturers. The FDA uses data generated via these systems to monitor medical products post-launch and take regulatory action if necessary. One study (Mostaghim et al., 2017) reported a rate of 0.68 safety-related label changes per drug per year (for non-expedited pathway drugs), based on a review of drugs approved by the FDA during the period 1997-2016. The most common change was to add a precaution to the drug label.

  Berger et al. (2017a) summarise the US experience with RWE and RWD for regulatory purposes in the diagram below (Figure 2). The diagram highlights the areas in which RWE is used occasionally, on a limited basis, and routinely, throughout approval and the post-launch period. Of note, the diagram shows that there is only limited use of RWE in confirmatory studies. These studies are often required after accelerated approval, but evidence suggests that only half of the required studies are completed three years post-approval. The quality of the studies that are undertaken has also been questioned, as study design is not stipulated by the FDA (Naci et al., 2017). Clearly this is one area in which high quality RWE could, in principle, play an important role.

Figure 2: Experience Utilizing RWD and RWE for Regulatory Purposes



Source: Berger et al., 2017(a)

[7] https://www.fda.gov/safety/fdassentinelinitiative/ucm2007250.htm
[8] https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm

- **HTA assessments and payer coverage decisions - initial decisions**: US payers often use epidemiological data, based in part on claims data, at this stage of decision making to generate estimates of the potential population they cover that could receive the treatment, and to estimate potential cost offsets. RWE can be used to estimate differing patient numbers depending, for example, on a drug's use in second or third line treatment. There is also the potential for RWE to inform the potential impact of any off-label use. This form of "innovation management" is often based on claims data or other epidemiological data.

  In making assessments of comparative effectiveness at launch, however, both HTA organizations like ICER and payers usually must rely largely on RCT or other controlled trial evidence generated for regulatory submission. Many Pharmacy and Therapeutics (P&T) Committees use evidence templates and conceptual approaches focusing on the RCT data of new drugs versus placebo, and perform indirect comparisons to active comparators only through qualitative means. Until very recently there has also rarely been patient-reported outcomes data available to supplement the information from pivotal trials.

  ICER's reports have always sought to emphasize comparative clinical effectiveness, and routinely uses network meta-analysis to conduct quantitative indirect comparisons when the patient populations and outcome measures are similar enough to make these analyses feasible. ICER has also begun working whenever possible with patient groups to analyse existing patient-reported data, or even to gather new information from patients to complement other data.

  ICER also routinely seeks RWE when available to guide the estimates for clinical and economic outcomes in its cost-effectiveness models. Almost always this RWE is only available for existing treatments and not for the emerging drug. Payers may also seek evidence from decision analytical modelling performed by the manufacturer (for example by requesting manufacturer submissions using the AMCP dossier format).

  The existence of RWE for existing treatment options creates a common dilemma when HTA groups and payers assess the comparative effectiveness of a new drug that has no RWE available yet. Even when RWE is viewed as providing important insights into the performance of existing treatments, HTA/payers must determine whether and how to use it to judge the comparative effectiveness of existing versus emerging treatment options. Is it better to compare only the pivotal trials of both existing and emerging treatments, or should the RWE on existing treatments be used somehow, even if the RWE comes from observational studies or pragmatic clinical trials with very different types of patients? This specific dilemma will be addressed further in section **Error! Reference source not found.** of this paper.

- **HTA assessments and payer coverage decisions – reassessments**: RWE gives HTA/payers the opportunity to reconsider coverage, formulary placement, and price/payment terms in light of how the products are performing in their relevant population. Formularies are updated regularly, and with competing product entry or the end of a contract period with the manufacturer for a particular product, this provides an opportunity for RWE to be worked into the review process by both HTA/payers and manufacturers. In addition to evidence on real-world safety and effectiveness, manufacturers can use RWE to provide

payers with evidence on dimensions of value (from either a clinical and economic perspective) that are not evaluated during drug development.

Interviewees suggested that evidence of real world effect is what matters to payers. Uses can be as broad as: trying to identify or rule out safety signals; measure adherence in order to evaluate whether a drug that is slightly more tolerable could be more effective in the real world if patients actually take it; establish effectiveness and value for money within the health plan's specific population (i.e. RWE can be a good test of external validity, as the product is evaluated within a more representative sample); or to establish effectiveness within sub-populations. This use of RWE for reassessment coverage decisions was highlighted and the most discussed by our expert interviewees.

- **Outcomes-based contracting**: The final area in which RWE is currently used by payers is in outcomes based contracting, in which payment is linked via rebate levels or some other mechanism to the demonstrated real-world outcomes of patients. Examples from the US include:

  i) Novartis recently announced an agreement with CMS where CMS will cover Kymriah® (CAR-T therapy) only if patients respond within the first month after treatment;

  ii) Merck agreed to provide rebate payments to Cigna and Prime Therapeutics for its MS drug Rebif if hospital visits were required due to relapses (QuintilesIMS, 2013).

  To date, outcomes-based agreements have not featured prominently in the US health care system, largely because of the difficulty of collecting the RWE that could support such agreements (Garrison et al. 2015). However, interest in outcomes-based contracting may be growing. Harvard Pilgrim Health Care, for example, signed a three-year contract in 2017 with AstraZeneca for therapies used to treat acute coronary disease and type 2 diabetes (Business Wire, 2017).

2.3.2. Related Initiatives in the US

In addition to these uses of RWE, there are various initiatives ongoing which are looking at how RWE can be used in decision making. We set out three:

- **The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the International Society for Pharmacoepidemiology (ISPE) joint taskforce:**
  This taskforce was created to make recommendations regarding good procedural practices that would enhance decision makers' confidence in evidence derived from RWD studies (Berger et al., 2017b). The papers produced by the taskforce (Berger et al., 2017b; Greenfield, 2017; Wang et al., 2017) cover topics such as study registration, replicability, stakeholder involvement, reporting and transparency in RWE studies. They suggest, amongst other things, that greater transparency in reporting could lead to "a substantial improvement in reproducibility, rigor and confidence [in RWE]" (Wang et al., 2017).
- **The NEW Drug Development Paradigms Initiative (NEWDIGS) WISDOM project:**
  NEWDIGS describes itself as a "think and do" tank that aims to "reliably and

sustainably deliver new, better, affordable therapeutics to the right patients faster". The aim of this work is to explore how new kinds of evidence (integrated with that from traditional RCTs) could impact regulatory and reimbursement decision making. The work includes "efficacy to effectiveness" (E2E) exercises, which aim to explore the gaps in evidence generation across various case-based scenarios. The aim is to ensure that evidence produced throughout development and into practice is "meaningful, valid, expedited, and transparent" (see NEWDIGS[9] for more details). This work is ongoing throughout 2017-2018.

- **The Center for Medical Technology Policy (CMTP) and Green Park Collaborative's RWE Decoder:**
  This is a tool to facilitate review and evaluation of RWE to enable decision makers to feel confident when making decisions based on RWE. It contains various modules in which the user inputs details of the studies, and produces a visual summary of available evidence according to relevance and rigor. It is available for use from the CMTP website[10].
- **PCORI Pragmatic Clinical Studies initiative**: Since PCORI (Patient-Centered Outcomes Research Institute) launched its Pragmatic Clinical Studies initiative in 2014 it has commissioned 28 projects, although none have yet reported results. Several major payers have also commissioned pragmatic clinical trials, for a recent example see Anthem[11], and there are also NIH and industry commissioned studies. We are not aware, however, of any *pre-launch* PCT for a drug being conducted in the USA.

3. CHALLENGES ASSOCIATED WITH THE USE OF RWE

It is clear from the current use of RWE in the US and elsewhere that some stakeholders see great value in RWE and are exploring ways to make use of these data sources. However, several challenges were identified in the literature review and in the interviews with experts, that have meant that acceptance of RWE is far from universal. A common reservation is around the 'quality' of RWE. In practice this comprises a variety of factors, which are discussed below.

3.1. Bias and confounding

Whilst observational RWE offers advantages against RCTs in terms of external validity, questions are often raised about internal validity. This is because observational analyses are inherently vulnerable to selection biases and confounding. Garrison et al. (2007) note that for RWE, "the most significant concern is the potential for bias". They further comment that "observational or database studies do not meet the methodological rigor of RCTs". This topic is well documented, and the debate is not repeated here.[12]

[9] https://newdigs.mit.edu/sites/default/files/documents/NEWDIGS%20WIS-DOM%20June%202017.pdf
[10] http://www.cmtpnet.org/resource-center/view/rwe-decoder/
[11] https://www.healthcore.com/anthem-healthcore-boehringer-ingelheim-initiate-worlds-largest-pragmatic-clinical-trial-study-people-living-copd-real-world-setting/
[12] For more information see: Carrao (2013); Grimes and Schulz (2002); Hill and Kleinbaum (2014); Norgaard et al. (2017).

Our interviewees confirmed that selection bias is a key potential challenge for observational RWE. However, they also indicated that selection biases are relatively well understood, and that techniques are available to adjust for them in many cases (for example by adjusting for covariates, matching, or using instrumental variables), or at least to identify and describe them.

Reporting bias was also highlighted as a potential problem. This occurs when some outcomes (or even whole datasets) are selectively revealed or withheld. A review conducted by McGauran et al. (2010) found evidence of reporting bias across multiple disease areas including: depression, bipolar disorder, schizophrenia, anxiety disorder, attention-deficit hyperactivity disorder, Alzheimer's disease, and many others, via the withholding of study data by manufacturers and regulatory agencies. The authors found that this bias had led to the overestimation of efficacy and the underestimation of safety risks. Such evidence of reporting bias damages trust between manufacturers and payers, and leads readers to question the validity of the evidence that is available.

In order to mitigate the impact of selection biases, RWE studies need to be rigorously designed and evaluated. Some checklists have been developed (see section 3.5 below), but consensus does not appear to have been achieved.

Many of the issues around non-reporting of studies arose in the case of RCTs and have been addressed. A similar approach could be made in the case of observational studies. A mandatory national registry for studies, such as is available for RCTs[13], could help mitigate this problem.

3.2. Incomplete data

A second challenge for the use of RWE is that of incomplete data. Datasets are vulnerable to systematic omissions or misclassification – one interviewee provided the example of mental health diagnosis codes in the US. It seems that, based on their own data, more people are taking treatments for mental health conditions than are classified as requiring these treatments, even when taking into account that some of these medications could be being used for other indications. This suggests that mental health diagnoses are systematically underreported within datasets. Another example is central line-associated bloodstream infection (CLABSI) which have been used as a publicly reported indicator of health care quality, leading to fears that cases may be intentionally underreported (Thompson et al., 2012). Clearly, analysis of data that has been intentionally manipulated would not be reliable.

In addition, there are often gaps in the data. Claims data may tell you that a patient had a test, but it cannot necessarily tell you the results. Electronic health records are of variable quality in terms of what is recorded and how interpretable it is. Part of the problem is that the information needed for payment (for which there is a legal or contractual requirement mandating accuracy), may not necessarily include data that are useful for RWE. And finally, when a patient joins a specific health plan, the payer is often blind to their medical history, which makes long-term follow up or adjustment for covariates very difficult. Interviewees reported that data gaps are particularly prevalent when relying on patients or physicians to submit their own data, rather than when it is proactively collected by researchers. Incomplete data can be imputed to some extent,

[13] Clinicaltrials.gov

but can also further exacerbate problems related to bias and confounding. Systematic misclassification is sometimes referred to as information bias.

In order to mitigate the challenge of incomplete data, national data repositories and strict guidelines for reporting to reduce data gaps could be helpful, although these bring challenges of their own. The ability to link datasets to one another would also be a useful way of not only filling in the gaps but also validating the data.

3.3. Data mining

Data mining occurs when analysts re-examine existing datasets to generate new information. Data mining is not inherently bad, and can be used to generate useful information. However, the concern in the context of RWE is that commercial organisations can continue to reanalyse datasets using different modelling approaches until one that delivers preferential outcomes is identified. One interviewee further commented that researchers have been known to collect additional information outside of the protocol, simply out of interest. This data can throw up questions that had not even been considered before, and thus opens the analysis up to data mining. Another interviewee commented that "Data mining is a concern because you can find whatever you want…The analyses can be easy to do, but [it can be] difficult to explain the results".

The comments above highlight the vulnerability of RWE to manipulation via repeat analyses with non-disclosure of unhelpful results, and underline the need for strict protocols, analysis plans, and well defined research questions.

3.4. Access to data

Sharing of data across different health care organizations is not common in the US, and this leads to gaps in the data (see section 3.2 for a discussion of incomplete data).

This challenge is exacerbated by legal frameworks that legally restrict data sharing and access to patient identifiable information. This in turn can reduce the ability of health care organizations to share data so that it can be linked across different datasets. Even if researchers are aware that relevant data exists, data governance may mean they are not able to access it, face delays in access, or can only access a subset of the data.

In an OHE Consulting Report, Cole and colleagues (2015) suggest that these problems arise because RWD is being used for purposes beyond those for which it was originally collected – to directly manage the care of the patient. As a result, legal frameworks are "playing catch-up" in order to respond to data demands which clearly benefit patients and society but in a different way.

A balance needs to be struck between protection of private information and informing real-world research. Cole et al. (2015) outline a framework for good data governance that supports a favourable environment for the creation of RWE. The framework covers collection of raw data, cleaning and managing the data, linkage and aggregation and access/use of data.

Clearly, appropriate governance arrangements for RWE are crucial to facilitate evidence collection, and to make the most of health care information and the role it can play in improving patient care.

3.5. Lack of universally accepted methodological standards

Many of the challenges outlined so far are exacerbated by a lack of universally accepted standards or principles for the design, conduct, analysis and/or reporting of RWE. This lack of consensus has meant that RWE is often not considered of high enough quality to be part of the body of evidence used to determine comparative effectiveness of different treatment options. This undercuts the potential value of the information that is produced, reducing the incentive to generate it.

One interviewee commented that: "The RWE area is evolving so rapidly that standards can't keep up with what we are able to do with the data." Another highlighted that this is particularly the case for the types of studies that span the gap between RCTs and traditional observational studies, such as pragmatic clinical trials, combining randomisation with data from routine clinical settings. Another commented that reporting for RWE is about 10 years or so behind reporting for RCTs. Reporting standards for RWE will most likely improve and this will help people be able to judge whether a RWE study has been designed and analysed appropriately. Finally, one interviewee felt that "lack of consensus has meant that RWE is often not considered to be high quality compared to trials, and is not considered high quality for evidence based medicine". Clearly the expert interviewees felt that the lack of established standards in this area is a key challenge.

There have been various efforts to set out best practices and standards for collecting and analysing RWE (NPC, 2017). These include:

- Tools developed by the Comparative Effectiveness Research Collaborative[14], which are designed to help decision makers evaluate various types of evidence;
- The Good ReseArch for Comparative Effectiveness (GRACE) Principles[15], and the associated checklist which provide good practice principles for observational studies;
- A series of papers published in the Journal of Comparative Effectiveness Research (Greenfield and Kaplan, 2012; Montori et al., 2012).

However, a review of nine standards/guidelines for observational studies undertaken by Morton et al. (2016) found a lack of agreement between the various sets of principles. This underscores the importance of establishing a common set of agreed principles for the conduct of RWE studies.

The ISPOR taskforce report on using RWD for coverage decisions (Garrison et al., 2007) distinguishes between good research *practices* for *collecting and reporting RWD* and good *process* in using RWD in coverage decisions. Good practices for collecting and reporting RWD include elements such as posing well-defined research questions, specifying time frames for the duration of data collection, and limiting sample sizes to the minimum necessary; whereas good process for using RWD in decision making is about making sure the decision and rationale are transparent, relevant and allows for stakeholder engagement. More recently the ISPOR/ISPE taskforce has published recommendations regarding good procedural practices that would enhance decision makers' confidence in evidence derived from RWD studies (Berger et al., 2017b) (see section 2.3.2).

[14] https://cercollaborative.org/global/default.aspx?RedirectURL=%2fhome%2fdefault.aspx
[15] https://www.graceprinciples.org/

In order for a set of standards to be useful, all stakeholder groups are required to 'buy into' them and agree that they offer the most suitable guidance for the development, conduct, analysis and/or reporting of RWE. Ideally, all studies can then be conceptualised, designed, conducted, analysed and reported according to this common set of standards. This would increase transparency, reliability, and trust in the results of RWE.

3.6. Lack of investigator expertise

Linked to the problem of lack of universally accepted standards is that of lack of investigator expertise. It is important that investigators understand RWD well in order to be able to interpret it properly and adjust for systematic omissions and confounding biases (see A1) appropriately.

The majority of our interviewees agreed that a lack of expertise in the analysis of RWD was a concern. One commented that "innocent misinterpretation" (for example misunderstanding relationships as causality) can happen when inexperienced people are working with large datasets, and that it is a big worry. Another interviewee commented that "Some of the methods for analysis are complicated, so you have to know what you are doing", and another "It is easy to come to the wrong conclusions if you don't understand the data well."

It is also important to note that a good understanding of the available dataset can aid conclusions around validity. One interviewee noted that some insurance data is audited and is therefore thought to be extremely reliable: for example, procedure codes and pharmacy data are audited because they form part of the claim, and thus incorrect reporting can amount to fraud. Diagnosis codes however, do not form part of the claim, thus are not audited and may be less reliable. This is perhaps less of a concern when the analyses are conducted by the manufacturers or payers, as they know their data well, but will become more problematic as the data becomes more widely available.

This perception that there is a lack of expertise (whether it is correct or not) is an important challenge because it erodes trust in RWE and undermines its conclusions.

Strong methodological guidance is required to ensure rigorous standards for analysis, and as RWD becomes more readily available, expertise will spread.

3.7. Obsolete evidence hierarchies

Traditional evidence hierarchies that promote RCTs as the gold standard for evidence generation were developed for a world without big data, and do not necessarily account for the potential for RWE to supplement our understanding of the safety and effectiveness of treatments in different populations. Indeed, RWE studies are more useful than trials to address key questions regarding durability of effect, generalizability, and long-term safety.[16]

One interviewee suggested that RWE provides an opportunity to revolutionise coverage assessments, not necessarily at launch, but over time as the flood of information becomes more interpretable and available. Manufacturers need to know how best to present this information to payers and other stakeholders (e.g. patients), and payers need to know that the information they are receiving is reliable. As the nature of the

[16] Regulatory decision makers do acknowledge this to some extent, through their reliance on RWE for post-launch safety monitoring.

information that we are able to gather evolves and improves, it is important to consider how we can integrate these into assessment processes to make sure we are making decisions based on the best available evidence.

3.8. Measures that can be taken today to address RWE limitations and challenges

3.8.1. Increase the quality and credibility of observational RWE studies

Some of the measures proposed by our interviewees are:

- A mandatory national registry for observational studies, such as is available for RCTs;
- National data repositories to reduce data gaps;
- Investment in the quality and consistency of EHRs in a way that would transform the potential value of RWE;
- Strict protocols, analysis plans, and well defined research questions.

Good practice guidelines could also steer analyses. They would encourage submission of the research protocol to national RWE registries, and could include a checklist against which all studies could be assessed. To support transparency, good practice guidelines could also include recommendations on the publication of the code(s) used to analyse the dataset(s). The ISPOR / ISPE Task Force papers provide a good starting point for a definitive set of principles and procedures (Berger et al., 2017b; Greenfield, 2017; Wang et al., 2017).

3.8.2. Establish effective governance arrangements that clarify how much data can be shared

A balance needs to be struck between protection of private information and informing real-world research. Clearly, appropriate governance arrangements for RWE are crucial to facilitate evidence collection, and to make the most of health care information and the role it can play in improving patient care. The FDA's Sentinel database uses a distributed network model. A key challenge is whether it can also be used to address post-launch effectiveness issues. Can a distributed network approach be used to assess effectiveness? The general question is the ability to move beyond one payer / system when undertaking RWE studies.

3.8.3. Focus RWE efforts on the development of pragmatic clinical trials

Many of the concerns raised about the use of RWE relate to concerns around the quality of RWD and validity of the results. Indeed, traditional evidence hierarchies that promote full RCTs as the gold standard for evidence generation were developed in a world without big data.

PCTs can offer a bridge between RCTs and observational RWE by combining randomization with elements of the real world setting. As such they can, in principle, provide reassurance to payers about quality, whilst also providing relevance to the

health care settings in which the drugs will be used. The PRECIS 2 website[17] provides a resource for planning how much a trial can or should move towards RWE using a pragmatic/explanatory continuum. The tool has nine domains, each of which are scored on a 1 (very explanatory) to 5 (very pragmatic) scale. These nine domains are illustrated in **Error! Reference source not found.** below. For a recent discussion of PCTs and PRECIS 2 see Ford and Norrie (2016).

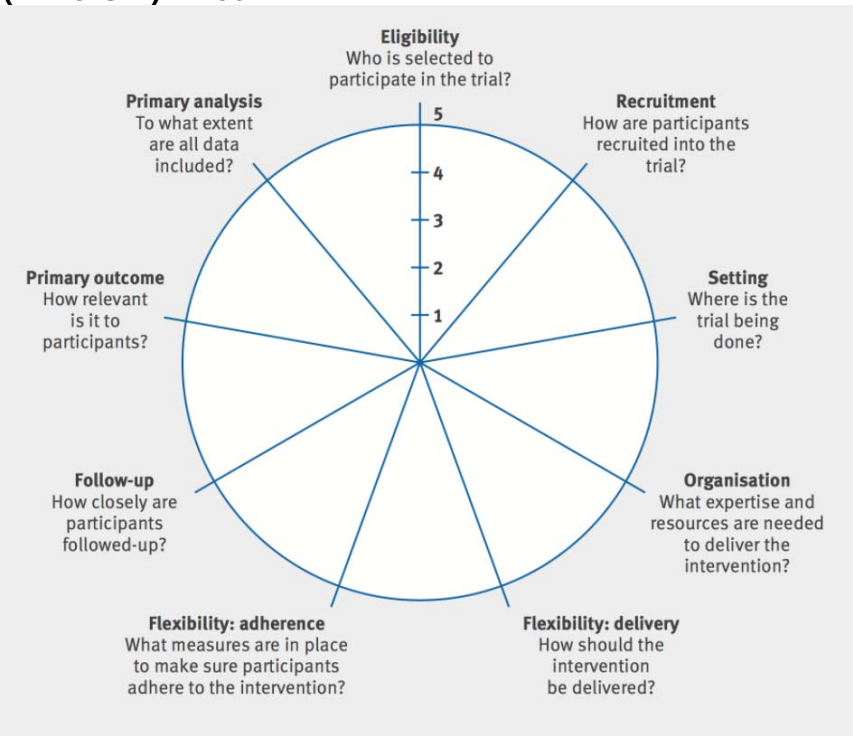
The benefits of PCTs are routed in their ability to pick the favoured elements of RCT and observational study designs. As such, they stand to be beneficial to manufacturers and payers, if the appropriate study designs can be agreed. The challenge in undertaking PCTs is often cost, and the need for tailored data collection. The experience of the Salford Lung Study (section 2.2.4) is one example of this. The ability to embed the data needed for PCTs in the routine data collection processes of health care systems is key to getting costs down.

Another challenge is limitations on undertaking PCTs in a pre-launch setting. The Salford Lung Study was a pre-launch PCT demonstrating that it can be done. Reluctance of regulators to accept PCTs is another factor cited by manufacturers. The EMA accepted the Salford Lung Study, and the CHMP gave approval in September 2017. The FDA also approved the drug in September 2017.

It is also worth noting that the NIH Health Care Systems Research Collaboratory is investing in the development of PCTs. They suggest that PCTs offer the potential for generating "actionable clinical evidence at a fraction of the typical cost and time needed to conduct a traditional trial". Their online resource, *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*[18], guides stakeholders though each phase of the PCT process, from the development of a clinical question to the dissemination of results. The NIH are currently evaluating applications for demonstration projects for PCTs which will enable and provide support for a number of PCTs through planning and implementation, with the goal of strengthening national capacity for this type of study.

[17] https://www.precis-2.org/
[18] http://www.rethinkingclinicaltrials.org/

Figure 3: The PRagmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2) wheel



Source: Adapted from Louden et al., 2015

Finally, and as noted earlier, PCORI also supports the use of PCTs[19] and private payers and manufacturers are also undertaking PCTs. PCORI is seeking to fund large PCTs that compare two or more alternative health care technologies, commenting that these studies must include large representative populations, and must support evaluation of potential differences in patient subgroups.

Going forwards, there is a strong case for greater use of PCTs and nested designs which seek to combine randomisation with real life settings. The ability to embed the data needed for PCTs in the routine data collection processes of health care systems is key to getting costs down. The official position of the FDA is unclear, but a recent paper by Sherman et al. (2016) and the discussions at the Duke-Margolis RWE Regulatory Framework Public Workshop, suggests that PCTs are under consideration.[20]

Manufacturers and payers/regulators should engage in early dialogue to agree *a priori* the types of study design that are most relevant to the technology and the question under evaluation. In many cases this could be a mix of RCTs and RWE, PCTs, or even nested designs in which one study informs part of another. It will be important to include regulators in these discussions, as this is the first external decision making point (i.e. not made by the manufacturers), and as such is often the key focus for evidence generation.

[19] https://www.pcori.org/research-results/research-we-support; https://www.pcori.org/research-results/pragmatic-clinical-studies

[20] https://healthpolicy.duke.edu/events/public-workshop-framework-regulatory-use-real-world-evidence

4. ADDITIONAL OPPORTUNITIES TO ENHANCE THE GENERATION AND USE OF RWE

Having outlined the current uses of RWE and the associated challenges, we now move on to identify the potential opportunities for the expansion of the use of RWE. Innovative study designs could revolutionise evidence generation, e.g. routine use of PCTs (Ford and Norrie, 2016), cohort multiple randomised controlled trials (Relton et al., 2010) or nested trials within cohorts. Randomisation remains an important tool for delivering high quality evidence, but this can be used in real world settings to provide more relevant evidence for payers. The potential to embed such evidence collection in routine data collection systems offers major benefits to payers.

4.1. Real time evidence-based medicine.

The biggest potential benefit of "Big Data" may be health systems' ability to combine this data with analysis that is translated into protocols and guidelines for health professionals that enable them to actively manage patients when entering key patient characteristics. Handheld devices can have software that, for example, stratifies patients and identifies the relevant set of interventions. This is an extension of analysis payers are already undertaking routinely to identify sub-sets of patients who are likely to face health deterioration and become intensive users of health care services and resources if they are not proactively managed.

4.2. Real time monitoring of patients.

Wearables and the Apple Research Kit, for example, will enable some data collection to become routine. These can be viewed as one or more of (i) tools that will reduce the costs of collecting evidence – along with the use of handheld devices by health professionals to input patient data; (ii) mechanisms for expanding evidence as to how patients are responding to drugs; or (iii) types of remote monitoring providing further opportunities for services designed to enable patients to better manage their medication, and alert health professionals as to when intervention is required.

4.3. Accelerated access to innovative therapies (adaptive pathways and coverage with evidence development)

The launch of new products under accelerated access paths (such as accelerated approval and breakthrough designation) means payers have at the point of initial drug listing decisions less evidence than for "traditional" new therapies. The ability to collect post-launch RWE is crucial for payers, both observational and through PCTs.

Considering accelerated pathways alongside improvements in study design (see Section 3.8.3), arguably provides a new paradigm for payers to get evidence on the comparative effectiveness and cost-effectiveness of the drugs within their health systems, in response to the greater use of accelerated access regulatory pathways.

4.4. Summary of key opportunities and challenges

A summary of all the key opportunities for both short-term and long-term improvement in the use of RWE presented in the sections above is presented in Table 1 on the following pages.

Table 1: Opportunities to realise the potential of RWE in coverage decisions

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
1. Improving current uses of RWE				
A. Evaluation of benefits and harms in a real world setting, including: • Effectiveness, safety, adherence • The durability of benefits and side effects over a longer period than the RCT • Exploring population subgroups in which clinical benefit is (likely to be) greatest • Gaining comparable evidence on the new drug and usual care • Use outside of the initial indication	Recent advances in data collection via claims databases and EHR provide richer data and more opportunities for analysis. Techniques for analysis need to keep up with the evolving data environment to ensure best use can be made of available data. For example, RWE can be used to expand upon previous studies to new indications. In Sweden, Xyrem was recommended for use in children who developed narcolepsy as a result of swine flu vaccine, despite not having a paediatric label (Quintiles IMS, 2013). To date, such use of RWE in the US has been limited (Berger et al., 2017c)	**Patients & payers** Such evaluations have the potential to increase value to the patient (clinical benefit) and value to the payer (cost-effectiveness). Therapies may also become available to wider patient groups. **Manufacturers** Manufacturers are able to ascertain what they should concentrate on for marketing approval and coverage, and assert competitive advantage. Data on use outside of the initial indication could be of particular importance if manufacturers are able to expand their eligible population without investing in additional trials.	**Prior to launch** Manufacturers are able to ascertain what they should concentrate on for marketing approval and coverage. **At Launch / Initial Coverage Decision** Payers are able to ascertain the group(s) for which value is greatest. This enables them to make the best use of available resources. **Post Launch** Evidence can support manufactures and payers in re-evaluating effectiveness in various groups and pricing decisions.	• Bias and confounding • Incomplete data • Data mining • Access to data • Lack of expertise • Lack of universally accepted standards • Persistent evidence hierarchies and reliance on RCTs

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
B. Evaluate outcomes that are not measured during standard development process.	This is an opportunity to expand the data set that is available at launch to include additional outcomes. As value assessments evolve to include elements of value outside of the typical health gains, this opportunity will become even more important. Some of these outcomes (for example productivity or adherence) could be measured within a typical RCT, but RWD is likely to be fairly different when generated outside of the controlled research setting. RWD on these outcomes is thus likely to be much more informative.	**Patients** Consider that if drug X has better adherence rates than drug Y, it could be much more effective in the real world, even if drug Y is shown to be marginally more effective in RCTs. This type of information is extremely value for payers, and can ensure that patients get access to the most effective medicines. **Payers** These outcomes are increasingly important for value assessments. The opportunity to evaluate them in the real world provides payers with the opportunity to consider all relevant factors **Manufacturers** Manufactures have the chance to establish holistic value.	**Post Launch** This is an opportunity to be explored post-launch and factored into reassessment decisions. Such data could also be useful for factoring into initial coverage decisions for subsequent products.	• Bias and confounding • Incomplete data • Data mining • Access to data • Lack of expertise • Lack of universally accepted standards.

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
C. Evaluation of prevalence of the condition for budget impact analyses and cost-effectiveness in a real world setting	As effectiveness and adherence estimates are refined with RWE, and resource use data is generated from real clinical scenarios, there is an opportunity to revise and refine budget impact and cost-effectiveness estimates. Additional elements of value could also be brought into the assessment at this stage (see row above). New methods of collecting and analysing large datasets will provide more opportunities for this type of investigation.	**Patients** Reliable estimations of value will increase overall health benefit of the formulary for patients. **Payers** Refined estimates of budget impact and value should increase value for money for the payer **Manufacturers** Such additional evidence gives manufacturers the opportunity to demonstrate the true value of their product in a real world setting.	**Pre-launch** Can inform pre-launch preparations and estimates of how many will be affected within a given payer's membership **At Launch** Prevalence information inform discussions about the importance of the condition and inform **Post-launch** This is an opportunity to be explored post-launch.	• Bias and confounding • Incomplete data • Data mining • Access to data • Lack of expertise • Lack of universally accepted standards

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
2. Key opportunities for the future				
D. Development and use of innovative study designs to revolutionise evidence generation, e.g. PCTs and nested trials within cohorts	Such study designs can combine benefits of collecting data from real world settings while incorporating best practice methods (i.e. randomisation methods from traditional RCTs). Early dialogue will be required to ensure that all stakeholders are aligned in their expectations about the evidence generated via these study designs.	**Patients** Innovative trial designs could expedite and improve the efficiency of evidence generation, thereby bringing more innovative therapies to patients and doing so faster than is currently achieved. **Payers** Payers can analyse the value of the product in a real-world setting. **Manufacturers** Manufacturers can demonstrate the full value of their product in a real-world setting.	**Pre-launch** This is an opportunity to be leveraged pre-launch to ensure that the best possible evidence package is available for the initial coverage decision. **Post launch** Evidence collection could continue into the post-launch phase and used for reassessment of the coverage decision.	• Lack of expertise • Lack of universally accepted standards • Cost

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
E. Real time evidence-based medicine	Rapid technological data accumulation and data interpretation offers the potential for continuous data analysis and reporting (Schneeweiss et al., 2015; Schneeweiss, 2014). This is supported by patient reported data gathered from personal apps and wearable devices. These provide different measures of functionality that we have not previously had. Schneeweiss (2014) provides the examples of live alerts telling doctors how many strokes have been averted due to recent policies, or even the likelihood that an individual patient will adhere to their medication. He notes that much of this analysis will need to be automated.	**Patients** This type of live evidence generation and dissemination could be particularly invaluable in improving decision making between doctors and patients. **Payers** It could also aid payers in achieving maximum value for money across the plan but ensuring the most appropriate use of medicines, even at the patient level. **Manufacturers** There is also a commercial opportunity here for the development of appropriate software to analyse and deliver this information.	**Post launch** Data would be required to feed into the real time assessments. This data could further feed into initial coverage decisions about subsequent products. This is an opportunity to move away from static decision making towards a more collaborative and dynamic system.	• Bias and confounding • Incomplete data • Data mining • Access to data • Lack of expertise

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
F. Real time monitoring of patients	This opportunity is closely linked with the real-time analysis of data (see above) and is made possible via the use of personal apps, wearable devices, and virtual communication. Real time monitoring allows a personalised approach to medicine, and could increase efficiency of the health system by reducing expensive hospital visits or consultations. Patient outcomes could be monitored and treatment adjusted accordingly.	**Patients** Patients stand to benefit from individualised care **Payers** Payers stand to benefit from an improvement in the likelihood of effective treatment. **Manufacturers** Manufacturers would be able to track outcomes that are important to patients and demonstrate value across a new range of elements.	**Post launch** This is most likely an opportunity to be explored post-launch, although it could be extended to real time monitoring of patients in RCTs and PCTs. Data collection would be continuous and would feed into more regular reassessments of coverage.	• Bias and confounding • Incomplete data • Data mining • Access to data • Lack of expertise

Opportunity	How does this build on current use?	Benefits by stakeholder group	Timing	Associated challenges
G. Accelerated access to innovative therapies (adaptive pathways and coverage with evidence development), linked to aligned, co-produced, real-time, real world data (Husereau et al., 2016)	RWE can be used to support submissions when the full evidence package preferred by decision makers is not available. The FDA has accepted RWE at the regulatory stage in areas that have high unmet medical need. There is a case for expanding this to therapies that show promise of clinical benefits but there is uncertainty in the results (as is done by the EMA). Payers in the US could also use a similar mechanism through which coverage is granted alongside requirements to collect further evidence.	**Patients** Accelerating access to innovative therapies has obvious benefits to patients. **Payers** Payers get the advantage of offering early access to therapies, whilst exercising caution. A clear exit strategy is required in case the therapy does not prove to be of benefit. **Manufacturers** It may also have commercial benefits, particularly smaller enterprises that lack the capital to fund large phase three trials. Such efficiencies can also reduce the cost of evidence generation meaning more therapies may have the opportunity to be brought to market.	**At Launch and Post-launch** This is an opportunity to be explored at launch of the product, whilst assessments continue post-launch.	• Bias and confounding • Lack of expertise • Lack of universally accepted standards • Persistent evidence hierarchies and reliance on RCTs

5. CONCLUDING REMARKS

Clearly our potential for rapid accumulation and analysis of data is increasing alongside technological advancements. These developments provide exciting opportunities for the use of RWE, yet important reservations remain, and overcoming challenges is likely to require dialogue and collaboration between multiple stakeholders, notably payers and manufacturers. Pragmatic clinical trials, adaptive pathways, and real time monitoring of patients all offer the potential for more efficient or applicable evaluations of health technologies, and are all within reach if such effective partnerships can be nurtured.

This paper has set out the key opportunities and challenges around the use of RWE, with the primary purpose of stimulating discussion at the 2017 ICER Policy Summit meeting. A separate paper is available that summarises the authors reflections and proposed ways forwards based on the discussions that were had at the meeting.

REFERENCES

Berger, M., Daniel, G., Frank, K., Hernandez, A., McClellan, M., Okun, S., Overhage, M., Platt, R., Romine, M., and Wilson, M., 2017a. A framework for regulatory use of real-world evidence. White Paper. The Duke-Margolis Centre for Health Policy. Available at https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf

Berger, M., Sox, H., Wilke, R., et al., 2017c. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value in Health*, 20, pp. 1003-8.

Berger, M., Mullins, D., Schneeweiss, and S, Wang, S., 2017c. Enhancing the Transparency and Credibility of Real World Evidence in Health Care Decision Making. *Value in Health*.

Business Wire, 2017. *Harvard Pilgrim Signs Outcomes-Based Contracts with AstraZeneca for Brilinta and Bydureon: Effectiveness in actual patients will be measured*. Available at: http://www.businesswire.com/news/home/20170530005665/en/Harvard-Pilgrim-Signs-Outcomes-Based-Contracts-AstraZeneca-Brilinta [Accessed November 2017]

Caldwell, D., Ades, A., Higgins, J., 2005. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*, Oct 15, 331(7521), pp.897-900

Cameron, C., Fireman, B., Hutton, B., et al., 2015. Network meta-analysis incorporating randomized controlled trials and non-randomized comparative cohort studies for assessing the safety and effectiveness of medical treatments: challenges and opportunities. *Syst Rev*, Nov, 4, pp.147.

Carrao, G., 2013. Building reliable evidence from real-world data: methods, cautiousness and recommendations. *Epidemiology, Biostatistics and Public Health*, 10 (3). Available at http://ebph.it/article/viewFile/8981/8192.

Cerreta, F., 2015. *EMA Adaptive Pathways Pilot*. Presentation at STAMP. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2015/11/WC500196727.pdf.

Cole, A., Garrison, L., Mestre-Ferrandiz, J., and Towse, A., 2015. *Data Governance Arrangements for Real-World Evidence*. OHE Consulting Report. Available at https://www.ohe.org/publications/data-governance-arrangements-real-world-evidence#.

Dias, S., Welton, N., Sutton, A., Ades, A., 2013. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making*, Jul, 33(5), pp.657-70.

Efthimiou, O., Mavridis, D. Debray, T. et al., 2017. Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med*, Apr 15, 36(8), pp.1210-1226.

EMA, 2015. *Guidance on Post-Authorisation Efficacy Studies.* Available from http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2015/03/WC500184245.pdf [Accessed October 2017]

EMA, 2016. Final report on the adaptive pathways pilot. Available from http://www.ema.europa.eu/docs/en_GB/document_library/Report/2016/08/WC500211526.pdf [Accessed March 2018]

FDA, 2017. *Use of Real-World Evidence to Support Regulatory Decisions-Making for Medical Devices: Guidance for Industry and Food and Drug Administration Staff.*

Document issued on August 31, 2017. Available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

Findlay, S., 2017. Why an FDA Drug Safety Monitoring System Is Failing. Available at: https://medshadow.org/your-meds/drug-safety-monitoring/ [Accessed November 2017]

Ford, I. and Norrie, J., 2016. Pragmatic Trials. *N Engl J Med*, 375, pp.454-463.

Garrison, L., Carlson, J., Bajaj, P., Towse, A., Neumann, P., Sullivan, S., Westrich, K., and Dubois, R., (2015). Private Sector Risk-Sharing Agreements in the United States: Trends, Barriers, and Prospects. *Am J Manag Care*, 21(9):632-640

Garrison, L., Neumann, P., Erickson, P., Marshall, D., and Mullins, D., 2007. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health*, 10 (5), pp. 326-325.

Gill, J.L., Avouac, B., Duncombe, R., Hutton, J., Jahnz-Royzk, K., Schramm, W., Spandonaro, F., Thomas, M., and Kanavos, P.G., 2016. *The Use of Real World Evidence in the European Context: An Analysis of Key Expert Opinion*. Europe research paper. London School of Economics.

Greenfield, S., Kaplan, S.H., 2012. Building useful evidence: changing the clinical research paradigm to account for comparative effectiveness research. *Journal of Comparative Research*, 1 (3). DOI: https://doi.org/10.2217/cer.12.23.

Greenfield, S., 2017. Making Real-World Evidence More Useful for Decision Making. *Value in Health*, 20, pp. 1023-24.

Griffiths, E.A., and Vadlamudi,. N.K., 2016. Not ready for the real world? The role of non-RCT evidence in health technology assessment. *Value in Health*, 19 (3), pp. A286. DOI: http://dx.doi.org/10.1016/j.jval.2016.03.760.

Grimes, D.A., and Schultz, K.F., 2002. Bias and causal associations in observational research. *The Lancet*, 359 (9302), pp. 248-252.

Hill, H., and Kleinbaum, D.G., 2014. *Bias in Observational Studies*. Published online in 2005 in Encyclopedia of Biostatistics. Republished in 2014 in Wiley StatsRef, DOI: 10.1002/9781118445112.

Husereau, D., Henshall, C., Sampietro-Colom, L., and Thomas, S., 2016. Changing Health Technology Assessment Paradigms. *International Journal of Technology Assessment in Health Care*, 32 (4), pp. 191-199.

Jarow, J., LaVange, L., and Woodcock, J., 2017. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. *Journal of the American Medical Association* (JAMA), 318 (8), pp. 703-4.

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

Lu, G., Ades, E., 2004. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*, Oct 30; 23(20), pp.3105-24

Makady, A., Ham, R.T., de Boer, A., Hillege, H., Klungel, O., and Goettsch, W., 2017, Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. *Value in Health*, 20 (4), pp. 520-532.

Messner, D., Mohr, P., Towse, A., 2015. Futurescapes: evidence expectations in the USA for comparative effectiveness research for drugs in 2020. *JCER*, Aug, 4(4), pp.385-400.

McGauran, N., Beate, W., Kreis, J., Schuler, Y.B., Kolsch, H., and Kaiser, T., 2010. Reporting bias in medical research – a narrative review. *BioMed Central*, 11 (37). Available at https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-11-37.

Montilla, S., Xoxi, E., Russo, P., Pani, L., 2015. Monitoring registries at Italian medicines agency: Fostering access, guaranteeing sustainability. *International Journal of Technology Assessment in Health Care*, 31 (4), pp. 210-2013.

Montori, V.M., Kim, S.P., Guyatt, G.H., and Shah, N.D., 2012. Which design for which question? An exploration toward a translation table for comparative effectiveness research. *Journal of Comparative Effectiveness Research*, vol. 1, no. 3, DOI: https://doi.org/10.2217/cer.12.24.

Morton, S.C., Costlow, A.R., Graff, J.S., and Dubois, R.W., 2016. Standards and guidelines for observational studies: quality is in the eye of the beholder. *Journal of Clinical Epidemiology*, vol. 71, pp. 3-10. DOI: http://dx.doi.org/10.1016/j.jclinepi.2015.10.014.

Mostagim, S.R., Gagne, J.J., and Kesselheim, A.S., 2017. Safety related label changes for new drugs after approval in the US through expedited regulatory pathways: retrospective cohort study. *BMJ* 2017, 358 :j3837.

Naci, H., Smalley, K., Kesselheim, A., 2017. Characteristics of Preapproval and Postapproval Studies for Drugs Granted Accelerated Approval by the US Food and Drug Administration. *JAMA,* Aug 15, 318(7) pp.626-636.

Nagar, R., Yuan, Q., Freitfeld, C., Santillana, M., Nojima, A., Chunara, R., and Brownstein, J.S., 2014. A Case Study of the New York City 2012-2013 Influenza Season With Daily Geocoded Twitter Data From Temporal and Spatiotemporal Perspectives, *Journal of Medical Internet Research*, 16 (10). DOI: 10.2196/jmir.3416.

NEWDIGS, n.d. PROSPECTUS: NEWDIGS WISDOM Project. MIT: Centre for Biomedical Innovation. Available from https://newdigs.mit.edu/sites/default/files/documents/NEWDIGS%20WISDOM%20June%202017.pdf [Accessed November 2017].

Norgaard, M., Ehrenstein, V., and Vandenbrouke, J.P., 2017. Confounding in observational studies based on large health care databases: problems and potential solutions – a primer for the clinician. *Journal of Clinical Epidemiology* (2017:9), pp. 185-193.

NPC, 2017. *Standards for Real-World Evidence*. Available at: http://www.npcnow.org/issues/evidence/standards-for-real-world-evidence [Accessed September 2017].

Pierce, C.E., Bouri, K., Pamer, C., Proestel, S., Rodriguez, H.W., Van Le, H., Freitfeld, C.C., Brownstein, J.S., Walderhaug, M., Edwards, I.R., and Dasgupta, N., 2017. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. *Drug Safety*, 40 (4), pp. 317-331. DOI: https://dx.doi.org/10.1007%2Fs40264-016-0491-0.

Quintiles IMS, 2013. *Case Studies on Real World Evidence Impacting Payer Decisions*. Available at http://imsbrogancapabilities.com/pdf/healthcare-real-world-evidence-1.pdf.

Relton, C., Torgerson, D., O'Cathain, A., Nicholl, J., 2010. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ*, 340, c1066.

Riley, R., Lambert, P., Abo-Zaid, G., 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, Feb, 5; 340:c221

Roche, N., Reddel, H.K., Agusti, A., Bateman, E.D., Krishnan, Martin, R.J., Papi, A., Postma, D., Thomas, M., Brusselle, G., Israel, E., Rand, C., Chisholm, A., and Price, D., 2013. Integrating real-life studies in the global therapeutic research framework, *The Lancet*, 1 (10), e29-e30. DOI: http://dx.doi.org/10.1016/S2213-2600(13)70199-1.

Salanti, G., Higgina, J., Ades, A., Ioannidis, J., 2008. Evaluation of networks of randomized trials. *Stat Methods Med Res*, Jun, 17(3), pp. 279-301

Salanti, G. 2012. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*, Jun,3(2), pp.80-97

Schneeweiss, S., 2014. Learning from Big Health Care Data. The New England Journal of Medicine, 370: 2161-2163. DOI: 10.1056/NEJMp1401111.

Schneeweiss, S., Shrank, W.H., Ruhl, M., and Maclure, M., 2015. Decision-Making Aligned with Rapid-Cycle Evaluation in Health Care. *International Journal of Technology Assessment in Health Care*, 31 (4), pp. 214-222.

Sherman, R.E., Anderson, S.A., Dal Pan, G.J., Gray, G.W., Gross, T., Hunter, N.L., LaVange, L., Marinac-Dabic, D., Marks, P.W., Robb, M.A., Shuren, J., Temple, R., Woodcock, J., Yue, L.Q., and Califf, R.M., 2016. Real-World Evidence - What Is It and What Can It Tell Us? *The New England Journal of Medicine*, 2016, 375:2293-2297. DOI:10.1056/NEJMsb1609216.

Soares, M.O., Dumville, J.C., Ades, A.E., Welton, N.J., 2013. Treatment comparisons for decision making: facing the problems of sparse and few data. J.Roy.Statist.Soc.A.

The Bipartisan Policy Center, 2016. *Using Real-World-Evidence to Accelerate Safe and Effective Cures: Advancing Medical Innovation for a Healthier America*. Available at https://cdn.bipartisanpolicy.org/wp-content/uploads/2016/06/BPC-Health-Innovation-Safe-Effective-Cures.pdf.

Verde, P., Ohmann, C., 2015. 2015. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Res Synth Methods*, Mar, 6(1), pp.45-62

Vestbo, J., Leather, D., Bakerly, N. et al., 2016. Effectiveness of Fluticasone Furoate–Vilanterol for COPD in Clinical Practice. *NEJM*, 357, pp.1253-1260.

Wang, S., Schneeweiss, S., Berger, M., et al., 2017. Reporting to Improve reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Value in Health*, 20, pp. 1009-22.

Woodcock, A., Vestbo, J., Bakerly N., et al. 2016. Effectiveness of fluticasone furoate plus vilanterol on asthma control in clinical practice: an open-label, parallel group, randomised controlled trial. Lancet, 390(10109), pp.2247-2255.

ANNEX: CHALLENGES ASSOCIATED WITH USE OF RWE

A1. Use of RWE in Performing Indirect Comparisons with Network Meta-Analysis (NMA)

One area of significant interest involves the use of RWE in the conduct of network meta-analysis (NMA), a statistical technique that combines direct, head-to-head evidence comparing treatments with "indirect" evidence (i.e., data obtained from studies linked by common comparators) (Caldwell et al., 2005; Lu and Ades, 2004).

Traditionally, NMAs have only included data from RCTs, but integrating additional data from high-quality comparative RWE may confer benefits to a NMA (Cameron et al. 2015: Efthimiou et al., 2017):

- First, the inclusion of these studies improves the density of a network by linking interventions or reinforcing existing links. For example, RCTs of newer interventions may have been assessed only versus placebo or standard care, whereas RWE may directly compare these treatments.
- Second, the inclusion of RWE increases the sample size, thereby improving the precision of estimates of treatment effect and increasing statistical power, which may help to identify clinically relevant subgroups. Increasing sample size is particularly beneficial for analyses of rare events as well as when networks of RCTs are sparse.
- Third, the RWE may link interventions in otherwise disconnected networks, thereby allowing comparisons of sets of treatments that otherwise would be infeasible.
- Fourth, including RWE in a NMA also improves the generalizability of results. Study populations from RWE tend to be more diverse than in RCTs, and outcomes are often measured over longer periods of follow-up. If conclusions drawn from RWE agree with those from RCTs, there may be more confidence in generalizing the results to real-world settings.

Nevertheless, before including RWE in a NMA, careful consideration must be given to the quality of the studies and available data. Poorly designed studies or analyses of RWE that do not appropriately control for differences between intervention groups should not be included, as they may produce biased estimates of treatment effects (see Chapter ☐ for a discussion of the challenges surrounding RWE).

Furthermore, RWE data are sometimes only available from studies in published aggregate form (e.g., mean age, mean improvement from baseline), with few details provided in terms of aspects of study design and/or data analysis intended to control for confounding. With access to only the published information, it is difficult to fully ascertain the quality of RWE. Providing systematic reviewers with access to individual patient data may help quell concerns and allow them to standardize analyses (Riley et al., 2010), but such access is typically restricted.

A key step in conducting a NMA, even when it consists of only RCT data, is to assess the similarity of patients, characteristics, interventions, and outcomes across the complete network of comparisons (Caldwell et al., 2005; Salanti, 2012). Although one benefit of including RWE in a NMA is the improved generalizability of results, if the patients come from broader populations in RWE and are systematically different from those in RCTs in

ways that affect the treatment effect estimates, then the results from a NMA will be biased. Similarly, systematic differences in the way outcomes were measured or assessed across RWE studies may also bias results. When there are loops of evidence in a network, the bias can manifest with discrepancies between the direct and indirect evidence (i.e., inconsistency) (Salanti et al., 2008). If inconsistencies can be attributed to the type of study (i.e., RCT versus RWE), then the data should be analyzed separately by study design. However, if patient-level data are available, it may be possible to reconcile the differences. For example, for each study, a subset of comparable patients can be generated from the patient-level data to ensure there are no systematic differences across the study populations. Outcomes can also be generated consistently (Riley et al., 2010).

However, if there are indications that included RWE studies, adequately adjust for confounding, and the patients, treatments, and outcomes across studies in the network are sufficiently similar, then including both RWE and RCT data in a NMA may be considered. The evidence should be analyzed separately by type of study initially, and any discrepancies should be further examined. If there are no discrepancies or they can be appropriately explained (e.g., by using techniques such as network meta-regression), then a NMA using RWE and RCTs may be performed. Methods should account for the different study designs, such as those summarized by the IMI GetReal working group (Efthimiou et al., 2017) and other research initiatives. We refer to Dias et al. (2013), Efthimiou et al. (2010), Soares et al. (2013) and Verde et al. (2015) for details on how to implement these approaches.

A2. Use of RWE for existing drugs when a new entrant has none

Key points raised in the expert interviews included:

- This (i.e. using RWE for existing drugs when a new entrant has none) happens a lot in practice and it is considered to be a useful way of generating information. Indeed, forecasts need to be made, and it does not make sense to ignore RWE if available, just because a competitor does not have it;
- If the existing and new drugs are very closely related, and particularly if the RCT data are similar, class effects can be explored (and adjustments can be made within these estimates for different strengths of data);
- Even where the drugs are less closely related, it could be useful to compare the RWE on the estimating treatment to the trial evidence (also for the existing treatment) and identify any key differences (for example in the population receiving the treatment, efficacy, adherence) and apply these learnings to the new treatment;
- Sensitivity analyses could also be useful for exploring uncertainties in the comparisons.

Overall the interviewees seemed to agree that all available evidence should be used, but that comparisons and inferences should be made with caution.