REAL-WORLD EVIDENCE FOR HEALTH TECHNOLOGY ASSESSMENT OF PHARMACEUTICALS: OPPORTUNITIES AND CHALLENGES

Amr Ahmed Mahmoud Abdelkader Makady

ISBN: 978-94-6182-886-6

Cover Image: Off Page, Amsterdam

Layout and print production: Off Page, Amsterdam

Chapter 2:
© 2017 International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. All rights reserved.

Chapter 3:
© 2017 International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. All rights reserved.

Chapter 4:
© The Author(s) 2017. Open Access.
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made

REAL-WORLD EVIDENCE FOR HEALTH TECHNOLOGY ASSESSMENT OF PHARMACEUTICALS: OPPORTUNITIES AND CHALLENGES

Real-World Evidence voor Health Technology Assessment van Geneesmiddelen: Kansen en Uitdagingen
(met een samenvatting in het Nederlands)

Proefschrift
ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen
op woensdag 9 mei 2018 des middags te 4.15 uur

door

Amr Ahmed Mahmoud Abdelkader Makady
geboren op 5 januari 1990 te Adan, Koeweit

Promotoren: Prof.dr. A. de Boer
Prof.dr. J. L. Hillege
Prof.dr. O.H. Klungel

Copromotor: Dr. W.G. Goettsch

TABLE OF CONTENTS

CHAPTER 1

Introduction

HEALTH, HEALTH SYSTEMS & THE ROLE OF HTA

Shortly after the end of the Second World War and the establishment of the World Health Organization (WHO) as a subsidiary of the United Nations, representatives from 61 member states convened to develop the first Constitution of the WHO (1). In this constitution, the WHO ventured to lay the fundamental principles of the organization and its aims. This constituted devising a definition for "health"; a complicated concept which may know different meanings across cultures. Eventually a holistic definition was devised, encompassing a state of complete well-being which extends to the physical, mental and social well-being of an individual (i.e., not merely the absence of ailment). Moreover, the constitution stipulated that health, encompassing such a state of complete well-being is a fundamental human right.

By ratifying this constitution, member states of the WHO thus commit to the provision of health to all their citizens. However, the achievement of good health on population level is a herculean task, requiring the development of healthcare systems that guarantee the delivery of a plethora of interventions, such as curative therapy for acute or chronic diseases and public health programs to raise awareness on preventive measures for diseases. In order to provide the reader with more context on the aims and properties of healthcare systems, we take note of the following excerpt from Garrido et al. (2) in Box 1.

Bearing in mind the aims of healthcare systems to provide good health to their respective populations, it may not come as a surprise that the financial resources needed to strive towards these aims are colossal. The Global Healthcare Expenditure Database (GHED) of the WHO states that member states dedicated $6.5 trillion U.S. dollars to healthcare systems in 2010 alone (3). To draw on a national example, the Dutch Ministry of Finance published a report in 2015 which estimated national healthcare expenditures in 2016 to reach €75 billion Euros, making them the government's second-largest annual financial expenditure (4). However, despite the fact that such figures may imply that immense resources are available to establish healthcare systems worldwide, one must bear in mind that these resources are not infinite. To the contrary, such resources have very tangible limits. To make

> "A health system consists of all the people and actions whose primary purpose is to improve health. This definition covers a variety of professions and institutions and a broad range of activities dedicated to the promotion, restoration and protection of health. Health systems encompass both individual and population services, in addition to activities aimed at influencing the policies and actions of other sectors, in an effort to address the social, environmental, and economic determinants of health … modern health systems generally pursue the fundamental goals of improving the health of a population, responding to the wishes and expectations of individuals, and providing financial protection against the costs of ill-health."

Box 1 - Hallmarks of health systems.

matters more complex, healthcare expenditures over the past decade have been increasing beyond the annual rate of gross domestic product (GDP) growth of the majority of countries (5).

As a consequence of rising healthcare costs and finite budgets, governments are constantly faced with challenging questions on how to allocate resources to achieve the greatest health gains for their citizens. In an attempt to provide a transparent and accountable approach to decision making related to healthcare policy on resource allocation, governments increasingly turn to Health Technology Assessment (HTA). Briefly defined, HTA pertains to the systematic evaluation of the properties and effects of health technologies (whether drugs, medical devices, surgical procedures or organizational aspects of health systems (6)), addressing their direct and intended effects, as well as their indirect and unintended consequences with the aim of informing decision making (7). In general, HTA is a policy analysis process including two components; firstly, an assessment of all available evidence relevant to the policy question at hand and secondly, an appraisal of the findings from the evidence to reach a decision (8). The prior of these components (i.e. assessment) is conventionally a scientific, robust process (8).

From an organizational perspective, the conduct of HTA to inform decision making for healthcare systems is often delegated to HTA agencies, each operating under different governance structures and different mandates. For example, HTA agencies could be independent, public advisory bodies operating at arm's length of national Ministries of Health (e.g. Zorginstituut Nederland (ZIN), the Netherlands) or private, not-for-profit entities (e.g. the Institute for Clinical and Economic Review (ICER), the United States of America). Moreover, HTA agencies' mandates could be limited to one domain of HTA (e.g. clinical effectiveness of health technologies; Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Germany) or encompass numerous tasks such as HTA, the development of clinical guidelines and ensuring good quality of healthcare (e.g. National Institute for Health and Care Excellence (NICE), the United Kingdom).

The properties of health technologies assessed through HTA may cover a broad range of domains including: legal issues, ethical issues, societal considerations, organizational aspects, clinical effectiveness, cost-effectiveness and safety. In recent years, the European network of HTA (EUnetHTA) has developed a comprehensive HTA model that systematically guides the conduct of HTA across nine domains, including those mentioned above (9). However, the timeframe within which decision makers are often obliged to reach a decision on the reimbursement of health technologies may necessitate a more abridged, rapid HTA based on core domains (10). Relative effectiveness of the intervention being investigated (i.e. the added therapeutic effect of an intervention compared to current standard of care) (11) and, increasingly, the cost-effectiveness of the intervention (i.e. the incremental costs and effects incurred by reimbursing the new intervention compared to standard of care) (12) are two examples of such core domains. In this thesis, the scope of HTA is restricted to the relative effectiveness- (REA) and cost-effectiveness assessment (CEA) of pharmaceutical drugs.

EVIDENCE-BASED MEDICINE & THE EFFICACY-EFFECTIVENESS GAP: WHY IS THIS AN ISSUE?

In order to ensure the scientific validity of conclusions reached during an assessment, HTA agencies often rely on the principles of Evidence-Based Medicine (EBM) (13) to conduct systematic reviews of available scientific literature, categorize the quality of evidence (e.g., based upon GRADE (14) recommendations) and its interpretation. When applying this methodology to REA and CEA, agencies tend to gravitate towards the prioritization of Randomized Controlled Trials (RCTs) or meta-analyses of RCTs in their evidence base since they are conventionally classified as the highest levels of evidence. Moreover, RCTs constitute the core source of evidence for marketing authorization applications of drugs submitted to regulatory agencies (15). As such, they are a form of relevant evidence that is conventionally available at the time HTA is conducted.

Randomized controlled clinical trials are considered the most valid method for assessing the intended and unintended effects of a drug. As applied to the medical field, RCTs comprise strict inclusion- and exclusion criteria of patients, randomized allocation of patients to an experimental drug and a control arm (e.g. placebo or active comparator) and strict monitoring and follow-up protocols of trial subjects. Therefore, RCTs are designed to address a very specific question, namely: is the drug efficacious when delivered under ideal conditions? (13)

On the other hand, from the perspective of a healthcare system, a drug is considered effective when there is sufficient evidence of benefit to patients when administered by physicians in routine clinical practice settings (i.e. the "real-world"). In the latter setting, patients are more heterogeneous and present with different comorbidities than patients in RCTs. Other aspects, such as the degree of clinical experience amongst healthcare professionals with the implementation of a new drug, may also impact the realized effects of the intervention. When deciding on resource allocations, decision makers thus have to consider whether a new drug can be as efficacious in the real-world as it has been demonstrated to be in the ideal context of RCTs. Moreover, in contrast to RCTs whereby the new drug is conventionally compared to one alterative, decision-makers require evidence on the new drug's comparative effects in relation to all available alternatives. As such, decision makers regularly face a different question to that addressed by RCTs, namely: does the drug work when delivered in routine clinical practice settings compared to current standard of care? (13) This is referred to as the relative effectiveness of the intervention.

Examples in scientific literature about the theoretical mismatch between the questions RCTs are designed to answer and the questions decision makers face are plentiful. As early as 1967, Schwartz et Lellouch alluded to the complementary insights provided by exploratory trials (i.e. observational trials) and confirmatory trials (i.e. RCTs) (16). In 1972, Cochrane stated that "Between measurements based on RCTs and benefit in the community there is a gulf which has been much under-estimated" (17). Throughout the past three decades, numerous advances in the field of pharmacoepidemiology have aimed to develop

methods for the analysis of evidence from alternative trial designs on drug effectiveness and safety in the broad clinical population (18;19). Researchers have also voiced their concerns over the need for relevant, alternative sources of evidence to assess the real-world effectiveness of drugs (20-22). In 2011, Eichler et al. concluded that a discrepancy does exist between the effects of drugs as demonstrated in RCTs and as realized in the real-world. They refer to this phenomenon as the "efficacy-effectiveness gap" (23). From a decision-maker's perspective, such a gap can be worrisome as it implies that evidence for REA and CEA is based on data from RCTs designed to provide alternative insights to the question at hand (13;22). Ultimately, this may result in healthcare resources not being optimally allocated, resulting in a loss in potential health benefits incurred to the population as a whole (2;13).

REAL-WORLD EVIDENCE: POTENTIAL COMPLEMENT TO RCTS? GROWING NEED & THE IMI-GETREAL PROJECT

As mentioned above, the recognition of the shortcomings associated with heavy reliance on evidence from RCTs for decision making in healthcare has long been evident in literature (6;24). On a different note, HTA agencies have recently witnessed an increase in submissions for the reimbursement of innovative, yet expensive, drugs with less evidence on efficacy and safety and therefore less evidence on relative effectiveness and cost-effectiveness of the drugs in question (25;26). Historical examples include drugs for orphan diseases and oncology drugs (27). Although such drugs potentially address unmet medical need for patients, the combination of their exorbitant prices along with significant uncertainties in their evidence base left decision makers with a dilemma: guaranteeing quick access to drugs for patients versus the need for post-authorization evidence to address uncertainties in the evidence base. Consequently, decision makers around the world began to resort to managed entry agreements (MEAs) (27;28). Within the general framework of MEAs (particularly coverage with evidence development (CED) schemes) (29), patient access to new drugs would be facilitated with the promise for additional evidence generation on (relative) effectiveness, cost-effectiveness and/or budget impact of the drug in clinical practice. By definition, the evidence generated under the auspices of such agreements would thus originate from non-RCT sources.

Arguably, these historical and contextual factors fueled discussions on the use of alternative data sources to complement RCT data in decision making on coverage and reimbursement in health policy journals. A seminal example is the paper by Garrison et al. (2007) which coined the term Real-World Data (RWD) to mean any data on the effects of a health intervention (clinical, economic or patient-reported) collected outside the context of RCTs (30). A number of potential sources of RWD were delineated, including electronic health records (EHRs), patient registries and administrative claims databases. Additionally, the term Real-World Evidence was coined to entail knowledge generated based on the synthesis of RWD. Since then, a wealth of literature abounded whereby authors focused on multiple aspects of the implementation of RWE in decision making: the potential uses

for RWE in informing decision making (31;32), the design of appropriate studies to generate RWE (22;33) and the analysis and interpretation of RWE (34).

In the midst of rising awareness on RWE amongst various stakeholders, the Innovative Medicines Initiative (IMI)-GetReal project was launched in February 2014 (35). The project was funded through IMI, a public-private partnership between the European Commission and European Federation for Pharmaceutical Industries and Associations (EFPIA). As such, the project consortium consisted of a broad range of stakeholders; from HTA agencies, to the European Medicines Agency (EMA), pharmaceutical industry, small-to-medium enterprises, academia and patient organizations. Over a period of 3 years, the IMI-GetReal consortium sought to explore the potential use of RWE to complement RCT data in effectiveness research throughout the lifecycle of drugs. A series of case studies were conducted to explore methods for using RWE to improve (relative) effectiveness estimates and to examine the acceptability of these methods amongst relevant stakeholders. Furthermore, best practices for evidence synthesis from RWE and/or RCTs were explored through literature reviews and case studies. Bearing in mind the authors' involvement in IMI-GetReal, the consortium's efforts influenced the scope of this thesis.

STATEMENT OF RESEARCH GAP & THESIS OBJECTIVE

Provided the relatively recent interest in RWE use in decision making in healthcare, and despite the abundance of literature on the topic published between 2007 and 2014, an apparent gap exists in knowledge regarding several aspects of the use of RWE in HTA. These include, but are not limited to: whether stakeholders have a unanimous understanding of what RWD is, whether RWE is used by HTA agencies, in which contexts RWE is used (e.g. for REA, CEA or MEA) and the practical and cultural obstacles associated with the use of RWE in HTA and decision making.

This thesis aims to address this gap by exploring the policies and practices of RWE use in HTA of drugs.

OUTLINE OF THESIS

The first section, "What is Real-World Data?", examines the definitions of RWD available in literature and amongst eight different stakeholder groups to provide a clear understanding of what it specifically constitutes. In the second section, "Policies for Real-World Evidence Use", a comparative review of the policies from 6 European HTA agencies on the use of RWE in decision making is conducted. The third section, "Real-World Evidence Use in Practice", examines if and how RWE is used in HTA practice. Chapter 4 explores if RWE is used in practice by 5 European agencies and compares the use of RWE across the 5 agencies. Chapters 5 and 6 present an in-depth analysis of experiences gained with the collection and use of RWE in HTA in the Netherlands on the implementation of Conditional Financing; a form of MEA (36). The fourth section, "Access to Real-World Evidence", summarizes the experiences regarding the access and use of RWE in all case studies conducted within IMI-GetReal

and discusses their implications on RWE use in decision making. The fifth section, "Novel Sources for Generating Real-World Evidence", explores the potential for generating RWE from novel sources (i.e. social media) as an alternative to, for example, EHR's and patient registries. Chapter 8 reviews applications of using social media to generate RWE on relative effectiveness of interventions. Meanwhile, Chapter 9 presents findings from a study whereby social media is used to collect data on patient perceptions on health-related quality-of-life (HRQoL) through health surveys. The final section, "General Discussion", summarizes the findings and discusses the challenges, opportunities and possible future approaches for the increased use of RWE in HTA and decision making.

REFERENCE LIST

1. Work Health Organisation. Constitution of WHO: Principles. 2017 [cited Aug 2017]. Available from: http://www.who.int/about/mission/en/.
2. Garrido MV, Gerhardus A, Röttingen JA, Busse R. Developing health technology assessment to address health care system needs. Health Policy 2010;94(3):196-202.
3. Work Health Organisation. Spending on healthcare: A global overview. 2017 April [cited Aug 2017]. Available from: http://www.who.int/mediacentre/factsheets/fs319/en/.
4. Rijksoverheid. Miljoenennota 2016. 15-9-2015.
5. OECD. What future for health spending? OECD Economic Department Policy Notes . 19-6-2013.
6. Banta HD, Behney CJ, Andrulid DP. Assessing the efficacy and safety of medical technologies. Washington: Office of Technology Assessment 1978.
7. Health Technology Assessment International. HTAi Glossary. 2017 [cited Aug 2017]. Available from: http://htaglossary.net/tiki-index.php?page=List+all+terms.
8. Drummond MF, Schwartz JS, Jänsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. Int J Technol Assess Health Care 2008;24(3):244-58.
9. EUnetHTA. HTA Core Model®. 2017 [cited Nov 2017]. Available from: http://www.eunethta.eu/hta-core-model.
10. Kleijnen S, George E, Goulden S, d'Andon A, Vitré P, Osińska B, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. Value Health 2012;15(6):954-60.
11. HLPF. High Level Pharmaceutical Forum 2005-2008. Final Conclusions and Recommendations. 2008 [cited Apr 2017].
12. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.
13. Luce BR, Drummond M, Jänsson B, Neumann PJ, Schwartz JS, Siebert UWE, et al. EBM, HTA, and CER: clearing the confusion. The Milbank Quarterly 2010;88(2):256-76.
14. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 24;336(7650):924-6.
15. Food and Drug Administration. The Drug Development Process. 2015 June 24 [cited Oct 2017]. Available from: https://www.fda.gov/ForPatients/Approvals/Drugs/default.htm.
16. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20(8):637-48.
17. Cochrane A. Effectiveness and efficiency: Random reflections on health services. 1972 June 1 Available from: https://www.nuffieldtrust.org.uk/research/effectiveness-and-efficiency-random-reflections-on-health-services.
18. Strom BL. Study Designs Available for Pharmacoepidemiologic Studies. Textbook of Pharmacoepidemiology. John Wiley & Sons Ltd; 2013. p. 17-29.
19. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BHC, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol 2004;57(12):1223-31.
20. Rothwell PM. External validity of randomised controlled trials: To whom do the results of this trial apply? The Lancet 2005;365(9453):82-93.
21. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. Clinical Medicine 2008;8(6):579-88.
22. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. J Clin Epidemiol 2010;63(10):1053-8.
23. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens B, et

al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 2011;10(7):495-506.

24. Lohr KN, Eleazer K, Mauskopf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. Health Policy 1998;46(1):1-19.
25. Lipska I, Hoekman J, McAuslane N, Leufkens HGM, H+Âvels AM. Does conditional approval for new oncology drugs in Europe lead to differences in health technology assessment decisions? Clin Pharmacol Ther 2015 Nov 1;98(5):489-91.
26. Garrison Jr LP, Carlson JJ, Bajaj PS, Towse A, Neumann PJ, Sullivan SD, et al. Private sector risk-sharing agreements in the United States: trends, barriers, and prospects. The American journal of managed care 2015;21(9):632-40.
27. Carlson JJ, Gries KS, Yeung K, Sullivan SD, Garrison LP. Current status and trends in performance-based risk-sharing arrangements between healthcare payers and medical product manufacturers. Applied health economics and health policy 2014;12(3):231-8.
28. Faulkner SD, Lee M, Qin D, Morrell L, Xoxi E, Sammarco A, et al. Pricing and reimbursement experiences and insights in the European Union and the United States: Lessons learned to approach adaptive payer pathways. Clin Pharmacol Ther 2016;100(6):730-42.
29. Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. Health Policy 2010;96(3):179-90.
30. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report. Value Health 2007;10(5):326-35.
31. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health 2012;15(2):217-30.
32. Pleil AM. Using Real World Data in Pharamcoeconomic Evaluations: Challenges, Opportunities and Approaches. 2013 [cited Jul 2015]. Available from: http://pharmacy.ucsd.edu/faculty/AppliedPEForum/docs/Andreas_M_Pleil.pdf.
33. ENCePP. The ENCePP Code of Conduct for Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies. 2010 [cited Jul 2015]. Available from: http://www.encepp.eu/documents/encepp_studies/ENCePP%20Code%20of%20Conduct_20100507.pdf.
34. Alemayehu D, Riaz Ali MPP, Alvir JM, Mardekian J, Mullins CD, Perfetto EM, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17(9 suppl.):S3-S37.
35. IMI-GetReal. IMI-GetReal: Overall objectives. 2017 [cited Dec 2016]. Available from: https://www.imi-getreal.eu/About-GetReal/Overall-objectives.
36. Garrison LP, Towse A, Briggs A, De Pouvourville G, Grueger J, Mohr PE, et al. Performance-based risk-sharing arrangements - good practices for design, implementation, and evaluation: report of the ISPOR good practices task force. Value Health 2013;16(5):703-19.

SECTION I

What is Real-World Data?

CHAPTER 2

What is Real-World Data? A review of definitions based on literature and stakeholder interviews

A. Makady[1,2], A. de Boer[2], H. Hillege[3], O. Klungel[2], W. Goettsch[1,2]
(on behalf of GetReal Work Package 1).

[1.] The National Healthcare Institute (ZIN), Diemen, the Netherlands;
[2.] Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, the Netherlands;
[3.] Department of Epidemiology, University Medical Centre Groningen, Groningen, the Netherlands.

Published in Value in Health. Citation is as follows: Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What Is Real-World Data (RWD)? A Review of Definitions Based on Literature and Stakeholder Interviews. Value in Health. 2017 Aug 20(7);858-865.

ABSTRACT

Background

Despite increasing recognition of the value of real-world data (RWD), consensus on the definition of RWD is lacking. We aimed to review definitions publicly available for RWD in order to shed light on similarities and differences between them.

Methods

A literature review and stakeholder interviews were used to compile data from 8 groups of stakeholders. Data from documents and interviews was subjected to coding analysis. Definitions identified were classified into 4 categories: 1-Data collected in a non-RCT setting, 2-Data collected in a non-interventional/non-controlled setting, 3-Data collected in a non-experimental setting and 4-Other (i.e. do not fit into three categories above). The frequency of definitions identified per category was recorded.

Results

53 documents and 20 interviews were assessed. 38 definitions were identified: 20 out of 38 definitions (53%) were category 1 definitions, 9 (24%) were category 2 definitions, 5 (13%) were category 3 definitions and 4 (11%) were category 4 definitions. Differences were identified between, and within, definition categories. For example, opinions differed on the aspects of intervention with which non-interventional/non-controlled settings should abide. No definitions were provided in 2 interviews or identified in 33 documents.

Conclusions

The majority of definitions defined RWD as data collected in a non-RCT setting. However, a considerable number of definitions diverged from this concept. Moreover, a significant number of authors and stakeholders did not have an official, institutional definition for RWD. Persisting variability in stakeholder definitions of RWD may lead to disparities among different stakeholders when discussing RWD use in decision-making.

INTRODUCTION

Randomised controlled clinical trials (RCT's) provide the ideal study design for demonstrating causality between the use of a specific medicine and intended and unintended effects under ideal conditions. In conventional RCT's conducted during Phase III drug development, patients are based on stringent inclusion and exclusion criteria and subsequently randomised to different treatment arms to counteract the influence for known and unknown confounders (1;2). Additionally, monitoring and follow-up procedures for trial subjects is often highly controlled (1;2).

The highly-selective populations examined within the setting of RCTs are often not comparable to the more heterogeneous populations in clinical practice where medicines are administered to patients with varying genetic make-ups, who present with different co-morbidities, or already receive different medications for other morbidities. Consequently, experimental medicines being presented for marketing authorisation are accompanied by data that provides efficacy and safety data with very high internal validity but whose results may not be easily generalizable to a broader, more heterogeneous population (2).This disparity of findings on the therapeutic efficacy of medicines from tightly-controlled RCT settings and the effectiveness of medicines in the real world has been previously defined by Eichler et al. as the "efficacy-effectiveness gap" (3).

Regulatory agencies are thus faced with the issue of making decisions based upon data with inherent uncertainties on the aspects of real-world effectiveness. Similarly, Health Technology Assessment (HTA) agencies and healthcare payers conventionally exploit RCT-generated evidence available at the time of initial reimbursement decisions to assess the relative effectiveness of new products. As a result, many stakeholders such as pharmaceutical industry, regulatory agencies, HTA agencies and payers have begun exploring options for the use of real-world data (RWD) as a complementary source to RCT data for establishing a more robust evidence base on the effectiveness of medicines, as well as the relative effectiveness compared to existing products in clinical practice (4;5).

Additionally, RWD is currently used during drug development to examine aspects such as the natural history of a disease, delineating treatment pathways in clinical practice, determining the costs and resource use associated with treatment interventions and determining outcomes related to comparator interventions (4;6). Such knowledge may inform aspects of early drug development such as clinical trial design or the comparative effectiveness of comparator treatments within a given indication.

Despite the increasing popularity of RWD collection and use for drug development, drug regulation and HTA, a certain degree of disparity remains among different stakeholders when it comes to thoroughly defining RWD (6). Therefore, this article aims to conduct a review of definitions for RWD available in literature and stakeholders' definitions of the term within the context of drug development, drug regulation and HTA of pharmaceutical products, in order to identify the similarities and differences between them. Additionally, the article will review which data sources stakeholders believe as being RWD and which study designs they

consider to generate RWD. Subsequently, the article will shed light on existing definitions for the term RWD developed by the International Society for Pharmacoeconomic and Health Outcomes Research (ISPOR) (7), the Association of the British Pharmaceutical Industry (ABPI) (8), the RAND Corporation (9) and the IMI-GetReal consortium (10) (see Table 1).

METHODS

Two qualitative methods were used to compile data from relevant stakeholders; a literature review and stakeholder interviews. Data compilation from eight stakeholder groups was performed, namely: Health Technology Assessment (HTA) agencies, the pharmaceutical industry, regulatory agencies, academia, healthcare providers, healthcare insurers/payers, patient organisations and initiatives using, or commissioning research on, RWD (e.g. ISPOR, Patient-Centered Outcomes Research Institute (PCORI)).

For the literature review, PubMed was used to search scientific literature from January 1st, 2005 to December 31st, 2016 (date of search). The search strategy used is presented in Figure i in the Appendix. To locate grey literature, websites belonging to 8 stakeholder groups were consulted (see Table i in the Appendix for a list of websites consulted). Search functions on stakeholder websites were used when available, using terms such as: "real world data", "real world evidence", "clinical effectiveness data" , "real world outcome", "comparative effectiveness" or "relative effectiveness". Search results from both scientific and grey literature were independently screened by 2 authors (AM & WG) according to pre-defined inclusion and exclusion criteria (Table ii in the Appendix). Any discrepancies for inclusion and exclusion of articles was resolved by consensus amongst the 2 authors.

Table 1- ISPOR, ABPI, RAND, and IMI-GetReal definitions for real-world data (RWD).

Term & Source	Definition
ISPOR, 2007	Data used for decision-making that are not collected in conventional RCTs.
ABPI, 2011	For the purposes of this guidance, RW data will refer to data obtained by any non-interventional methodology that describes what is happening in normal clinical practice.
RAND, 2014	Real-world data (RWD) is an umbrella term for different types of healthcare data that are not collected in conventional randomised controlled trials. RWD in the healthcare sector comes from various sources and includes patient data, data from clinicians, hospital data, data from payers and social data.
IMI-GetReal, 2015)	An umbrella term for data regarding the effects of health interventions (e.g. benefit, risk, resource use, etc) that are not collected in the context of conventional randomised controlled trials. Instead, real world data (RWD) is collected both prospectively and retrospectively from observations of routine clinical practice. Data collected include, but are not limited to, clinical and economic outcomes, patient reported outcomes (PRO) and health-related quality of life (HRQoL). RWD can be obtained from many sources including patient registries, electronic medical records, and observational studies.

A standardised data abstraction form was created in Microsoft Excel and used to locate information in the documents selected after screening. Data elements included in the data abstraction form were: author(s), publication year, the type of document, definition(s) of RWD provided, and data sources considered as RWD and study designs considered to generate RWD (e.g. claims databases, observational studies, respectively). Two authors (AM &WG) extracted data independently from the selected documents. Any discrepancies in the extracted data were resolved by consensus amongst the 2 authors.

With regards to stakeholder interviews, stakeholders from the 8 previously-mentioned groups were selectively sampled based on seniority and function, with a preference for senior representatives involved in work on RWD use within their respective organisations. Information for identifying representatives was retrieved from stakeholder websites and/or the authors' professional network. All representatives were approached by e-mail using a standardised invitation to participate in semi-structured interviews. In order to increase the validity of stakeholder views, participants were provided the freedom to invite colleagues they deemed relevant to take part in the interviews. Tailored questionnaires were developed for each stakeholder group and sent to stakeholders who agreed to participate 2 weeks prior to the interview to guide discussions (see Figures ii to iv in the Appendix for examples of questionnaires sent to 3 stakeholders groups). Interviews were conducted, recorded and subsequently transcribed for further analysis.

The sampling of stakeholders and interview protocols were compared to recommendations in the consolidated criteria for reporting qualitative studies (COREQ) (11) to ensure good quality. The COREQ checklist provides guidance for explicit and comprehensive reporting of qualitative studies employing interviews and focus groups.

It is important to note that the interviews were conducted as part of a larger study on policies and perspectives on RWD (6), thus the scope of questions posed during the interviews extended beyond the definition of RWD. However, all questionnaires included the following three questions:

1. What is your understanding of the term real-world data (RWD)?
2. Could you provide your own specific definition for RWD?
3. Is RWD routinely collected/used in the context of stakeholder-specific activities and if so, what type of RWD?

This allowed for the standardised collection of data on stakeholders' definitions of RWD, data sources they consider to be RWD and study designs they consider to generate RWD.

Data extracted from documents selected from the literature review and transcripts of stakeholder interviews were subjected to a coding analysis using MaxQDA software 11.0. Following the grounded theory approach in qualitative research (12), data was iteratively assessed by 2 authors (AM & WG) independently to identify repeating themes and tag them using codes. Any discrepancies in codes created were resolved by consensus between the 2 authors. Subsequently, the codes of repeating themes were iteratively refined and grouped into categories. The categories generated formed the categories for RWD definitions and

RWD sources for subsequent analyses. The final coding scheme developed was discussed amongst all authors to ensure consensus. The scheme generated was:

- Categories of RWD definitions
  - Category 1: Data collected in a non-RCT setting (i.e. all health data except that collected in the setting of a conventional phase III RCT setting)
  - Category 2: Data collected in a non-interventional/ non-controlled setting (i.e. data collected without interference with treatment assignment, and/or patient monitoring/follow-up, and/or selection of study population)
  - Category 3: Data collected in a non-experimental setting (i.e. in a setting where the investigator has no control over any of the conditions and no *de novo* data collection occurs based on a pre-established study protocol)
  - Category 4: Other (i.e. none of the above)
- Categories of RWD sources
  - Category A: Data sources (e.g. claims databases, registries)
  - Category B: Study designs that generate RWD (e.g. observational studies, pragmatic clinical trials)

For category 1 (data collected in a non-RCT setting), the term "RCT" referred to the design of a conventional phase III RCT which involves: implementation of inclusion/exclusion criteria for trial subjects, randomisation of subjects to different treatment arms, and consistent monitoring and follow-up procedures for trial subjects and implicit *de novo* data collection. This interpretation of the term "RCT" corresponds to several sources in scientific literature (1;2;10;13).

For category 2 (data collected in a non-interventional/ non-controlled setting), the terms "non-interventional/ non-controlled" referred to a setting wherein the investigator may not be able to interfere with one, or more, of the following aspects: treatment assignment, monitoring and follow-up procedures or inclusion/exclusion criteria. *De novo* data collection may, or may not, occur in this setting. While the authors are aware that several non-identical definitions already exist to define intervention in clinical trials (10;14;15), the interpretation of the term "non-interventional" for the category developed here depended on definitions available from the compiled data.

For category 3 (data collected in a non-experimental setting) the term "non-experimental" referred to a setting in which the investigator cannot alter any of the factors or conditions observed in the study and as such no *de novo* data collection occurs other than data collected in routine clinical practice. This interpretation of the term "non-experimental" corresponds to several sources in scientific literature (10;13;15).

It is important to note that categories 1 to 3 above are not mutually exclusive. For example, data collected in a non-interventional/non-controlled setting is theoretically equivalent to data collected in a non-RCT setting. However, not all data collected in a non-RCT setting is collected in a non-interventional/ non-controlled manner. Similarly, all data collected in a non-experimental setting is theoretically equivalent to that from

a non-interventional/ non-controlled setting but not vice versa. Therefore, there are subtle qualitative differences between the categories which have implications on defining RWD. This will be elaborated upon in the discussion section below.

ANALYSIS

Each RWD definition identified was classified into one of the 4 definition categories (1 to 4) created. The number of definitions per definition category was recorded. Additionally, definitions in each category were qualitatively analysed to highlight differences within, and between, the categories.

Each RWD source identified was classified into one of the 2 sources categories (A and B) created. The number and type of sources per category was recorded.

A sub-analysis was performed for definitions provided by 3 stakeholder groups which are directly involved with RWD collection or appraisal to determine drug effectiveness: the pharmaceutical industry, regulatory agencies and HTA agencies. Definitions identified were compared both within and between the three stakeholder groups.

RESULTS

Initially, the PubMed search yielded 496 hits while the grey literature search yielded 66 hits. Of the 562 total hits, 509 were excluded due to the following reasons: document did not focus on RWD use in pharmaceutical drug development, regulation or HTA (n=490), was not published in English (n=7), was not in one of the document formats outlined in the inclusion criteria (n=6), focused solely on data analysis or evidence synthesis (n=5) or comprised only a summary/abstract (n=1) (see Figures v and vi in the Appendix for PRISMA diagrams of document in- and exclusion from PubMed and grey literature searches, respectively). Eventually, 53 documents were selected (see Table iii in the Appendix for a list of included documents).

Twenty stakeholders from the 8 stakeholder groups agreed to participate (see Table iv in the Appendix for a list of interviews conducted). Eight of the 20 interviews included at least 2 representatives per stakeholder, and 2 included 3 representatives per stakeholder.

In total, 20 definitions were identified in literature documents and 18 definitions were provided in interviews. No definitions were identified in 33 documents nor provided in 2 interviews; 1 interviewee stated not to be familiar with the term at all and the second indicated they cannot provide a definition for RWD. Twenty of the 38 definitions identified (53%) were category 1 definitions. Nine of 38 (24%) were category 2 definitions. Five of the 38 (13%) definitions were category 3 definitions. Four of the 38 (11%) were category 4 definitions; these either provided definitions too general to fit in one of categories 1 to 3, or had defined the concept of "real-world trials", rather than RWD. For an overview of the total number of definitions identified per category, see Figure 1. For examples of definitions identified per category from literature documents and interviews, see Table 2.

For category 2 definitions, it was not always clearly stated what authors and stakeholders perceived as non-interventional or non-controlled settings. According to some, non-

interventional data collection related specifically to the researcher not interfering with treatment assignment and patient management and follow-up (see citations for Pleil et al. and Initiative B in Table 2). Others focussed on another aspect of intervention, namely the selection of study population. One stakeholder believed that RWD should be collected from the population in clinical practice without the implementation of any inclusion or exclusion criteria for selection of patients, while another implied that there might be a selection of study population albeit based on less stringent criteria than those of a RCT (see citations for HTA Agency B and HTA Agency C in Table 2). Another stakeholder focussed on patient randomisation as a criterion for intervention, stating that RWD should thus be collected in a setting where no randomisation of patients occurs (see the citation for Initiative B in Table 2). Meanwhile, other stakeholders cited pragmatic clinical trials (PCT's) and large simple trials (LST's) within their definitions of RWD, despite the fact that both study designs involve randomisation of patients between treatment arms.

The 5 RWD data sources cited most in literature documents and interviews were: registries (18 documents, 7 interviews), Electronic Health Records (EHR's) (16 documents, 6 interviews), claims databases (12 documents, 4 interviews), administrative data (6 documents, 4 interviews) and patient-reported outcomes (10 documents). Meanwhile, the 3 study designs mentioned on more than 5 occasions were: observational studies (22 documents, 6 interviews), PCT's (16 documents, 6 interviews) and post-marketing studies (5 documents, 2 interviews). For a list of the different data sources and study designs retrieved from documents and interviews, as well as the frequency of their mention, see Table 3.

All 4 pharmaceutical industry stakeholders interviewed defined RWD as health data collected in a non-RCT setting (i.e. category 1). Of the 3 regulatory stakeholders interviewed,

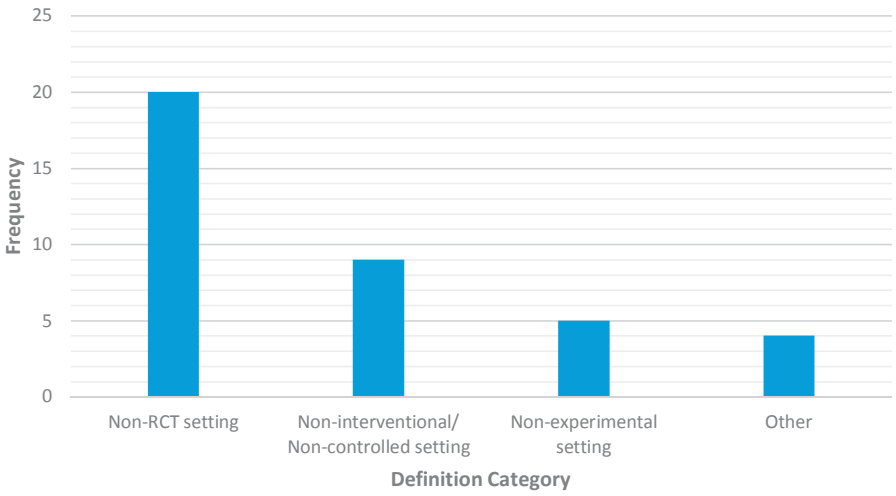


Figure 1 - Overview of the total number of definitions classified under each of the definition categories created.

Table 2 - Examples of definitions from literature documents and stakeholder interviews which have been classified under the 4 definition categories created.

Category of Definition	Citations from literature documents	Citations from stakeholder interviews
Data collected in a non-RCT setting	"We settled on a definition that reflects data used for decision-making that are not collected in conventional RCTs." (7)	"RWD to us means any health record information that is not collected as part of a strict clinical trial (an RCT). So that the physicians and the patients are acting as they are in normal clinical practice. It is more observational in nature." – Pharmaceutical Industry A
Data collected in a non-interventional/ non-controlled setting	"In general, real-world data are observations of effects based on what happens after a prescriptive (treatment) decision is made where the researcher does not, or cannot, control who gets what treatment and does not, or cannot, control the medical management of the patient beyond observing outcomes" (29)	"[RWD is] observational data without blinding and no specific inclusion or exclusion criteria." – HTA Agency B "To us the term RWD is about the scientific process. So we think of it as a step that is not done in a RCT... [RWD] more closely matches the population who will be receiving the drug, or is actually derived from that population. So it is something that was not done in the controlled condition." – HTA Agency C "RWD is data that is generated from the delivery of healthcare in non-controlled settings. Non-controlled settings will generally imply the lack of random assignment." – Initiative B
Data collected in a non-experimental setting	"With RWD, we mean data that are not collected under experimental conditions, but data generated in routine care." (30)	"RWD is data collected from daily clinical practice. This means that it is not collected in a protocol-driven way. Any additional procedures that are conducted because of a research protocol endanger the "real world" aspect of the data." – Regulatory Agency A

Table 2 - continued

Category of Definition	Citations from literature documents	Citations from stakeholder interviews
Other	"RWT's are heretofore ill-defined as a class and, when conducting literature searches, appear to include a large design spectrum ranging from uncontrolled studies or NROTs (stand-alone or follow-up of RCTs' to properly randomised trials that differ only in a few aspects from conventional phase 3 trials. Their stated objective includes the term 'effectiveness' as opposed to 'efficacy', implying that assessment of benefit or risk is taking place in a setting closer to real world clinical practice…" (2)	"RWD are data about effectiveness of treatments collected in the real world. This can be in the setting of a pragmatic trial, collecting for example evidence over time, or a setting where data is collected by health professionals. It can be done retrospectively or prospectively." – Patient Organisation A

Table 3 - List and frequency of occurrence of real-world data (RWD) sources and study types that generate RWD retrieved from literature documents and stakeholder interviews.

Data Sources	**Literature**	**Interviews**	**Total**
Registries	17	8	25
Electronic Health Records (EHR's)	16	6	22
Claims databases	12	4	16
Administrative databases	6	4	10
Patient-Reported Outcomes (PRO's)	10	-	10
Health Surveys	4	2	6
Hospital data	3	3	6
Electronic health data	2	3	5
Clinicians	1	2	3
Payers	1	2	3
Social media	3	-	3
Patient charts	2	-	2
Pharmacy data	2	-	1
Clinical databases	1	-	1
Study Designs	**Literature**	**Interviews**	**Total**
Observational studies	22	6	28
Pragmatic Clinical Tirals (PCT's)	16	6	22
Post-marketing studies	5	2	7
Supplements to RCT's	3	-	3
Drug utilisation studies	1	-	1
Large Simple Trials (LST's)	1	-	1

1 defined RWD as data collected in a non-RCT setting (category 1), while the remaining 2 defined RWD as data collected in a non-experimental setting (category 3). Of the 5 HTA stakeholders interviewed, 2 defined RWD as data collected in a non-RCT setting (category 1), 2 as data collected in a non-interventional/non-controlled setting (category 2) and 1 was unable to provide a definition. Importantly, only 4 of the 12 stakeholders in the sub-analysis had an official, institutional definition for RWD.

DISCUSSION

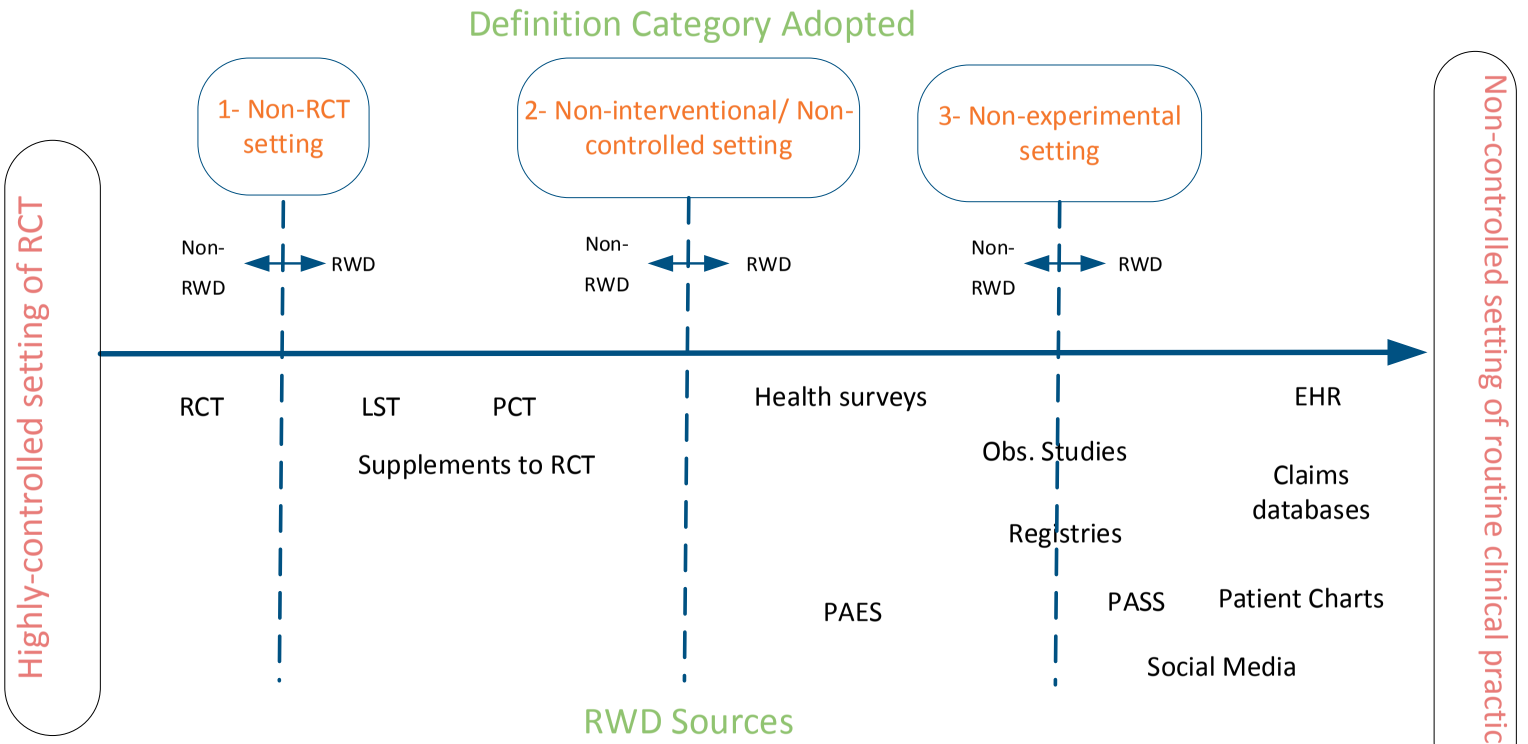
Stakeholders' perception of the value of RWD in enriching evidence on the effectiveness of medications has been steadily increasing. This can be observed in guidelines of HTA agencies which now conventionally include sections on the use of data from non-RCT's (16-18), documents produced by regulatory agencies on post-marketing effectiveness and safety studies (19;20), as well as referral of industry stakeholders to their use of RWD in product development (8;21;22). However, consensus on the value of RWD is contrasted by a lower degree of consensus on what RWD precisely constitutes. As a result, disparity arises amongst stakeholders regarding the definition of what RWD is, the data sources considered as RWD and study designs that generate RWD.

Although results demonstrate that RWD is perceived as health data that is not collected in the setting of a RCT in the majority of cases (20 of 38 cases), this perception is not unanimous. In addition to the qualitative differences between the 3 categories of RWD definitions, critical disparities emerge between definitions of the same category, namely in the category of RWD as data from non-interventional/ non-controlled setting (category 2). Stakeholders differed, and sometimes disagreed, on whether the intervention pertains to treatment assignment, patient monitoring and follow-up, or selection of the study population. This demonstrates, moreover, that some stakeholders may have an incorrect understanding of which aspects of a clinical trial study design classify as an intervention. According to the European Commission, for example, intervention is defined as the researcher's control of treatment assignment or the implementation of additional diagnostic or monitoring procedures (14). This implies that the implementation of selection criteria for the study population would not qualify as an intervention, according to the Commission's definition.

Results presented above indicate that observational studies, registries, EHR's, PCT's and claims databases were the RWD sources most mentioned, respectively. The discussion points mentioned in the previous paragraphs notwithstanding, this would imply that some degree of consensus exists regarding these RWD sources. However, observational studies, the most recurrent example, featured in only 22 of 52 literature documents and 6 of 19 interviews. Looking beyond the 5 most common types, stark controversy exists on whether supplements to RCT's classify as RWD; although the seminal paper by Garrison et al (7) and other literature documents included supplements to RCT's in their list of RWD types, other stakeholders explicitly stated in interviews that they do not consider them as RWD. Therefore, it may be argued that consensus on what sources constitute RWD is also weak.

Moreover, different stakeholders cited data sources and study designs interchangeably as RWD. Although this is not theoretically incorrect, it can lead to disparities between various stakeholders when discussing which sources of data qualify as RWD. For example, depending on their design observational studies may be regarded as interventional or non-interventional by different stakeholders thus more, or less, representative of RWD. Similarly, whether a registry qualifies as a source for RWD mainly depends on the protocol used for data collection: is data collection purely observational of routine care, or is there intervention in the form of additional quality-of-life surveys for included patients?

Bearing the previous points in mind, it would seem that from the perspective of RWD definitions, data sources and study designs identified in literature documents and stakeholder interviews, a spectrum of data exists where on one end, the highly-controlled, randomised setting of the RCT lies (least representative of RWD) and on the other end the non-experimental setting of EHR's, where no intervention is implemented by the investigator and no extra data is collected other than that from routine clinical practice (most representative of RWD). Other data sources and study designs such as PCT´s, observational studies and registries fall between both ends of the spectrum (see Figure 2). Whether such data sources and study designs qualify as RWD is subsequently determined



Legend: RWD = Real-World Data; RCT = Randomised controlled Clinical Trial; LST = Large Simple Trial; PCT = Pragmatic Clinical Trial; PAES: Post-Authorisation Efficacy Study; PASS = Post-Authorisation Safety Studies; Obs. Studies = Observational studies; EHR = Electronic Health Record.

Figure 2 - Data spectrum in relation to RWD definition categories.

by the categories stakeholders adopt for defining RWD. These categories, set in order of least representative of RWD to most representative of RWD, are: all data collected in a non-RCT setting (category 1), data collected in a non-interventional/non-controlled setting (category 2), and data collected in a non-experimental setting (category 3).

If one were to adopt category 1, all data sources/ study designs other than RCT would qualify as RWD; from PCT's to claims databases and EHR's. If one were to adopt category 2, only observational studies whose protocols do not interfere with treatment assignment, patient follow-up, or study population selection would qualify as RWD sources. This would effectively exclude PCT's, LST's and some observational study designs. Finally, if one were to adopt category 3, only data sources such as claims databases and EHR's would qualify as RWD (please see Figure 2 for a diagrammatic representation). Therefore, the choice of categories for defining RWD has direct implications for the types of data and study designs that subsequently classify as RWD.

Several definitions for RWD have been developed over the past years by dedicated taskforces, the seminal examples being provided by ISPOR (7), ABPI (8), RAND (9) and IMI-GetReal (10). The ISPOR definition, developed by a dedicated taskforce, formed the starting point for subsequent ones by ABPI, RAND corporation and IMI-GetReal and succinctly stated that RWD referred to data collected outside the setting of a conventional RCT. To the authors' knowledge, definitions proposed by the ABPI and the RAND corporation were developed by similar taskforces within the respective institutions through internal rounds of discussions. The recent definition developed by IMI-GetReal underwent internal rounds of review within the consortium, as well as external procedures of public consultation, whereby all stakeholders from the wider community were able to provide their opinions on the proposed definition. Eventually, a comprehensive definition was agreed upon by multiple stakeholders which included elements from the ISPOR, ABPI and RAND versions on: the concept of RWD, the domains of information RWD can inform and the types of data which constitute RWD. Moreover, the consortium introduced the term Real-World Studies (RWS) to shed light on the types of study designs which generate RWD, thereby distinguishing these from data sources (10).

Definitions developed by these institutions may provide a starting point for discussions amongst the wider community to achieve consensus on what RWD constitutes. This is particularly important when different stakeholders with differing mandates attempt to discuss the use of RWD in decision-making within the context of drug development, drug regulation and HTA of pharmaceutical products. However, RWD definitions developed by ISPOR, ABPI, RAND and IMI-GetReal, were rarely cited in literature documents and stakeholder interviews. Moreover, several documents either proposed their own definition or lacked one entirely (21;23-26). In addition to this, a significant number of stakeholders interviewed from the pharmaceutical industry, regulatory agencies and HTA agencies did not have an official, institutional definition of RWD nor had adopted any of the definitions mentioned above.

Strengths

Several steps were taken to ensure good research practice during data compilation and analysis. Within the literature review performed on academic and grey literature, the in- and exclusion of documents and subsequent data extraction from selected documents were conducted independently by 2 authors and all discrepancies resolved by consensus. Within stakeholder interviews, the sampling of stakeholders and interview protocols were compared to recommendations in the consolidated criteria for reporting qualitative studies (COREQ) (11) to ensure good quality. Moreover, coding analysis of data extracted from literature documents and interview transcripts was performed independently by 2 authors and all discrepancies resolved by consensus. Finally, categories developed for RWD definitions and sources of RWD based on results of the coding analysis were discussed amongst all authors to ensure consensus.

Two methods were used to compile data needed to achieve the aims of this article, namely a literature review and stakeholder interviews. This provided multiple sources from which the authors could triangulate data on definitions of RWD based on two well-acknowledged qualitative research methods. Moreover, the selection of stakeholders from 8 diverse groups for the grey literature search and interviews helped ensure that a comprehensive view of definitions currently used by relevant stakeholders was available.

Limitations

In order to capture the full perspective of a stakeholder's view on RWD, a representative sample within an organisation should be interviewed. Therefore, it can be argued that stakeholder interviews conducted were insufficient to gather stakeholder perspectives comprehensively. We attempted to account for this by selectively sampling stakeholders, explicitly offering stakeholders approached the opportunity to invite colleagues they deemed relevant to participate in the interviews, and by interviewing more than 1 person per institute. Eventually, 8 of the 20 interviews included at least 2 representatives per stakeholder, and 2 of 20 included 3 representatives.

Definitions provided in documents and interviews varied in length and degree of detail thus implying that the extent of familiarity and experience of different stakeholders with RWD varied. For example, while some were quite detailed, citing a definition of the concept and several data sources, other stakeholders indicated that they were unfamiliar with the term. The degree of variance in length and level of detail provided in different definitions was not analysed in this article, since the aim was not to compare the quality of definitions provided. Instead, the focus of this article was on providing an overview of available definitions of RWD and qualitative differences between them.

Criteria used for defining RCT's to create category 1 of RWD definitions ("Data collected in a non-RCT setting") may present an inherent limitation when trying to conceptualise the placement of certain data sources within the categories created. For example, gene therapy trials, often conducted as open-label, single-arm trials, do not fall under the adopted definition of RCT's. Meanwhile, they are also not non-interventional trials. Another example

relates to open-label extension of RCT's which conventionally precede long-term post-authorisation studies. Such open-label extensions studies are neither RCT's nor non-interventional studies. A final example relates to PCT's and LST's; such trial designs feature randomisation yet implement broader inclusion/exclusion criteria and outcome measures more relevant for clinical practice. As such, they are neither RCT's nor non-interventional studies. In accordance with the grounded theory approach, the criteria for RCT's adopted to develop category 1 were directly elucidated from the data compiled from literature documents and stakeholder interviews. This alludes to a dichotomous attitude amongst stakeholders towards the difference between RCT's and "non-RCT's". The authors of this paper do not favour such a dichotomous representation and have subsequently developed the notion of a data spectrum demonstrated in figure 2 to re-assert the idea that a wide spectrum of data is generated both within RCT settings and non-RCT settings.

CONCLUSION

Stakeholders' acknowledgement of the potential value of RWD in throughout the product lifecycle is increasing. However, despite awareness of the promise RWD brings, disparities persist regarding what RWD precisely is, the types of data sources considered as RWD and study designs generating RWD. Despite the fact that most documents and stakeholders defined RWD as data not collected in the context of a RCT, this perception was not unanimous. Other definitions identified differed and, often, contradicted one another. Moreover, a significant number of authors and stakeholders do not have an official, institutional definition for RWD nor have adopted definitions developed by ISPOR, ABPI, RAND or IMI-GetReal.

From the perspective of RWD definitions, data sources and study designs identified in literature documents and stakeholder interviews, a spectrum of data exists where on one end, the highly-controlled, randomised setting of the RCT lies (least representative of RWD) and on the other end the non-experimental setting of EHR's, where no intervention is implemented by the investigator and no extra data is collected other than that from routine clinical practice (most representative of RWD). All stakeholders concede that data generated by RCT's is not RWD. On the other hand, the question whether health data originating from other data sources or study designs within such a spectrum qualifies as RWD depends on varying categories adopted by stakeholders in their definitions.

In order to ensure that future work involving the collection or use of RWD for drug development, drug regulation and HTA delivers the greatest value to the widest audience, we should move towards developing a common understanding amongst stakeholders of what RWD precisely means, the types of information domains it may inform, the types of data sources which qualify as RWD and study designs which generate RWD. Definitions developed by previous initiatives such as ISPOR, ABPI, RAND and IMI-GetReal provide a good starting point for discussions amongst the wider community to do so.

REFERENCE LIST

1. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol 2009;62(5):499-505.
2. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. J Clin Epidemiol 2010;63(10):1053-8.
3. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, Rowland M, Schneider CK, Bloechl-Daum B. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 2011;10(7):495-506.
4. Berger ML, Lipset C, Gutteridge A, Axelsen K, Subedi P, Madigan D. Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines. Value Health 2015;18(1):127-30.
5. Alemayehu D, Riaz Ali MPP, Alvir JM, Mardekian J, Mullins CD, Perfetto EM, Sanchez RJ, Subedi P, Willke RJ. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17(9 suppl.):S3-S37.
6. Makady A. Review of Policies and Perspectives on Real-World Data. 2015;Available from: URL: http://tinyurl.com/p2qq9xh
7. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report. Value Health 2007;10(5):326-35.
8. ABPI. Demonstrating Value with Real World Data: A practical guide. 2011;Available from: URL: http://www.abpi.org.uk/our-work/library/guidelines/Pages/real-world-data.aspx
9. Celine M, Robin E, Horvath V et al. Health and Healthcare: Assessing the Real World Data Policy Landscape in Europe. Santa Monica, CA: RAND Corporation, 2014. http://www.rand.org/pubs/research_reports/RR544.html.
10. Makady A. IMI-GetReal Glossary. Goettsch W., Willemsen A., Jonsson P, Hummel N, Nordon C, editors. 2015. IMI-GetReal.
11. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-57.
12. Corbin JM, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21.
13. Makady A. GetReal. Glossary of definitions of common terms. 2015;Available from: URL: http://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary.pdf
14. European Commission. Directive 2001/20/ec of the European Parliament and of the Council on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Official Journal of the European Union 2001;Available from: URL: http://ec.europa.eu/health/files/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf
15. Last JM, International Epidemiological Association. A dictionary of epidemiology. 141 ed. Oxford Univ Press; 2001.
16. Delwel G.O. Leideraad voor Uitkomstenonderzoek. 2008. Zorginstituut Nederland.
17. Haute Autorité de Santé. Les études post-inscription sur les technologies de santé (médicaments, dispositifs, médicaux et actes). 2011;Available from: URL: http://tinyurl.com/orw8cw2
18. Faria R, Alava MH, Manca A, Wailoo AJ. NICE DSU Technical Support Document 17: The Use of Observation Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data. 2015.

19. European Medicines Agency. European Medicines Agency post-authorisation procedural advice for users of the centralised procedure. 2015;Available from: URL: http://tinyurl.com/ph7aa8k

20. European Medicines Agency. Delegated Act on Post-authorisation Efficacy Studies. 2013;Available from: URL: http://ec.europa.eu/health/files/pharmacovigilance/2012_11_28_pc_paes.pdf

21. Sanofi. Main Sanofi positions on CSR topics. 2013;Available from: URL: http://tinyurl.com/qxewnu4

22. Novartis. Leaders in Clinical Trial Data Transparency. 2014;Available from: URL: https://www.novartis.com/sites/www.novartis.com/files/leaders-in-clinical-trial-data-transparency.pdf

23. Foltz D, Ferrara L, Volkommer R, Turisco F. Real-World Data Research: A case for action. 2013;Available from: URL: http://assets1.csc.com/health_services/downloads/CSC_Real_World_Data_Research.pdf

24. Doležal T. Real world data in Czech Republic. 2008;Available from: URL: http://tinyurl.com/o3lvm4o

25. van Staa TP, Klungel OH. Background Paper 8.4 Real-life data and learning from practice to advance innovation. 2013;Available from: URL: http://www.who.int/medicines/areas/priority_medicines/BP8_4Data.pdf

26. Merck. Merck and Israel's Maccabi Healthcare to Leverage Unique Real-World Database to Inform Novel Health Approaches. 2013; Available from: URL: http://tinyurl.com/nnuffxn

27. Machin D., Campbell M.J., Walters S.J. Medical Statistics: A textbook for Health Sciences. Fourth ed. 2007.

28. European Commission. DIRECTIVE 2001/20/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Official Journal of the European Union . 2001.

29. Pleil AM. Using Real World Data in Pharamcoeconomic Evaluations: Challenges, Opportunities and Approaches. 2013;Available from: URL: http://pharmacy.ucsd.edu/faculty/AppliedPEForum/docs/Andreas_M_Pleil.pdf

30. Romio S, Sturkenboom M, Corrao G. Real-world data from the health decision maker perspective. What are we talking about? Epidemiol Biostat Public Health 2013;10(3).

SECTION II

Policies for Real-World Evidence Use

CHAPTER 3

Policies for Use of Real-World Data in Health Technology Assessment (HTA): A comparative study of 6 HTA agencies

A. Makady[1,2], R. ten Ham[2], A. de Boer[2], H. Hillege[3], O. Klungel, W. Goettsch[1,2] (on behalf of GetReal Workpackage 1)

[1] The National Healthcare Institute (ZIN), the Netherlands
[2] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, the Netherlands
[3] Department of Epidemiology, University Medical Centre Groningen, Groningen, the Netherlands

Published in Value in Health. Citations is as follows: Makady A, ten Ham R, de Boer A, Hillege H, Klungel O, Goettsch W. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. Value in Health. 2017 Apr 30;20(4):520-32.

ABSTRACT

Randomised Controlled Trials (RCT) provide robust data on the efficacy of interventions, rather than effectiveness. Health Technology Assessment (HTA) agencies worldwide are thus exploring whether Real-World Data (RWD) may provide alternative sources of data on effectiveness of interventions. Presently, an overview of HTA agencies' policies for RWD use in relative effectiveness assessments (REA) is lacking. This study aimed to review policies of 6 European HTA agencies on RWD use in REA of drugs. A literature review and stakeholder interviews were conducted to collect information on RWD policies for 6 agencies: TLV (Sweden), NICE (U.K.), IQWiG (Germany), HAS (France), AIFA (Italy) and ZIN (Netherlands). The following contexts for RWD use in REA of drugs were reviewed: initial reimbursement discussions, pharmacoeconomic analysis and conditional reimbursement schemes. We identified 13 policy documents and 9 academic publications, and conducted 6 interviews. Policies for RWD use in REA of drugs notably differed across contexts. Moreover, policies differed between HTA agencies. Such variations might discourage the use of RWD for HTA. To facilitate the use of RWD for HTA across Europe, more alignment of policies seems necessary. Recent EUnetHTA papers and project proposals may provide a starting point to achieve this.

INTRODUCTION

In light of rising healthcare costs and the introduction of innovative, yet expensive, pharmaceutical products, Health Technology Assessment (HTA) agencies are seeking robust methods for relative effectiveness assessments (REA) of drugs in routine clinical practice. The relative effectiveness of an intervention is defined as: "The extent to which an intervention does more good than harm, when compared to one or more intervention alternatives for achieving the desired results and when provided under the routine setting of health care practice (i.e. real-world setting)." (1)

Conventionally, data on treatment effects for drugs is collected within the context of randomised controlled trials (RCT's), whereby a selected, homogenous group of patients is randomly assigned to either the experimental drug or a comparator (e.g. placebo or active comparator) under highly controlled conditions. This study design is ideal to demonstrate the efficacy of a drug, due to its ability to minimise problems with confounding, information bias and selection bias. However, once a drug gains marketing authorisation, it is administered to a heterogeneous patient group in routine clinical practice whereby patients present with differing co-morbidities, co-medication and genetic profiles. Consequently, it is challenging to extrapolate results from RCT's to drug effects in clinical practice (2).

Due to limitations associated with the use of RCT-generated efficacy data to predict the relative effectiveness of drugs, HTA agencies worldwide are currently exploring the possibilities for using Real-World Data (RWD) to supplement and enrich the evidence for REA of drugs. Examples of national and international collaborations exploring these possibilities include the Patient-Centered Outcomes Research Initiative (PCORI) and the Innovative Medicines Initiative GetReal Consortium (IMI-GetReal). IMI-GetReal is a 3-year project aiming at investigating policies and methodologies for the collection and use of RWD in drug development and assessment. It combines a broad array of stakeholders across Europe to collaborate on developing a policy framework for RWD use and good practices for its integration in the evidence base.

Additionally, HTA agencies are exploring the use of evidence development strategies that provide effectiveness research data earlier during drug development in the framework of medicine adaptive pathways to patients (MAPP's)3. One example, the IMI-ADAPT SMART project, is a 3-year project enabling a platform for multiple-stakeholder discussions on questions relating to implementation of MAPP's activities in the European setting. Moreover, numerous publications have highlighted the growing need for RWD use in HTA decision-making to inform: clinical effectiveness parameters, natural history of disease, adherence to treatment and health-related quality of life, or information on demand and supply constraints for health economic evaluations in specific settings (4-9).

Research conducted by IMI-GetReal identified three contexts within which RWD is currently being used for REA of drugs: as supplementary input for initial REA after market authorisation, as input for pharmacoeconomic analyses (PEA) and for the re-assessment of relative effectiveness within conditional reimbursement schemes (CRS) (8). However,

an overview of the similarities and differences between different HTA agencies' policies for the use of RWD in the three contexts mentioned above seems to be lacking. Provided recent efforts and growing interest for the harmonisation of HTA activities across Europe (e.g. as demonstrated by activities of the European network of HTA (EUnetHTA)), an initial comparison of policies for RWD use by HTA agencies across a number of European jurisdictions may provide a good starting point for further discussions on harmonisation of policies on this topic.

Therefore, this article aims to review the policies of 6 HTA agencies in Europe on RWD use in REA of drugs. More specifically, the article considers agencies' policies regarding RWD accepted or requested, as well as policies for the appraisal of RWD in the following three contexts: initial reimbursement discussions (IRD), PEA and CRS. It is important to note that this article does not aim to provide a comprehensive overview of RWD policies of HTA agencies in all 29 European jurisdictions but rather to present a comparison across several relevant jurisdictions in Europe.

METHODS

Six European HTA agencies were selected for analysis: the Dental and Pharmaceutical Benefits Agency (TLV, Sweden), the National Institute for Health and Care Excellence (NICE, the United Kingdom), the Institute for Quality and Efficiency in Healthcare (IQWiG, Germany), Haute Autorité de Santé (HAS, France), the Italian Medicines Agency (AIFA, Italy) and the National Healthcare Institute (ZIN, the Netherlands). HTA agencies within France, Germany, Italy and the U.K. were selected since they represent the 4 largest European jurisdictions (the so co-called "Big Four"); jurisdictions bearing most influence on European policies on several aspects, including health (10-12). Meanwhile HTA agencies in Sweden and the Netherlands were selected due to their pioneering roles, both historically and currently, in cutting-edge European HTA projects, such as the European network of HTA (EUnetHTA) (13). To ensure that all relevant information on agencies' policies on RWD use in REA of drugs was collected, 3 methods were used to retrieve information: a review of agencies' guidelines and policy papers, a review of academic publications by HTA affiliates on RWD use in REA of drugs and semi-structured interviews with representatives from the selected agencies.

Firstly, the websites of the 6 HTA agencies were searched for guidelines and policy papers within the three contexts: IRD, PEA and CRS. Documents were included if they were published in English, German, French or Dutch. Secondly, a search for academic articles published by agency affiliates relating to RWD use in REA of drugs was conducted in Medline using the PubMed interface (see the Appendix for the search strategy). To minimise chances of missing relevant literature, a time span of 10 years was selected. Articles were included if: they were published between January 1st 2006 and June 21st 2016 (date of search), explicitly discussed the use of RWD in REA of drugs, were published in English, German, French or Dutch (Swedish and Italian documents were excluded since the study authors do not master

these two languages) and comprised more than an abstract. Articles were excluded if they did not meet all inclusion criteria. Documents retrieved from agency websites and PubMed searches were evaluated independently by two authors. Any disagreements regarding inclusion or exclusion of articles were resolved by consensus.

Thirdly, semi-structured interviews were conducted with representatives from the 6 HTA agencies. Representatives were selectively sampled based on seniority and function, with a preference for senior HTA assessors and Research & Development senior officers. Information for identifying representatives was retrieved from agency websites and/or the authors' professional network. All representatives were approached by e-mail using a standardised invitation. A standardised questionnaire was sent to all representatives who agreed to participate 2 weeks prior to the interview to guide discussions (see Figure i in the Appendix). In order to increase the validity of stakeholder views, participants were provided the freedom to invite colleagues they deemed relevant to take part in the interviews. Interviews were conducted, recorded and subsequently transcribed for further analysis. The sampling of representatives and interview protocols were compared to the consolidated criteria for reporting qualitative studies (COREQ) to ensure good quality (14).

It is important to note that the interviews were conducted as part of a broader review of stakeholder policies and perspectives on RWD (8). Therefore, the scope of questions posed in interviews extended beyond the aims of this research.

A standardised coding scheme was developed using MaxQDA 11.0 software to extract data from all compiled documents and transcripts on two aspects; RWD accepted or requested, as well as the appraisal of RWD for REA of drugs within IRD, PEA and CRS (see Figure 1). The scheme was developed by iterative assessment of included documents and interview transcripts, in accordance with the directed content analysis approach for qualitative research (15). Two authors independently performed data abstraction and coding. Any discrepancies were resolved by consensus.

The results from the coding analysis of the compiled documents and transcripts reported in this manuscript were subsequently verified with the interviewed representatives of all 6 agencies to ensure factual correctness.

For the purpose of this article, we based our definition for RWD on the IMI-GetReal definition: "An umbrella term for data regarding the effects of health interventions (e.g. safety, effectiveness, resource use, etc) that are not collected in the context of highly-controlled RCT's. Instead, RWD can either be primary research data collected in a manner which reflects how interventions would be used in routine clinical practice or secondary research data derived from routinely collected data. Data collected include, but are not limited to, clinical and economic outcomes, patient-reported outcomes and health-related quality of life. RWD can be obtained from many sources including patient registries, electronic medical records, and claims databases." (16)

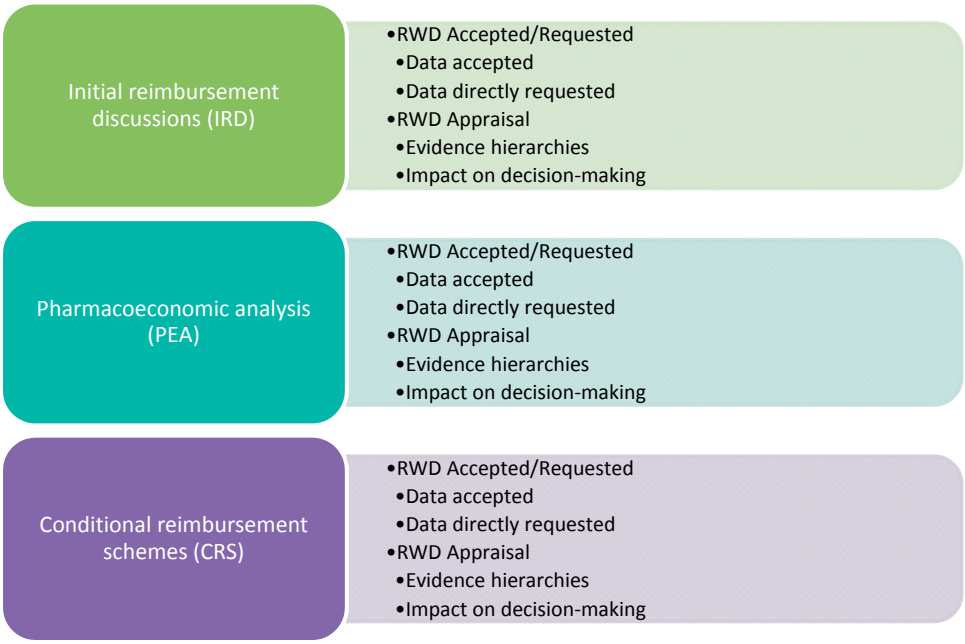


Figure 1 – Coding scheme developed to conduct coding analysis.

RESULTS

The search for guidelines and policy papers on RWD use on agency websites yielded 13 documents (see Table 1). All 6 agencies had guidance and policy papers available for IRD, 5 agencies for PEA and 3 agencies for CRS. The number and nature of documents varied per institute. Some agencies had separate guidelines for IRD and PEA (e.g. TLV, HAS, ZIN), whereas others combined both in one document (e.g. NICE).

The Pubmed search initially yielded 284 hits; 9 were selected for further analysis and 275 were excluded because they did not meet all inclusion criteria (see Figure 2 for PRISMA diagram on article selection). Of the 9 selected articles, 1 involved affiliates from several HTA agencies (17), 4 were specific to AIFA affiliates (18-21), 3 were specific to NICE affiliates (22-24) and 1 was specific to an HAS affiliate (25) (see Table 1).

Of the 9 agency representatives approached across the 6 agencies, all agreed to participate (response rate= 100%). For 2 of the 6 agencies, 1 additional colleague was invited by the approached representatives to participate in the interview. Two interviews included 1 agency participant, 3 included 2 agency participants and 1 included 3 agency participants (see Table 1).

In total, 22 documents and 6 interview transcripts were included in the analysis.

Table 1 - List of policy documents, guidelines and academic publications retrieved, as well as the number of interview participants and transcript reference per agency.

HTA Agency	Policy Papers & Guidelines	Academic Publications	Number of Interview Participants & Transcript Reference
TLV	Guide for companies when applying for subsidies and pricing for pharmaceutical products (21) General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2) (28) The Swedish Pharmaceutical Reimbursement System (33)	-	1 participant Transcript reference: 'a'
NICE	Guide to the methods of technology appraisal 2013 (22) NICE DSU Technical Support Document 17: The Use of Observational Data to Inform the Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data (26)	Evidence Requirements for Reimbursements of Pharmaceuticals Across Europe (12) Methodological Challenges in Evaluating the Value of Registries (18) Evidence Informed Decision Making: The Use of "Colloquial Evidence" at NICE (17) How real-world data compensate for scarce evidence in HTA (19) How to improve the quality of evidence when new treatments are funded conditional on collecting evidence of effectiveness and safety (20)	3 participants Transcript reference: 'b'
IQWiG	Allgemeine Methoden version 4.2 (23)	-	1 participant Transcript reference: 'c'

Table 1 - continued)

HTA Agency	Policy Papers & Guidelines	Academic Publications	Number of Interview Participants & Transcript Reference
HAS	General Method for Assessing Health Technologies (24) Choices in Methods for Economic Evaluation (29) Les études post-inscription sure les technologies de santé (médicaments, dispositifs médicaux et actes) (31)	-	2 participants Transcript reference: 'd'
AIFA	-	Evidence Requirements for Reimbursements of Pharmaceuticals Across Europe (12) New Perspective and new challenges in clinical trial regulation in Italy (13) Feasibility and challenges of independent research on drugs: the Italian Medicines Agency (AIFA) experience (15) The Italian post-marketing registries (14) The nationwide Osmed Health-Db database: a tool to support health-care decision-making and real-world evidence generation (16)	2 participants Transcript reference: 'e'
ZIN	Beoordeling stand van de wetenschap en praktijk (25) Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg (30) Leideraad voor Uitkomstenonderzoek (27) Procedure voorwaardelijke toelating geneeskundige zorg 2015 (32)	Evidence Requirements for Reimbursements of Pharmaceuticals Across Europe (12)	2 participants Transcript reference: 'f'

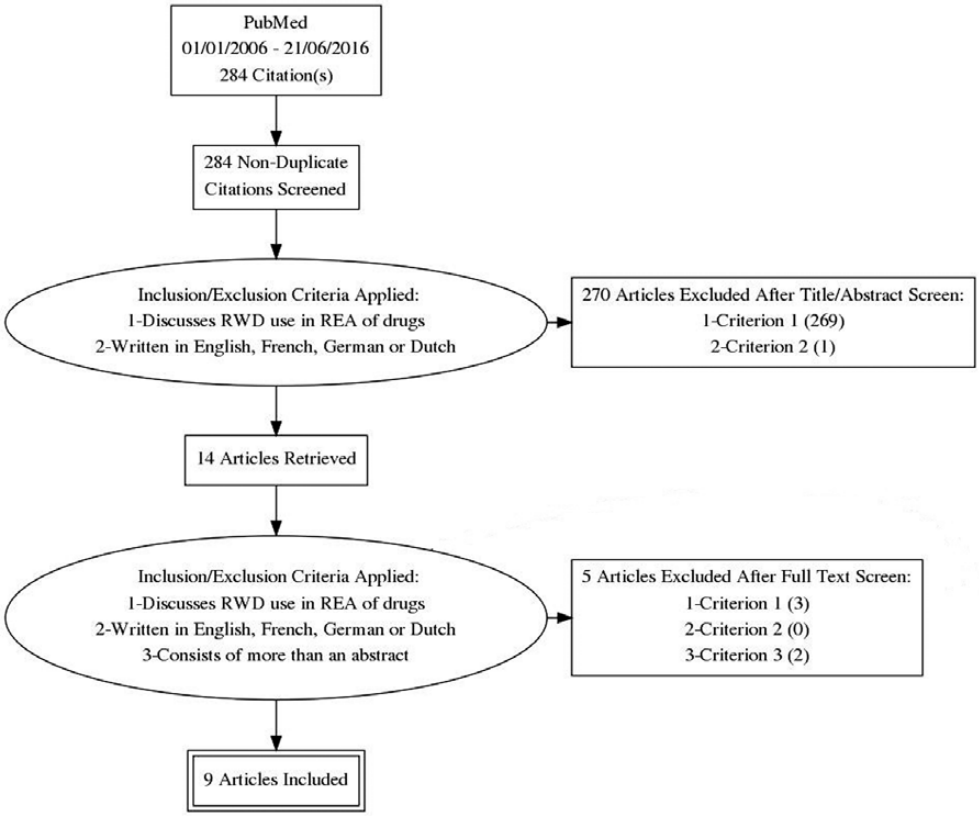


Figure 2 - PRISMA diagram of inclusion & exclusion of articles retrieved through the PubMed search.

Initial Reimbursement Discussions

All HTA agencies accept all available evidence on the drug undergoing REA, which implicitly includes RWD (26-30;a-f). Agencies do not specify which sources of RWD nor which methodologies for RWD collection the applicant should resort to (26, 27, 29, 30;a,b,d-f). However, several do provide suggestions for specific RWD sources as well as preliminary guidance on the suitability of these sources to answering different scientific questions (27, 28, 31, 32;b,c,f).

Agencies iterate that RWD may be used to demonstrate treatment effects of the assessed drug but only under specific circumstances. For example, RWD may be used in the absence of RCT evidence on drug efficacy (27, 28, 30;b,f). In the absence of RCT data on head-to-head comparisons between treatments, RWD may be drawn upon to provide information on estimates of effectiveness to enable indirect treatment comparisons (27, 30;b,f). Finally, RWD may be used to supplement RCT data on treatment effects if data on specific subpopulations or long-term follow-up is lacking (27, 30;b,f). In all situations mentioned above, agencies require an explicit justification

why RWD was used and a clear discussion of the biases associated with the RWD used and its consequences on treatment effect estimates (26-30;a-c,e).

Moreover, 3 agency guidelines iterate that RWD may be used to provide information on aspects other than treatment effect, such as: epidemiological data (e.g. incidence and prevalence), resource use data and cost data (27, 28, 30).

All agencies adopt similar hierarchies of evidence in accordance with principles of evidence-based medicines (26-30). Adopted hierarchies unanimously place sources of RWD on a lower level of quality and reliability than RCT's. Consequently, agencies iterate that RWD may be used to confirm or supplement, rather than substitute, findings on causal treatment effects demonstrated by RCT's (27-30;b-d,f). Thus, conclusions on treatment effects derived from RWD are generally regarded as more circumspect than RCT-derived conclusions by decision-making committees; examples of quotes to this effect can be found in table 2. However, 2 agencies explicitly recognise limitations associated with strictly adopting evidence hierarchies in guidelines and state that such hierarchies should not preclude the exclusion of valuable non-RCT evidence from decision-making (27, 30;b,f).

Agencies differ on the acceptability and impact of RWD on decision-making in cases where RCT data is sparse, for example for orphan diseases: several state that non-RCT data could be resorted to for decision-making in these cases (27, 30;a,b,f), whereas one states that resorting to non-RCT data presents a greater risk to validity of conclusions thus should be avoided (28;c). Examples of quotes demonstrating agencies' disparity of views on this issue can be found in Table 2.

Table 3 presents a summary of policies on RWD accepted or requested and RWD appraisal in the context of IRD per agency.

Pharmacoeconomic Analysis

Contrary to the first context, RWD is directly requested by 5 HTA agencies for various aspects of PEA (the 6th agency does not conduct PEA). More specifically, agencies recommend that epidemiological data (e.g. incidence and prevalence), direct and indirect costs, and resource use in routine practice be collected from national RWD sources (e.g. claims databases, registries, hospital databases) (27, 28, 32-34;b,e,f). Other aspects of the evaluation like adherence to treatment and compliance, can also be collected from RWD sources such as registries, databases, ad-hoc studies or epidemiological surveys (34).

Several agencies specify that treatment effects used for modelling relative effectiveness should primarily be based upon results from RCT's (27, 32, 34, 35;b,d,f). Alternatively, RWD may provide complementary evidence on treatment effects (27, 30, 34;b,d,f), be used to valuate the health effects over time in the form of utilities (27, 32, 35) or provide data on transition probabilities between different disease states in pharmacoeconomic models (27, 32).

With regards to RWD appraisal in PEA, the use of RWD for epidemiological data, direct and indirect costs, resource use in routine practice, adherence and compliance to

Table 2 - Examples of interview quotes on RWD use in IRD and CRS.

Context for RWD Use (IRD or CRS)	Topic	Quotation A	Quotation B
IRD	Appraisal of RWD vs. RCT data for treatment effect estimates in general	"There is this red flag in there. If you use non randomized and non-controlled evidence, you have to be more careful, more circumspect about the relative treatment effect drawn from those studies. Ideally you should use more than one independent source of such evidence, as a back-up." (b)	"Of course we accept those data. We are forced by law to accept those data but we don't have to conclude the benefit from such data." (c)
IRD	RWD use to inform treatment effect estimates for orphan diseases	"Yes, RWD certainly plays a role in orphan diseases since RCT's are difficult to conduct in that area. In this case, patient registries may be the most ideal source for RWD." (f)	"… we would then need a registry with a very, very, very, high quality. In terms of having all patients in the registry, no selection criteria and no selection bias. We could imagine that we would only then accept these registry analyses for very rare diseases, but not in general." (c)
CRS	Use of RWD generated within CRS for decision-making	"So, we are used to using that kind of data, though we know the bias and the problems that are related to the robustness of that kind of [RWD] data. For the re-evaluation for the pricing and reimbursement of the product, this kind of data are robust enough for the analysis that we need to do for the reevaluation of pricing and reimbursement of the product." (e)	"You can't really rely on it. You can use the RWD as a a confirmation of the expectation you have on initial assessment and the data for the first line population you have, and the data you have had already of the post-hoc subgroup analysis. So it is used as a confirmation of previous conclusions." (f)

treatment is largely accepted by HTA agencies. However, for relative treatment effects, the same hierarchies of evidence apply as in the context of IRD, implying that RWD is conventionally placed on a lower quality level (27, 28, 32, 34, 35;b,d,f). Therefore,

Table 3 - Summary of policies on RWD accepted or requested and the appraisal of RWD in the context of IRD per agency.

	RWD Accepted/Requested			RWD Appraisal		
HTA Agency	**RWD Accepted**	**RWD to inform treatment effects**	**RWD to inform other parameters**	**Hierarchy of evidence adopted**	**Conclusions on treatment effects based on RWD regarded as circumspect**	**Conclusions on treatment effects based on RWD possible in exceptional circumstances (e.g. orphan diseases)**
TLV	Yes	Under specific circumstances	Not mentioned	Yes; with regards to evidence for treatment effects	Yes	Yes
NICE	Yes	Under specific circumstances	epidemiological data (e.g. incidence and prevalence), resource use data & cost data	Yes*; with regards to evidence for treatment effects	Yes	Yes
IQWiG	Yes	Under specific circumstances	epidemiological data (e.g. incidence and prevalence), resource use data	Yes; with regards to evidence for treatment effects	Yes	No
HAS	Yes	Under specific circumstances	Not mentioned	Yes; with regards to evidence for treatment effects	Yes	Not mentioned
AIFA	Yes	Under specific circumstances	Not mentioned	Yes; with regards to evidence for treatment effects	Yes	Not mentioned
ZIN	Yes	Under specific circumstances	epidemiological data (e.g. incidence and prevalence), resource use data & cost data	Yes*; with regards to evidence for treatment effects	Yes	Yes

*However, agency explicitly recognises limitations associated with strictly adopting evidence hierarchies in guidelines and states that such hierarchies should not preclude the exclusion of valuable non-RCT evidence from decision-making.

conclusions for relative treatment effects based on RWD are considered as being more circumspect (27, 28, 32, 34, 35;b-d,f).

Table 4 presents a summary of policies on RWD accepted or requested and RWD appraisal in the context of PEA per agency.

Conditional Reimbursement Schemes

Three of the 6 HTA agencies implement CRS (19, 20, 36, 37;d-f). A fourth agency stated briefly that reimbursement can be conditionally offered to allow an applicant time to procure more RWD on long-term effects (38;a). Meanwhile, a fifth agency recently announced the establishment of a CRS scheme for oncologic drugs (39;b). However, it remains unclear whether the latter two schemes constitute ones as established as those outlined by the other 3 agencies (a,b).

Only 1 of the 3 agencies clearly defined criteria for the selection of candidates for CRS and a procedure to do so (32, 37;f).

The purposes for RWD collection for CRS differed between the 3 agencies. For the first agency, a product is nominated for conditional reimbursement on two conditions; that it is highly innovative and data on its effectiveness is highly sparse at initial assessment. Therefore, the purpose for data collection is focused primarily on demonstrating effectiveness, with a preference for RCT data and a supplementary role for RWD (32, 37;f). For the second agency, a contract is drawn up between the agency and an applicant to conduct post-marketing studies that aim to answer questions raised during initial assessment. These questions may relate equally to issues of effectiveness and/or cost-effectiveness of the drug in national clinical practice and a preference is made for RWD rather than RCT data (36;d). For the last agency, recommendations to set up post-marketing studies are similarly based upon questions raised during initial assessment with a preference for RWD. However, the use of study results for the last agency varies; they can be used to inform re-assessment of effectiveness and/or cost-effectiveness in clinical practice, but may also be used for re-pricing discussions (19, 20;e).

These principle differences notwithstanding, 2 agencies follow the same procedure for conditional reimbursement. Firstly, gaps in evidence presented in submissions for initial reimbursement discussions are systematically identified by the agencies. Secondly, the agencies request that the applicant develop a study protocol to collect the RWD needed to inform such gaps, implying that RWD collected for each drug candidate is highly case-specific. Both agencies provide methodological guidance to applicants on which study designs to choose in order to answer the scientific questions raised during initial assessment. This guidance also includes detailed examples of existing national RWD sources that may be used to answer specific questions (32, 36, 37). Thirdly, the applicant's study protocol(s) are reviewed by independent committees to judge their scientific quality and feasibility. Once relevant adjustments are made to the protocol(s), a contract is drawn up between the agency and applicant within which the study protocol and date for submitting additional evidence is specified. Further adaptations to the study protocol by the applicant

Table 4 - Summary of policies on RWD accepted or requested and the appraisal of RWD in the context of PEA per agency.

	RWD Accepted/Requested			RWD Appraisal		
HTA Agency	**RWD Recommended**	**RWD to inform treatment effects**	**RWD to inform other parameters**	**Hierarchy of evidence adopted**	**Conclusions on treatment effects based on RWD regarded as circumspect**	**Conclusions on other parameters based on RWD regarded as reliable**
TLV	Yes	Under specific circumstances	epidemiological data (e.g. incidence & prevalence), costs (direct & indirect) and resource use	Yes; with regards to evidence for treatment effects	Yes	Yes
NICE	Yes	Under specific circumstances	epidemiological data (e.g. incidence & prevalence), costs (direct & indirect) and resource use	Yes*; specifically with regards to relative treatment effects	Yes	Yes
IQWiG	N/A	N/A	N/A	N/A	N/A	N/A
HAS	Yes	Under specific circumstances	epidemiological data (e.g. incidence & prevalence), costs (direct & indirect), resource use, adherence and compliance	Yes; with regards to evidence for treatment effects	Yes	Yes
AIFA	Yes	Under specific circumstances	epidemiological data (e.g. incidence & prevalence), costs (direct & indirect) and resource use	Yes; with regards to evidence for treatment effects	Yes	Yes

Table 4 - continued

	RWD Accepted/Requested			RWD Appraisal		
HTA Agency	**RWD Recommended**	**RWD to inform treatment effects**	**RWD to inform other parameters**	**Hierarchy of evidence adopted**	**Conclusions on treatment effects based on RWD regarded as circumspect**	**Conclusions on other parameters based on RWD regarded as reliable**
ZIN	Yes	Under specific circumstances	epidemiological data (e.g. incidence & prevalence), costs (direct & indirect) and resource use	Yes*; with regards to evidence for treatment effects	Yes	Yes

are possible but only after consultation with the agency (32, 36, 37). It is unclear whether the same procedure also applies for CRS implemented by the third agency.

Unlike the first 2 agencies, which lay the burden of RWD collection on applicants, the third agency often actively participates in, or initiates its own, product or indication registries (18-20, 36, 37;d,e,f).

All 3 agencies require that the studies implemented deliver data of adequate quality and robustness to answer questions identified during initial assessment (32, 36, 37;d,e,f). Moreover, 2 agencies require that the study eventually conducted adhere strictly to the protocol agreed upon by all parties. This is to ensure that the scientific quality and outcomes of the study remain valuable for decision-making. If these conditions are met, results generated by the studies would form the basis for decision-making during re-assessment (32, 36, 37;d,e,f). However, quotes from interviews shed light on varying acceptability of resulted generated from such studies for decision-making practice (see table 2). Moreover, there was no guidance on the impact of RWD on decision-making if conclusions for treatment effects based on RWD contradict those from RCT-based evidence.

Table 5 presents a summary of policies on RWD use in the context of CRS per agency.

Similarities and differences in policies for RWD accepted or requested and RWD appraisal within IRD, PEA and CRS are presented in Table 6.

Table 5 - Summary of policies on RWD use in the context of CRS per agency

HTA Agency	CRS implemented?	CRS Aims	CRS Procedure	Preference for RWD	Involvement in collection of RWD?	Preference for RWD	Impact of RWD on decision-making
TLV	No*	N/A	N/A	N/A	N/A	N/A	N/A
NICE	No*	N/A	N/A	N/A	N/A	N/A	N/A
IQWiG	No	N/A	N/A	N/A	N/A	N/A	N/A
HAS	Yes	Effectiveness and/or cost-effectiveness	1- Identification of evidence gap 2- Consultation on study design 3- Decision-making based on results	Yes	No	Yes	Conditional on whether data delivered sufficiently addresses evidence gap highlighted and adherence to agreed-upon study protocol.
AIFA	Yes	Effectiveness, cost-effectiveness and/or price re-negotiations	Not mentioned	Yes	Yes	Yes	Conditional on whether data delivered sufficiently addresses evidence gap highlighted.
ZIN	Yes	Effectiveness	1- Identification of evidence gap 2- Consultation on study design 3- Decision-making based on results	No; in first instance RCT data with RWD as supplementary	No	No; in first instance RCT data with RWD as supplementary evidence	Conditional on whether data delivered sufficiently addresses evidence gap highlighted and adherence to agreed-upon study protocol.

Table 6 - Summary of similarities and differences in policies for RWD accepted or requested and RWD appraisal within the three contexts.

Context	RWD Accepted/Requested	RWD Appraisal
Initial reimbursement discussions (IRD)	**Summary of commonalities:** • All sources of data are welcomed in submissions. Implies RWD also welcome. • Treatment Effects: RWD can be used to inform on treatment effects when RCT evidence is absent on specific head-to-head comparisons. However, biases related to RWD must be explored and documented. • Other Domains: RWD can be used to provide evidence on epidemiological data, natural history of disease or resource use data. • Agencies do not specify which kind of RWD should be collected nor methods for collection. However, the choice of which RWD and collection methods should be justifiable given the scientific questions at hand. **Summary of differences:** • One agency recently published a comprehensive list of RWD used in technology appraisals, detailing that comparative individual patient-level data (IPD), non-comparative IPD and aggregated data have been used in decision-making. In addition, the document included detailed guidance on statistical methods for use of RWD in submissions. • No significant differences further.	**Summary of commonalities:** • All agencies adopt evidence hierarchies in accordance with evidence-based medicine. Hierarchies consistently rank RWD at a lower quality level than RCT data. • Impact of RWD on decision-making differs according to contextual factors: › Conclusions regarding causal effects that are based on RWD will be regarded as more circumspect. › RWD can be used to supplement/confirm RCT-based conclusions on treatment effects › For some agencies, impact of RWD may be higher in cases where RCT are difficult to conduct (e.g. rare diseases) • Lack of clarity on RWD impact in the case of conflicting evidence (c.f. RCT) **Summary of differences:** • 2 agencies explicitly recognise limitations in adhering to strict evidence hierarchies in guidelines by stating that such hierarchies should not preclude the exclusion of valuable non-RCT evidence from decision-making. One agency advises against deviating from evidence hierarchies when considering evidence inclusion for decision-making. • 2 agencies stipulate that in cases where RCT data is sparse (esp. orphan diseases), RWD may be the only source of data available thus could be used for decision-making. Contrastingly, 1 agency stipulates that the circumstance of small patient populations (e.g. orphan diseases) does not necessitate deviance from the principles of evidence hierarchies.

Table 6 - continued

Context	RWD Accepted/Requested	RWD Appraisal
Pharmacoeconomic analysis (PEA)	**Summary of commonalities:** • RWD is directly requested by HTA agencies for PEA. • Treatment Effects: RWD can be used to inform on treatment effects when RCT evidence is absent on specific head-to-head comparisons. However, biases related to RWD must be explored and documented. • Costs & Resource Use Data: national RWD is the preferred source for costs data (direct, and indirect) and resource use data. • Other Domains: RWD can be used to provide data on quality of life, adherence, epidemiological data, transition probabilities for models. **Summary of differences:** No significant differences.	**Summary of commonalities:** • RWD use to inform parameters other than treatment effects is largely accepted. • Hierarchies of evidence adopted by HTA agencies consistently rank RWD at a lower quality level than RCT data. • Impact of RWD on decision-making differs according to contextual factors: › Conclusions regarding causal effects that are based on RWD will be regarded as more circumspect. › RWD can be used to supplement/confirm RCT-based conclusions on treatment effects **Summary of differences:** No significant differences.

Table 6 - continued

Context	RWD Accepted/Requested	RWD Appraisal
Conditional reimbursement schemes (CRS)	**Summary of commonalities:** • RWD requested in any scheme is case-specific but follows similar processes for 2 agencies: 1- Identification of evidence gaps during initial reimbursement discussions 2- Assessment of study proposal to collect data for scientific quality, feasibility and relevance. 3- Agreement on study protocol and date of collected RWD delivery for re-assessment of relative effectiveness. • Agencies provide practical guidance for applicants on: › Which scientific questions different study designs can and cannot answer. › Existing national RWD sources and relevance for providing specific information. **Summary of differences:** • Only 3 of the 6 HTA agencies implement CRS. • Differing aims of CRS (effectiveness vs. cost-effectiveness vs. re-pricing discussions) influence the type of data requested within each scheme. • The degree of guidance available for applicants varies between the three agencies: 2 agencies have guidelines to this effect, yet 1 does not. • One agency often actively participates in, or initiates its own product or indication registries. The remaining 2 agencies lay burden of data collection on applicant.	**Summary of commonalities:** • The impact of RWD collected rests on the following conditions: › That applicants take practical guidance available into consideration when designing the study protocol; › That the research conducted delivers the answers to evidence gaps identified; › That the research conducted adheres to the protocol agreed upon by all parties. **Summary of differences:** No significant differences.

DISCUSSION

Policies for RWD accepted or requested and RWD appraisal for REA of drugs adopted by the 6 agencies differed between the 3 contexts analysed. For example, while RWD use for IRD was accepted but not explicitly recommended, its use was recommended by agencies for PEA and CRS. RWD may provide evidence on numerous parameters of REA: (relative) treatment effects, epidemiological data, resource use data and cost data.

Policies for RWD accepted or requested and RWD appraisal for REA of drugs differed between the 6 agencies within the same contexts. An important example relates to RWD use to provide data on treatment effects for IRD in situations where it may be difficult to conduct RCT's (e.g. orphan diseases). While some agencies deem this acceptable, others explicitly advise against it. Similarly, policies for CRS differed whereby the aims of the 3 agencies' schemes, procedures for conducting CRS, as well as agencies' involvement in RWD collection in CRS schemes varied.

Inter-context policy variations may be an issue if the effectiveness and pharmacoeconomic components of HTA dossiers submitted to an agency are examined by two different assessors who subsequently appraise the RWD differently. Another compounding factor presents itself in agencies that offer a possibility for CRS, since the manner with which these different assessors would be required to appraise RWD in the effectiveness and pharmacoeconomic components of a specific dossier will inevitably depend on whether the dossier is submitted as a standard dossier or as a candidate for CRS. Bearing these points in mind, one can argue that standardising the implementation of policies on RWD use for decision-making in practice may be difficult within any single HTA agency.

Meanwhile, variations between agencies' policies may present marketing authorization holders (MAH) with a multitude of challenging questions when developing strategies for evidence generation across the product lifecycle (8, 40, 41). For instance, in the context of CRS, MAH's would need to question whether their product qualifies as a candidate for CRS in the different countries; whether questions raised by the various agencies would overlap or differ; and consider if one study would suffice to collect the RWD needed for all 3 agencies.

Hierarchies of evidence adopted by HTA agencies prominently featured in documents and interviews transcripts assessed. Several agencies implement such hierarchies through tools for classification of evidence quality (e.g. GRADE) (42). Although evidence hierarchies have well-established roots in evidence-based medicine, it is debatable whether they are applicable to the concept of RWD use for HTA. Conventionally, hierarchies automatically downgrade all RWD without exploring the subtle differences between the advantages, disadvantages and relevance of different RWD sources (e.g. patient registries or claims databases). More importantly, such evidence hierarchies do not address the differences in the type of insights provided by RCT data (efficacy data with high internal validity) and different forms of RWD (long-term data on safety and effectiveness from registry data, resource use data from claims databases or patient-reported outcomes from pragmatic

clinical trials (PCT)). An increasing body of literature also refers to the relevance of using data from PCT's for more generalizable and translatable evidence on real-world outcomes (43-45) yet guidance on this topic was not always reflected in agency guidelines. This can result in excluding valuable evidence in decision-making. Furthermore, some agencies may abandon the rigid framework of evidence hierarchies due to pragmatic reasons (e.g. in situations where RCT's are difficult to conduct or for CRS), and others even provide methodological guidance on such aspects (27, 31, 32, 36). Therefore, it may be necessary for HTA agencies to consider how implementation of rigid evidence hierarchies could be adapted to enable effective use of RWD in decision-making processes.

The lack of harmonisation of policies for RWD use in REA of drugs may discourage MAH's from collecting or analysing RWD for HTA purposes (8, 40, 41). Therefore, it may be useful for HTA agencies within Europe to align policies on RWD and provide guidance on practical aspects of RWD collection and analysis. This is especially important in light of the increasing trend of new (oncology or orphan) drugs granted conditional marketing authorisation based on phase II data or surrogate outcomes, rather than phase III RCT data (46-48). A harmonised set of policies on RWD use for HTA would provide MAH's with the ability to plan alternative evidence generation pathways which rely less on RCT trials, and more on real-world studies; the latter theoretically yielding outcomes more relevant for HTA purposes (49-52). The European network for HTA (EUnetHTA) may provide a platform for discussions on aligning RWD policies. EUnetHTA has recently published position papers on additional (non-RCT) evidence generation for REA and is finalising proposals for pilot projects which will address some of the issues mentioned above (53-55).

In addition to studying differences in policies for RWD use in REA of drugs between different contexts and agencies, determining if differences extend to the implementation of these policies in practice is important. When asked if their agency accepts or requests RWD, one HTA representative stated: "Of course we accept those data. We are forced by law to accept those data but we don't have to conclude the benefit from such data." This implies that RWD has quite a low impact on decision-making within that agency, in contrast to others. When representatives from 2 of the 3 agencies implementing CRS were asked about the impact of RWD in decision-making at re-assessment, they displayed contradicting views. One stated: "You can't really rely on them. You can use the RWD as a confirmation of the expectation you have at initial discussions..", whereas the other stated: "For the re-evaluation of pricing and reimbursement of the product, that kind of data are robust enough." Therefore, the reality of how RWD is used in practice may differ from policies and should be the focus for future research.

Strengths

To ensure that all available information on RWD policies was gathered for all 6 HTA agencies, a mixed-methods approach was used that included a review of agency websites, academic literature and stakeholder interviews. This minimised the probability

of important information being excluded from analysis. Moreover, the selection of documents for analysis, data abstraction, and coding was conducted independently by two different authors.

Limitations

Although 6 European HTA agencies were included, this does not automatically mean that we provided a representative overview of all European policies on RWD use in REA of drugs. The agencies selected only represent those vested within the "Big Four" jurisdictions and 2 agencies with pioneering roles in cutting-edge European HTA initiatives. However, considering the novelty of the topic on RWD use in REA of drugs and the impact of the agencies and jurisdictions included, this sample was deemed as relevant for an initial policy analysis on RWD use in REA of drugs in Europe.

The information available for analysis varied between agencies. Language capabilities of the involved researchers meant that Swedish and Italian documents were excluded from the analysis. As a result, valuable information from documents written by TLV or AIFA may have been overlooked. Moreover, not all agencies published guidelines that specifically focus on the use of RWD in REA. On the other hand, gathering information from several sources through agency website searches, the PubMed search and stakeholder interviews ensured that the impact of excluded information was minimal. Furthermore, TLV published numerous English guidelines on REA (26, 33, 38) and AIFA affiliates published several English academic articles on RWD use in Italian practice (17-21).

It can be argued that data gathered during interviews may only reflect the interviewees' opinion, rather than represent the agencies' official position. We attempted to account for this through selective sampling of participants, providing all approached participants with the opportunity of inviting colleagues they deemed relevant to the interview and by interviewing more than one person per institute. Additionally, information provided during interviews was compared with that from policy documents and academic publications to ensure alignment between data sources.

CONCLUSION

Individual agencies' policies regarding RWD accepted or requested and appraisal of RWD for REA of drugs vary notably across the three contexts assessed: IRD, PEA and CRS. Additionally, differences are present between each agencies' policies on RWD use for IRD, PEA and CRS. For example, the manner by which RWD is appraised for decision-making varies within any given agency, being largely acceptable for numerous PEA parameters and CRS but not for informing treatment effects for IRD. Moreover, the existence of CRS schemes, as well as the manner for the implementation of RWD use in CRS schemes is different within the agencies examined.

The lack of harmonisation of policies on RWD use for REA of drugs may present MAH's with a multitude of challenging questions when they consider collecting and using RWD for

HTA purposes. As a result, MAH's may be discouraged to use RWD for HTA. Therefore, HTA agencies within Europe may collaborate to align policies on RWD and provide guidance on practical aspects of RWD collection and analysis. Recently published position papers and future project proposals by EUnetHTA may provide a starting point for discussions and a suitable platform for HTA agencies to achieve this.

REFERENCE LIST

1. European Commission, "Directive 2001/20/ec of the European Parliament and of the Council," 2015. Available from http://ec.europa.eu/health/files/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf.
2. Hans Georg Eichler et al., "Bridging the Efficacy-Effectiveness Gap: a Regulator's Perspective on Addressing Variability of Drug Response," Nature Reviews Drug Discovery 10, no. 7 (2011): 495-506.
3. H. G. Eichler et al., "Adaptive Licensing: Taking the Next Step in the Evolution of Drug Approval," Clin Pharmacol Ther 91, no. 3 (2012): 426-437.
4. Silvana Romio, Miriam Sturkenboom, and Giovanni Corrao, "Real-World Data From the Health Decision Maker Perspective. What Are We Talking About?," Epidemiology, Biostatistics and Public Health 10, no. 3 (2013).
5. Louis P. Garrison et al., "Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report," Value in Health 10, no. 5 (2007): 326-335.
6. Vicki Fung et al., "Using Medicare Data for Comparative Effectiveness Research Opportunities and Challenges," Am J Manag Care 17, no. 7 (2011): 488.
7. L. Annemans, M. Aristides, and M. Kubin, "Real-Life Data: A growing Need," 2015. Available from http://www.ispor.org/news/articles/oct07/rld.asp.
8. A. Makady, "Review of Policies and Perspectives on Real-World Data," 2015. Available from http://tinyurl.com/p2qq9xh.
9. Anna Vassall et al., "Incorporating Demand and Supply Constraints into Economic Evaluations in LowGÇÉIncome and MiddleGÇÉIncome Countries," Health Economics 25, no. S1 (2016): 95-115.
10. OECD, "OECD Composite Leading Indicators: Reference Turning Points and Component Series," 2016. Available from http://www.oecd.org/std/leading-indicators/oecdcomposite leading indicatorsreference turning points and component series.htm.
11. Peter Debaere, EU Coordination in International Institutions: Policy & Process in Gx Forums 2015 ed.: Palgrave Macmillan, 2015).
12. Knud Erik Jorgensen and Katie Verlin Laatikainen, Routledge handbook on the European Union and international institutions: performance, policy, power: Routledge, 2012).
13. EUnetHTA, "EUnetHTA JA3 (2016-2020) Lead Partners," 2016. Available from http://eunethta.eu/activities/joint-action-3/jointaction31/eunethta-joint-action-3-2016-2020#involved_organisations.
14. Allison Tong, Peter Sainsbury, and Jonathan Craig, "Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-Item Checklist for Interviews and Focus Groups," International Journal for Quality in Health Care 19, no. 6 (2007): 349-357.
15. Hsiu Fang Hsieh and Sarah E. Shannon, "Three Approaches to Qualitative Content Analysis," Qualitative Health Research 15, no. 9 (2005): 1277-1288.
16. A. Makady, "GetReal. Updated glossary of definitions of common terms. " 2015. Available from https://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary%20-%20Update.pdf.
17. Oyinlola Oyebode et al., "EVIDENCE REQUIREMENTS FOR REIMBURSEMENT OF PHARMACEUTICALS ACROSS EUROPE," International Journal of Technology Assessment in Health Care 31, no. 1-2 (2015): 59-67.
18. Carlo Tomino, "New Perspective and New Challenges in Clinical Trial Regulation in Italy," Annali dell'Istituto superiore di sanitá 47, no. 1 (2011): 19-21.
19. Entela Xoxi et al., "The Italian Post-Marketing Registries," Pharmaceutical Programming 5, no. 1/2 (2012): 57-60.
20. S. Bonini et al., "Feasibility and Challenges of Independent Research on Drugs: the Italian Medicines Agency (AIFA) Experience,"

European Journal of Clinical Investigation 40, no. 1 (2009): 69-86.

21. L. Degli Esposti et al., "The Nationwide Osmed Health-Db Database. A Tool To Support Healthcare Decision-Making And Real-World Evidence Generation," Value in Health 18, no. 7 (2015): A693.
22. Jayne Taylor et al., "METHODOLOGICAL CHALLENGES IN EVALUATING THE VALUE OF REGISTERS," International Journal of Technology Assessment in Health Care 30, no. 01 (2014): 28-33.
23. Elisabeth George, "How Real-World Data Compensate for Scarce Evidence in HTA," Zeitschrift für Evidenz, Fortbildung und Qualit+ñt im Gesundheitswesen (2016).
24. David Glynn et al., "How to Improve the Quality of Evidence When New Treatments Are Funded Conditional on Collecting Evidence of Effectiveness and Safety," Journal of health services research & policy 21, no. 2 (2016): 71-72.
25. Maryse Lapeyre-Mestre et al., "Études Pharmaco-Épidémiologiques: Quels Niveaux De Preuve Et Comment Les Atteindre?," Th+®rapie 68, no. 4 (2013): 241-246.
26. Tandvårds- och läkemedelsförmånsverket, "Guide for companies when applying for subsidies and pricing for pharmaceutical products.,".
27. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. National Institute for Health and Care Excellence. 3-7-2015.
28. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. 22-4-0015.
29. Haute Autorité de Santé, "General Method for Assessing Health Technologies," 2015. Available from http://www.has-sante.fr/portail/upload/docs/application/pdf/general_method_eval_techno.pdf.
30. Zorginstituut Nederland. Beoordeling stand van de wetenschap en praktijk. 2015. Zorginstituut Nederland.
31. National Institute for Health and Care Excellence. The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data. 2015. National Institute for Health and Care Excellence.
32. College voor Zorgverzekeringen, "Leideraad voor Uitkomstenonderzoek,".
33. Tandvårds- och läkemedelsförmånsverket. General guidelines for economic evaluations from the Pharmaceutical Benefits Board. 24-4-2003. Tandvårds- och Läkemedelsförmånsverket.
34. Haute Autorité de Santé. Choices in Methods for Economic Evaluation. 2012. Haute Autorité de Santé.
35. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 17-11-2015. Zorginstituut Nederland.
36. Haute Autorité de Santé, "Les études post-inscription sur les technologies de santé (médicaments, dispositifs, médicaux et actes)," 2015. Available from http://tinyurl.com/orw8cw2.
37. Zorginstituut Nederland. Procedure voorwaardelijke toelating geneeskundige zorg 2015. 19-6-2015. Zorginstituut Nederland.
38. Tandvårds- och läkemedelsförmånsverket. The Swedish Pharmaceutical Reimbursement System. 2007. Tandvårds- och läkemedelsförmånsverket.
39. NHS England, "The Cancer Drugs Fund: Transition to the new model," Available from https://www.england.nhs.uk/ourwork/cancer/cdf/#.
40. Jorge Mestre-Ferrandiz et al., "The Current Drug Development Paradigm: Responding to US and European Demands for Evidence of Comparative Effectiveness and Relative Effectiveness," Office of Health Economics (2014).
41. The Academy of Medical Sciences, "Real world evidence: Summary of a joint meeting held on 17 September 2015 by the Academy of Medical Sciences and the Association of the British Pharmaceutical Industry," 2016. Available from https://www.acmedsci.ac.uk/viewFile/56cab22108cf9.pdf.

42. Gordon H. Guyatt et al., "GRADE: an Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations," BMJ : British Medical Journal 336, no. 7650 (2008): 924-926.

43. Geonnotti K Wang W Peikes D, "Using Pragmatic Clinical Trials to Test the Effectiveness of Patient-Centered Medical Home Models in Real-World Settings," Agency for Healthcare Research and Quality, 2013).

44. Anthony Delitto, "Pragmatic Clinical Trials: Implementation Opportunity, or Just Another Fad?," Physical Therapy 96, no. 2 (2016): 137.

45. Kirsty Loudon et al., "The PRECIS-2 Tool: Designing Trials That Are Fit for Purpose," (2015).

46. I. Lipska et al., "Does Conditional Approval for New Oncology Drugs in Europe Lead to Differences in Health Technology Assessment Decisions?," Clinical Pharmacology & Therapeutics 98, no. 5 (2015): 489-491.

47. J. Hoekman et al., "Use of the Conditional Marketing Authorization Pathway for Oncology Medicines in Europe," Clinical Pharmacology & Therapeutics 98, no. 5 (2015): 534-541.

48. A. Kaaniche, A. Troubat, and A. Sherwood, "Does Conditional Marketing Authorisation Influence Market Access in France, England, and Germany?," Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research 18, no. 7 (2015): A559.

49. Nick Freemantle and Thomas Strack, "Real-World Effectiveness of New Medicines Should Be Evaluated by Appropriately Designed Clinical Trials," Journal of Clinical Epidemiology 63, no. 10 (2010): 1053-1058.

50. ABPI, "Demonstrating Value with Real World Data: A practical guide," 2015. Available from http://www.abpi.org.uk/our-work/library/guidelines/Pages/real-world-data.aspx.

51. D. Foltz et al., "Real-World Data Research: A case for action," 2015. Available from http://assets1.csc.com/health_services/downloads/CSC_Real_World_Data_Research.pdf.

52. Marc L. Berger et al., "Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines," Value in Health 18, no. 1 (2015): 127-130.

53. EUnetHTA, "Position paper on how to best formulate research recommendations for primary research arising from HTA reports," Available from http://eunethta.eu/sites/5026.fedimbo.belgium.be/files/EUnetHTA%20Position%20Paper%20on%20research%20recommendations.pdf.

54. EUnetHTA, "Position paper on how to decide on the appropriate study design for primary research arising from HTA reports," Available from http://eunethta.eu/sites/5026.fedimbo.belgium.be/files/EUnetHTA%20Position%20Paper%20on%20study%20design_0.pdf.

55. EUnetHTA, "Core Protocol Pilot for Additional Evidence Generation," Available from http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/news-attachments/eunethta_core_protocol_pilot_for_aeg.pdf.

SECTION III

Real-World Evidence Use in Practice

CHAPTER 4

Using Real-World Data in Health Technology Assessment (HTA) Practice: A comparative study of 5 HTA agencies

A. Makady[1,2], A. van Veelen[2], Jonsson P.[3], Moseley O.[4], A. D'Andon[5], A. de Boer[2], H. Hillege[6], O. Klungel[2], W. Goettsch[1,2].

[1.] The National Healthcare Institute (ZIN), Eekholt 4, 1112 XH, Diemen, the Netherlands.
[2.] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Universiteitsweg 99, 3584 CE, Utrecht, the Netherlands.
[3.] The National Institute for Health and Care Excellence (NICE), Level 1A, City Tower, Piccadilly Plaza, Manchester, M1 4BT, Manchester, the United Kingdom.
[4.] The Scottish Medicines Consortium (SMC), Healthcare Improvement Scotland (HIS), Delta House (8th floor), 50 West Nile Street, G1 2NP, Glasgow, Scotland.
[5.] La Haute Autorité de Santé (HAS), 5 avenue du Stade de France, 93218 Saint-Denis La Plaine Cedex, Paris, France.
[6.] Department of Epidemiology, University Medical Centre Groningen, Broerstraat 5 9712 CP, Groningen, the Netherlands.

Published in PharmacoEconomics. Citation is as follows: Makady A, van Veelen A, Jonsson P, Moseley O, D'Andon A, de Boer A, Hillege H, Klungel O, Goettsch W. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. PharmacoEconomics. 2017 Dec 6:1-0.

ABSTRACT

Background

Reimbursement decisions are conventionally based on evidence from randomised controlled trials (RCTs) which often have high internal validity but low external validity. Real-world data (RWD) may provide complimentary evidence for relative effectiveness assessments (REAs) and cost-effectiveness assessments (CEAs). This study examines whether RWD is incorporated in Health Technology Assessment (HTA) of melanoma drugs by European HTA agencies, differences in RWD use between agencies and across time.

Methods

HTA reports published between 01.01.2011 and 31.12.2016 were retrieved from websites of agencies representing 5 jurisdictions: England(NICE), Scotland (SMC), France (HAS), Germany (IQWiG) and the Netherlands (ZIN). A standardized data-extraction form was used to extract information on RWD inclusion for both REAs and CEAs.

Results

Fifty-two reports were retrieved. All 52 reports contained REAs; CEAs were present in 25. RWD was included in 28 of 52 REAs (54%); mainly to estimate melanoma prevalence. RWD was included in 22 of 25 (88%) of CEAs; mainly to extrapolate long-term effectiveness and/or identify drug-related costs drugs. Differences emerged between agencies regarding RWD use in REAs; ZIN and IQWiG cited RWD for evidence on prevalence whereas NICE, SMC and HAS additionally cited RWD use for drug effectiveness. No visible trend for RWD use in REAs and CEAs over time was observed.

Conclusion

In general, RWD inclusion was higher in CEAs than REAs. It was mostly used to estimate melanoma prevalence in REAs or to predict long-term effectiveness in CEAs. Differences emerged between agencies' use of RWD. However, no visible trends for RWD use over time were observed.

INTRODUCTION

Melanoma is the most serious and fatal form of skin cancer(1). Its incidence has been increasing, largely caused by increased exposure to ultra-violet radiation (1-3). Primary tumours are most often removed by surgical excision. However, after tumour metastasis, surgical excision is often no longer feasible, and pharmacotherapy becomes the remaining option (1, 4). According to literature, before 2011, dacarbazine was the standard chemotherapeutic of choice for the treatment of metastatic (or non-operable) melanoma (henceforth melanoma)(5, 6). Since 2011, multiple drugs for the treatment of melanoma have entered the market representing 4 novel mechanisms of action, thereby increasing treatment options substantially (1, 7).

Regulatory approval of new therapeutics in Europe is centralized, with decisions being issued by the European Commission(8). However, each European jurisdiction decides nationally on drug reimbursement and pricing, conventionally based on assessments and appraisals of available evidence conducted by national Health Technology Assessment (HTA) agencies. These involve relative effectiveness assessments (REAs), sometimes in combination with cost-effectiveness assessments (CEAs), based on evidence submitted by the marketing authorisation holders of drugs. For the purposes of this article, we define REAs as assessments that examine the extent to which an intervention does more good than harm, when compared to one or more alternative interventions for achieving the desired results and when provided under the routine setting of health care practice(9, 10). Meanwhile, CEAs examine the relationship between relative effects and the respective costs of implementing the intervention versus its comparators (11).

Evidence on drug effectiveness informing HTA submissions is conventionally derived from randomised clinical trials (RCTs) (12). Due to their design characteristics, RCTs have a high degree of internal validity, making them a good fit to demonstrate causality (13-15). However, due patient randomisation, inclusion and exclusion criteria and regulated follow-up protocols, the external validity of RCTs is relatively low (14-17). Consequently, extrapolation of drug efficacy to drug effectiveness in clinical practice is difficult. This discrepancy is frequently referred to as the efficacy–effectiveness gap (13). Therefore, despite recent advances in melanoma drugs and their potential additional benefit to patients, HTA agencies still face challenges in interpreting results of REAs and CEAs that rely on evidence from RCTs due to factors such as the large heterogeneity of patients in clinical practice compared to RCT populations and the lack of head-to-head comparisons in RCTs.

Real-world data (RWD), defined here as data collected outside the setting of RCTs (14, 15), could theoretically be used to inform effectiveness estimates of novel or existing drugs in clinical practice, thereby supporting RCT evidence. RWD can be derived from numerous sources, including disease registries, observational studies and electronic health records (14, 15). Due to specific characteristics of RWD (e.g. non-randomised treatment allocation, longer patient follow-up and broader patient populations), it may provide a more generalizable picture of treatment effects in clinical practice (18). In contrast, using RWD for

decision making presents new methodological and analytical challenges. For example, due to non-randomized treatment allocation, confounding in estimated treatment effects may occur due to an imbalance in the potential known and unknown confounders in the groups of patients being compared (18). Moreover, other practical aspects such as missing data in RWD sources and the lack of interoperability across RWD sources with different database infrastructures may affect the quality of data present or may complicate research across different datasets, respectively (18). Some statistical methods have been developed in an attempt to address a number of issues cited here such as propensity scoring techniques and instrumental variable techniques (to address confounding) or multiple imputation methods (to address missing data) (19-21).However, these techniques come with their own assumptions and limitations (19, 21). A subsequent question remains whether and how one should combine RWD with RCT data for REA and CEA for HTA purposes (22).In brief, although RWD may potentially supply much-needed insights on the effectiveness and cost-effectiveness of new drugs in practice, its incorporation into analyses and subsequent decision making for HTA is not clear-cut.

Currently, RWD is used in drug development to examine natural history of diseases, delineating clinical treatment pathways, determining costs and resource use associated with treatments and examining health outcomes associated with comparators (23).Previous research has demonstrated that policies on RWD assessment and appraisal in decision making vary between HTA agencies and depend on the context of use (i.e. whether for REAs or CEAs) (23). This study aims to examine the use of RWD in HTA practice. Specifically, it examines whether RWD is included in REAs and CEAs of melanoma drugs and the appraisal of RWD for its intended purposes by 5 HTA agencies in Europe.

METHODS

Methods used were comparable to those presented in the study by Kleijnen et al (8). A retrospective, comparative analysis of HTA reports (henceforth reports) on melanoma drugs was performed. Six HTA agencies representing 6 European jurisdictions were selected for inclusion, since they make full reports publicly available: NICE - England; SMC - Scotland; HAS - France; IQWiG – Germany; AOTMiT – Poland; and ZIN - the Netherlands. However, due to the authors' inability to read Polish reports, the study proceeded with 5 agencies.

HTA reports on 7 new melanoma drugs (ipilimumab, vemurafenib, dabrafenib, cobimetinib, trametinib, nivolumab and pembrolizumab) were retrieved from agency websites. Inclusion criteria were: a melanoma indication, publication dates between the 1st of January 2011 and the 31st of December 2016 and availability of at least 3 reports, published by 3 different agencies, per drug. The latter criterion ensured that the majority of included agencies had conducted assessments for each drug. Each resubmission or addendum was categorized as a new report.

Data extraction from compiled reports was performed independently by AM and AvV using a standardized data-extraction form containing open-ended and closed questions

(DEF; see Appendix 1). Inclusion of RWD in REAs and CEAs was examined separately. When RWD was included, 2 aspects were examined: the reason for inclusion (i.e. the parameter(s) it informed) and the source of RWD. Subsequently, agencies' appraisals of the validity of RWD use and the sources chosen for the intended parameter (henceforth RWD appraisal) was examined by identifying corresponding statements in reports and scoring them using the following algorithm:

- Positive: statement identifying a positive opinion on validity of RWD use and source.
- Negative: statement identifying a negative opinion on validity of RWD use and source.
- Neutral: statement identifying a neutral opinion on validity of RWD use and source.
- Unknown: statement that cannot clearly be identified as positive, negative or neutral.
- Not identified: no statement regarding appraisal despite RWD inclusion in the assessment.

To measure agreement within data extraction and scoring performed by AM & AvV, the inter-rater reliability (IRR) was calculated twice in 2 different rounds. In each round, authors independently extracted data from 4 randomly-selected reports (see Appendix 2 for reports per round). Authors' extraction for closed questions were compared using the Fleiss' kappa method, whereby a score of 0 indicates poor agreement and a score of 1 indicates perfect agreement (24). Authors' extraction for open-ended questions was compared by a third, independent researcher. Once IRR was established, the remaining reports were equally divided amongst both authors.

To verify whether data extracted from reports on RWD inclusion, RWD appraisal scoring and results of analyses accurately reflect practice in the agencies included, a panel of 5 senior assessors representing the 5 respective agencies was consulted (see Appendix 3 for panel members). The data extracted from reports of HTA agencies and results of the analyses mentioned below were mailed to the panel members. Panel members then indicated if, for example, reports were missing from the dataset, whether data for specific questions of the data extraction form was missing and where to find it in reports, as well as their feedback on the results of analyses. Panel members subsequently received a copy of the modified dataset and analyses results for a final check.

ANALYSIS

The frequency of RWD inclusion in REAs and CEAs was recorded separately. Subsequently, the parameter(s) for which RWD was used and the frequency thereof were recorded. Then, the source(s) of RWD used per parameter and the frequency thereof were recorded. It is important to note that the authors registered the nature of the source as cited in the reports e.g. "SEER registry data" was recorded as "registry", whereas "MELODY observational study" was recorded as "observational study". However, the authors are aware of overlap between the definitions of registries and observational studies (14).

In addition to the general analysis mentioned above, potential variation in RWD use amongst the 5 agencies was examined by comparing RWD inclusion in REAs and CEAs per agency.

Finally, an analysis of RWD inclusion in REAs and CEAs combined for all compiled reports per publication year was performed to examine potential changes in RWD inclusion over time.

RESULTS

65 reports were identified for the 7 drugs on agencies' websites. Of the 65 reports, 52 were indicated for melanoma. All 52 were published between the 1st of January 2011 and the 31st of December 2016. NICE, HAS, and IQWiG published at least one report for all 7 drugs, allowing for the inclusion of all 52 reports (see Appendix 4 for full list). The distribution of reports across the agencies was as follows: ZIN (n=2), HAS (n=8), NICE (n=10), SMC (n=13) and IQWiG (n=19). All reports included REAs. However, IQWiG and HAS reports did not include CEAs. In total, 25 CEAs were located in reports of NICE, SMC and ZIN. It is important to note that ZIN reports entailed initial assessments as part of conditional reimbursement schemes. As such, they included sections beyond REAs and CEAs, such as outcomes research proposals for prospective RWD collection. However, for this study only the REAs and CEAs were included.

The IRR was calculated twice and improved from 0.60 in the first round to 0.80 in the second round, corresponding to substantial agreement between AM & AvV (24).

RWD was included in 28/52 (54%) REAs and was mainly used to estimate melanoma prevalence and/or incidence (28/28 REAs). Additionally, RWD was used to estimate the effectiveness (7/28) and safety (6/28) of the new drug. The majority of the RWD included for estimation of melanoma prevalence/incidence originated from registries. Additionally, national statistics databases, data from observational studies and claims databases were used. RWD included for effectiveness or safety was mainly derived from observational studies and/or non-randomized phase I/II studies. For a detailed summary of the frequency of RWD use per parameter and RWD source, see Table 1. For a detailed summary of the studies used to provide RWD on effectiveness and safety, see Table S1 in Appendix 5.

RWD was included in 22/25 (88%) CEAs and was primarily used to extrapolate effectiveness of the new drug beyond RCT trial duration to estimate its long-term effectiveness (21/22 CEAs). Additionally, RWD was included to estimate costs associated with drugs (12/22), estimate resource use (8/22) and determine utilities using quality of life information (4/22). All CEAs that included RWD to estimate long-term effectiveness derived data from registries. In some reports, this was further supported by RWD from national statistics databases. In that case, registry data was used to extrapolate overall survival until a specific time-point beyond trial duration (e.g. 10 or 15 years), while national statistics data was used to extrapolate overall survival from that point forwards until the end of the model's time horizon. Costs were estimated using data from claims databases, observational studies

or cost-of-illness studies. Data sources used for resource use and quality-of-life parameters are presented in Table 1.

Figure 1 shows the outcome of RWD appraisal in REAs and CEAs. For 16 of 49 (33%) and 27 of 58 (32%) parameters for which RWD was used in REAs and CEAs respectively, no appraisal statements could be identified. Meanwhile, appraisal statements identified in REAs or CEAs indicated that appraisal outcome was mostly unknown (25/49 (51%) and 18/58 (31%) parameters, respectively) or negative (6/49 (12%) and 9/58 (16%) parameters, respectively). The negative appraisal of RWD in REAs was primarily caused by decision-

Table 1 - Parameters for which Real-World Data (RWD) is included and RWD sources used per parameter (including frequency).

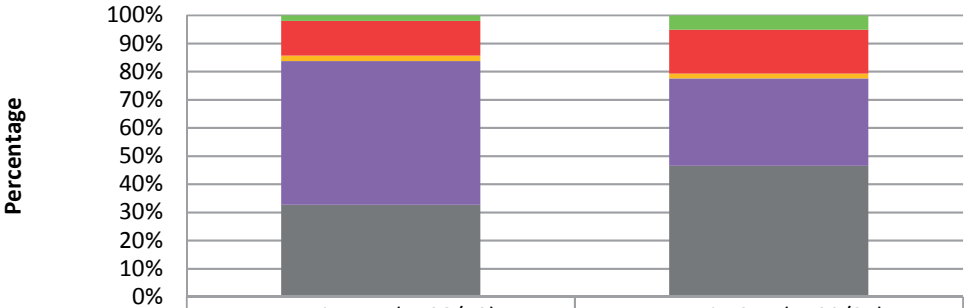
Relative effectiveness assessment			Cost-effectiveness assessment		
Reason for inclusion	**Frequency**	**Source**	**Reason for inclusion**	**Frequency**	**Source**
Prevalence/ incidence	29	Registry (n=22) National statistics database (n=9) Observational study (n=6) Claims database (n=2)	Long-term effectiveness	21	Registry (n=21) National statistics database (n=12)
			Costs	12	Claims database (n=10) Observational study (n=4) Cost-of-illness study (n=1)
Effectiveness	7	Observational study (n=6) Non-randomized Phase I / II trial (n=6) Registry (n=1)			
			Resource use	8	Observational study (n=7) Claims database (n=4) Registry (n=1)
Safety	6	Non-randomized Phase I / II trial (n=4) Observational study (n=3)			
			Quality-of-Life data	4	Quality-of-life study (n=3) Registry (n=1)

makers' perceptions of the low reliability of RWD use from observational studies to estimate clinical effectiveness due to biases associated with observational data. Similarly, the negative appraisal of RWD in CEAs was primarily due to decision-makers' uncertainties regarding extrapolations of long-term effectiveness. However, in some reports it was difficult to discern whether these uncertainties solely pertained to the nature of RWD and its associated biases or in combination with the statistical methods applied for extrapolation of long-term effects.

Inclusion of RWD in REAs differed between the 5 agencies. For example, NICE reports cited RWD in 10/10 (100%) REAs, while SMC reports cited RWD in 3/10 (33%) (Figure 2). ZIN and IQWiG mainly cited RWD for estimating melanoma prevalence, while NICE, SMC and HAS cited RWD use for the estimation of effectiveness and/or safety more frequently. In contrast, no notable differences were found in RWD inclusion in CEAs; inclusion was above 75% for all 3 agencies (Figure 3). However, RWD cited in ZIN CEAs mainly pertained to drug costs and quality-of-life data, whereas that in NICE and SMC reports mainly pertained to long-term effectiveness and resource use estimates.

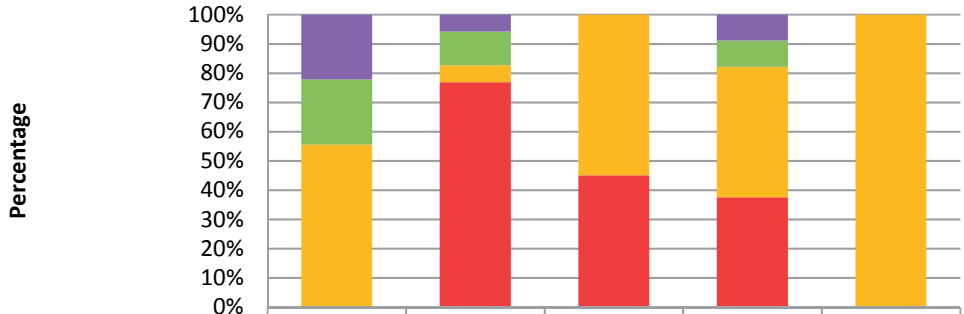
The inclusion of RWD over time in REAs and CEAs combined varied per year, ranging from 1/1 reports (100%) in 2011 to 17/28 reports (61%) in 2016 (Figure 4). The inclusion of RWD in REAs and CEAs over time is shown separately in Figures S1 and S2 in Appendix 5. No trend was visible for RWD inclusion in REAs. The inclusion of RWD in CEAs exceeded 75% in all years (2011-2016), displaying no visible variation in trend.

In the current study only 2 of the 52 reports were initial assessment reports within conditional reimbursement schemes (CRS), namely those published by ZIN. However,



	RWD in REA (n=28/52)	RWD in CEA (n=22/25)
Positive	2%	5%
Negative	12%	16%
Neutral	2%	2%
Unknown	51%	31%
Not identified	33%	47%

Figure 1 - Appraisal of the validity of RWD use and sources chosen when included in REAs and CEAs



	NICE (n=10)	SMC (n=13)	IQWiG (n=19)	HAS (n=8)	ZIN (n=2)
Safety	22%	6%	0%	9%	0%
Effectiveness	22%	12%	0%	9%	0%
Prevalence/incidence	56%	6%	58%	45%	100%
Not included	0%	77%	47%	38%	0%

Figure 2 - Inclusion of RWD in REAs and the reasons for inclusion per agency

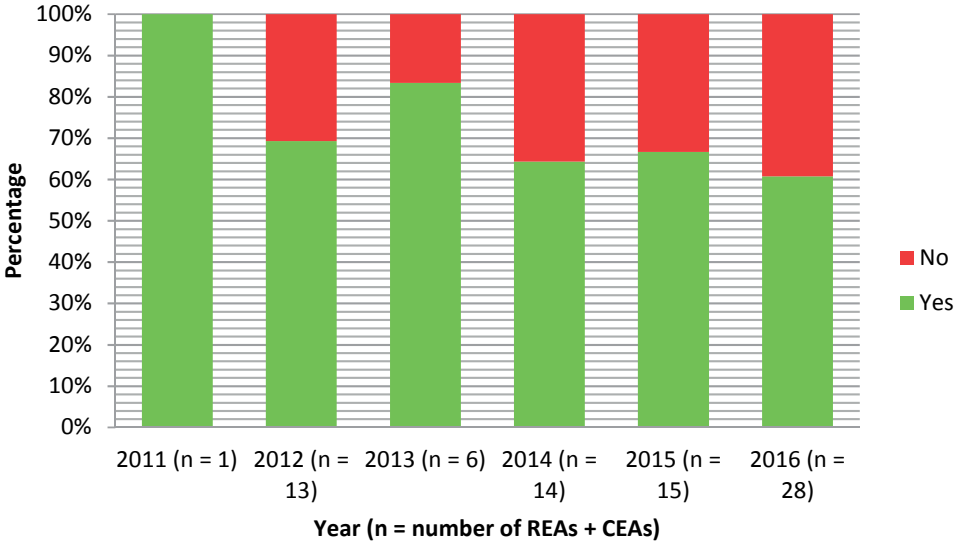
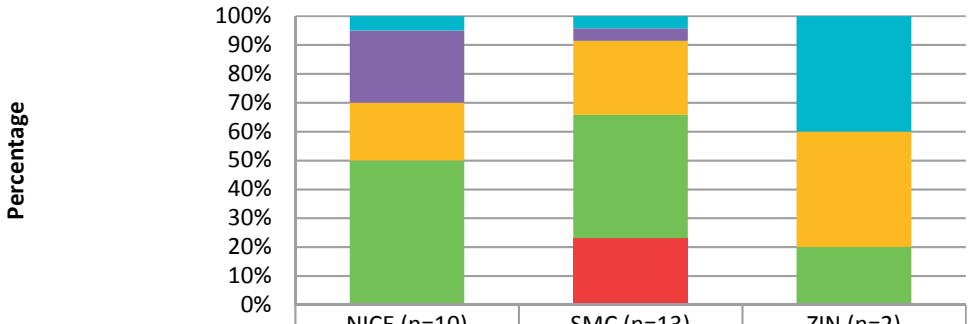


Figure 3 - Inclusion of RWD in CEAs across the 3 agencies and reasons for inclusion per agency

the respective re-assessment reports have not yet been published. We will return to the implications of this in the discussion section below.



	NICE (n=10)	SMC (n=13)	ZIN (n=2)
Quality-of-life data	5%	4%	40%
Resource Use	25%	4%	0%
Costs	20%	26%	40%
Long-term effectiveness	50%	43%	20%
Not included	0%	23%	0%

Figure 4 - Inclusion of RWD in REAs and CEAs (combined) over time

DISCUSSION

This study examined the extent with which RWD was included and its appraisal in HTA reports of 7 melanoma drugs from 5 different agencies. Results demonstrate an overall difference in RWD inclusion between REAs and CEAs, whereby inclusion is more common in CEAs (88%) than REAs (54%). RWD included mainly informed melanoma prevalence and/or incidence in REAs and long-term effectiveness and costs in CEAs. Sources of RWD used to inform those parameters varied and included: registries, observational studies, national statistics databases and claims databases. Statements on RWD appraisal were often not found in REAs and CEAs. When identified, the nature of appraisal statements was mostly unknown or negative. Reasons for negative appraisals were manifold, often relating to decision-makers' awareness of biases associated with RWD as well as the statistical approaches used to incorporate it in effectiveness estimates.

Inclusion of RWD in REAs varied somewhat between agencies. In contrast, little variation in RWD inclusion in CEAs was observed. Analysis of differences in RWD inclusion in both REAs and CEAs over time revealed no identifiable trends between 2011 and 2016. However, analyses between agencies and across time were complicated by the varying number of total reports per agency and per year, as well as the fact that not all agencies conducted CEAs. Therefore, interpretation of differences in RWD use between agencies and across time must be made with caution.

The findings summarised above coincide well with results from a previous review of policies on RWD use amongst 6 HTA agencies (4 of which were included in this study) thus indicating that current RWD use in practice is in line with policies (23). The review examined

policies on RWD use in REAs, CEAs and conditional reimbursement schemes (CRS), concluding that policies differed somewhat between the different agencies, and differed markedly depending on the context analysed. For example, agencies' policies iterate that RWD use is welcome in REAs to provide incidence or prevalence data but that RCT's remain the preferred source for data on effectiveness estimates of drugs. Consequently, RWD use for effectiveness is more likely to be negatively appraised in REAs. Meanwhile, policies iterate that RWD inclusion in CEAs is largely accepted, and even demanded for specific parameters (e.g. treatment costs and resource use). However, policies also iterate that RCT's remain the preferred source for relative effectiveness estimates in CEAs.

In the past ten years, RWD use in drug development and healthcare decision-making has gained increasing attention both in scientific literature and grey literature (25). Moreover, a multitude of initiatives have explored possibilities for incorporating RWD in decision-making. Examples include the International Society for Pharmacoeconomic and Outcomes Research (ISPOR) Task Force on RWD (15), the Patient-Centred Outcomes Research Institute (PCORI) and the Innovative Medicines Initiative GetReal Consortium (IMI-GetReal) (26). Based on findings from this study, it may be argued that despite increased attention, little has changed with regards to the role for RWD in HTA practice. For example, RWD inclusion in reports did not increase proportionally over time. In fact, the rate of RWD inclusion was lowest in 2016.

These results raise the question why RWD currently plays a relatively minor role in HTA, especially for parameters relating to drug effectiveness. A possible reason could be the lack of robust RWD available at the time of initial HTA assessments. Since these assessments take place soon after regulatory approval of a drug, there might be insufficient time for marketing authorisation holders to collect RWD through registries or observational studies. Another factor could be the absence of guidance on systematic approaches for the inclusion, analysis and interpretation of RWD for HTA purposes. Moreover, HTA agencies have only recently begun collaborating on strengthening understanding of appropriate study designs for generating RWD and developing further analytic methods for synthesis of RWD from different sources through initiatives such as IMI-GetReal and the European Network of HTA (EUnetHTA) (27). Further dialogue amongst HTA agencies is necessary to ensure that the product of these ongoing collaborations will be deemed useful by decision-makers.

One potential source of RWD not found in the results of this study are pragmatic clinical trials (PCTs). Several design elements of PCTs imply that they may represent the ideal balance between RCTs and RWD: they often include a broader patient population than RCTs, a broader set of outcome measures than RCTs, are embedded in the setting of routine clinical practice and may include initial randomization followed by cross-over between arms based on interim analyses (14, 28). The advantages of PCTs use in HTA decision-making may seem straightforward at first sight. However, the design of such trials is fraught with many strategic choices that may impact the generalizability of results for different settings such

as the selection of participating hospitals/ clinical centres and the choice of comparators and outcome measures (28). The implementation of PCTs in practice is also associated with numerous challenges such as operationalization of the intervention within routine clinical practice, data management across sites and monitoring across sites (28, 29). Moreover, not all stakeholders unanimously agree that PCTs qualify as RWD; previous research has shown that a considerable number of stakeholders define RWD strictly as data generated without any intervention by researchers on treatment assignment, inclusion/exclusion criteria and patient monitoring protocols (30). This is often not the case with PCTs whereby a pre-specified study protocol details such aspects of researcher intervention. The authors are aware that the balance between the internal and external generalizability of a study is difficult to achieve and that PCTs include a broad spectrum of design choices that make such studies more or less representative of RWD (28). On the other hand, the authors also believe that PCTs may offer a valuable source of RWD whose potential for decision making in HTA should be further explored.

With regards to pharmacoeconomic analysis for CEA, one could argue that quantitative methods for modelling and sensitivity analyses may address some of the issues associated with the efficacy-effectiveness gap, potentially supplanting the need for RWD. For example, techniques such as bootstrapping and probabilistic sensitivity analyses (PSA) may help shed light on the impact that different effectiveness estimates can have on the incremental cost-effectiveness ratio (ICER) (11, 31). On the other hand, a counter-argument is that the underlying distributions used to randomly sample effectiveness parameters in PSA are based on numerous assumptions and RCT data which may arguably also not be representative of drug effectiveness in the clinical population. Meanwhile, guidelines for health economic models increasingly require the use of a lifetime horizon in health economic analyses (31-33) and given the reality that it is neither ethical nor feasible to conduct long-term RCTs, one could argue that the need for RWD to provide data on (long-term) effectiveness in a heterogeneous clinical population remains crucial for HTA purposes. In order to provide a robust answer to the question whether current modelling methods and sensitivity analyses could supplant the need for RWD, quantitative research is required to bring to light the predictive validity of outputs from health economic models and sensitivity analyses (34). Although this is beyond the scope of this study, we recommend future pursuits on this topic.

Theoretically, CRS provide an ideal context for incorporating RWD in HTA. The value of RWD generated in CRSs would play a critical role in the re-assessment of drugs (e.g. to confirm previous efficacy estimates, cost-effectiveness ICER estimates or budget impact). According to previous research, policies for CRS implemented by 3 agencies indicated that RWD is largely accepted within this context, provided data collection and analysis abide by pre-defined conditions (23). In the current study only 2 of the 52 reports were initial assessment reports within CRS, namely those published by ZIN. However, the respective re-assessment reports have not yet been published. Moreover, HAS reports examined were not

part of CRS implemented in France. As such, the potential role of RWD in melanoma reports within CRS could not be assessed. To our knowledge, work is ongoing within ZIN and HAS to re-assess melanoma drugs using RWD. Therefore, provided no similar study on RWD inclusion and appraisal within CRS across HTA agencies has been performed, this should be the focus of future research once re-assessment reports are published.

Strengths

The study included all 52 reports from 5 HTA agencies' websites in the analyses, corresponding to the total number of reports published up until and including the 31st of December 2016. Inclusion of all reports for all 5 agencies minimised the chances of missing relevant information.

The IRR between the 2 authors responsible for data extraction and scoring was measured twice based on a randomly selected set of reports. In doing so, authors minimised the probability that results reached were a consequence of inter-author differences in extraction and scoring.

Findings generated by this study were presented to an HTA panel, consisting of 5 senior assessors representing all 5 agencies included, to verify whether the results accurately represent practice within their agency thus improving their plausibility.

Limitations

The inclusion of reports published by the Polish HTA agency (AOTMiT) could not be achieved due to the authors' inability to read Polish reports. Inclusion of AOTMiT's reports may have provided insights on RWD use by an HTA agency within Eastern Europe, thus arguably a more informative overview of RWD use in HTA practice across Europe. The authors identified a study by Wilk et al. on RWD use by AOTMiT (35), which reported increasing use in practice. However, since the study examined different disease areas and included reports within a different time period, its results are not easily comparable to those presented here. Moreover, the authors recognize that the issue of RWD use in HTA extends beyond HTA in Europe. Therefore, future research should aim to include HTA agencies from outside Europe (e.g. Canada (CADTH) and Australia (PBAC)).

The comparison of RWD inclusion and RWD appraisal between the 5 agencies and over time was complicated by the varying number of reports published per agency, per year, and procedural differences in practice between agencies. For example, almost ten times more reports were retrieved for IQWiG than for ZIN. Furthermore, not all agencies included automatically conduct CEAs as part of their HTA process; only NICE, SMC and ZIN included CEAs in their reports. Moreover, one panel member (PJ) indicated that some evidence (including RWD), assessed by NICE for REAs and CEAs is not explicitly mentioned in the final guidance document. However, it is provided in the more detailed evidence package that is considered by the decision makers. This may lead to a possible underestimation of the role of RWD in decision making. In an attempt to address these shortcomings, authors

respectively: included all melanoma reports published per agency, explicitly distinguished between REAs and CEAs in analyses, registered all cases where appraisal statements were not identified, and only considered published evidence for all agencies.

This study represents spin-off work from the IMI-GetReal case study on metastatic melanoma (4). Given the considerable number of new, yet expensive, drugs that have recently become available for the treatment of metastatic melanoma in the previous years based largely on (short-term) efficacy data, the case study team had hypothesized that the use of RWD to demonstrate the (long-term) value of drugs in clinical practice for HTA purposes in this indication would be pertinent. On the other hand, the focus on this disease area could arguably hinder generalizability of results to others whereby RWD use may also be relevant. Future research should therefore aim to investigate RWD inclusion and its appraisal in HTA reports in other disease areas or across multiple disease areas, thus increasing generalizability of results to broader HTA practice.

CONCLUSIONS

In general, RWD was more often included in CEAs than in REAs of HTA reports. The main reason for inclusion in REA was prevalence and/or incidence of melanoma and in CEA for extrapolating long-term effectiveness of new drugs. If RWD was included in reports, statements regarding its appraisal were often not identified. When identified, appraisal outcome was mostly unknown or negative. These results correspond with findings from a previously-performed policy review.

Inclusion of RWD in REAs differed between the 5 agencies; some citing RWD only for prevalence and/or incidence and others for drug effectiveness and safety. Meanwhile, no distinguishable trend in total RWD inclusion over time was found. However these results should be interpreted with caution, owing to differences in practices between agencies and varying numbers of reports published per year.

Future research should aim to explore RWD inclusion and appraisal within CRS implemented by different HTA agencies, which provide an ideal context for RWD use in HTA practice, and across multiple disease indications.

REFERENCE LIST

1. Liu Y, Sheikh MS. Melanoma: molecular pathogenesis and therapeutic management. Molecular and cellular pharmacology 2014; 6: 228.
2. Kumar P, Clark M. Clinical Medcine.2012.
3. Azoury C, Lange R. Epidemiology, risk factors, prevention, and early detection of melanoma. Surgical Clinics of North America 2014; 94: 945-962.
4. Makady A, Kalf R, W. G, Lees M. Deliverable D1.6 WP1 Case Study Review: Metastatic Melanoma. In. 2016.
5. Agarwala SS. Current systemic therapy for metastatic melanoma. Expert review of anticancer therapy 2009; 9: 587-595.
6. Julia F, Thomas L, Dumontet C, Dalle S. Targeted therapies in metastatic melanoma: toward a clinical breakthrough? Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents) 2010; 10: 661-665.
7. Rubio-Rodríguez D, Blanco SDD, Pérez M, Rubio-Terrés C. Cost-Effectiveness of Drug Treatments for Advanced Melanoma: A Systematic Literature Review. PharmacoEconomics 2017; 1-15.
8. Kleijnen S, Lipska I, Alves TL et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Annals of Oncology 2016; 27: 1768-1775.
9. HLPF. High Level Pharmaceutical Forum 2005-2008. Final Conclusions and Recommendations. In. 2008.
10. Kleijnen S, George E, Goulden S et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. Value in health 2012; 15: 954-960.
11. Drummond MF, Sculpher MJ, Claxton K et al. Methods for the economic evaluation of health care programmes. Oxford university press,2015.
12. Heintz E, Gerber-Grote A, Ghabri S et al. Is there a European view on health economic evaluations? Results from a synopsis of methodological guidelines used in the EUnetHTA partner countries. Pharmacoeconomics 2016; 34: 59-76.
13. Eichler HG, Abadie E, Breckenridge A et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nature Reviews Drug Discovery 2011; 10: 495-506.
14. Makady A, Goettsch W, Hummel N et al. GetReal. Updated glossary of definitions of common terms. In W. G, A. W, Jonsson P et al. (eds). IMI-GetReal 2016.
15. Garrison LP, Neumann PJ, Erickson P et al. Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report. Value in Health 2007; 10: 326-335.
16. Alemayehu D, Mardekian J. Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research. Journal of Managed Care Pharmacy 2011; 17: S16-S21.
17. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. Journal of Clinical Epidemiology 2010; 63: 1053-1058.
18. Alemayehu D, Riaz Ali MPP, Alvir JM et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011; 17: S3-S37.
19. Klungel OH, Martens EP, Psaty BM et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. Journal of clinical epidemiology 2004; 57: 1223-1231.
20. Schmidt AF, Klungel OH, Groenwold RH, Consortium G. Adjusting for confounding in early postlaunch settings: going beyond logistic regression models. Epidemiology 2016; 27: 133-142.
21. Myrtveit I, Stensrud E, Olsson UH. Analyzing data sets with missing data: An empirical evaluation of imputation methods and

likelihood-based methods. IEEE Transactions on Software Engineering 2001; 27: 999-1013.

22. Hummel N, Debray T, Didden E-M et al. Work Package 4 Methodological guidance, recommendations and illustrative case studies for (network) meta-analysis and modelling to predict real-world effectiveness using individual participant and/or aggregate data. 2016.
23. Makady A, ten Ham R, de Boer A et al. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. Value in Health 2017; 20: 520-532.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 159-174.
25. Makady A, W. G. Review of Policies and Perspectives on Real-World Data. In. IMI-GetReal 2015.
26. IMI-GetReal. Overall objectives. In. 2017.
27. EUnetHTA. Work Package 5 - Life cycle approach to improve Evidence Generation. In. 2017.
28. Zuidgeest M, Goetz I, Groenwold R et al. Pragmatic trials and real world evidence–an introduction to the series. Journal of Clinical Epidemiology 2017.
29. Heim N, van Stel HF, Ettema RG et al. HELP! Problems in executing a pragmatic, randomized, stepped wedge trial on the Hospital Elder Life Program to prevent delirium in older patients. Trials 2017; 18: 220.
30. Makady A, de Boer A, Hillege H et al. What Is Real-World Data (RWD)? A Review of Definitions Based on Literature and Stakeholder Interviews. Value in Health 2017.
31. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Handbooks in Health Economic E,2006.
32. Nederland Z. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. In. Zorginstituut Nederland 2015.
33. Excellence NIfHaC. Guide to the methods of technology appraisal. In. National Institute for Health and Care Excellence 2013.
34. Karnon J, Afzali HH. Predictive Validation and the Re-Analysis of Cost-Effectiveness: Do We Dare to Tread? PharmacoEconomics 2017; 35: 1111-1112.
35. Wilk N, Skrzekowska-Baran I, Wierzbicka N et al. Adoption of Real World Evidence in decision-making processes on public funding of drugs in Poland. Journal of Health Policy & Outcomes Research 2015; 2.

CHAPTER 5

Implementing Managed Entry Agreements in Practice: The Dutch reality check

A. Makady[1,2], A. van Veelen[2], A. de Boer[2], H. Hillege[3], O. Klungel[2], W. Goettsch[1,2]

[1] The National Healthcare Institute (ZIN), Diemen, the Netherlands
[2] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, the Netherlands
[3] Department of Epidemiology, University Medical Centre Groningen, Groningen, the Netherlands

Currently under review in Health Policy.

ABSTRACT

Background

Conditional financing (CF) of expensive hospital drugs was applied in the Netherlands between 2006 and 2012; a 4-year coverage with evidence development (CED) framework for expensive hospital drugs. This study aims to evaluate the CF framework, focusing on Health Technology Assessment (HTA) procedures.

Methods

Using a standardised data extraction form, researchers independently extracted information on procedural, methodological and decision-making aspects from HTA reports of drugs selected for CF.

Results

Forty-nine drugs were chosen for CF, of which 12 underwent the full procedure. The procedure extended beyond the envisioned 4 years period for 11/12 drugs. Outcomes research conducted provided insufficient scientific data for 5/12 drugs. After re-assessment, continuation of reimbursement was advised for 10/12 drugs, with 6 necessitating yet additional conditions for evidence generation. Notably, advice to discontinue reimbursement for 2/12 drugs has not yet been implemented in Dutch healthcare practice.

Conclusions

Theoretically, CF provided an option for quick but conditional access to drugs. However, numerous aspects related to the design and implementation of CF negatively affected its value in practice. Future CED schemes should aim to incorporate learnings from the CF example to increase their impact in healthcare practice.

INTRODUCTION

In an era of rising healthcare expenditures due to the advent of innovative, yet expensive drugs, it is suggested that managed entry agreements (MEAs) may provide healthcare payers and insurers with a flexible policy framework that incorporates both early access to drugs and additional evidence generation (1). The use of MEAs as policy tools to address this dilemma has increased globally (2). MEAs can be described as "*arrangements between drug manufacturers and payers or providers that -ensure access to coverage or reimbursement of a drug or medical technology under specified conditions"* (1).Three different categories of MEAs are defined based on issues they address: (i) managing budget impact, (ii) managing uncertainty relating to clinical and/or cost-effectiveness, and (iii) managing utilization to optimize performance (3).However, numerous challenges are associated with their design and implementation, leading to topics of ongoing debate (2;3).

In 2005, the Netherlands encountered the issue of unequal access to the then innovative, yet expensive trastuzumab as adjuvant therapy for the treatment of early breast cancer with HER2+ over-expression (4). Access varied significantly between hospitals in different provinces leading to the so-called "ZIP code healthcare" phenomenon and public outcry (4). To address this, the Dutch National Healthcare Authority (NZa) was asked by the Dutch Ministry of Health to implement two policy frameworks between 2006 and 2012 for the conditional financing (CF) of expensive drugs and orphan drugs administered within the hospital setting, respectively. These policy frameworks were linked to the development of a MEA, specifically a coverage with evidence development (CED) framework (1;5).

The National Healthcare Institute (ZIN; formerly known as the Healthcare Insurance Board (CVZ)), the national Health Technology Assessment (HTA) authority, was responsible for the implementation of CF and issuing eventual advice on reimbursement on behalf of the NZa. According to ZIN guidelines (6;7), drugs nominated for CF would be included in an initial assessment (T=0 years) comprising the following components: therapeutic value, cost-effectiveness, budget impact analysis, and assessment of the outcomes research proposal (preferably including a value of information analysis)(6). Inclusion of drugs in CF was only warranted if 3 criteria were met: a budget impact above €2.5 million/year, a proven additional therapeutic value in comparison to available comparator treatments, and a well-defined proposal for outcomes research to address uncertainties regarding appropriate use (AU) and cost-effectiveness (CE) in routine practice. Subsequently, marketing authorisation holders, in collaboration with hospitals, clinicians and clinician societies would implement the proposed outcomes research to collect real-world evidence (RWE) on AU and CE in routine practice throughout a period of 3 years, which was eventually extended to 4 years. Hospitals administering the selected expensive or orphan drugs were funded for 80% and 100% of their drug expenditures through the basic healthcare package, respectively.

After the 4-year period, ZIN would conduct a re-assessment (T=4) of drugs comprising the following elements: therapeutic value, appropriate use, cost-effectiveness and budget impact. Finally, an appraisal of all available evidence at T=4 would be performed to advise

on the reimbursement of drugs based on 4 criteria: necessity, clinical effectiveness, cost-effectiveness and implementability within the healthcare system (7). The Scientific Advisory Committee (WAR; hereafter Assessment Committee) was responsible for the assessment of evidence at T=0 and T=4. Meanwhile, the appraisal of evidence at T=4 based on the 4 criteria was conducted by the Insured Package Advisory Committee (ACP; hereafter Appraisal Committee). See Figure 1 for a process chart of the CF scheme.

To our knowledge, no systematic evaluation of CF in the Netherlands has been conducted since its inclusion stopped in 2012. Therefore, this article aims to evaluate experiences gained with the implementation of CF to date by reviewing HTA reports. In doing so, the authors endeavour to provide empirical insights to inform ongoing discussions on the implementation of MEAs in practice.

METHODS

To generate an overview of all drugs in the CF scheme, documents listing notifications of report assessments per year and announcements of assessment statuses were compiled from 2006 to 2017 from the ZIN website (www.zorginstituutnederland.nl). This period corresponds to the date of CF scheme implementation (01.01.2006) and the last available document (15.05.2017). For each notifications document, all assessments registered under CF were collected. For each drug the trade name, active ingredient, registered indication and status of the assessment were compiled. Duplicate entries for each drug were removed from the different documents based on a combination of the trade name, active ingredient, registered indication and status of assessment.

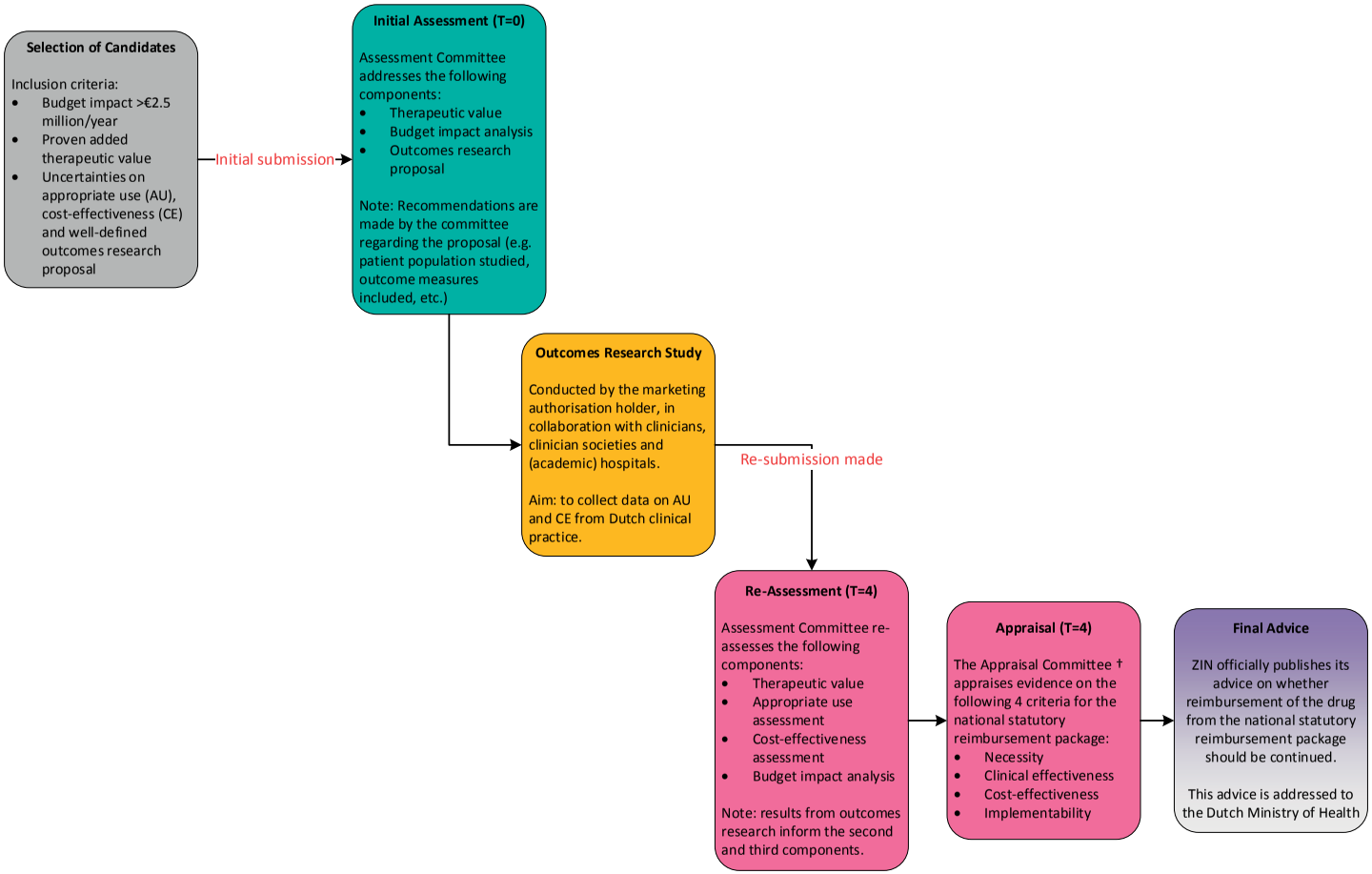
To subsequently evaluate the CF scheme, the authors used a three-pronged approach based on procedural, methodological and decision-making aspects outlined below. The authors are aware of other MEA analysis frameworks proposed in literature (1;5;8) but refer to the fact that such frameworks aim to classify the taxonomy of MEAs and recommend best practices for their design, rather than to retrospectively analyse their implementation thoroughly within a particular context. Therefore, in order to best address the research question at hand, the authors opted for the use of an alternative, tailored approach.

A. Procedural aspects

Procedural aspects related to whether due procedure had been followed in the implementation of CF as per ZIN guidelines, specifically:

- Whether T=0 and T=4 assessments were conducted for all CF drugs. If not, whether reasons for not conducting T=4 assessments were transparently communicated.
- Whether the time span between published T=0 and T=4 reports for drugs that underwent the full procedure (hereafter finalized drugs) equalled 4 years.
- Whether all components of the T=0 and T=4 reports for finalized drugs were present.
- Whether the relevant committees were consulted throughout the procedure.

Figure 1 – Process chart of the CF scheme.



† It is important to note that in some cases, the final appraisal of evidence in relation to the 4 package criteria was performed by the Assessment Committee, rather than the Appraisal Committee. This occurred for drugs whereby appraisal was relatively straight-forward (i.e. evidence at T=4 on all 4 criteria indicated a positive opinion on continued reimbursement). However, in cases where evidence may have led to a negative opinion on continued reimbursement, the Appraisal Committee was consulted.

To address the latter three points, T=0 and T=4 reports for finalized drugs were collected from the ZIN website. Subsequently, a standardised data extraction form was used to retrieve the following information from reports: date of publication of T=0 and T=4 reports, components of T=0 and T=4 reports present and the committees consulted.

B. Methodological aspects

Methodological aspects related to the assessment of evidence at T=0 and T=4, specifically:

Quantity of critical commentary provided by the Assessment Committee on the outcomes research proposals (T=0), appropriate use assessments (T=4) and cost-effectiveness assessments (T=4) of finalized drugs. For the purposes of this analysis, a critical comment was defined as a recorded instance in a report whereby the Assessment Committee provided an objective critique on a specific element of the evidence being assessed.

- Whether recommendations made by the Assessment Committee on the proposed outcomes research at T=0 were incorporated in the outcomes research implemented.
- Whether the Assessment Committee at T=4 deemed the evidence collected and its analysis to be of sufficient scientific quality to provide conclusions for the questions formulated at T=0.

Specific attention was given to these aspects because they represent the core aims of the CF scheme (i.e. to prospectively design studies to collect RWE on AU and CE) in comparison to conventional HTA performed by ZIN.

A standardised data extraction form was used to retrieve information from the respective T=0 and T=4 report components. The quantity of critical commentary was recorded per finalized drug and per section of the report components. For a list of the specific elements for which commentary was collected, see Tables A to C in the appendix. Analysis on implementation of recommendations made on outcomes research proposals was performed by comparing critical comments at T=0 to those at T=4. The committee's conclusions on the scientific quality of the evidence submitted at T=4 was analysed in respective statements in T=4 reports.

C. Decision-making aspects

Decision-making aspects related to the appraisal of evidence presented at T=4 and the final reimbursement advice, specifically:

- The nature of conclusions made by the Assessment Committee on AU and CE based on evidence submitted at T=4.
- The nature of the Appraisal Committee's advice on reimbursement at T=4 in relation to the 4 package criteria.
- The final advice published by ZIN on reimbursement of finalized drugs.

Specific attention was provided to decision-making on AU, CE & reimbursement advice because they represent the core aims of the CF scheme in comparison to conventional HTA performed by ZIN.

A standardised data extraction form was used to retrieve information on the following aspects from T=4 report components: the Assessment Committee's conclusions on AU & CE, the Appraisal Committee's advice on reimbursement and official ZIN advice on reimbursement.

All data extraction steps and analyses mentioned above for procedural, methodological and decision-making aspects were conducted independently by 2 authors (AM and AvV). Any discrepancies in data extracted and analyses were resolved by consensus amongst the 2 authors.

RESULTS

A. Procedural aspects

Forty-nine drugs were nominated for CF, of which 24 were excluded after T=0 assessments. The drugs were excluded because the expected budget impact at T=4 was below €2.5 million/year or the expected added therapeutic value at T=4 was diminished (e.g. due to emergence of equally effective comparator products; 22/24), or the drugs were transferred to an alternative national scheme on orphan drugs ("monitoring of orphan disease products"(9)) (2/24). Twenty-five drugs remained in the CF scheme. Information could only be retrieved in the public domain for 12/25 drugs, which have been finalized with subsequent publication of official ZIN advice. For information on the status of the remaining 13/25 drugs, authors were obliged to retrieve information from assessors within ZIN. For these 13 drugs, re-assessments are ongoing (5/13) or pending (e.g. due to extended deadlines allowing for extra data collection to supplement inadequate datasets; 8/13). See Figure A in the Appendix for a flowchart of drugs in CF and Table 1 for a list of finalized drugs.

For 11/12 finalized drugs, the elapsed time period between publication of the T=0 and T=4 reports extended beyond 4 years; ranging from 3.99 years (trastuzumab) to 7.58 years (natalizumab), with an average of 5.93 years per drug (Table 2).

The availability of report components for T=0 and T=4 reports of finalized drugs varied (Table D- Appendix). For all 12 T=0 reports, therapeutic value assessments and cost-effectiveness assessments were present. Contrastingly, none of the outcome research proposals contained value of information (VoI) analyses at T=0 despite guideline recommendations. However, it was mentioned (internal communication, WG) that for one drug VoI analysis was included in the submission file but later deemed unusable due to an incorrect choice of comparator treatment. For T=4 reports, therapeutic value assessments, cost-effectiveness assessments and budget impact analyses were present in all 12 reports and appropriate use assessments were present in 11/12 reports.

Table 1 – List of finalized drugs' trade names, active ingredients and indications.

Trade Name	Active Ingredient	Indication
Myozyme®	alglucosidase alpha	Pompe disease (glycogen storage disease type II).
Replagal®	agalsidase alpha	Fabry´s disease (alpha-galactosidase A deficiency).
Fabrazyme®	agalsidase beta	Fabry´s disease (alpha-galactosidase A deficiency).
Soliris®	eculizumab	Paroxysmal nocturnal hemoglobinuria (PNH).
MabThera	rituximab	Severe, active rheumatoid arthritis after failure to respond to at least 1 TNF/alpha blocker.
Tysabri®	natalizumab	Highly active relapsing remitting multiple sclerosis (RRMS).
Herceptin®	trastuzumab	Adjuvant therapy for the treatment of early breast cancer with increased HER2+ expression.
Xolair®	omalizumab	Add-on therapy for severe, persistent allergic asthma.
Vfend®	voriconazol	Serious, invasive aspergillosis.
Lucentis®	ranibizumab	Wet, age-related macular degeneration.
Metvix®	methyl aminolevulinate	Actinic keratosis.
Alimta®	pemetrexed	Metastatic non-small cell lung cancer (NSCLC).

Table 2 - T=0 and T=4 report publication dates and total elapsed time per finalized drug.

Finalized drug	Date of completion & publication of T=0 assessment	Date of completion and publication of T=4 re-assessment	Duration of procedure for conditional financing
alglucosidase alpha	24-07-2006	23-01-2012	5,50
agalsidase alpha	21-05-2007	27-02-2012	4,77
agalsidase beta	21-05-2007	27-02-2012	4,77
eculizumab	25-02-2008	18-03-2013	5,06
rituximab	25-09-2006	30-06-2014	7,51
natalizumab	18-12-2006	14-07-2014	7,58
trastuzumab	03-07-2010	30-06-2014	3,99
omalizumab	23-05-2006	02-07-2012	6,11
voriconazol	17-12-2007	30-06-2014	6,54
ranibizumab	23-04-2007	13-08-2012	5,31
methyl aminolevulinate	28-04-2008	23-03-2015	6,98
pemetrexed	22-06-2009	18-07-2016	7,07

As per ZIN guidelines, the Assessment Committee was consulted for all T=0 and T=4 assessments and for conclusions on AU and CE at T=4. However, contrary to guidelines, the Assessment Committee also performed the appraisal of evidence at T=4 in relation to the 4 package criteria for 5/12 drugs (Table E – Appendix). This occurred for drugs whereby appraisal was relatively straight-forward (i.e. evidence at T=4 on all 4 package criteria indicated a positive opinion on continued reimbursement). However, for the remaining

7/12 drugs where evidence may have led to a negative advice, the Appraisal Committee was consulted.

B. Methodological aspects

The total number of critical comments made by the Assessment Committee addressing the components outcomes research proposals (T=0), appropriate use assessments (T=4) and cost-effectiveness assessments (T=4) for all finalized drugs varied. In total, 68/249 (27%) comments related to outcomes research proposals at T=0 and were mostly directed at the proposed cost-effectiveness model (14/68; 21%) and the selected outcome measures for clinical effect (13/68; 19%).

The majority of critical comments were posed at T=4, of which 58/249 (23%) related to appropriate use assessments and 123/249 (49%) related to cost-effectiveness assessments. Commentary provided at T=4 on appropriate use assessment was mostly directed at quality of life information collected (12/58; 21%), clinical effectiveness outcome measures included (10/58; 17%) and the studied patient population (9/58; 16%). Finally, critical commentary provided at T=4 on cost-effectiveness assessment was mostly directed at costs outcomes for which information was gathered (20/123; 16%), the presented model structure (15/123; 12%) and clinical effectiveness outcomes measured (14/123; 11%).

The total number of critical comments for T=0 and T=4 combined varied considerably between finalized drugs (Figure B - Appendix). For example, pemetrexed incurred the least number of comments at T=0 and T=4 combined (2/249; 0.01%), whereas rituximab incurred the most (54/249;19%).

Recommendations made at T=0 on the outcomes research proposal were fully implemented in studies conducted for 5/12 finalized drugs. For 6/12 finalized drugs, recommendations were only partially implemented. Moreover, the number of recommendations that were not incorporated varied. For example, for rituximab 6/7 (86%) of recommendations made were not incorporated, compared to only 2/11 (18%) for methylaminolevulinate (Figure C – Appendix). Due to the absence of an outcomes research proposal for trastuzumab, this analysis could not be conducted for this drug.

The Assessment Committee concluded that evidence submitted at re-assessment (T=4) and its analysis was of sufficient scientific quality to assess AU in Dutch clinical practice for 9/12 (75%) of finalized drugs and inadequate for 3/12 (25%) of drugs. Meanwhile, the committee concluded that evidence submitted at re-assessment and its analysis was of sufficient scientific quality to assess CE in Dutch clinical practice for 7/12 (58%) of finalized drugs and inadequate for 5/12 (42%) of drugs.

C. Decision-making aspects

For 8/9 drugs with sufficient evidence on AU, the Assessment Committee concluded that they were used appropriately in clinical practice; for the last drug (eculizumab), the committee concluded that the drug was administered to a broader patient population

than intended. Meanwhile, the committee stated that it could not reach conclusions on AU for drugs for which the submitted evidence was insufficient.

Four of the 7 drugs with sufficient evidence on CE were indicated for orphan diseases, whereby high incremental cost-effectiveness ratios (ICERs) led to the Assessment Committee concluding that the ICERS were above the threshold value of €80,000/QALY and delegating further discussions in relation to other societal considerations to the Appraisal Committee. For 2/7 drugs, the committee concluded that the ICERs presented were below the threshold and substantiated by the evidence submitted. For the last drug (pemetrexed), the committee concluded that despite the low probability (10-40%) of the drug being cost-effective at the threshold, impending expiry of its patent and emergence of generic products would improve its cost-effectiveness in the near future.

On the other hand, for 4/5 drugs with inadequate evidence on CE, the Assessment Committee concluded that the ICERs presented were not substantiated by the evidence thus no conclusions could be reached on their CE in practice. For the final drug (rituximab), the committee concluded that additional data collection was unnecessary due to diminished added therapeutic value and costs which are comparable to a novel comparator treatment, both factors thereby minimising the risk for incurring high ICER's.

The Assessment Committee went on to appraise all evidence at T=4 in relation to the 4 package criteria (necessity, clinical effectiveness, cost-effectiveness and implementability in the healthcare system) for 5/12 drugs; for 4/5 drugs, continued reimbursement from the basic healthcare package was advised. For the final drug (natalizumab), the committee advised to postpone the decision on discontinuation of reimbursement until further evidence becomes available from a separate initiative (Round Table on Multiple Sclerosis) (10). Meanwhile, the Appraisal Committee appraised evidence at T=4 for 7/12 drugs. For 5/7 drugs, continued reimbursement was advised based on additional conditions. Such conditions varied based on which NZa framework the drug belonged to and on a case-by-case basis.

For 3 of the 4 orphan drugs (alglucosidase alpha, agalsidase alpha and agalsidase beta), conditions included the need for exceptional financing of orphan drugs outside the basic healthcare package, tailored policies on development and pricing of orphan drugs, the establishment of necessary patient registries to monitor real-world outcomes and bundling of clinical expertise to ensure AU. Conditions for expensive drugs varied per case. For omalizumab, the committee argued for a pragmatic solution in the form of a Pay-for-Performance scheme to avoid its exclusion from the basic healthcare package. Meanwhile, the committee advised clinician societies to update clinical guidelines to clearly specify criteria for patients who qualify for treatment with methylaminolevulinate, thereby avoiding over-prescription (e.g. due to low compliance amongst patients using comparator treatments leading to apparent non-response). For 2/7 drugs (eculizumab and ranibizumab), the Appraisal Committee advised to discontinue reimbursement.

Based upon the assessment and appraisal of evidence at re-assessment (T=4) by the respective committees, ZIN issued their final advice to continue reimbursement for

4/12 (33%) finalized drugs, continue reimbursement based on additional conditions for 6/12 (50%) finalized drugs, and discontinue reimbursement for 2/12 (17%) drugs (Table 3). Additional conditions for the reimbursement of 6/12 drugs were similar to, albeit more extensive, than those cited by the committees.

For a detailed summary of all decision-making aspects described above per drug, see Tables F and G in the Appendix.

DISCUSSION

Of the 49 drugs nominated for CF, 25 remained in the scheme, of which 12 underwent the full procedure. Only 1 drug was completed within the envisioned 4-year period. Published T=0 and T=4 reports did not consistently include all necessary components. Contrary to procedures outlined in guidelines, appraisal of evidence at T=4 was conducted by the Assessment Committee for almost half of the drugs. Critical commentary provided by the Assessment Committee on the outcomes research proposal (T=0), appropriate use assessment (T=4) and cost-effectiveness assessment (T=4) varied considerably per finalized drug. Recommendations provided on the outcomes research proposal were fully implemented for less than half of finalized drugs, with a varying percentage of unaddressed recommendations for the remaining drugs. At T=4, the Assessment Committee concluded that evidence generated through outcomes research was of insufficient quality to answer a third of research questions defined at T=0. Eventually, based on advice of its committees, ZIN advised to continue reimbursement for 10 drugs, of which 6 with additional conditions, and to discontinue reimbursement for 2.

In light of results summarised above, one may question whether some design aspects of CF, an example of a CED framework, were fit for its envisioned purpose. For example, only 1 drug had been processed within the envisioned 4-year time window. Although reasons for failure to timely processing of the remaining drugs are not directly apparent from the extracted data for this study, they may relate to a myriad of factors, including the time needed to set up registries required for data collection, to compile and evaluate data generated from outcome studies, and subsequently to assess and appraise the evidence generated (11;12). In Italy for instance, extensive resources were invested in setting up necessary infrastructures to collect fit-for-purpose data over many years (13). Moreover, one may wonder whether a 4-year period is applicable to all indications for which the finalized drugs were approved; the assessment of mortality outcomes with the use of voriconazol for serious, invasive aspergillosis (an acute, life-threatening condition) requires shorter follow-up than for pemetrexed for non-small cell lung cancer. The use of tailored approaches for determining required time-frames to answer the questions raised at T=0, rather than a fixed 4-year window, may provide a more intuitive design.

Importantly, for a third of research questions defined at T=0, insufficient evidence was generated through the implemented outcome research studies to reach grounded conclusions at T=4. Moreover, for half of the finalized drugs, reimbursement was continued

Table 3 - Summary of ZIN advice on reimbursement for all finalized drugs.

Finalized drug	ZIN advice on reimbursement from the basic healthcare package	Extra conditions specified
alglucosidase alpha	Keep drug in basic healthcare package based on certain conditions.	• Temporarily continue reimbursement of the drug from the basic healthcare package. • Develop a separate financial framework for drugs for orphan diseases. • Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases. • Negotiate price negotiations with the marketing authorisation holder (MAH). • Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification). • Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria. • Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.
agalsidase alpha	Keep drug in basic healthcare package based on certain conditions.	• Temporarily continue reimbursement of the drug from the basic healthcare package. • Develop a separate financial framework for drugs for orphan diseases. • Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases. • Negotiate price negotiations with the marketing authorisation holder (MAH). • Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification). • Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria. • Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.

Table 3 - continued

Finalized drug	ZIN advice on reimbursement from the basic healthcare package	Extra conditions specified
agalsidase beta	Keep drug in basic healthcare package based on certain conditions.	• Temporarily continue reimbursement of the drug from the basic healthcare package. • Develop a separate financial framework for drugs for orphan diseases. • Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases. • Negotiate price negotiations with the marketing authorisation holder (MAH). • Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification). • Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria. • Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.
eculizumab	Remove drug from basic healthcare package.	N/A
rituximab	Keep drug in basic healthcare package	N/A
natalizumab	Keep drug in basic healthcare package based on certain conditions.	ZIN postpones its final decision for removal of this drug from the basic healthcare package until results from the [separate] Round Table on Multiple Sclerosis are presented.
trastuzumab	Keep drug in basic healthcare package	N/A
omalizumab	Keep drug in basic healthcare package based on certain conditions.	To guarantee continued reimbursement, the marketing authorisation holder (MAH) should sign Pay-for-Performance (PfPO) agreements with all hospitals whereby the drug will be prescribed. In the case of defaults on PfPO agreements (e.g. due to lack of cooperation from individual hospitals or no refunds to hospitals based on outcomes), ZIN will advise for the removal of this drug from the basic healthcare package.

Table 3 - continued

Finalized drug	ZIN advice on reimbursement from the basic healthcare package	Extra conditions specified
voriconazol	Keep drug in basic healthcare package	N/A
ranibizumab	Remove drug from basic healthcare package.	N/A
methyl aminolevulinate	Keep drug in basic healthcare package based on certain conditions.	ZIN requests the clinicians' societies to update the clinical guideline, in order to clarify and specify the criteria for treatment with methylaminolevulinate thus ensuring that the implementation of such criteria becomes feasible in practice.
pemetrexed	Keep drug in basic healthcare package	N/A

based on yet further evidence generation to address remaining uncertainties. Once again, although the potential reasons behind such a finding are not directly apparent from the data extracted, literature alludes to numerous reasons such as challenges with analysing and interpreting RWE generated (14;15). It may also be that the lack of full incorporation of recommendations on the proposed outcomes research contributed to this. Two safeguards proposed in ZIN guidelines may have prevented such shortcomings in hindsight. Firstly, the conduct of VoI analyses at T=0 to highlight the feasibility and intrinsic value of data collection for specific parameters within the timelines projected. Secondly the mid-term reporting of outcomes research progress and interim results between T=0 and T=4 (specifically at T=1 & T=3) may have led to more timely decisions regarding continuation, adjustment or termination of the CF procedure for drugs, thereby avoiding waste of valuable time and money for all stakeholders involved. Unfortunately, both recommendations (VoI and interim reporting) were published in December 2008, more than 2 years after the start date of the CF scheme (6). By then, T=0 assessments for the majority of finalized drugs had already been completed. Nevertheless, both design aspects may be essential for future design of MEAs (particularly CEDs), as has been iterated in previous literature (16).

Another shortcoming is the absence of an *a priori* strategy for the implementation of CF outputs in the actual healthcare setting. To the authors' knowledge, it was not specified in guidelines beforehand how advice officially issued by ZIN on reimbursement of CF drugs from the basic healthcare package would or should be implemented by the responsible external stakeholders in the Dutch healthcare setting for their respective tasks. For example, it is known that ranibizumab has not been removed from the basic healthcare package by the Ministry of Health to date, and it remains unknown if the appropriate use of voriconazol has been improved through the modification of clinical guidelines as per ZIN advice. Previous experiences in Germany allude to difficulties associated with removing medicines from national reimbursement packages or limiting physicians´ choice in treatment prescription (17). Contrastingly, one successful story is that of omalizumab, whereby a Pay-for-Performance scheme was initiated jointly by ZIN, the Ministry of Health, the marketing authorisation holder, patient organisations and participating hospitals as per the advice of ZIN's Appraisal Committee. However, it would be burdensome and discouraging to all parties to first implement a CED scheme, only to follow up with a Pay-for-Performance scheme for each drug (2;18). Moreover, implementing pay-for-performance schemes incurs other practical considerations relating to retrieving costs from responsible parties, as experienced in Italy (13). Therefore, provided the diversity of stakeholders active within the Dutch healthcare setting, the complexity of interactions between their mandates and stakeholders' differing interests, more attention should have been paid to establishing *a priori* strategies on how CF outputs would and should be implemented in practice by different stakeholders.

The emergence of innovative, yet expensive medications is occurring rapidly. Moreover, a notable trend amongst novel oncology treatments relates to conditional marketing based on less conclusive evidence on safety or efficacy (e.g. phase I/II studies) within the context

of accelerated/conditional approval pathways (19). Consequently, HTA agencies and payers increasingly encounter submissions with more uncertainties on aspects such as long-term health outcomes and effectiveness in clinical practice. Meanwhile, an increasing global interest in medicines adaptive pathways to patients (MAPP's), whereby an iterative approach to evidence generation is adopted for products throughout their lifetime, reasserts the increasing dependence on MEAs for both HTA and regulatory decision-making (20). However, the design and implementation of MEAs, particularly CEDs, remains complicated (2;13;17;18;20). One may argue that without systematic evaluations of established MEAs, novel schemes are likely to suffer similar caveats as previous ones. To counter this potential risk, knowledge regarding the successes, failures, strengths and weaknesses of established MEAs should be the focus of future research, in order to avoid repeating historical mistakes when setting up new schemes within the Netherlands and elsewhere.

Limitations

The evaluation scheme developed and implemented by the authors for this study is a novel one. The authors are aware of other MEA analysis frameworks proposed in literature (1;5;8) but refer to the fact that such frameworks aim to classify the taxonomy of MEAs and recommend best practices for their design, rather than to retrospectively analyse their implementation within a particular context. Therefore, in order to best address the research question at hand, the authors opted for the use of an alternative, tailored approach.

In the assessment of methodological aspects, the authors examined the quantity of critical commentary provided by the Assessment Committee on appropriate use assessments and cost-effectiveness assessments. Although this provided insights as to which elements may have been most controversial during the re-assessment of submitted evidence, the qualitative nature of comments provided have not been separately addressed to determine their impact on evidence appraisal. For example, in appropriate use assessments of the finalized drugs, we noted that 9 critical comments were provided on patient populations examined in outcomes research studies. Bearing in mind that research questions on AU hinge on the generalizability of the examined study population to the Dutch clinical population, such comments may have had a more prominent role in the final appraisal of evidence compared to other comments. In an attempt to address this limitation, the authors examined both the Assessment Committee's conclusions on the scientific quality of the evidence submitted for AU and CE, as well as its final conclusions on AU and CE. In doing so, the authors were able to discern which aspects influenced the Committee's conclusions on AU and CE the most.

Finally, this study presents an analysis of reports as a means to determine experiences gained in implementing CF. However, this research question additionally warrants alternative methods (e.g. stakeholder interviews) to gather information on the experiences gained by the wide array of stakeholders involved in implementing CF. In doing so, numerous findings could be brought to light which may not be part of HTA reports analysed. This is currently the topic of ongoing research by the authors.

CONCLUSION

In principle, CF may provide a valuable MEA framework, guaranteeing patient access to innovative agreements while simultaneously obliging responsible parties to collect RWE on appropriate use and cost-effectiveness to address uncertainties, thereby informing decision-making at re-assessment. However, a variety of shortcomings related to procedural, methodological and decision-making aspects may have affected its value in practice. Such shortcomings have been echoed in available literature on MEAs implemented in other jurisdictions.

This study illustrates an attempt to systematically evaluate CF in order to inform ongoing international discussions on the design and implementation of future MEA schemes. However, provided the continuing onslaught of innovative, yet expensive drugs and HTA agencies´ and payers' increasing reliance on MEAs, further research on experiences gained with other MEAs is critical to inform the design of better schemes in the future.

REFERENCE LIST

1. Carlson JJ, Sullivan SD, Garrison LP, et al. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. Health Policy 2010;96:179-90.
2. Carlson JJ, Gries KS, Yeung K, et al. Current status and trends in performance-based risk-sharing arrangements between healthcare payers and medical product manufacturers. Applied health economics and health policy 2014;12:231-8.
3. Klemp M, Frønsdal KB, Facey K. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care 2011;27:77-83.
4. Klappe-Sabadi G, Jansman FGA, Honkoop HA, et al. Regionale verschillen in voorschrijfgedrag trastuzumab opnieuw getoetst. PW Wetenschappelijk Platform 2[7]. 2008. 2008.
5. Garrison LP, Towse A, Briggs A, et al. Performance-based risk-sharing arrangements - good practices for design, implementation, and evaluation: report of the ISPOR good practices task force. Value Health 2013;16:703-19.
6. College voor Zorgverzekeringen. Leidraad voor Uitkomstenonderzoek 'ten behoeve van de beoordeling doelmatigheid intramurale geneesmiddelen'. 2008. College voor Zorgverzekeringen. 22-10-2015.
7. Zorginstituut Nederland, mw.dr.G.Ligtenberg. Voorwaardelijke toelating/financiering van zorg. 6-4-2012.
8. McCabe CJ, Stafinski T, Edlin R, et al. Access with evidence development schemes. Pharmacoecon 2010;28:143-52.
9. Link AJ, Pasman P. Pakketbeheer Weesgeneesmiddelen. 26-10-2015. Zorginstituut Nederland.
10. Zorginstituut Nederland. Fingolimod (Gilenya): Summary of Recommendations. 8-7-2016.
11. Alemayehu D, Riaz Ali MPP, Alvir JM, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17:S3-S37.
12. Alemayehu D, Mardekian J. Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research. J Manag Care Pharm 2011;17:S16-S21.
13. Navarria A, Drago V, Gozzo L, et al. Do the current performance-based schemes in Italy really work?GÇ£Success FeeGÇØ: A novel measure for cost-containment of drug expenditure. Value Health 2015;18:131-6.
14. Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: lessons learned from the PHAROS registry. Pharmacoecon 2015;33:551-60.
15. Mohseninejad L, van Gils C, Uyl-de Groot CA, et al. Evaluation of patient registries supporting reimbursement decisions: the case of oxaliplatin for treatment of stage III colon cancer. Value Health 2015;18:84-90.
16. Walker S, Sculpher M, Claxton K, et al. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. Value Health 2012;15:570-9.
17. Drug Commission of the German Medical Association. Opinion on the "cost-sharing initiatives" and "risk-sharing agreements" between pharmaceutical manufacturers and Health and Hospital. 2008.[cited 2017 April 14]; Available from: http://www.akdae.de/Stellungnahmen/Weitere/20080508.pdf
18. Garrison Jr LP, Carlson JJ, Bajaj PS, et al. Private sector risk-sharing agreements in the United States: trends, barriers, and prospects. The American journal of managed care 2015;21:632-40.
19. Lipska I, Hoekman J, McAuslane N, et al. Does conditional approval for new oncology drugs in Europe lead to differences in health technology assessment decisions? Clin Pharmacol Ther 2015;98:489-91.

20. Faulkner SD, Lee M, Qin D, et al. Pricing and reimbursement experiences and insights in the European Union and the United States: Lessons learned to approach adaptive payer pathways. Clin Pharmacol Ther 2016;100:730-42.

CHAPTER 6

Conditional Financing of Drugs in the Netherlands: Past, Present and Future. Results from Stakeholder Interviews

A. Makady[1,2], S. van Acker[3], H. Nijmeijer[4], A. de Boer[2], H. Hillege[5], O. Klungel[2], W. Goettsch[1,2]

[1] The National Healthcare Institute (ZIN), Diemen, the Netherlands
[2] Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, the Netherlands
[3] Vrije Universiteit, Amsterdam, the Netherlands
[4] Radboud University Medical Centre, Nijmegen, the Netherlands
[5] Department of Epidemiology, University Medical Centre Groningen, Groningen, the Netherlands

Article prepared for submission.

ABSTRACT

Background

Conditional financing (CF) of hospital drugs was implemented in the Netherlands as a form of Managed Entry Agreements (MEAs) between 2006 and 2012. CF was a 4-year process comprising three stages: initial health technology assessment of the drug (HTA; T=0), conduct of outcomes research studies and re-assessment of the drug (T=4). This article aims to analyze stakeholder experiences in implementing CF in practice.

Methods

Public and private stakeholders were approached for participation in stakeholder interviews using standardized e-mail invitations. An interview guide was developed to guide discussions which covered the following topics: perceived aim of CF, functioning of CF, impact of CF and conclusions on CF and future perspectives. Extensive summaries were generated for each interview and subsequently used for directed content analysis.

Results & Conclusion

Thirty stakeholders were interviewed. Differences emerged amongst the stakeholders on the perceived aim of CF. Conversely, there was some agreement amongst stakeholders on the shortcomings in the functioning of CF, the positive impact of CF on the Dutch healthcare setting and improvement points for CF. Despite stakeholders' belief that CF only partially met its aims, if not at all, there was agreement on the need for new policy to address the same aims of CF in the future. However, stakeholders diverged on whether CF should be improved based on learnings identified and re-introduced into practice or replaced with new policy schemes.

INTRODUCTION

Provided that healthcare budgets are finite, decision makers consistently face difficult questions regarding the allocation of resources within the healthcare system. According to the Organization for Economic Co-operation and Development (OECD), pharmaceutical drug expenditure in accounts for an average of 16.9% of total healthcare expenditures across 31 OECD member countries; in some countries exceeding 50% of expenditures (1). Literature also alludes to increased drug expenditure in the future, partly due to an increased trend in the emergence of innovative, yet expensive drugs (2). Consequently, policy makers have been attempting to control drug expenditure through a wide array of policy instruments (e.g. preference systems for generic drugs or co-payment mechanisms) (2;3).

One policy instrument comprises managed entry agreements (MEAs). Briefly defined, MEAs are "arrangements between drug manufacturers and payers or providers that ensure access to coverage or reimbursement of a drug or medical technology under specified conditions" (4). Several forms of MEAs exist, each addressing different policy questions. One form, coverage with evidence development (CED) schemes, includes mechanisms to address uncertainties in clinical and/or cost-effectiveness of drugs through the generation of (real-world) evidence (4). A notable advantage of CED schemes seemed to be their capability to resolve the dilemma between quick patient access to drugs and the collection of additional data to resolve potential uncertainties in the evidence base. However, despite these perceived advantages, it remains questionable whether they can deliver on their promises in practice (5;6).

In 2005, public outcry in the Netherlands ensued due to unequal access to the then innovative, yet expensive drug, trastuzumab (7). Inequality in access led to so-called "ZIP code healthcare", whereby patient access to trastuzumab varied from 25% in some provinces to 75% in others (7). Between 2006 and 2012, the Netherlands Healthcare Authority (Nederlandse Zorgautoriteit; NZa) devised two policy frameworks to facilitate conditional financing (CF) of expensive and orphan drugs in hospitals, respectively, from the national healthcare insurance package (henceforth reimbursement package) (8). The implementation of these frameworks in the form of a CED scheme was subsequently delegated to the National Healthcare Institute (ZIN; formerly CVZ), the national health technology assessment (HTA) agency. Drugs qualifying for CF had to meet three criteria: have a budget impact higher than €2.5 million per year, have a proven added therapeutic value and there needed to be uncertainties regarding the appropriate use and/or cost-effectiveness of the drugs in Dutch clinical practice (9).

The CF process comprised three main stages: initial HTA (T=0), conduct of outcomes research and re-assessment (T=4)(see Figure 1 on page 62). Various stakeholders were involved at each phase of the process. For example, ZIN was responsible for the assessment of evidence submitted for HTA at T=0 and T=4 and for providing feedback on outcomes research proposals at T=0. Meanwhile, the marketing authorization holder (MAH; i.e.

pharmaceutical industry), was responsible for preparing submissions for both T=0 and T=4 and submitting an outcomes research study proposal to address uncertainties identified at T=0. Other stakeholders involved in CF included: public policy bodies (e.g. NZa), healthcare insurers, medical specialists societies, academic and/or private hospitals and patient organizations. Please see Table 1 in the Appendix for full details of the roles of different stakeholders throughout the CF process.

Despite being one of the first MEAs implemented in Europe, no policy evaluation of CF has been conducted since the inclusion of the last drugs in 2012. HTA dossiers produced at T=0 and T=4 for all CF drugs were recently analyzed to assess procedural, methodological and decision-making aspects of the scheme (10). The current study aims to evaluate stakeholders' experiences in implementing CF in practice.

METHODS

Data Collection

Data collection was conducted in 2 phases (see further details below). In the first phase, data was collected from public organizations involved in designing and/or implementing policy and conducting research. These stakeholders were: the NZa, the Ministry of Health (VWS), the Netherlands Organization for Health Research and Development (ZonMW), members of the scientific assessment committee of ZIN ("Wetenschappelijk Adviesraad"; henceforth Assessment Committee), members of the Insurance Package Advisory Committee of ZIN ("Adviescommissie Pakket"; henceforth Appraisal Committee), senior advisors at ZIN (e.g. the secretariat of drug assessors), pharmacotherapeutic assessors and pharmacoeconomic assessors at ZIN. In the second phase, data was collected from the remaining stakeholders involved in CF, namely: pharmaceutical industry, healthcare insurers, medical specialists societies, academic and/or private hospitals and patient organizations.

The authors used purposeful- and snowballing sampling to select stakeholders to approach for participation (11). Firstly, the authors selected stakeholders that were directly involved in implementing CF in practice. Secondly, those who agreed to participate were provided the opportunity to recommend other stakeholders during the interviews. The specific stakeholder representatives approached were sampled based on seniority and function, with a preference for senior representatives with a history of direct involvement in CF. All stakeholder representatives were approached using a standardized e-mail invitation (see Figure 1 in the Appendix for the invitation e-mail). Data saturation was discussed amongst authors and provided grounds for determining the final number of interviews conducted.

An interview guide was developed for stakeholder interviews. The guide covered the following topics:

- Perceived aims of CF (i.e. which purpose it served)
- Perceived functioning of CF (i.e. in relation to procedural, methodological and decision-making aspects; definitions for these aspects correspond to those implemented in the first study on HTA dossiers (10))

- Impact of the CF scheme (i.e. its positive and negative effects on the Dutch healthcare setting)
- Conclusions and future perspectives (i.e. if CF achieved its aims, improvement points for CF and if CF-like schemes should be stopped, re-introduced or replaced)

The interview guide included both open- and closed questions. An initial version of the guide was piloted in the first 3 interviews. Based on feedback from interviewees, minor adjustments were made to the guides. See Figure 2 in the Appendix for the final interview guide used.

A preference was made for face-to-face interviews. If these were infeasible, telephone interviews were held. Stakeholders were asked if interviews could be audio-recorded. Field notes were also taken during the interviews. AM and HN conducted interviews in phase 1 between 04.07.2016 and 06.11.2016. Meanwhile, AM and SA conducted interviews in phase 2 between 24.03.2017 and 10.05.2017.

Based on audio recordings and/or field notes, extensive summaries (3-4 pages) were made. The summaries were sent to interviewees for a member check. The summaries were subsequently edited based on the feedback received and sent to the interviewees for final approval.

Data Analysis

Directed content analysis was conducted on the extensive summaries generated using MaxQDA software version 11.0 (VERBI Software GmbH, Location: Bismarckstraße 10-12 10625 Berlin Germany) (12). The empty coding tree was structured to reflect the topics of the interview guide mentioned above. In November 2016, AM and HN conducted the content analysis for interviews from phase 1. In May 2017, AM and SA conducted the content analysis for interviews from phase 2. Each author coded half of the interview summaries themselves and reviewed the coding performed by the other author for the remaining interviews. Any discrepancies in codes generated between the authors were resolved by consensus. Finally, the separate coding trees generated by the analysis of interviews from phases 1 and 2 were combined in August 2017 by AM and SA.

Due to the large number of codes generated for open-ended questions for three of the topics in the interview guide (perceived functioning of CF, impact of CF and conclusions and future perspectives), the authors selected the codes mentioned by at least a quarter of stakeholders (≥25%) for further descriptive analyses. Illustrative quotes were cited to clarify the meaning of the themes included in the analyses.

Additionally, a comparative sub-analysis was conducted for answers to both open- and closed questions provided in phase 1 and phase 2 for four topics: perceived functioning of CF, impact of CF, conclusions and future perspectives. This enabled a qualitative comparison of themes that the two stakeholder groups deemed relevant for the different topics.

The methods used for this study were compared to the consolidated criteria for reporting qualitative research (COREQ) 32-item checklist (see Table 4 in the Appendix) (13).

RESULTS

Study sample

Stakeholders approached in phase 1 comprised representatives from non-ZIN public bodies (n=3), the Assessment Committee (n=2), the Appraisal Committee (n=3), senior advisers at ZIN (n=4), pharmacotherapeutic assessors (n=4) and pharmacoeconomic assessors (n=2). Stakeholders approached in phase 2 comprised representatives from pharmaceutical companies (n=5), healthcare insurers (n=3), medical specialists societies (n=3), academic/private hospitals (n=3) and patient organizations (n=3). All representatives approached agreed to participate in the interviews (response rate 100%). Eventually, 35 representatives spanning 30 stakeholders were included.

Thirty interviews were conducted between 04.07.2016 and 10.05.2017; 14 for phase 1 and 16 for phase 2. Three interviews included 2 or more interviewees. Twenty-five interviews were held face-to-face and 5 over the telephone. Each interview lasted between 60 and 90 minutes. Audio recordings were made for 29 interviews; one stakeholder refused to have the interview recorded.

For a summary of the study sample, see Table 1.

For the full coding tree developed, see Figures 3a-3d in the Appendix.

Perceived aims of CF

The majority of stakeholders (17/30; 55%) indicated that the aim of CF was to strike a balance between quick patient access to drugs and the promise for additional evidence generation. Meanwhile, 4/30 (13%) stakeholders indicated that CF only aimed to promote early access to drugs and 4/30 (13%) argued that it was merely a mechanism to control healthcare expenditure. Finally, 6/30 (19%) believed that CF had other aims. For example, one stakeholder indicated that CF aimed to provide a controlled environment whereby new drugs could be experimented with in clinical practice, in order to determine their additional value, based on clear agreements on treatment criteria[HO1].

See Figure 2 for an overview of the perceived aims.

Perceived functioning of CF

Procedural aspects

With regards to procedural aspects of CF, 27/30 (90%) of the stakeholders indicated to have doubts towards the envisioned 4-year timeframe. For example, stakeholders indicated that for some indications (e.g. acute diseases), 4 years may be sufficient to collect meaningful data whereas for other indications (e.g. chronic diseases or orphan diseases), much longer follow-up would be needed[MS3,PI3]. Moreover, stakeholders emphasized the extensive time needed to practically set up registries for data collection; a process lasting at least a year for specific drugs[PI3] thus leaving less time for actual data collection.

Another main theme mentioned by 15/30 (50%) of the stakeholders pertained to the ambiguity regarding the roles of different stakeholders throughout the CF process.

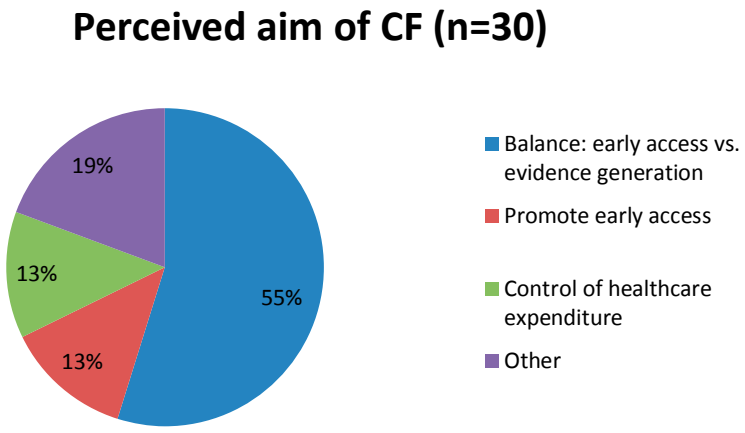


Figure 2 – Stakeholder views on perceived aim of conditional financing (CF).

For example, it was not obvious which stakeholder was responsible for communicating the results of evidence generated through outcomes research to ZIN at T=4 (i.e. pharmaceutical industry as dossier holders or medical specialists who collected the data)[ZA1]. Another stakeholder was of the opinion that while procedures at T=0 were clear, much was unknown about the process and roles of different stakeholders at T=4[MS2].

Thirteen stakeholders (13/30; 43%) referred to a design flaw in the CF procedure, namely the disregard of the relationship between the division of roles and conflicting interests of stakeholders. Pharmaceutical industry and medical specialists, tasked with financing and implementing outcomes research after T=0, respectively, may not have been intrinsically inclined to collect robust evidence that could indicate that the drugs are not cost-effective[ZW2]. If the data would lead to that conclusion, there would have been reason to remove the drug from the reimbursement package. For industry, this would result in a loss of revenue, whilst for the medical specialist, it would mean that patients would likely stop receiving their treatment[ZW2].Therefore, stakeholders mentioned that once reimbursement was granted from the reimbursement package at T=0, the incentives structure to generate evidence drastically shifted amongst stakeholders[PE2].

Another less frequently mentioned theme was the (lack of) mechanisms embedded in CF for the monitoring of progress throughout the procedure (8/30 stakeholders; 27%). For example, in none of the guidelines was an interim time point scheduled for a mid-term review of the progress in the outcomes research study (e.g. at T=1 year or T=3 years); ZIN was also not provided the authority to enforce such mid-term reviews[ZS4]. Stakeholders also iterated that the lack of monitoring meant that errors encountered at T=4 (e.g. regarding data collection or analysis) could no longer be retrospectively corrected[HI2]. Additionally, some stakeholders (8/30 27%) referred to the mechanisms for financing outcomes research studies, which was mainly paid for by the pharmaceutical industry. Stakeholders argued that this negatively impacted the independence of research conducted[PE3,ZA3,FG1]. Other

Table 1 – Summary of the stakeholders interviewed, number of interviewees per interview, interview code, date of interview, manner of interview (face-to-face vs. telephone) and whether the interview was recorded.

Stakeholder Group	Stakeholder	Number of Interviewees	Interview Code	Date of Interview	Manner of Interview	Interview recorded (yes/no)
External Public Policy/ Research Bodies	Nederlandze Zorgautoriteit (NZa)	1	PE1	18.07.2016	Face-to-face	Yes
	Ministerie voor Volksgezondheid, Welzijn en Sport (VWS)	1	PE2	04.07.2016	Face-to-face	Yes
	De Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie (ZonMW)	1	PE3	20.07.0216	Face-to-face	Yes
ZIN - Assessment Committee	Committee member	1	ZW1	20.07.2016	Telephone	Yes
	Committee member	1	ZW2	11.07.2016	Face-to-face	Yes
ZIN – Appraisal Committee	Committee member	1	ZA1	07.09.2016	Face-to-face	Yes
	Committee member	1	ZA2	12.07.2016	Face-to-face	Yes
	Committee member	1	ZA3	14.07.2016	Face-to-face	Yes
ZIN – Senior Advisers	Senior adviser	1	ZS1	06.09.2016	Face-to-face	Yes
	Senior adviser	1	ZS2	01.09.2016	Face-to-face	Yes
	Senior adviser	1	ZS3	31.08.2016	Face-to-face	Yes
	Senior adviser	1	ZS4	30.08.2016	Face-to-face	Yes
ZIN – Drug Assessors	Pharmacotherapeutic assessors	4	FG1	27.10.2016	Face-to-face	Yes
	Pharmacoeconomic assessors	2	FG2	16.11.2016	Face-to-face	Yes
Pharmaceutical Industry	Janssen Pharmaceuticals B.V.	1	PI1	13.04.2017	Face-to-face	Yes
	Novartis Pharma B.V.	2	PI2	19.04.2017	Telephone	No
	Bristol-Myers Squibb Pharmaceuticals B.V.	1	PI3	19.04.2017	Face-to-face	Yes
	Vereniging voor Innovatieve Geneesmiddelen (VIG)	1	PI4	20.04.2017	Face-to-face	Yes

Table 1 - continued

Stakeholder Group	Stakeholder	Number of Interviewees	Interview Code	Date of Interview	Manner of Interview	Interview recorded (yes/no)
Healthcare Insurers	Zorgverzekeringen VGZ	1	HI1	01.05.2017	Telephone	Yes
	Menzis	1	HI2	04.04.2017	Telephone	Yes
	Zorgverzekeraars Nederland (ZN)	1	HI3	24.04.2017	Face-to-face	Yes
Medical Specialists Societies	Stichting Werkgroep Antibioticabeleid (SWAB)	1	MS1	24.03.2017	Face-to-face	Yes
	Stichting Oncologische Samenwerking (SONCOS)	1	MS2	09.05.2017	Face-to-face	Yes
	Integraal Kankercentrum Nederland (iKNL)	1	MS3	10.05.2017	Face-to-face	Yes
Private/ Academic Hospitals	Nederlandse Federatie van Universitair Medische Centra (NFU)	1	HO1	21.04.2017	Face-to-face	Yes
	Het Academisch Medisch Centrum (AMC)	1	HO2	03.05.2017	Face-to-face	Yes
	Nederlandse Vereniging van Ziekenhuizen (NVZ)	1	HO3	25.04.2017	Telephone	Yes
Patient Organizations	Vereniging Volwassen, Kinderen en Stofwisselingsziekten (VKS)	1	PO1	04.04.2017	Face-to-face	Yes
	Nederlandse Federatie voor Kankerpatientorganisaties (NFK)	1	PO2	08.05.2017	Face-to-face	Yes
	Longfonds	1	PO3	03.04.2017	Face-to-face	Yes

Abbreviations: ZIN: Zorginstituut Nederland

stakeholders indicated that financing structures varied per registry, as was the case with governance structures for registries[PI3].

Methodological aspects

According to 12/30 (40%) of the stakeholders, there was no clear methodological guidance and/or consensus with regards to the design of outcomes research studies conducted between T=0 and T=4. For example, methodological guidance on study design was only finalized by ZIN in 2008, 2 years after the start of CF[ZS4]. Moreover, at the time of development of CF drugs, there was often limited medical knowledge on the disease areas for which CF drugs were developed. Therefore, consensus on core outcome sets that are relevant to the drugs in question was difficult to reach[PI1,ZS3]. This often led to an inflated list of parameters for which data needed to be collected that were, in hindsight, of little relevance to the policy question[ZW2,ZS3,ZS4,PI1,PI3,MS3].

Furthermore, 10/30 (33%) of stakeholders indicated that the quality of outcomes research studies conducted was generally poor. Recurring problems in these studies were the absence of a control group or that the intervention and control groups were not comparable. In the latter case, patients who did not wish to be treated with the new drug automatically became the control group, leading to potential selection bias[FG1]. Other aspects such as low patient recruitment and fragmented data collection in practice also impacted study quality[ZS2].

Stakeholders also emphasized the impact of rapid changes in clinical practice on the relevance of evidence generated through outcomes research studies (10/30; 33%). Oncology was mentioned as a primary example of a disease area where new drugs are introduced at a rapid pace. As a result, the drugs that may have been due for investigation in second-line treatment at T=0 became standard first-line treatments within the duration of the outcomes research study[PE3,ZS2,FG1,HI3]. Moreover, different combinations of oncology treatments also became introduced after study designs for monotherapies were finalized at T=0[PO2].

Some stakeholders (8/30; (27%) mentioned the perceived focus on excessive data collection on non-critical parameters. This supplements points previously mentioned above on the lack of consensus on methodological issues, particularly with regards to core health outcomes sets and clinical parameters sets.

Decision-making aspects

Half of all stakeholders interviewed (15/30; 50%) stated that external factors had a significant effect on the advice issued by ZIN at T=4 on drug reimbursement. The main examples whereby political pressure played such a role were alglucosidase alpha and agalsidase alpha- and beta for the treatment of Pompe's and Fabry's diseases, respectively. One stakeholder recalled that in the summer of 2012, ZIN was mentioned in national news on a daily basis due to its preliminary advice to remove these orphan drugs from

the reimbursement package[ZS1]. Pressure from patient organizations and medical societies on decision makers was also high so as to not deny patients access to the treatments[FG2,HI3,HO2].

Many stakeholders (14/30 (43%) expressed the opinion that the outcomes research studies conducted as part of CF contributed little to decision making at T=4. In fact, several stakeholders indicated that uncertainties were rarely diminished at T=4 in comparison to T=0; particularly with regards to cost-effectiveness analyses[PE2,FG2]. In general, this was the result of the methodological limitations of the studies cited above[FG2,PI2,MS3] and/or skepticism regarding the use of real-world evidence (RWE) in decision making[ZA3].

Another theme referred to by stakeholders (12/30; 40%) was the impossibility of removing drugs from the reimbursement package at T=4, even if ZIN's advice recommended to do so. The legal implications associated with drug removal were often deemed too large to attempt the feat[PE1]. Another stakeholder referred to the fact that their negotiation power and their argumentation to discontinue drug reimbursement was highly compromised at T=4[HI3].

A third of stakeholders (10/30; 33%) stated that for many of the drugs included in CF, conclusions at T=4 on appropriate use and cost-effectiveness were predictable based on insights at T=0. For example, incremental cost-effectiveness ratios (ICERs) for some drugs were so far above the reference value of €80.000 per quality-adjusted life-year (QALY) (14) that no amount of additional evidence could have proven that their use in practice was cost-effective[ZW2,FG2,HO2]. Other stakeholders stated that while outcomes research studies provided useful, product-specific insights, they did not deliver any new conclusions for decision making[PI2].

Impact of CF scheme

Several themes were identified in relation to the positive effects of CF in the Dutch healthcare setting. Firstly, 16/30 (53%) stakeholders stated that as a result of CF, cost-effectiveness of drugs and the displacement of healthcare due to exorbitant drug expenditures became topics of societal discussion. In other words, awareness was created amongst all stakeholders (including the general public) on the sustainability of the healthcare system in light of high drug prices[PE2,ZA2,HI1,MS2].

Secondly, 12/30 (40%) stakeholders stated that valuable RWE was generated, particularly on the appropriate use of drugs in clinical practice. In some instances, this allowed medical societies to develop start- and stop criteria for treatments[PE3,FG1,PI1].

Thirdly, 10/30 stakeholders (33%) were of the opinion that CF delivered valuable experiences from a policy perspective. The shortcomings encountered in CF implementation would provide concrete recommendations for the design of better future schemes[ZS3,ZS4,HI1,HI2]. Two stakeholders asserted that such learnings have already been applied for the design of ongoing MEAs ("Voorwaardelijke Toelating" (9))[FG2,HI1] and for the value-based assessment of drugs that came after the CF scheme (e.g. eculizumab and pertuzumab)[ZS3].

Finally, some stakeholders identified that CF increased awareness amongst all stakeholders for the need for multi-stakeholder collaboration to make MEAs successful

(8/30; 27%). This collaboration spans all stakeholder groups mentioned above and extends into several aspects of MEAs, including: scheme design, (outcomes research) study design, data collection, discussions on added value and decision making[ZS1,FG2,PI1,MS3].

None of the themes identified regarding the negative effects of CF met the inclusion criterion (i.e. were mentioned by <25% of stakeholders).

Conclusions and Future Perspectives

When asked if CF had achieved its perceived aims, 15/30 (50%) of the stakeholders answered "No", 15/30 (50%) answered "Partially" and 0/30 (0%) answered "Yes" (see Figure 3). For those who answered "Partially" (n=15), the responses were divided as follows: 8/15 indicated that the goal of early patient access to drugs was met, 3/15 indicated that the goal of (real-world) evidence generation was met and 4/15 indicated other aspects (e.g. that CF fulfilled its aims only for specific drugs[PO3]).

Two main themes were identified with regards to improvement points for the CF scheme. Firstly, 11/30 (37%) of the stakeholders emphasized the need for consensus amongst all stakeholders on the aims and importance of the scheme, as well as a recognition of the importance of inter-stakeholder collaboration to achieve these aims. These diverging interests impact stakeholders' perception of the aim and relevance of CF-like schemes, implying that consensus on scheme aims is vital from the start[ZA1,ZA2]. In light of this, collaborative efforts on designing and implementing CF-like schemes are also essential[PI1,PI2,HI2,PO3].

Secondly, 10/30 (33%) of the stakeholders emphasized the need for a framework whereby the underlying incentives structure can ensure that different stakeholders take up their responsibilities and be held accountable if such responsibilities are not met. For example, stakeholders indicated that CF included no sanctions mechanisms, a fact that greatly impacted the outcomes of the scheme[ZS3]. Moreover, they thought that CF drugs should not

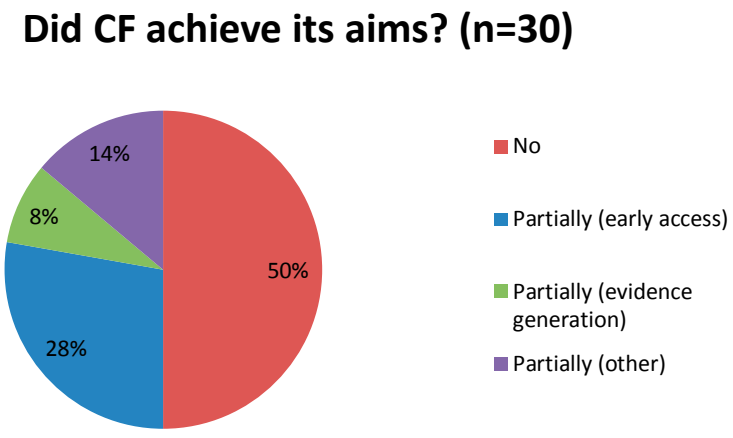


Figure 3 - Stakeholder views on achievement of aims of conditional financing (CF).

have been financed from the reimbursement package, but rather from a temporary funding source[FG1]. Knowing that drug availability is thus temporary, stakeholders responsible for data collection would thus be better incentivized to do so[FG1,PI4]. Furthermore, they indicated that conditions of obligatory inclusion of patients in outcomes research in return for access to the drug should have been considered[FG1,ZA1]. In CF, patient inclusion was done on a voluntary basis, leading to many under-powered studies and selection bias[FG1].

Other topics mentioned by stakeholders (9/30; 30%) on the improvement of the CF scheme were the need for better governance structures and distribution of roles, but also the need for monitoring procedures and pre-defined time points for progress reviews. Additionally, some stakeholders (9/30; 30%) were of the opinion that definitive conclusions should be formulated regarding highly unfavorable ICERs at T=0 (i.e. those disproportionately higher than the reference value of €80.000/QALY). Subsequently, decision makers could choose to reach measures to reduce the budget impact of these drugs[PE2] or enforce strict reference values for ICERs[MS1,MS2].

When asked how to proceed with CF in the future, 11/30 (37%) stakeholders suggested to replace CF with a scheme that resembles adaptive pathways; a scientific concept for medicine development and data generation whereby an iterative approach to evidence generation is adopted for drugs throughout their lifetime (15;16). Meanwhile, 9/30 (30%) suggested to replace CF with other new policies such as adaptive pricing or the use of electronic health records (EHRs) for evidence generation. Additionally, some stakeholders (8/30, 27%) suggested to improve CF based on the points mentioned above during interviews then subsequently re-introducing it. Finally, a few stakeholders (2/30, 7%) suggested to stop all forms of CEDs; in their opinion, such schemes do not work in practice[ZA3,HI3]. See Figure 4 for views on future perspectives in relation to CF.

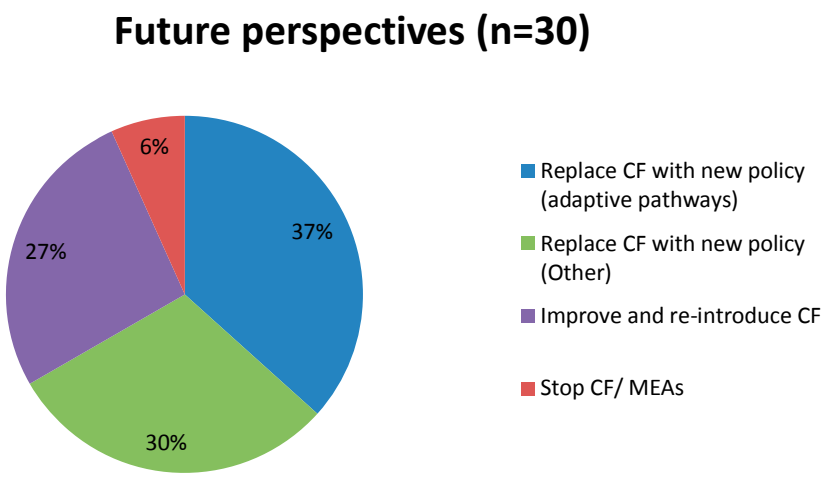


Figure 4 – Stakeholder views on future perspectives.
Abbreviations: CF: Conditional Financing; MEA: Managed Entry Agreement.

See Table 3 in the Appendix for an overview of all themes identified for the perceived functioning of CF, impact of CF, and improvement points for CF. Additionally, see Table 4 in the Appendix for illustrative quotes per theme.

Comparative sub-analysis (phase 1 and phase 2 stakeholders)

Responses to closed questions did not vary significantly between both groups; namely the questions on whether CF had achieved its aims and how to proceed with CF in the future (Figures 4 and 5 in the Appendix, respectively). Similarly, themes identified for open questions did not vary significantly. Several unique themes were identified which supplement the findings above. These themes can be found in Table 5 of the appendix.

DISCUSSION

This study examined experiences of stakeholders in implementing CF in Dutch practice. Results indicated that stakeholders had different perceptions of the aim of the CF scheme. Moreover, stakeholders highlighted numerous shortcomings in how the CF scheme functioned with regards to procedural, methodological and decision-making aspects (e.g. the 4-year timeframe, poor methodological quality of outcomes research studies and external political influence on advice at T=4, respectively). In contrast to this, stakeholders mentioned several positive effects of CF on the Dutch healthcare setting (e.g. open public discourse on cost-effectiveness of drugs and displacement of healthcare). Half of the stakeholders thought that CF had partially achieved it aims while the other half believed it had not achieved its aims. The majority of stakeholders indicated that CF should either be improved and re-introduced or replaced with new policy (e.g. adaptive pathways).

Some of the findings summarized above correspond to those from the first study on HTA dossiers analysis (10). One example relates to stakeholders' critique on the 4-year timeframe for CF often being too short and findings from dossiers indicating that only one CF drug was completed within the envisioned period. Moreover, stakeholders indicated that outcomes research studies were often of low methodological quality thus of little relevance to decision making. Meanwhile, the dossiers analysis indicated that the studies provided inadequate evidence for almost half of the research questions on cost-effectiveness. Finally, stakeholders' emphasis of the impact of external factors on decision making at T=4 correspond to findings from the dossiers analysis indicating that only couple of all CF drugs eventually received a negative reimbursement advice at T=4. On the other hand, stakeholder interviews also provided complementary insights on topics that could not be addressed through HTA dossiers analysis, such as stakeholders' perceived impact of the CF scheme, conclusions on CF and future perspectives.

Healthcare systems worldwide include a wide array of different stakeholders, each with their differing mandates and a complex network of interactions amongst them. As a result, MEAs present different trade-offs for each stakeholder in relation to their specific interests. Consequently, from a governance perspective, there is a critical need for clear frameworks

that entail stakeholders' roles, responsibilities, incentives and sanctions (18;19). To begin with, stakeholders' perceptions of the scheme aims, thus their own gains, greatly matter; in Germany, similar schemes failed due to clinicians perceiving them as posing limitations on their prescribing choices (20). Meanwhile in Italy, it still remains nearly impossible to reclaim costs for non-responder patients from pay-for-performance schemes (6), possibly due to the absence of sanctions mechanisms. Previous experiences from England also point to problems arising from the absence of "exit strategies" (21). Although such concepts on governance may seem quite elementary, their importance cannot be underestimated provided their recurrence both in this study and in literature.

In particular, the implementation of CEDs poses additional challenges relating to infrastructure for (real-world) data collection and subsequent data analysis for decision making. The current model for creating *ad hoc* product- or disease registries for separate research questions may be unsustainable due to various reasons, including: costs, administrative burden of extra data registration and data accessibility for research (5;22;23). Meanwhile, major investments are needed to establish (digital) systems for data collection and analysis, whether through paper-based questionnaires or EHRs (6;21). In light of stakeholders' comments above on the financing of outcomes research studies, it would be difficult to specify which stakeholders should be responsible for financing the establishment of information technology (IT) infrastructures for implementing EHRs. Even with the necessary infrastructure in place, healthcare professionals in clinical practice would need to be trained to use such IT systems, requiring both financial and time investments on their behalf. Provided the high workload experienced by healthcare professionals in general, it may thus be difficult to commit to such investments. Finally, the availability of data within EHRs does not automatically guarantee access to data for analysis purposes, as illustrated by numerous examples in literature (23-25).

Another important challenge is the analysis of real-world data (RWD) and interpretation of real-world evidence (RWE). Findings above allude to skepticism amongst decision makers in basing decisions on RWE. Furthermore, numerous articles refer to the methodological difficulties associated with analyzing RWD and using RWE in decision making (26-28). From a methodological perspective, many advances have been made in the analysis of RWD; both alone or in combination with RCT data (29-31). Moreover, recent guidelines issued by combined efforts of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and International Society for PharmacoEpidemiology (ISPE) provide an example of clear guidance on good procedures for the conduct and reporting of RWD studies to increase decision makers' confidence in the validity of RWE (32;33). However, implementing state-of-the-art methodology for RWD analysis requires extensive training in pharmacoepidemiology and biostatistics; implying yet again the need for considerable financial and time investments for the training of personnel conducting the analyses (e.g. pharmaceutical industry) or interpreting the results (e.g. HTA agencies). As a consequence of factors discussed above, decision makers in both public and private stakeholders still have little experience in incorporating RWE in current processes (23).

However, despite the challenges, most stakeholders encouraged the development of new CEDs to address dilemmas encountered in decision making on the reimbursement of drugs. Literature also alludes to increasing trends in conditional marketing authorizations with relatively larger uncertainties in evidence for HTA (particularly for oncology and orphan drugs) (34) and increasing trends in MEAs use (5;22). Arguably, the implementation of EHRs may provide the means for routine collection of patient-level data in practice that is of sufficient detail for advanced analyses, delivering more robust information for decision-making in the future. Ideally, evidence generated through the routine analysis of EHRs would subsequently facilitate iterative HTA of drugs within the context of adaptive pathways. Examples in literature on similar concepts currently exist whereby principles of artificial intelligence are applied to existing EHRs to generate personalized patient clinical pathways in practice (35;36). Though the authors are aware of the challenges associated with such approaches, it is our hope that they could be incorporated into HTA in the future, in order to inform the design of better CEDs.

Strengths

In this study, standardized methodology was implemented for identifying stakeholders, approaching stakeholder representatives and conducting interviews. Furthermore, content analysis for each stakeholder group was conducted by two authors with all discrepancies addressed by consensus-seeking amongst authors. Moreover, the quality of the research conducted and subsequent reporting thereof was compared to recommendations of the COREQ checklist for validation of both aspects.

Finally, the interview analyses conducted provide complementary findings to those from the previous study on HTA dossiers analysis (10). Together, the studies provide a thorough and systematic evaluation of experiences gained with the implementation of CF in Dutch practice.

Limitations

Although all relevant stakeholder groups were involved, we could not include all individual stakeholders involved with CF in the interviews for this study. However, the authors used several sampling methods (purposive sampling and snowballing) to ensure that a comprehensive range of stakeholders were included, spanning different stakeholder groups. Moreover, data saturation was discussed amongst authors and provided grounds for limiting the number of interviews.

The threshold implemented to select and include themes from content analysis, using at least 25% of the stakeholders, is not standard. However, the authors are not aware of the existence of standard thresholds for such criteria in literature on qualitative methods. Moreover, illustrative quotes cited in the Appendix cover additional themes that may not have met the 25% threshold implemented.

Finally, this study represents a policy evaluation of a national CED in an attempt to inform the development of future schemes. Ideally, the scope of this study would thus

include MEAs implemented in other countries (e.g. Italy (6), France (37), Sweden (37), the United States (5) and the United Kingdom (21)). However, the placement of the authors within Dutch institutions provided extensive access to national stakeholders thus allowing for a thorough, systematic analysis of CF. Such access may not be equally facilitated in other settings. Provided the complexity of designing and implementing CEDs, we therefore encourage further research on experiences gained in the implementation of MEAs (including CEDs) in other countries to provide complementary learnings for the design of future schemes.

CONCLUSION

This study provides insights on stakeholders' experiences in implementing CF in Dutch practice, an example of MEAs (namely a CED scheme). Results demonstrate differences amongst the stakeholders on the perceived aim of CF. Conversely, there is some agreement amongst stakeholders on the shortcomings in the functioning of CF (i.e. relating to procedural, methodological and decision-making aspects), the positive impact of CF on the Dutch healthcare setting and improvement points for CF. Despite stakeholders' belief that CF only partially met its aims, if not at all, there is still agreement on the need for new policy to address the same aims of CF in the future. However, stakeholders diverge on whether CF should be improved based on learnings identified and re-introduced into practice or replaced with new policy schemes.

This study was conducted with the aim of informing ongoing international discussions on the design and implementation of future MEA schemes. Provided the onslaught of innovative, yet expensive drugs and increasing trends of MEAs use by HTA agencies and payers, further research on experiences gained with other MEAs is thus critical to inform the design of better schemes in the future.

REFERENCE LIST

1. OECD. OECD Data: Pharmaceutical Spending. 2016.[cited 2017 December 28]; Available from: https://data.oecd.org/healthres/pharmaceutical-spending.htm
2. Iyn-Hyang L, Karen B, Catherine H, et al. International experience in controlling pharmaceutical expenditure: influencing patients and providers and regulating industry - a systematic review. Journal of health services research & policy 2014;20:52-9.
3. Godman B, Shrank W, Wettermark B, et al. Use of generics: a critical cost containment measure for all healthcare professionals in Europe? Pharmaceuticals 2010;3:2470-94.
4. Carlson JJ, Sullivan SD, Garrison LP, et al. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. Health Policy 2010;96:179-90.
5. Garrison Jr LP, Carlson JJ, Bajaj PS, et al. Private sector risk-sharing agreements in the United States: trends, barriers, and prospects. The American journal of managed care 2015;21:632-40.
6. Navarria A, Drago V, Gozzo L, et al. Do the current performance-based schemes in Italy really work? "Success Fee": A novel measure for cost-containment of drug expenditure. Value Health 2015;18:131-6.
7. Klappe-Sabadi G, Jansman FGA, Honkoop HA, et al. Regionale verschillen in voorschrijfgedrag trastuzumab opnieuw getoetst. PW Wetenschappelijk Platform 2[7]. 2008. 2008.
8. Nederlandse Zorgautoriteit (NZa). BELEIDSREGEL BR/CU-2017 Dure Geneesmiddelen. 2011.
9. Zorginstituut Nederland, mw.dr.G.Ligtenberg. Voorwaardelijke toelating/financiering van zorg. 6-4-2012.
10. Makady A, van Veelen A, De Boer A, et al. Implementing Managed Entry Agreement in Practice: the Dutch reality check. (UNDER REVIEW). Health Policy 2018.
11. Palinkas LA, Horwitz SM, Green CA, et al. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health and Mental Health Services Research 2015;42:533-44.
12. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15:1277-88.
13. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349-57.
14. Zorginstituut Nederland. Beoordeling geneesmiddelen procedure, https:// www.zorginstituutnederland.nl /pakket/ werkwijze+ pakketbeheer/ beoordeling+ geneesmiddelen/ procedure. 2016.
15. European Medicines Agency. Adaptive Pathways. 2017.[cited 2017 October 30]; Available from: http://www.ema.europa.eu/ ema /index.jsp?curl=pages/ regulation/ general/ general_content_000601.jsp
16. Bouvy JC, Jonsson P, Longson C, et al. Health technology assessment in the context of adaptive pathways for medicines in Europe: challenges and opportunities. Clin Pharmacol Ther 2016;100:594-7.
17. Zorginstituut Nederland. Beoordeling stand van de wetenschap en praktijk. 2015. Zorginstituut Nederland.
18. Bevir M. Governance: A very short introduction OUP Oxford, 2012.
19. Duit A, Galaz V. Governance and Complexity: Emerging Issues for Governance Theory. Governance 2008;21:311-35.
20. Drug Commission of the German Medical Association. Opinion on the "cost-sharing initiatives" and "risk-sharing agreements" between pharmaceutical manufacturers and Health and Hospital. 2008.[cited 2017 April 14]; Available from: http://www.akdae.de/Stellungnahmen/Weitere/20080508.pdf
21. Faulkner SD, Lee M, Qin D, et al. Pricing and reimbursement experiences and insights in

the European Union and the United States: Lessons learned to approach adaptive payer pathways. Clin Pharmacol Ther 2016;100:730-42.

22. Carlson JJ, Gries KS, Yeung K, et al. Current status and trends in performance-based risk-sharing arrangements between healthcare payers and medical product manufacturers. Applied health economics and health policy 2014;12:231-8.
23. Makady A, Stegenga H, Ciaglia A, et al. Practical implications of using real-world evidence (RWE) in comparative effectiveness research: learnings from IMI-GetReal. 2017. Future Medicine.
24. van Staa TP, Goldacre B, Gulliford M, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. BMJ 2012;344:e55.
25. CTTI. Registry Trials Project Multi-Stakeholder Expert Meeting: Summary of the meeting held March 30th, 2016. 2016.
26. Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: lessons learned from the PHAROS registry. Pharmacoecon 2015;33:551-60.
27. Mohseninejad L, van Gils C, Uyl-de Groot CA, et al. Evaluation of patient registries supporting reimbursement decisions: the case of oxaliplatin for treatment of stage III colon cancer. Value Health 2015;18:84-90.
28. Alemayehu D, Riaz Ali MPP, Alvir JM, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17:S3-S37.
29. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol 2004;57:1223-31.
30. Garbe E, Suissa S. Pharmacoepidemiology. In: Ahrens W, Pigeot I, eds., Handbook of Epidemiology. New York, NY: Springer New York, 2014. p. 1875-925.
31. Efthimiou O, Mavridis D, Debray T, et al. Combining randomized and non-randomized evidence in network meta-analysis. Stat Med 2017;36:1210-26.
32. Berger ML, Sox H, Willke RJ, et al. 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. Pharmacoepidemiol Drug Saf 2017;26:1033-9.
33. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1. 0. Pharmacoepidemiol Drug Saf 2017;26:1018-32.
34. Kleijnen S, Lipska I, Alves TL, et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Ann Oncol 2016;mdw233.
35. Zhang Y, Padman R, Patel N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. Journal of biomedical informatics 2015;58:186-97.
36. Wang HQ, Li JS, Zhang YF, et al. Creating personalised clinical pathways by semantic interoperability with electronic health records. Artif Intell Med 2013;58:81-9.
37. Makady A, ten Ham R, de Boer A, et al. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. Value Health 2017.

SECTION IV

Access to Real-World Evidence

CHAPTER 7

Practical Implications of Using Real-World Evidence in Comparative Effectiveness Research: Learnings from IMI-GetReal

A. Makady[1,2], H. Stegenga[3], A. Ciaglia[4], T.P.A. Debray[5,6], M. Lees[7], M. Happich[8], B. Ryll[9,10], K. Abrams[11], R. Thwaites[12], S. Garner[3], P. Jonsson[3], W. Goettsch[1,2]
(on behalf of GetReal Work Packages 1 & 4)

[1] The National Healthcare Institute (ZIN), Diemen, the Netherlands;
[2] Department of Pharmacoepidemiology & Clinical Pharmacotherapy, Utrecht University, Utrecht, the Netherlands;
[3] The National Institute for Health and Care Excellence (NICE), London, the United Kingdom;
[4] International Alliance of Patients' Organizations, London, the United Kingdom
[5] Julius Center for Health Sciences and Primary Care. University Medical Centre Utrecht, Utrecht, the Netherlands;
[6] Cochrane Netherlands, University Medical Center Utrecht, Utrecht, Netherlands
[7] Bristol-Myers Squibb, Paris, France;
[8] Eli Lilly, Hamburg, Germany;
[9] Melanoma Patient Network Europe, Uppsala, Sweden;
[10] Uppsala University, Uppsala, Sweden;
[11] Department of Health Science, University of Leicester, Leicester, the United Kingdom
[12] Takeda, London, the United Kingdom

Published in the Journal of Comparative Effectiveness Research. Citation is as follows: Makady A, Stegenga H, Ciaglia A, Debray TP, Lees M, Happich M, Ryll B, Abrams K, Thwaites R, Garner S, Jonsson P. Practical implications of using real-world evidence (RWE) in comparative effectiveness research: learnings from IMI-GetReal. Journal of Comparative Effectiveness Research. 2017 Aug 6(6).

ABSTRACT

In light of increasing attention towards the use of Real-World Evidence (RWE) in decision making in recent years (1), this commentary aims to reflect on the experiences gained in accessing and using RWE for Comparative Effectiveness Research (CER) as part of the Innovative Medicines Initiative GetReal Consortium (IMI-GetReal) (2) and discuss their implications for RWE use in decision-making. For the purposes of this commentary, we define RWE as evidence generated based on health data collected outside the context of RCTs (3). Meanwhile, we define Comparative Effectiveness Research (CER) as the conduct and/or synthesis of research comparing different benefits and harms of alternative interventions and strategies to prevent, diagnose, treat, and monitor health conditions in routine clinical practice (i.e. the real-world setting) (4). The equivalent term for CER as used in the European context of Health Technology Assessment (HTA) and decision making is Relative Effectiveness Assessment (REA).

WHY IS REAL-WORLD EVIDENCE (RWE) RELEVANT FOR COMPARATIVE EFFECTIVENESS RESEARCH (CER)?

Traditionally, Randomised Controlled Clinical Trials (RCTs) are considered the established method for providing information pertaining to the efficacy and safety of health interventions. However, the highly-controlled conditions characteristic of RCTs may not always accurately represent clinical practice (5). Whereas RCT patient populations are highly selected and homogenous, with a protocol-driven patient-follow up, patient populations seen in clinical practice are typically heterogeneous and often present with co-morbidities. Moreover, RCTs often have short follow-up durations, preventing the detection of rare- or long-term adverse events of interventions. Surrogate endpoints measured in RCTs, such as progression-free survival in oncology patients, may also be less relevant to decision-making than overall survival. Furthermore, clinical practice may vary on a regional or national level. These differences can lead to a discrepancy between the observed efficacy of interventions in RCTs and their effectiveness in clinical practice, a phenomenon often referred to as the efficacy-effectiveness gap (6;7).

Real-World Evidence (RWE), broadly defined as evidence generated based on health data collected outside the context of RCTs (3), may help identify, quantify and address this efficacy-effectiveness gap in treatment effects where needed. For example, RWE may supplement RCT data to improve estimates of treatment effects in the real-world setting through evidence synthesis or the use of predictive modelling techniques. RWE may also provide information on parameters not examined in clinical trials, such as adherence to treatment, rare adverse events, and resource use in clinical settings. The insights provided from RWE may have significant implications for drug developers, regulators and particularly, Health Technology Assessment (HTA) agencies and payers whose decisions rely on evidence of comparative effectiveness (1).

WHY IS INDIVIDUAL PATIENT-LEVEL DATA (IPD) IMPORTANT WHEN USING RWE IN CER?

Methodologies for the statistical analysis, synthesis and critical appraisal of RWE have developed considerably in the past 20 years, including formal checklists for assessing risk of bias, propensity scoring techniques, instrumental variable analyses, multivariable regression analyses, and advanced meta-analysis methods (8-11). These methods can begin to address a number of important shortcomings with RWE that are particularly problematic for its use in comparative effectiveness research (CER), such as the lack of randomisation of patients which can result in a lack of comparability between treatment groups, the presence of missing observations on relevant patient outcomes or covariates, and the presence of confounders. Therefore, the implementation of such methods can be used to increase the robustness of estimates derived from analyses using RWE in a number of CER scenarios.

Summary data (also known as aggregate data), whether from RCTs or RWE, such as estimates of comparative treatment effect are often of limited value for CER. For instance,

a major drawback of aggregate data (AD) is the limited ability to explore individual patient characteristics which may influence or confound treatment outcomes. Therefore, in order to conduct robust CER that can inform decision-making, HTA agencies and payers often require more sophisticated analyses to be conducted whereby researchers can adjust for individual patient characteristics to generate more accurate estimates of effectiveness (1;12). Several strategies may be employed by research teams to do so and will be discussed further below. Importantly, such strategies require analyses based on IPD, whether by the researchers themselves, or an alternative party.

WHAT WERE IMI-GETREAL'S EXPERIENCES IN ACCESSING IPD THROUGHOUT CASE STUDIES?

The Innovative Medicines Initiative (IMI)-GetReal project was a 3-year project exploring the use of RWE to improve drug effectiveness research throughout the lifecycle of drug development. The project was a public-private partnership with a multi-stakeholder constituency including industry, regulators, HTA agencies, academia and patient organisations. In total, the project comprised 5 Work Packages (WP1-5), each addressing different objectives (2). Work conducted for 2 work packages (WP1 & 4) involved attempts to access and use IPD from RWE and/or RCTs. Work Package 1 (WP1) conducted a series of case studies in multiple disease areas, aiming to explore methods for using RWE to improve effectiveness estimates and to examine the acceptability of these methods amongst relevant stakeholders through stakeholder workshops. Meanwhile, Work Package 4 (WP4) explored best practices for evidence synthesis from RWE and/or RCTs through literature reviews and a series of case studies. Some of the methods explored included extrapolation of long-term outcomes beyond trial durations, enrichment of Network Meta-Analyses (NMAs) with RWE and generalization of RCT results to real-world populations through propensity scoring techniques. Together, the case study teams, each jointly co-lead by a public and industry partner, sought to access individual patient-level data (IPD) from both RWE repositories and RCTs to conduct these analyses.

In total, 7 case studies were conducted as part of WP1 and WP4 work, spanning multiple disease indications and each lasting approximately 1.5 years. RWE repositories approached included 12 indication registries and 8 observational studies. Eventually, case study co-leads managed to secure access to IPD from 4/12 registries and 3/8 observational studies, indicating that IPD access from RWE sources succeeded in only 35% of cases. On the other hand, IPD was requested from 43 RCTs and granted in 41/43 studies, indicating that IPD retrieval from RCTs exceeded 95%.

Experiences encountered with accessing IPD from RWE repositories varied per case study. A positive example relates to a combined WP1/4 case study whereby co-leads secured access to IPD from registries in 2 different countries. Moreover, the registries actively informed the case study team of upcoming data updates (9;13). In 4 instances across different case studies, registry holders and observational study authors initially indicated

their willingness to provide access to IPD. However, they eventually communicated that their datasets were not research-ready within project timelines due to an extensive need for cleaning & trimming (9;13;14). A negative example relates to a WP1 case study, whereby one registry refused to discuss possibilities for collaboration upfront, due to being approached by an industry co-lead (14). Meanwhile, prolonged negotiations lasting 16 months with another registry were abandoned when representatives iterated that access to IPD would be refused until all PhD students associated with the registry completed their dissertations, in fear that they would otherwise lose ownership of findings based on the data (14). The same registry indicated earlier in negotiations that access to a tailored portion of IPD based on the research proposal submitted could also be bought for a fee. However, the considerable amount of this fee (surpassing €100,000) acted as a direct barrier to IPD retrieval.

In summary, IMI-GetReal's experiences in accessing IPD from RWE repositories were disparate. In general, only a third of all requests for IPD access from RWE repositories submitted across all case studies were successful. For half of the case studies, IPD was accessed from registries and observational studies. Furthermore, co-leads iterated that data sharing agreements and structures did not pose considerable problems for those case studies. However, for the remaining case studies, access to IPD was denied. Reasons for inaccessibility mostly related to datasets not being research-ready within project timelines or unwillingness to share data. These reasons raise important questions regarding general competence in generating data sets of sufficient quality to be readily available for research, as well as data ownership, respectively.

As an alternative to accessing IPD, case study teams explored options for using aggregate data (AD) from registries and observational studies. To do so, case study teams either requested that registries run pre-scripted analyses on IPD and report the aggregate results back to the team (14;15), or attempted to use AD as reported in literature (9;14;15). The AD retrieved from both approaches was subsequently used in several ways to perform CER, for example by simulating patient-level data or as direct input for effect estimates in NMA models. Access to AD through both approaches was relatively easier. Importantly, AD generated by pre-scripted analyses on IPD provided relevant insights for conducting CER (e.g. by illustrating the distribution of covariates within patient populations thus allowing for more accurate simulations of the original patient population). However, this approach requires considerable expertise to implement and relies heavily on cooperation from registry holders to run the requested analyses. On the other hand, the absence of information on patient covariates within AD retrieved from literature limited the robustness of health outcomes estimates generated from such data. Therefore, although AD can be easily obtained from literature, it is often of limited usefulness, mostly lending itself to descriptive statistical analyses rather than to analysis of treatment effects across different settings and populations (9;14).

Another point worth noting is that although accessibility of IPD from RWE repositories was a prominent issue encountered in using RWE for CER in IMI-GetReal case studies, it was not the only one (9;13). For example, in order to make use of IPD

accessed, the case study teams often had to invest considerable time and effort in making datasets research-ready (e.g. by trimming the dataset or imputing missing data values). Occasionally, observational studies only investigated treatment patterns, rather than treatment outcomes, making them of little use to analyses involving head-to-head comparisons of effectiveness. Moreover, where treatment outcomes were recorded, varying definitions of the outcome measures across different studies often complicated the synthesis of IPD from RWE and RCT sources. These issues raise additional methodological and practical concerns in applying RWE to CER, some of which have been addressed in scientific literature and should be considered by all stakeholders attempting to undertake similar efforts (10;16). However, in subsequent sections we focus on the issue of accessibility to IPD from RWE repositories and its implications for using RWE for CER and decision-making.

WHAT ARE THE CONSEQUENCES OF INACCESSIBILITY TO IPD FROM RWE REPOSITORIES ON ITS POTENTIAL USE FOR DECISION-MAKING IN HEALTHCARE?

IMI-GetReal case study workshops demonstrated considerable variability in external stakeholders' views on the acceptability of RWE use in CER and subsequent decision-making. The reasons behind such controversy are multi-factorial, yet generally hinged on two inter-related aspects: a lack of trust in the robustness of findings based on RWE compared with RCT data, as well as a lack of experience with using RWE in currently available methods to address questions relating to (comparative) drug effectiveness. Numerous ongoing initiatives aim to address the former aspect through guideline development on topics including: good practices to ensure data quality and standardised core outcomes datasets within registries to inform CER (17;18), statistical analysis of RWE (19-21) and the reporting of results from observational studies (22;23). On the other hand, the latter aspect implies a lack of published examples exploring advanced methods for RWE use in CER and subsequent feedback on these methods from relevant decision-makers.

Despite the multi-stakeholder nature of IMI-GetReal case study teams, adherence to application procedures for data access, as well as the necessary disclaimers to registry owners and study authors approached, accessibility to IPD from RWE repositories proved to be challenging. Consequently, insufficient data were available to thoroughly explore novel methods for RWE use in almost half of the case studies. More importantly, the consortium's experience with inaccessibility of IPD RWE for research purposes was echoed by many external stakeholders present in stakeholder workshops, implying that access to IPD RWE remains a persistent issue beyond the IMI-GetReal consortium. Arguably, this inaccessibility to IPD RWE both contributes to the lack of concrete examples demonstrating the potential added value of RWE use in CER and the wide lack of trust among decision-makers regarding the robustness of findings based on RWE.

WHAT ARE POTENTIAL SOLUTIONS TO ADDRESSING ISSUES FACED WITH ACCESS TO IPD FROM RWE REPOSITORIES IN THE FUTURE?

Bearing in mind that CER aims to shed light on the ideal implementation of healthcare interventions to achieve maximum societal benefits, inaccessibility to IPD from RWE repositories adversely affects society as a whole. Moreover, as RWE is generated by patients within routine healthcare, it is essential that patients benefit from the use of this data; increased accessibility to IPD RWE to improve CER and decision-making should benefit all patients, not just those who control access to such data. Consequently, the dynamic, multi-stakeholder nature of the healthcare sector warrants a collaborative approach to solving issues pertaining to governance of RWE repositories, including accessibility to IPD.

An important aspect to enable collaborative efforts is a general understanding amongst all stakeholders of the patient-centred goals behind healthcare in general, as well as RWE collection and analysis to improve healthcare. In this regard, the role RWE can play in pursuing patient-centred goals will be best understood if healthcare stakeholders make a strong commitment to involve all key actors in setting-up and developing procedures to enable access to registries. This requires that, contrary to current practice, all relevant stakeholders participate in steering committees of these registries, whereby a spirit of joint action is crucial for success.

Furthermore, registries are currently set up based on undisclosed contracts, leading to situations where it is difficult to deduce why accessibility is difficult and which stakeholders are involved in deciding on data requests. Therefore, making such contracts transparent is another important step to increase clarity in the wider community about governance issues such as data ownership, gate keepers for data access, funding sources and conflicts of interests.

Developments on other fronts may provide additional potential solutions. For example, the EU Clinical Trial Directive was recently established, whereby sponsors of RCTs conducted for marketing authorisation applications agreed to provide access to all patient-level clinical reports of trial subjects online (24). Presently, no equivalent initiative exists for the publication of similar IPD for RWE generated through observational studies and may be a worthwhile endeavour for the future. However, bearing in mind that patient-level data is subject to strict privacy rules, such endeavours should not preclude the review of research protocols by relevant committees to guarantee that such data is not misused and that the scientific rigor of analyses exploiting the data is guaranteed through transparent publication of the analysis protocols.

Another example relates to the FDA-Sentinel initiative, whereby external researchers can send standardised data queries to multiple nodes of a de-centralised network of participating databases (25). In this model, databases can opt-in or out of the sentinel initiative without having to relinquish complete access to their IPD, yet still run external research queries. The main advantage of such a model is its ability to circumvent sensitivities

relating to full-fledged access to IPD while delivering the required information for furthering scientific pursuits. This approach towards remote data querying has demonstrated potential in IMI-GetReal case studies (14;15). Moreover, similar frameworks have been implemented in other fields of research, such as DataSHIELD to conduct international research as part of the Healthy Obese Project and the Environmental Core Project (BioSHaRE-EU) (26). Other initiatives exploring such frameworks include the IMI-Big Data for Better Outcomes (IMI-BD4BO) (27). Arguably, equivalent systems for existing registries would bring RWE use in CER a long way.

In conclusion, the current state of accessibility to RWE experienced during IMI-GetReal case studies and stakeholders beyond the consortium poses a considerable barrier to furthering RWE use in CER and healthcare decision-making. Bearing in mind that such data is generated by patients in clinical practice, this barrier diminishes the potential benefit of using RWE to provide critical insights on the effectiveness of treatments for all patients in real practice; insights that RCTs are often not designed to provide. An array of potential solutions lend themselves to overcoming this persistent inaccessibility to RWE and maximising societal gain from its use in CER. However, the choice regarding which path to take, addressing trade-offs associated with such a choice, as well as its implementation, requires a collaborative effort spanning all relevant stakeholders; from decision-makers, to industry and patient representatives.

REFERENCE LIST

1. Makady A, Goettsch W. Review of Policies and Perspectives on Real-World Data. 2015. [cited 2017 April 6];
2. IMI-GetReal. IMI-GetReal: Overall objectives. 2017.[cited 2016 December 13]; Available from: https://www.imi-getreal.eu/About-GetReal/Overall-objectives
3. Garrison LP, Neumann PJ, Erickson P, et al. Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report. Value Health 2007;10:326-35.
4. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med 2009;151:203-5.
5. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. J Clin Epidemiol 2010;63:1053-8.
6. Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 2011;10:495-506.
7. Nordon C, Karcher H, Groenwold RH, et al. The Efficacy-Effectiveness Gap: Historical Background and Current Conceptualization. Value Health 2016;19:75-81.
8. Garbe E, Suissa S. Pharmacoepidemiology. In: Ahrens W, Pigeot I, eds., Handbook of Epidemiology. New York, NY: Springer New York, 2014. p. 1875-925.
9. Hummel N, Debray TPA, Didden EM, et al. Methodological guidance, recommendations and illustrative case studies for (network) meta-analysis and modelling to predict real-world effectiveness using individual participant and/or aggregate data. 2017.[cited 2017 May 20]; Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/2017-03-30%20-%20WP4%20-%20Methodological%20guidance%2C%20recommendations%20and%20illustrative%20case%20studies.pdf
10. Efthimiou O, Mavridis D, Debray T, et al. Combining randomized and non-randomized evidence in network meta-analysis. Stat Med 2017;36:1210-26.
11. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.
12. Debray TP, Schuit E, Efthimiou O, et al. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? Stat Methods Med Res 2016;0962280216660741.
13. Ericson A, Gsteiger S. Deliverable D1.6 WP1 Case Study Review: Rheumatoid Arthritis. 2016.[cited 2017 March 24]; Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.6%20Report%20-%20Rheumatoid%20Arthritis_webversion.pdf
14. Makady A, Kalf R, Goettsch W., Lees M. Deliverable D1.6 WP1 Case Study Review: Metastatic Melanoma. 2016.[cited 2017 March 24]; Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20WP1%20deliverables/Deliverable%201.6%20Report%20-%20Rheumatoid%20Arthritis_webversion.pdf
15. NICE, Novartis, University of Leicester. Deliverable 1.6 (Multiple Sclerosis Case Study). 2016.Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.6%20Report%20-%20multiple%20sclerosis%20case%20study_webversion.pdf
16. Alemayehu D, Riaz Ali MPP, Alvir JM, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17:S3-S37.
17. EUnetHTA. Work Package 5 - Life cycle approach to improve Evidence Generation. 2017.Available from: http://eunethta.eu/activities/eunethta-joint-action-3-2016-20/work-package-5-life-cycle-approach-improve-evidence-gener

18. PARENT JA. The Cross-Border Patient Registries Initiative. 2016.[cited 2017 March 24]; Available from: http://patientregistries.eu/

19. Cox E, Martin BC, Van Staa T, et al. Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force ReportGÇöPart II. Value Health 2009;12:1053-61.

20. Hall GC, Sauer B, Bourke A, et al. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf 2012;21:1-10.

21. OHDSI. Observation Health Data Sciences and Informatics. 2017.Available from: https://www.ohdsi.org/

22. ENCePP. The ENCePP Code of Conduct for Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies. 2010. [cited 2015 July 3]; Available from: http://www.encepp.eu/documents/encepp_studies/ENCePP%20Code%20of%20Conduct_20100507.pdf

23. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. International Journal of Surgery 2014;12:1495-9.

24. European Commission. Clinical trials - Regulation EU No 536/2014. 2017.[cited 2017 March 24]; Available from: https://ec.europa.eu/health/human-use/clinical-trials/regulation_en#ct4

25. McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. Pharmacoepidemiol Drug Saf 2012;21:18-22.

26. Gaye A, Marcon Y, Isaeva J, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014;43:1929-44.

27. Innovative Medicines Initiative. Big data, medicines safety and respiratory disease focus of €93 million IMI Call for proposals. 2015.[cited 2017 May 28]; Available from: http://www.imi.europa.eu/content/imi-2-call-6-launch

SECTION V

Novel Sources for Generating Real-World Evidence

CHAPTER 8

Use of Social Media in the Assessment of Relative Effectiveness: An explorative review with examples from oncology. Hopeful or hopeless?

R. Kalf[1], A. Makady[1,2], R. ten Ham[2], K. Meijboom[3], W. Goettsch[1,2]
(on behalf of GetReal Work Package 1)

[1] The National Health Care Institute, Eekholt 4,
1112 XH Diemen, the Netherlands.
[2] Department of Pharmacoepidemiology and Clinical Pharmacology,
Utrecht University, Utrecht, the Netherlands.
[3] VU University Amsterdam, Department of Health Sciences,
Amsterdam, the Netherlands

Accepted for publication in the Journal of Medical Internet Research

ABSTRACT

Background

An element of health technology assessment constitutes assessing the clinical effectiveness of drugs, generally called relative effectiveness assessment (REA). Little evidence from the real world is available directly after market access, therefore randomized controlled trials are used to obtain information for REA. However, there is growing interest in using real-world data (RWD) for REA. Social media may provide a source of RWD.

Objective

We assessed the extent to which social media-generated health data has provided insights for REA.

Methods

An explorative literature review was conducted following PRISMA guidelines to identify examples in oncology where health data were collected using social media. Scientific and grey literature published between January 2010 and June 2016 was identified by four reviewers, who independently screened studies for eligibility and extracted data. A descriptive qualitative analysis was performed.

Results

Of 1032 articles identified, 8 were included: 4 articles identified adverse events in response to cancer treatment, 3 articles disseminated quality of life (QoL) surveys, and 1 study assessed the occurrence of disease-specific symptoms. Several strengths of social media-generated health data were highlighted in the articles, such as efficient collection of patient experiences and recruiting patients with rare diseases. Conversely, limitations included validation of authenticity and presence of information and selection bias.

Conclusions

Social media may provide a potential source of RWD for REA, particularly on aspects such as adverse events, symptom occurrence, QoL, and adherence behaviour. This potential has not yet been fully realised and the degree of usefulness for REA should be further explored.

INTRODUCTION

Within the context of rising health care costs, limited budgets, and the onslaught of innovative yet expensive medications, the value of health technology assessment (HTA) for decision-makers, regulators, pharmaceutical companies and patients is becoming increasingly important. HTA is defined as 'the systematic evaluation of the properties and effects of a health technology' (1). Health technologies are defined as 'interventions developed to prevent, diagnose or treat medical conditions, promote health, provide rehabilitation, or organize health care delivery (2). An important element of HTA is relative effectiveness, i.e. the extent to which an intervention – provided under routine clinical conditions – does more good than harm in comparison to one or more alternatives (1). Traditionally, a relative effectiveness assessment (REA) conducted directly after market authorization of a new drug is extrapolated using health outcomes (e.g. mortality) obtained from randomized controlled trials (RCTs). However, the tightly-controlled conditions and highly selective patient groups within RCTs may result in findings that are not generalizable to routine clinical settings where patients are more heterogeneous. In routine practice, pregnant women, children, elderly people and patients with comorbidities may eventually receive the new drugs examined in RCTs, while these patient populations are generally excluded from such RCTs. Therefore, researchers may additionally resort to real-world data (RWD) as a supplementary source of evidence to assess relative effectiveness. Real-world data can be defined as 'an umbrella term for data regarding the effects of health interventions that are not collected in the context of conventional randomized controlled trials' (1). Patient registries and electronic health records are established examples of RWD sources, but another potential source of RWD may be social media.

Social media are often used by patients as a source to search for information on their health conditions, share their experiences and find social support (3,4). For example, many patients use Twitter to stay up to date with the latest health care developments and increase their knowledge on their disease, while Facebook is more often used for social support and exchanging experiences (3). Social media users who have a chronic condition are more likely to use the internet for such purposes than are healthy social media users (5). By assessing the content viewed, generated and exchanged by patients via social media, a considerable amount of information on patient perspectives and experiences can be gathered.

Although social media have been used for different aspects of research, such as patient recruitment (6-8), dissemination of interventions (9,10) and education (11), little is known about its contribution to REA. In 2008 a study showed that blogs could be used to collect patient experiences regarding diabetes and diabetes management to provide information for HTA by enhancing the evidence available in published literature (12). More recently, a number of pharmaceutical companies have begun to make use of social media to gain insight into patient perspectives on adverse events (AEs) (13,14) and to assess their switching behaviours (15). Similarly, the Association of the British Pharmaceutical Industry (ABPI) has published guidelines on best practices for the monitoring and management of

AEs via such sources (16). Moreover, the Food and Drug Administration (FDA) is increasingly focusing on the use of health data from social media by collaborating with PatientsLikeMe; a platform where patients can share their health data online to gain insight into patient perspectives on adverse events (17,18). In light of these initiatives, it may become possible for health data reported by patients on social media to contribute to the REA of new therapies.

The aim of this article is to assess the extent to which health data generated from social media have provided insights for REA. We conducted an explorative review to identify examples in oncology where health data were collected using social media. Oncology was chosen due to the considerable number of innovative drugs being developed at a rapid pace in this area. For example, the European Medicines Agency reported in 2015 that one-third of the medicines with a new active substance recommended for market access were for cancer treatment (19). As mentioned earlier, REAs of drugs are traditionally based on health outcomes such as overall survival and progression-free survival. However, in light of often marginal differences in overall survival and progression-free survival for oncological drugs, information on AEs, adherence and quality of life are becoming even more important in REA (20). Collecting these aspects from RCTs can be difficult, therefore other data sources such as social media may be useful. For the purposes of this explorative review, social media were defined as 'a group of Internet-based applications that allow the creation and exchange of user-generated content' (21).

METHODS

An explorative review was performed based on the PRISMA guidelines (22). To identify scientific literature, a search for peer-reviewed published articles was carried out in MEDLINE via the PubMed interface for the period between 1 January 2010 and 28 June 2016 on June 28th 2016. The following search query was used: *(Facebook(tiab) OR Twitter(tiab) OR blog(tiab) OR blogging(mesh) OR "social media"(tiab) OR ehealth(tiab) OR e-health(tiab) OR "online community"(tiab) OR "online communities"(tiab) OR "online patient"(tiab) OR "health data"(tiab) OR (online (tiab) AND research(tiab) AND platform*(tiab)) OR (personal*(tiab) AND health(tiab) AND record*(tiab)) OR (online(tiab) AND patient(tiab) AND communit*(tiab)) OR (online(tiab) AND data(tiab) AND shar*(tiab))) AND (oncolog*(tiab) OR cancer(tiab) OR carcinoma(tiab) OR metast*(tiab) OR neoplasms(mesh) OR melanoma(tiab) OR tumor(tiab) OR tumour(tiab))*. The reference lists from the literature, which were included based on title and abstract, were hand-searched to identify additional literature.

A Google search was conducted in July and August 2016 to identify grey literature, such as relevant websites, by combining the following keywords: 'social media, online patient, online research platform, relative effectiveness, health research, effectiveness research, pharmacovigilance, adherence, and/or to measure quality of life'. Before each search, the history of the browser was cleared to ensure findings would not be influenced by previous search queries. Due to the vast amount of websites retrieved through the Google

search, only websites that collect health data online, focus on patient-reported outcomes, or provide online information on drugs and conditions were deemed relevant for further analysis. The selection of relevant websites was also based on consensus between the authors RK and RtH. These websites were hand-searched to identify grey literature by browsing through the website in search of relevant reports or documents and by using the following keywords: 'social media, internet, Facebook, Twitter, pharmacovigilance and/or health research'. These keywords were different from those used for the Google search due to the character of the platform (i.e. a Google search is inherently different from searching a website). The following websites were included: PatientsLikeMe, Microsoft HealthVault, Dossio, CureTogether, WhatNext, MyGly, Drug Information Association, WEB-RADR, National Patient-Centered Clinical Research Network, College ter Beoordeling van Geneesmiddelen, Handle My Health, European Alliance for Personalised Medicine, Lareb, WHO Monitoring Centre for Pharmacovigilance Uppsala, PEW Research Center, Social Media Research Foundation, Treato, MediGuard, Healthy.me, and iVitality.

The review was conducted by four reviewers (RK, AM, RtH and KM) and the resulting literature was independently screened by the reviewers for eligibility. The titles and abstracts from scientific literature were assessed by RK, AM and KM, while grey literature was assessed by RK and RtH . Literature was considered eligible for inclusion when it was: 1) published between 1 January 2010 and 28 June 2016, 2) available in English, 3) examples were provided where social media were used to collect health data, 4) literature focused on cancer or cancer treatment, and 5) literature was either a peer-reviewed original research article or a report that was available in the public domain. We excluded literature that did not meet all inclusion criteria. Relevant full articles and reports were retrieved and reviewed for inclusion.

Two reviewers (RK and AM) independently extracted data from all included articles and reports using a predefined data abstraction form. Information on study characteristics (e.g. study design, study period, type of social media used), and the strengths, limitations and acceptability of using social media to generate health data were extracted. Disagreements in data extracted were resolved by consensus amongst RK and AM.

A descriptive qualitative analysis of the extracted data was carried out, since the topics, methods and outcomes of included literature were notably diverse.

RESULTS

A total of 1032 citations were identified from scientific literature (n=879), a hand search of reference lists from scientific literature (n=56), and grey literature (n=97). From these, a total of 988 citations were excluded based on title or abstract, additionally 9 duplicates were excluded. Of the 35 full scientific publications and documents assessed, 27 were excluded: 15 citations did not provide an example of health data collection, 9 were not oncology-specific, and 3 provided insufficient information on the collection of health data. Data were abstracted from a total of 8 scientific publications (Figure 1).

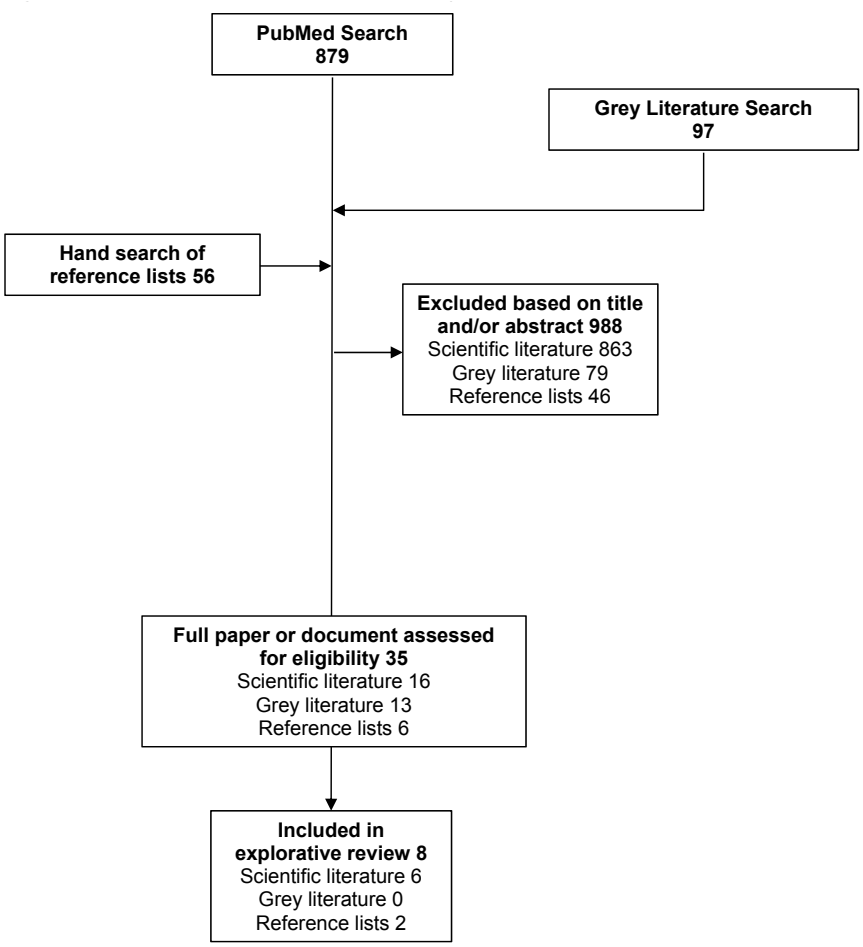


Figure 1 - Flowchart of the literature review process

Table 1 provides an overview of the 8 scientific publications included. Different types of cancer and medications were assessed in each of the publications. The main focus of all 8 articles was testing the feasibility and added value of generating health data from social media, such as AEs, QoL, adherence, symptom occurrence and experience from social media.

Table 2 shows that publications differed substantially in study design, study period, the number of posts analysed and the number of respondents included in the analysis. Forum topics and discussions were assessed in 4 papers, in 2 studies a survey was posted on the Facebook page of either a patient community or support group, in 1 study Twitter conversations were assessed and in 1 study an online patient platform was used to disseminate a survey. Of the 8 studies, a total of 4 studies collected health data on AEs (23,24,27,29). Another 3 studies collected health data on quality of life (QoL) (25,26,30).

Table 1 - Overview of included scientific publications.

Study	Aim	Cancer Type	Drug
Beusterien 2013 (23)	To better understand patient experience with CCC in the real world setting	Colorectal cancer	Chemo-therapeutic agents
Freifeld 2014 (24)	To evaluate the level of concordance between Twitter posts mentioning AE-like reactions and spontaneous reports received by a regulatory agency	N/A	Methotrexate[a]
van der Heijden 2016 (25)	To investigate whether we could use crowdsourcing via Facebook and online surveys for medical research purposes on PVNS	Pigmented villonodular synovitis	N/A
McCarrier 2016 (26)	To explore the feasibility of using social media-based patient networks to gather qualitative data on patient-reported outcome concepts relevant to CLL	Chronic lymphocytic leukaemia	N/A
Mao 2013 (27)	To understand frequency and content of AE's and associated adherence behaviours discussed by breast cancer patients related to using AI	Breast Cancer	Aromatase inhibitors
Marshall 2015 (28)	To identify and examine symptom patterns generated by data extracted from a breast cancer forum, and compare these findings to an analysis of symptoms reported by breast cancer survivors enrolled in a research study and who responded to a symptom checklist	Breast Cancer	N/A
Pages 2014 (29)	To describe the characteristics of AE's reported by patients exposed to OAN agents in an online discussion, and compare these with those reported by health professionals as recorded in the French pharmacovigilance database	Cancer	OAN agents
Zaid 2014 (30)	To determine the feasibility of using social media to perform cross-sectional epidemiologic and QoL research on patients with rare gynaecologic tumours	Neuroendocrine carcinoma of the cervix	N/A

CCC, colorectal cancer chemotherapies ; AE's, adverse events; N/A, not applicable; PVNS, pigmented villonodular synovitis; CLL, chronic lymphocytic leukaemia; AI, aromatase inhibitors; OAN, oral antineoplastic; QoL, quality of life.

[a] This study assessed adverse events reported in social media for a total of 23 drugs and 4 vaccines, including 1 drug (methotrexate) specific for oncology.

Each study used different QoL instruments, such as the Concerns About Recurrence Scale scores (30), and short form-36 health survey (25). Finally, 1 study focused on identifying symptom (co-) occurrence (28). In addition to the aforementioned main outcome measures, van der Heijden et al., McCarrier et al., and Zaid et al. (25,26,30) collected data on socio-demographic factors and disease specific characteristics. Furthermore, Beusterien et al. collected health data on physical functioning and emotional impacts (23), and Mao et al. collected information on adherence by mapping decisions about continuing or stopping treatment (27).

Table 2 - Study characteristics of included scientific publications that use social media to collect health data.

Study	Study design	Study period	Posts analysed	Number of respondents	Type of social media used to collect health data	Type of health data collected
Beusterien 2013 (23)	Cross-sectional	52 days	1522 posts	264	2 disease specific forums	AE's, physical functioning & emotional impacts
Freifeld 2014 (24)	Retrospective	7 months	6,900,000 posts	N/A	Twitter	AE's
van der Heijden 2016 (25)	Prospective	70 months	N/A	272	Facebook (patient community)	Socio-demographic factors, disease-specific characteristics[a], functional outcome, and QoL
McCarrier 2016 (26)	Cross-sectional	4 months	N/A	50	Online patient platform	Socio-demographic factors, disease-specific characteristics[b], experience of symptoms, perceptions about treatment, and QoL
Mao 2013 (27)	Retrospective	8 years	1,235,400 posts	N/A	12 disease- specific forums	AE's and adherence
Marshall 2015 (28)	Retrospective	8 years	50,426 posts	12,991	1 disease- specific forum	Symptom occurrence, co-occurrence, and similarity index of 25 pre-selected symptoms.
Pages 2014 (29)	Retrospective	1 year	111 posts	66	5 health forums	AE's
Zaid 2014 (30)	Cross-sectional	30 days	N/A	57	Facebook (support group)	Socio-demographic factors, disease-specific characteristics[c], and QoL

AE's, adverse events; N/A, not applicable; QoL, quality of life.

[a] Disease-specific characteristics include clinical presentation, findings on imaging and biopsy material, type and localization of disease, surgical and adjuvant treatment, local recurrences, and post-operative complications.

[b] Disease-specific characteristics include self-reported current CLL stage, performance status, and past and current treatment.

[c] Disease-specific characteristics include clinical presentation, initial work-up, treatments, past and current disease status, follow-up, and recurrence pattern.

The four publications that used forums to collect health data varied substantially in the explanation for their forum selection (Table 3). For example, Beusterien et al. used two search engines and two different computers for their forum search which they repeated every other day for two weeks. Additionally, they used selection criteria to include the 2 forums (i.e. site active >5 years, >12,000 posts on forum, >20 individuals currently browsing, and >10 new posts per day) (23). Meanwhile, Marshall et al. selected one forum without clarifying selection criteria for the selected forum (28). The other four publications, making use of Twitter, Facebook or an online patient platform, selected this social media platform due to the access of a large volume of health data (24) or access to a patient community (25,26,30). Regarding the use of automated processes to collect health data from social media, two publications specifically indicated to have used a web crawler (27,28) and one publication made use of the Twitter application programming interface (24). Two of the included publications indicated to have collected all the forum posts related to search terms without specifically indicating the collection method used (23,29) and three publications used the social media platform to distribute a survey (25,26,30). Automated techniques were used by Freifeld et al., Mao et al. and Marshall et al. to analyse the health data collected (24,27,28). For example, a Natural Language Processing (NLP) semi-automated classifier was used to identify AEs (24), or data mining algorithms were used to identify symptoms (28). The remaining five publications made use of content analysis (23,26), descriptive and/or quantitative analysis (e.g. chi-squared test) (25,30), or labelled forum posts manually (29).

In Table 4 the strengths and limitations of health data generated through social media that were identified in the 8 included publications are presented. Five publications identified the ability to assess patient perspectives as an important strength (23,24,27-29). The ability to access patients who have rare diseases and/or are distributed over wide geographic areas was considered a major strength by 5 publications (25-28,30). Furthermore, Freifeld et al., Marshall et al. and Pages et al. emphasized that social media should complement conventional (pharmacovigilance) methods, since a difference between results from social media and conventional methods may be present (24,28,29). For example, patients were shown to report different AEs compared to health professionals who traditionally provide this information (29). Other strengths identified included the efficient collection of patient-reported outcomes (23), the short time-period needed to survey patients (28,30), and the identification of new or unlabelled AEs (29).

Limitations of social media-generated health data mainly focused on validating authenticity, selection bias, information bias, and the inability to actively probe patients for responses. Validating authenticity focuses on the difficulty of verifying the accuracy of information provided via social media (25,28), such as verifying whether posters actually have the disease (26,30) or are indeed on the drugs (23,26) they discuss. Regarding selection bias, publications reported differences in the patient population that use social media compared to those who do not; for example, patients using social media are conventionally more highly educated (23, 28), are more likely to be female (25,26), may have a different

Table 3 - Selection of social media platform and use of automated techniques by included literature that use social media to collect health data.

Study	Clear explanation for selection of social media platform	Web crawler used for collecting social media health data	Automated technique used for analysis of health data
Beusterien 2013 (23)	Yes	No	No
Freifeld 2014 (24)	Yes	No[a]	Yes
van der Heijden 2016 (25)	Yes	No[b]	No
McCarrier 2016 (26)	Yes	No[b]	No
Mao 2013 (27)	Yes	Yes	Yes
Marshall 2015 (28)	No	Yes	Yes
Pages 2014 (29)	Yes	No	No
Zaid 2014 (30)	Yes	No[b]	No

[a] The Twitter application programming interface (API) was used to identify relevant tweets.
[b] A survey was distributed via the social media platform.

symptom experience (27), and are generally younger (26,28,30). With regards to information bias, Freifeld et al. and Pages et al. reported duplication of posts (24, 29), Mao et al. reported multiple posts by the same patients (27), and Freifeld et al. indicated that patients may not identify AEs correctly (24). Finally, several publications mentioned the inability of using social media to actively probe patients for responses (23, 26, 28). For example patients may use alternative wording than that which researchers anticipate, which could lead to misclassifying symptom experiences (28).

With regards to the acceptability of using social media to generate health data, Pages et al. indicated that pharmaceutical companies are already using this type of data to gather information on AEs from patient perspectives (29). Furthermore, Beusterien et al. indicated that in patient-reported outcomes research, patient perspectives are commonly accepted with regards to disease and treatment impact (23), and both Freifeld et al. and van der Heijden et al. noted the importance of insights into the patient perspective provided by social media research for regulatory authorities (24,25). However, Freifeld et al. was also cautious on the use of social media to generate health data (24). Reasons for their caution was the need to still establish its role in pharmacovigilance as social media are not yet used in routine surveillance. In addition, they indicated that data acquisition from social media and automation thereof need to be improved.

Table 4 - Strengths and limitations specific to the use of social media to generate health data

Study	Strengths	Limitations
Beusterien 2013 (23)	Patient perspective; Efficient and comprehensive collection of PROMS;	Validating authenticity: Selection bias; No active probing of patient responses; Incomplete information of sample;
Freifeld 2014 (24)	Patient perspective; Complementary to pharmacovigilance; Rapid information on AEs;	Information bias; Volume of posts; Noisy data;
van der Heijden 2016 (25)	Access to patients with rare diseases; Collection of PROMS; Convenient to fill in; Long-term follow-up	Validating authenticity; Selection bias; Low participation rate;
McCarrier 2016 (26)	Alternative approaches to qualitative data collection; Support development of PRO instruments; Access to patients with rare diseases; Motivated patients; Lower costs per enrolled patient	Validating authenticity; Selection bias; No active probing of patient responses; Not achieving concept saturation; Larger sample sizes needed;
Mao 2013 (27)	Patient perspective; Access to patients distributed over wide geographic areas; Increased generalizability due to more diverse patient population; Observed frequency key AEs reflected those reported in traditional studies	Selection bias; Information bias; Frequency data is not an indication of prevalence AEs;
Marshall 2016 (28)	Vast quantities of data; Easily accessible information; Short time-period; Access to patients with rare diseases; Low costs; Patient perspective; Complementary to traditional studies	Validating authenticity; Selection bias; Noisy data; No active probing of patient responses; Incomplete information of sample; Data quality or format inadequate; Ethical considerations; Misinterpretation of posts;
Pages 2014 (29)	Patient perspective; Complementary to pharmacovigilance; Identification new/unlabelled AEs	Information bias;
Zaid 2014 (30)	Access to patients with rare diseases and that are distributed over wide geographic areas; Short time-period; Motivated patients;	Validating authenticity; Selection bias;

PROMS, patient-reported outcome measures; AEs, adverse events; PRO, patient-reported outcomes

DISCUSSION

This explorative review demonstrates that, within the field of oncology, social media could be used for assessing AEs by collecting health data from forums and to evaluate QoL via Facebook or online patient platforms. Social media provides an opportunity to efficiently assess patient perspectives and collect health data from patients with rare diseases that are distributed over wide geographic areas. However, validating the authenticity of health data from social media is difficult, and is prone to selection and information bias. Furthermore, this type of data should be used complementary to traditional forms of research.

Arguably, the results found in this review on social media-generated data in oncology may not be generalizable to other fields of medicine. However, many studies conducted in fields of medicine other than oncology similarly focused on identifying AEs (31-37), suggesting our results are at least partially generalizable. Although little is known about assessing QoL via social media in other fields of medicine, there is potential for this mode of health data collection since QoL is often difficult to measure in RCTs and observational studies (20). Finally, as our results show, another aspect of relative effectiveness that may be assessed via social media is treatment-switching and adherence behaviour. A few pharmaceutical companies have been assessing this aspect already, thus demonstrating its potential (14,15,38). Given the possibility of social media to generate data on AEs, QoL, and treatment-switching and adherence behaviour, there is a great potential for social media-generated health data to enrich REA by incorporating information on these aspects.

One caveat of using social media to collect health data that requires special attention is the lack of clear methodological guidance. Standardized approaches to collecting health data from social media are necessary to ensure comparability and reproducibility between studies. For example, posts may either be extracted manually or by automated processes. The interpretation of these posts could also be done manually or by automated processes. However, some argue that automated processes may be unable to successfully interpret sarcasm in text posted on social media (24), while others argue that automated natural language processing could assist in analysing the vast amounts of data available on social media (34,39,40). Another methodological issue involves the use of correct search terms, as posts may include misspellings, non-medical terms, and slang (24,34,41). Additionally, several studies reported important methodological limitations to consider when assessing data from social media, which include validating authenticity (e.g. posts may be not genuine) (42-44), selection bias (e.g. social media users may differ in age, gender, ethnicity and physical location compared to non-users) (41,43,44) and information bias (e.g. patients may be taking a specific drug but fail to report the drug or its effects) (42, 44). To manage these methodological limitations it is important to systematically assess the risk of bias in order to determine the quality of the health data collected via social media. Extracting relevant health data from social media may be difficult and challenging due to the issues described above. Clear and uniform methodological guidance may improve the extraction, interpretation and subsequent use of social media to collect health data.

An additional caveat that may hamper the use of social media for collecting health data for REA is the perceived risk of easy manipulation. A recent example of manipulation in social media was the circulation of fake news on social media during the 2016 elections in the United States of America (45-47). These kind of examples affects the ability of social media users to discern what is true and correct information. However, although manipulation may occur, many still use social media to find information and to exchange experiences. Therefore, harnessing and analysing the vast amount of health data available on social media remains important.

Although caveats can be recognized in the use of social media-generated health data, the added value of collecting information on patients' perspectives and experiences towards relative effectiveness (e.g. AEs, quality of life, switching-behaviour) should be highlighted. For example, health data collected via social media may uncover AEs that occur after long-term use of new drugs, or they may detect AEs earlier compared to traditional methods (43,48), or provide insights that are not available in published literature (e.g. diabetes patient experiences with laser therapy) (12). Additionally, social media may be a better source to identify AEs that are mild or symptom-related compared to more traditional methods (43). However, health data collected via social media should be used in conjunction with traditional methods to ensure the collection of a comprehensive overview of aspects than can provide information for REA.

Important for the comprehensiveness of this review is that we assessed both academic and grey literature, which minimizes the possibility of missing important insights. Additionally, we ensured the quality of the review through data abstraction conducted by two authors, which allowed a better substantiation of deductions made.

One limitation of this review was the focus on oncology, which may have resulted in missing literature on other aspects related to REA that could potentially be collected using social media. For example, PatientsLikeMe, an online patient platform that allows patients to share health data and/or exchange experiences on conditions and medications, published a few studies on the effectiveness of off-label drug use(42,49). Additionally, PatientsLikeMe published a study focused on assessing the impact of menopause on disease severity in patients with multiple sclerosis.(50) These types of data may contribute to providing information for REA. The focus on oncology in this review was deemed appropriate since many new drugs are developed in the field of oncology, studies that assess these new drugs can be small and incomplete, and the European Medicines Agency and the European network for Health Technology Assessment are also putting focus on the assessment of oncological drugs.

A second limitation relates to the search strategy employed in this explorative review. Firstly, the broad definition of social media that was used in this review may not allow for differentiating between passively collecting data (e.g. by collecting posts from a forum) and actively collecting data (e.g. by posting a survey on Facebook). There may be a difference in the information available from passively collecting information that patients discuss and

post on social media, compared to actively posing questions to these patients in a survey. Secondly, by employing one database for our scientific and grey literature search we may have missed studies published in relevant journals that are not indexed by PubMed or grey literature that was not identified by the Google search engine. To overcome this limitation to some extent, we hand-searched the reference lists of included studies, based on title and abstract, and identified a few articles that had not been captured in the PubMed and Google search.

CONCLUSION

Social media may be a potential source of RWD for REA, particularly on aspects such as AEs, occurrence of disease-specific symptoms, adherence behaviour, and QoL. This potential has not yet been fully realised due to methodological limitations that accompany social media-generated health data, such as information bias and selection bias, as well as the limited acceptability of such data. However, the degree of usefulness of such data for relative effectiveness should be further explored. Moreover, methodological guidelines and tools should be developed to address the limitations mentioned above.

REFERENCE LIST

1. Makady A, Goettsch W, on behalf of GetReal Workpackage 1. IMI-GetReal: Glossary of definitions of common terms. 2015.
2. HTAi. HTA Glossary: Health Technology. http://htaglossary.net/health+technology. Archived at: http://www.webcitation.org/6uV8kwKnY
3. Antheunis M, Tates K, Nieboer T. Patients' and health professionals' use of social media in health care: Motives, barriers and expectations. Patient Educ Couns 2013;92(3):426-31. PMID:23899831
4. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients Tweet? Examining the Twitter use of cancer patients in Japan. J Med Internet Res 2014;16(5):e137. PMID: 24867458
5. Pew Research Center. 2013. The diagnosis difference. http://www.pewinternet.org/2013/11/26/the-diagnosis-difference/. Archived at: http://www.webcitation.org/6q0wYe4XI
6. Kapp J, Peters C, Oliver D. Research recruitment using facebook advertising: Big potential, big challenges. J Cancer Educ 2013;28(1):134-7. PMID:23292877
7. Khatri C, Chapman S, Glasbey J, Kelly M, Nepogodiev D, Bhangu A, Fitzgerald JE, on behalf of the STARSurg Committee. Social media and internet driven study recruitment: Evaluating a new model for promoting collaborator engagement and participation. PLoS ONE 2015;10(3). PMID:25775005
8. Lane T, Armin J, Gordon J. Online recruitment methods for web-based and mobile health studies: A review of the literature. J Med Internet Res 2015;17(7):e183. PMID:26202991
9. Cavallo D, Chou W, McQueen A, Ramirez A, Riley W. Cancer prevention and control interventions using social media: User-generated approaches. Cancer Epidemiol Biomarkers Prev 2014;23(9):1953-6. PMID:25103820
10. Laranjo L, Arguel A, Neves A, Gallagher AM, Kaplan R, Mortimer N, Mendes GA, Lau AYS. The influence of social networking sites on health behavior change: A systematic review and meta-analysis. J Am Med Inform Assoc 2015;22(1):243-56. PMID:25005606
11. Dizon D, Graham D, Thompson M, Johnson LJ, Johnston C, Fisch MJ. Practical guidance: The use of social media in oncology practice. J Oncol Pract 2012;8(5):e114-24. PMID:23277774
12. Street JM, Braunack-Mayer AJ, Facey K, Ashcroft RE, Hiller JE. Virtual community consultation? Using the literature and weblogs to link community perspectives and health technology assessment. Health Expectations 2008;11(2):189-200. PMID: 18430153
13. Powell G, Seifert H, Reblin T, Burstein PJ, Blowers J, Menius JA, Painter JL, Thomas M, Pierce CE, Rodriguez HW, Brownstein JS, Freifeld CC, Bell HG, Dasgupta N. Social media listening for routine post-marketing safety surveillance. Drug Saf 2016;39(5):443-54. PMID:26798054
14. Sukkar E. 2015. Searching social networks to detect adverse reactions. The Pharmaceutical Journal. http://www.pharmaceutical-journal.com/news-and-analysis/features/searching-social-networks-to-detect-adverse-reactions/20067624.article . Archived at: http://www.webcitation.org/6q0wwYsjE
15. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: Application of social media content analysis to health outcomes research. J Med Internet Res 2016;18(3):e62. PMID:26987964
16. ABPI Pharmacovigilance Expert Network. Guidance notes on the management of adverse events and product complaints from digital media. 2013.
17. PatientsLikeMe. 2015. PatientsLikeMe and the FDA Sign Research Collaboration Agreement. http://blog.patientslikeme.com/2015/06/15/patientslikeme-and-the-fda-sign-research-collaboration-agreement/. Archived at: http://www.webcitation.org/6q0x3uFXv

18. Heywood B. Advancing knowledge from patient experience. 2016.
19. European Medicines Agency. Annual Report 2015. 2016.
20. Kleijnen S, Lipska I, Alves T, Meijboom K, Elsada A, Vervölgyi V, d'Andon A, Timoney A, Leufkens HG, de Boer A, Goettsch WG. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Ann Oncol 2016;27(9):1768-75. PMID:27329251
21. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Business Horizons 2010;53(1):59-68. DOI:10.1016/j.bushor.2009.09.003
22. Moher D, Liberati A, Tetzlaff J, Altman D, the PRISMA Statement. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. J Clin Epidemiol 2009;62(10):1006-12. PMID:20171303
23. Beusterien K, Tsay S, Gholizadeh S, Su Y. Real-world experience with colorectal cancer chemotherapies: patient web forum analysis. Ecancermedicalscience 2013;7:361. PMID:24143155
24. Freifeld C, Brownstein J, Menone C, Bao W, Filice R, Kass-Hout T, Dasgupta N. Digital drug safety surveillance: Monitoring pharmaceutical products in Twitter. Drug Saf 2014;37(5):343-50. PMID:24777653
25. van der Heijden L, Piner S, van de Sande M. Pigmented villonodular synovitis: A crowdsourcing study of two hundred and seventy two patients. Int Orthop 2016;40(20):2459-68. PMID:27169531
26. McCarrier K, Bull S, Fleming S, Simacek K, Wicks P, Cella D, Pierson R. Concept elicitation within patient-powered research networks: A feasibility study in chronic lymphocytic leukemia. Value Health 2016;19(1):42-52. PMID:26797235
27. Mao J, Chung A, Benton A, Hill S, Ungar L, Leonard CE, Hennessy S, Holmes JH. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol and Drug Saf 2013;22(3):256-62. PMID:23322591
28. Marshall S, Yang C, Ping Q, Zhao M, Avis N, Ip E. Symptom clusters in women with breast cancer: An analysis of data from social media and a research study. Qual Life Res 2016;25(3):547-57. PMID:26476836
29. Pages A, Bondon-Guitton E, Montastruc J, Bagheri H. Undesirable effects related to oral antineoplastic drugs: Comparison between patients' internet narratives and a national pharmacovigilance database. Drug Saf 2014;37(8):629-637. PMID:25027671
30. Zaid T, Burzawa J, Basen-Engquist K, Bodurka DC, Ramondetta LM, Brown J, Frumovitz M. Use of social media to conduct a cross-sectional epidemiologic and qualtity of life survey of patients with neuroendocrine carcinoma of the cervix: A feasibility survey. Gynecol Oncol 2014;132(1):149-53. PMID:24145111
31. Abou Taam M, Rossard C, Cantaloube L, Bouscaren N, Roche G, Pochard L, Montastruc F, Herxheimert A, Montastruc JL, Bagheri H. Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator(®)) withdrawal in France. J Clin Pharm Ther 2014;39(1):53-5. PMID:24304185
32. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, Jaulent M, Beyens M, Burgun A, Bousquet C. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res 2015;17(7):e171. PMID:26163365
33. Liu X, Chen H. Identifying Adverse Drug Events from Patient Social Media: A Case Study for Diabetes. Intelligent Systems IEEE 2015;30(3):44-51. DOI:10.1109/MIS.2015.7
34. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015;54:202-12. PMID:25720841
35. Topaz M, Lai K, Dhopeshwarkar N, Seger DL, Sa'adon R, Goss F, Rozenblum R, Zhou L. Clinicians' reports in electronic health records versus patients' concerns in social media: A pilot study of adverse drug reactions of aspirin and atorvastatin. Drug Saf 2016;39(3):241-50. PMID:26715498

36. Wu H, Fang H, Stanhope S. Exploiting Online Discussions to Discover Unrecognized Drug Side Effects. Methods Inf Med 2013;52(2):152-159. PMID:23450374

37. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication informations on the Internet. J Med Internet Res 2011;13(3):e53. PMID:21807607

38. Petersen C. 2015. Concrete Examples of Social Listening's Impact on Pharma. https://www.pm360online.com/concrete-examples-of-social-listenings-impact-on-pharma/. Archived at: http://www.webcitation.org/6q8ERpEcP

39. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L. How noisy social media text, how diffrnt social media sources? Proceedings of the 6th International Joint Conference on Nautral Language Processing (IJCNLP), 2013, Nagoya, Japan.

40. Reyes A, Rosso P, Buscaldi D. From humor recognition to irony detection: The figurative language of social media. Data & Knowledge Engineering 2012; 74:1-12. doi: 10.1016/j.datak.2012.02.005

41. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. Br J Clin Pharmacol 2015;80(4):910-20. PMID:26147850

42. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: Analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6. PMID:21252034

43. Golder S, Norman G, Loke Y. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015;80(4):878-88. PMID:26271492

44. Leng H. Methodological issues in using data from social networking sites. Cyberpsychol Behav Soc Netw 2013;16(9):686-9. PMID:23679568

45. Allcott H, Gentzkow M. Social Media and Fake News in the 2016 Election. Stanford University. 2017.

46. Silverman C. 2016. This Analysis Shows How Viral Face Election News Stories Outperformed Real News On Facebook. https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.ymmRRxdMPk#.mgpLLo5NM9 . Archived at: http://www.webcitation.org/6q8Eg1VLZ

47. Perrott K. 2016. 'Fake new' on social media influences US election voters, experts say. http://www.abc.net.au/news/2016-11-14/fake-news-would-have-influenced-us-election-experts-say/8024660 . Archived at: http://www.webcitation.org/6q8F0ScwQ

48. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, van Le H, Freifeld CC, Brownstein JS, Walderhaug M, Edwards IR, Dasgupta N. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: An analysis of recent FDA safety alerts. Drug Saf 2017; 40(4):317-31. PMID: 28044249

49. Wicks P, Vaughan T, Massagli M, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotechnol 2011;29(5):411-4. PMID:21516084

50. Bove R, Healy B, Secor E, Vaughan T, Katic B, Chitnis T, Wicks P, de Jager PL. Patients report worse MS symptoms after menopause: Findings from an online cohort. Mult Scler Relat Disord 2015;4(1):18-24. PMID:25787049

CHAPTER 9

Social Media as a Tool for Assessing Patient Perspectives on Quality of Life in Metastatic Melanoma: a feasibility study

A. Makady[1,2*], R. Kalf[1,2*], B. Ryll[3,4], G. Spurrier[3], A. de Boer[2], H. Hillege[5], O. Klungel[2]; W. Goettsch[1,2] (on behalf of GetReal Work Package 1).

[1] The National Healthcare Institute (ZIN), Diemen, the Netherlands.
[2] Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, the Netherlands.
[3] Melanoma Patient Network Europe, Uppsala, Sweden
[4] Uppsala University, Uppsala, Sweden.
[5] Department of Epidemiology, University Medical Centre Groningen, Groningen, the Netherlands.

*R. Kalf and A. Makady contributed equally to this work

Currently under review in Health and Quality of Life Outcomes.

ABSTRACT

Background

Development of innovative drugs for melanoma is occurring rapidly. Incremental gains in overall survival amongst innovative products may be difficult to measure in clinical trials, and their use may be associated with increased toxicity profiles. Therefore, Health Technology Assessment (HTA) agencies increasingly require information on Health-Related Quality of Life (HRQoL) for the assessment of such drugs. This study explores the feasibility of social media use to assess patient perspectives on HRQoL in melanoma, and whether current cancer- and melanoma-specific HRQoL questionnaires represent these perspectives.

Methods

A survey was distributed on social media channels of Melanoma Patient Network Europe to assess melanoma patients' perspectives regarding HRQoL. Two researchers independently conducted content analysis to identify key themes, which were subsequently compared to questions from three HRQoL questionnaires (e.g. European Organisation for Research and Treatment of Cancer Module for Melanoma (EORTC-QLQ-MEL38) and Functional Assessment of Cancer Treatment - Melanoma (FACT-M)).

Results

In total, 72 patients and 17 carers completed the survey. Patients indicated that family, having a normal life, and enjoying life were the three most important aspects of HRQoL. Carers indicated that being capable, having manageable adverse events, and being pain-free were the three most important aspects of HRQoL for patients. Respondents seem to find some questions from HRQoL questionnaires relevant (e.g. 'Have you felt able to carry on with things as normal?') and others less relevant (e.g. 'Have you had swelling near your melanoma site?'). Additionally, wording may differ between patients and HRQoL questionnaires, whereby patients generally use a more positive tone.

Conclusions

Social media may provide a valuable tool in assessing patient perspectives regarding HRQoL. However, differences seem to emerge between patient and carer perspectives. Additionally, cancer- and melanoma-specific HRQoL questionnaires do not seem to correlate fully with patient perspectives.

INTRODUCTION

Health Technology Assessment (HTA) entails the systematic evaluation of the properties and effects of health technologies, addressing their direct and intended effects, as well as their indirect and unintended consequences with the aim of informing decision-making (1). Such HTA assessments can encompass several aspects of the implementation of health technologies in clinical practice such as their relative effectiveness, cost-effectiveness or appropriate use (1). Relative effectiveness is defined as the extent to which an intervention does more good than harm, when compared to one or more alternative interventions for achieving the desired results and when provided under the routine setting of health care practice (2).

Within the field of oncology, the development of innovative yet expensive therapeutic drugs is occurring at a rapid pace. An illustrative example can be provided in the field of metastatic melanoma whereby 7 drugs exhibiting numerous new mechanisms of action have gained market authorisation in the past 5 years (3;4). One positive consequence of the increased number of treatments has been the general prolongation of overall survival of metastatic melanoma patients (3;4). However, provided that incremental gains in overall survival associated with innovative products may be difficult to measure in the context of clinical trials, and the toxicity profiles associated with their use may be considerable, HTA agencies increasingly require information on Health-Related Quality of Life (HRQoL) experienced by patients during the prolonged periods of survival as a means to assess the added value of innovative products within relative effectiveness assessments (REA) (5-7).

Conventionally, HRQoL of patients is measured using validated questionnaires which can either be generic (e.g. EuroQol 5 Dimensions; EQ-5D) (8), disease-specific (e.g. European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ-30) or Functional Assessment of Cancer Treatment - Melanoma (FACT-M)) (9;10) or even include individualised measures (11). From an HTA perspective, HRQoL acquired using generic questionnaires can enable comparison of healthcare gains across different disease areas (e.g. oncology vs. chronic pulmonary diseases). Meanwhile, data acquired through disease-specific questionnaires can help to distinguish between HRQoL experienced at different stages of a particular disease (e.g. metastatic melanoma), thus possibly identifying medical need per disease stage. Therefore, HRQoL data can contribute to HTA in several ways, whether as primary or secondary health outcomes for relative effectiveness data or as sources for utility values used in cost-utility analyses of new oncologic treatments (5-7).

However, despite HTA guidance encouraging the collection of HRQoL data for HTA submissions, it is seldom included in submissions. Recent research has shown that HRQoL data features in only a third of HTA submissions for oncologic treatments in 6 different European jurisdictions and that its impact on decision-making is generally low due to a number of reasons, including its sheer scarcity (5). Moreover, it can be argued that available validated HRQoL questionnaires, whether generic or disease-specific, do not accurately

represent the needs of patients, possibly contributing to the generally low completion rates of such questionnaires by patients (12-14).

The Innovative Medicines Initiative (IMI)-GetReal initiative is a 3-year public-private partnership exploring the use of Real-World Evidence (RWE) in early drug development and drug assessment (15). Within Work Package 1 of the GetReal consortium, a series of case studies have been conducted exploring this aim. One example is the case study on metastatic melanoma whereby the potential use of social media as a new source of RWE for HTA was investigated within a pilot literature review (12). This research demonstrated the potential value of using social media to inform several parameters of HTA in oncology, including: adverse events (12;16-18), treatment adherence (17) and HRQoL (12;19-21).

Building upon results from this pilot review, this article aims to explore the use social media as a tool to gather melanoma patients´ perspectives on HRQoL. More specifically, this article will: assess the comparability of the melanoma patient population accessible via social media with the general melanoma population, evaluate what melanoma patients and carers perceive as important in relation to HRQoL and compare this to validated cancer-specific HRQoL questionnaires, and assess whether current melanoma-specific HRQoL questionnaires represent melanoma patients' perspectives on HRQoL. It is important to emphasize that this is a feasibility study, aiming to advance the science of using social media to gain insights on patients' perspectives on HRQoL, rather than conducting a robust quantitative analysis to answer pre-defined hypotheses based on data collected through social media.

METHODS

Members of Melanoma Patient Network Europe (MPNE)(22), an established patient network for melanoma patients and their carers, were approached via multiple social media channels of MPNE to anonymously complete a web-based survey. An announcement with a brief description of the survey goals and link to the survey was posted on the private MPNE Facebook group, MPNE LinkedIn group, and MPNE twitter account. Members of MPNE were also approached by sending a single e-mail to the MPNE mailing list and by posting the announcement and link to the survey on the website of MPNE. Respondents were eligible for inclusion in the study if they self-reported a diagnosis of melanoma on the online survey or reported to be carer of a melanoma patient.

The web-based survey included 25 items (see Appendix 2). Socio-demographic and clinical characteristics were collected, including gender, country of residence, age, educational level, years since melanoma diagnosis, stage of disease, and treatments received. Patient and carer perspectives on HRQoL were collected by 1) assessing which aspects are most important to patients or carers regarding melanoma patients' HRQoL and 2) assessing the relevance of questions from the EORTC-QLQ-C30 questionnaire to respondents. Survey questions on patient or carer perspectives on HRQoL included:

- What is HRQoL in melanoma for you?

- Name the 3 things that deteriorate your/ the melanoma patient's HRQoL today?
- Name the single thing that would improve your/ the melanoma patient's HRQoL right now?

The survey was conducted using Survey Monkey, and once a member clicked on the survey link it was presented on a separate screen. The survey was open for 30 days from January 8th 2016. Two reminders of the ongoing survey were posted on MPNE's private Facebook group, LinkedIn group, and Twitter account throughout the 30 day period. By completing the survey, respondents gave their informed consent for this study.

ANALYSIS

To assess the comparability and generalizability of our study population to the general melanoma population, we compared the socio-demographic variables (e.g. gender, age, and educational level) with values reported by Bay et al., Eriksson et al, and the EORTC reference values (23-25).

In order to evaluate what patients and carers regard important for HRQoL, two researchers independently performed inductive content analysis on the responses to the open-ended questions posed in the survey(26). Content analysis allows for the organisation and cataloguing of respondent's descriptions of key aspects regarding melanoma patient views on HRQoL. Assigned codes and the grouping of similar codes were reviewed by both researchers and any discrepancies in coding were resolved by consensus. A wordcloud was created on the basis of the total frequency with which generated codes were cited (either by patients or carers) as the most important aspect to the patients' HRQoL (Figure 1).

To assess the extent to which current cancer-specific HRQoL questionnaires represent melanoma patients' perspectives on HRQoL, respondents were asked about the relevance of questions from the EORTC-QLQ-C30. Furthermore, a qualitative analysis was performed based on key aspects identified during the content analysis to assess the relevance of questions in two melanoma-specific HRQoL questionnaires to respondents, namely the EORTC Module for Melanoma (EORTC-MEL-38) and FACT-M.

Subgroup analyses were performed for patients and carers separately and were stratified by stage when possible. All data were coded, stored and analyzed using R version 4.00.03.05 (27).

RESULTS

A total of 96 respondents filled in the web-based survey. Of these 70 indicated to be patients, 17 were carers of a melanoma patient, 2 indicated to be both patients and carers, and 7 did not report either and were therefore excluded from the analyses. The 2 respondents indicating to be both patients and carers were included in the analysis as patients only. All

analyses were stratified by stage for patients, however this was not possible for carers due to the small number of respondents.

Most respondents accessed the survey via Facebook (77%), Twitter was used to a lesser extent (2%). Some respondents accessed the survey via the MPNE website (9%) or the MPNE mailing list (1%). Finally, 11% of respondents indicated to have used other online channels, such as the Berlin Support Group, Melanoma Romania Association, and Dutch Melanoma Association Forum.

The socio-demographic characteristics of the study population are shown in Table 1. Respondents were mostly female (70%), between 35 and 64 (82%), were university graduates or higher (64%), and originated from the United Kingdom (50%). The paper of Bay et al. showed that approximately 50% of patients with melanoma in Denmark were female (25); when stratified it was shown that 74% of patients in our study population were female whereas 53% of carers in our study population were female. The distribution of age in our study population was similar for most stages of melanoma and between carers and patients, except for patients with stage III where half of the respondents indicated to be between 55 and 64. The EORTC reference values also showed that the distribution of age was similar between stage I & II and stage III & IV, and were comparable to the age distribution found in our study population (24). Compared to the educational level reported in the paper by Eriksson et al., where only 25% of melanoma patients in Sweden had a high education (e.g. a college degree or higher), our study sample was more highly educated with 64% having an university degree or higher (23).

Table 2 shows that patients who responded to the survey represented all stages of melanoma, and that most patients had stage IV melanoma (39%). Of the carers who responded to the survey, most cared for a patient who had stage IV melanoma (71%), while none of the carers indicated to care for a patient with stage II melanoma. More than 60% of the patients with stage II, III, and IV melanoma indicated to have been diagnosed more than 2 years ago, compared to 44% of patients with stage I. A total of 65% of the carers indicated to take care of a patient who had been diagnosed more than 2 years ago. Most patients with stage II, III and IV melanoma as well as the carers who responded to the survey indicated that the patient had been diagnosed with cutaneous melanoma. Cutaneous melanoma had been diagnosed in 44% of patients with stage I melanoma, while 50% had been diagnosed with ocular, uveal or choroidal melanoma, and 6% of these patients didn't know the type of melanoma they had been diagnosed with. Most patients with stage I, II, or III melanoma indicated to be unfamiliar with any mutations present in their tumour, compared to 46% of patients with stage IV melanoma and 53% of carers who indicated that a BRAF mutation had been found in the tumour. Surgery was the treatment most often received by patients according to 92% of patients and carers in this study.

The overall HRQoL values reported by our study population are comparable to that reported by the EORTC as reference values for melanoma patients (Table 3) (24). Both stratified and non-stratified overall HRQoL values were similar, indicating that the HRQoL in our study population is similar to that in the general melanoma population.

Table 1 - Socio-demographic characteristics of the study population

	Patients (n=72)					Reference Patient Population*				
	Stage I (n=18)	**Stage II (n=10)**	**Stage III (n=16)**	**Stage IV (n=28)**	**Carers (n=17)**		**Stage I**	**Stage II**	**Stage III**	**Stage IV**
Gender:										
Male	17%	30%	25%	32%	47%	Male	43%	52%	56%	58%
Female	83%	70%	75%	68%	53%	Female	57%	48%	44%	42%
Age:							**Stage I&II**	**Stage III&IV**		
18-24	-	-	-	4%	6%	<40	25%			
25-34	6%	10%	13%	4%	12%	40-49	23%	27%		
35-44	22%	20%	13%	29%	24%	50-59	24%	26%		
45-54	28%	30%	25%	36%	29%	60-69	22%	23%		
55-64	33%	30%	50%	18%	24%	70-79	5%	18%		
65-74	6%	10%	-	7%	6%	80+	1%	6%		
>75	6%	-	-	4%	-		0%	0%		
								0%		
Highest educational level:							**All Patients**	**Female**	**Male**	
Did not attend school	-	-	-	-	-	Low education†	36%	36%		
Finished school after primary school	-	-	-	-	-	Intermediate education	36%	39%	39%	
Graduated from secondary school	11%	20%	25%	7%	12%		39%	25%	25%	
Graduated from college	22%	10%	31%	21%	29%	High education	25%			
Graduated with university degree level	50%	60%	38%	50%	35%					
Higher degree or doctorate	17%	10%	6%	21%	24%					

Table 1 - continued

	Patients (n=72)					Reference Patient Population*			
	Stage I (n=18)	**Stage II (n=10)**	**Stage III (n=16)**	**Stage IV (n=28)**	**Carers (n=17)**	**Stage I**	**Stage II**	**Stage III**	**Stage IV**
Country of Residence:									
Belgium	-	-	-	11%	6%				
France	-	-	-	4%	6%				
Germany	17%	10%	13%	-	-				
Ireland	-	-	6%	11%	-				
Italy	-	-	6%	-	-				
Netherlands	-	-	6%	7%	24%				
Norway	6%	20%	13%	7%	-				
Other‡	17%	-	6%	4%	6%				
Romania	11%	10%	-	7%	18%				
UK	50%	60%	50%	50%	41%				

*The reference patient population was based on (25) for gender, (24) for age, and (23) for educational level; - no respondents ticked this answer (e.g. 0%); †In this reference it is given that in Sweden low education corresponds to mandatory school, intermediate to high school, and high to college/university; ‡5 respondents originated from the USA and 1 respondent from Serbia.

Table 2 - Clinical characteristics of the study population

	Patients (n=72)				
	Stage I (n=18)	**Stage II (n=10)+**	**Stage III (n=16)***	**Stage IV (n=28)**	**Carers (n=17)Ŧ**
Stage of Melanoma					
Stage 1	25%	NA	NA	NA	6%
Stage 2	NA	14%	NA	NA	-
Stage 3	NA	NA	22%	NA	24%
Stage 4	NA	NA	NA	39%	71%
Melanoma diagnosis:					
< 1 month ago	-	-	-	-	-
1-3 months ago	6%	-	-	-	-
3-6 months ago	17%	-	7%	4%	-
6-12 months ago	28%	10%	13%	11%	6%
1-2 years ago	6%	20%	7%	25%	29%
2-5 years ago	22%	50%	40%	43%	24%
> 5 years ago	22%	20%	33%	18%	41%
Type of Melanoma:					
Cutaneous melanoma	44%	70%	57%	64%	62%
Ocular/ Uveal/ Choroidal melanoma	50%	10%	7%	18%	12%
Acral melanoma	-	-	-	-	6%
Mucosal melanoma	-	-	-	-	-
I don't know	6%	20%	36%	18%	19%
Melanoma mutations:					
BRAF mutant	11%	20%	27%	46%	53%
BRAF wild-type	-	-	-	18%	18%
NRAS mutant	-	-	-	7%	-
c-kit mutant	-	-	-	4%	6%
GNAQ/GNA11	-	-	7%	4%	-
I don't know	78%	60%	67%	11%	18%
None	6%	10%	-	4%	6%
Other‡	6%	10%	-	7%	-
Treatments received					
Surgery	89%	90%	94%	89%	100%
Radiotherapy	39%	20%	13%	39%	26%
Chemotherapy	-	11%	6%	25%	21%
Immune Therapies	-	-	25%	81%	56%
Targeted Therapies	-	-	6%	27%	50%

+ The total number of respondents on treatments received (chemotherapy) was 9; * The total number of respondents on melanoma diagnosis is 15, on type of melanoma is 14, melanoma mutation is 15, and on treatments received is 15; † Carers provided disease specific characteristics for the patient they care(d) for; ‡Other melanoma mutations mentioned by 4 respondents were mutations in chromosome 3, 6 and/or 8; Ŧ The total number of respondents on type of melanoma is 16, on treatments received is 16; - no respondents ticked this answer (e.g. 0%); NA: Not Applicable

Figure 1 shows the most important aspects mentioned by patients and carers in our study sample that are important to the patients' HRQoL. 'Family' was the single most important aspect while the second most important aspect in HRQoL was 'Normal Life', implying that patients find it highly important to lead a normal life while being ill.

Table 3 - Overall quality of life in the study population compared to the EORTC reference value for the EORTC QLQ-C30

	Overall quality of life						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Respondents: Stage I & II (n=28)	0%	0%	7%	11%	21%	36%	25%
EORTC Reference Value: Stage I & II	0%	1%	6%	16%	28%	31%	20%
Respondents: Stage III & IV (n=44)	2%	2%	11%	11%	27%	25%	20%
EORTC Reference Value: Stage III & IV	2%	2%	6%	18%	30%	26%	18%
Respondents: All patients (n=72)	1%	1%	10%	11%	25%	29%	22%
EORTC Reference Value: All Stages	2%	2%	8%	17%	28%	27%	16%

Figure 1 - Key aspects patients find important in QoL

In Table 4 a more detailed analysis is shown whereby the top 10 most important aspects on HRQoL were stratified by disease stage for patients, and for patients and carers separately. It can be seen that patients themselves rated 'Family' as one of the most important aspects in their HRQoL, together with 'Good Care' by patients with stage I melanoma, 'Fear' by patients with stage II melanoma, 'Worry' by melanoma patients with stage III, and 'Good medicines' and 'Normal Life' by patients with stage IV melanoma. Carers indicate that 'Capability', 'No Adverse Events', and 'Pain free' were the most important aspects to patients' HRQoL. According to carers 'Family' was the second most important aspect in a patients' HRQoL, which showed that patients and carers may have a different perspective regarding what is most important in patients' HRQoL.

As part of the survey respondents were asked to rate the relevance of several questions originating from the EORTC-QLQ-C30 to HRQoL (see Table 5 and Appendix Table 1). It can be seen that in our study sample patients with a different disease stage found different questions relevant to their HRQoL, and that carers also seemed to rate the relevance of questions differently than patients. For example, the question in the EORTC-QLQ-C30 regarding 'Trouble doing strenuous activities' did not seem to be relevant (at all) or did not apply to the majority of patients with stage I and II melanoma, while approximately 50% of stage III and IV melanoma patients found this a relevant question. Another example

Table 4 - Top 10 aspects patients and carers indicate that are important in patients' HRQoL

Patients (n=72)				
Stage I (n=18)	**Stage II (n=10)**	**Stage III (n=16)**	**Stage IV (n=28)**	**Carers (n=17)**
Family*	Family*	Family*	Family*	Capability*
Good care*	Fear*	Worry*	Good medicines*	No AEs*
Finances†	Enjoy life	Normal life	Normal life*	Pain free*
Normal life†	Capability†	Therapy burden	Capability†	Drug effectiveness†
Support†	Good doctors†	Counselling†	Enjoy life†	Family†
Enjoy Life†	Good health†	Enjoy life†	Support†	Normal life†
Access to medicines‡	Normal life†	Good care‡	Good care	Access to medicines‡
Fear‡	Pain free†	Good doctors‡	Good health‡	Cure‡
Good doctors‡	Relapse†	Not to worry‡	Good information‡	Finances‡
Capability∓	Worry†	Pain free‡	Access to medicines∓	Good care‡
Friends∓			Friends∓	Good health‡
Good health∓			Pain free∓	Uncertainty‡
No anxiety∓				
Patient network∓				
Positive mood∓				
Work∓				

* The same number of respondents reported this aspect to be important in their HRQoL; † The same number of respondents reported this aspect to be important in their HRQoL; ‡ The same number of respondents reported this aspect to be important in their HRQoL; ∓ The same number of respondents reported this aspect to be important in their HRQoL

showed that the question 'Have you had pain?' was rated as not relevant or does not apply to the majority of stage II melanoma patients, while more than 50% of stage III melanoma patients rated this question as (very) relevant. Also 60% of carers rated this question as (very) relevant.

Table 6 indicates that some questions in the EORTC-QLQ-MEL38 and FACT-M were relevant to our study population, while other questions seemed less relevant. For example, one question in the EORTC-QLQ-MEL38 focused on patients being given enough time to think about the treatment options available. However, patients seemed to be more interested in discussing access to adequate and clear information on treatment options. Additionally, wording of questions posed in HRQoL questionnaires may differ from how patients interpret these questions. For example, questions regarding pain at the melanoma site, surgical site or headaches posed in the HRQoL questionnaires seemed to be aspects of pain that our study population did not focus on. Instead, respondents discussed pain in more general terms (e.g. future pains or experiencing pain). Additionally, while 14 of the 89 respondents discussed pain, 33 respondents focused more on being pain free as important for their HRQoL.

DISCUSSION

In this study, the feasibility of using social media as a means to collect patient and carer perspectives on HRQoL was explored. Within the 30 days during which the survey was posted, 89 full responses were received. The majority of respondents accessed the survey via Facebook, with a smaller proportion doing so via Twitter, the MPNE website or other online channels. Respondents resembled the general melanoma population in some aspects (e.g. melanoma stage distribution, average HRQoL) but not others (e.g. gender distribution, educational level, geographic spread).

Based on responses compiled, patients with different stages of melanoma and carers identify and rank aspects important to HRQoL differently. For example, for all melanoma patients "Family" was one of the number one most important aspect, whereas for carers this was cited as one of the second most important. Some aspects among patients were stage-specific, such as "Fear" and "Good medicines" for stages II and IV, respectively. Patients of different stages and carers also rated the relevance of questions posed in EORTC QLQ-C30 differently, such as questions relating to trouble doing strenuous activities or occurrence of pain. Moreover, qualitative analysis of responses received on questions in melanoma-specific EORTC QLQ-MEL38 and FACT-M questionnaires revealed that some were relevant and others less relevant to our study population. Examples of less relevant questions include those on time to ponder treatment options, as well as those on pain near the melanoma site.

Social media may provide a quick and time-efficient manner to assemble valuable data on patient perspectives on HRQoL. Responses can be gathered within a short period of time and from audiences not usually included in randomised controlled clinical trials (RCTs)

Table 5 - Relevance of questions from the EORTC QLQ-C30 questionnaire in our study population (a subsample).

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Trouble doing strenuous activities	Stage I (n=17)	18	24	6	12	12	29
	Stage II (n=10)	40	10	20	-	-	30
	Stage III (n=16)	19	12	-	25	31	12
	Stage IV (n=28)	18	7	18	29	18	11
	Carers (n=19)	11	5	21	11	32	21
Trouble taking a long walk	Stage I (n=17)	12	24	6	6	24	29
	Stage II (n=10)	40	30	-	-	-	30
	Stage III (n=16)	19	6	12	6	38	19
	Stage IV (n=28)	18	7	18	29	18	11
	Carers (n=19)	5	5	21	21	21	26
Trouble taking a short walk outside the house	Stage I (n=18)	28	22	-	-	17	33
	Stage II (n=10)	40	20	-	-	10	30
	Stage III (n=16)	25	19	6	19	12	19
	Stage IV (n=28)	18	25	14	14	7	21
	Carers (n=19)	17	17	-	17	22	28
Need to stay in bed or a chair during the day	Stage I (n=17)	18	24	12	6	-	41
	Stage II (n=10)	50	10	-	10	-	30
	Stage III (n=16)	25	12	6	31	6	19
	Stage IV (n=28)	25	21	11	14	11	18
	Carers (n=19)	11	21	11	11	21	26

Table 5 - continued

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Need help with eating, dressing, washing yourself or using the toilet	Stage I (n=17)	18	18	12	-	6	47
	Stage II (n=10)	70	-	-	-	-	30
	Stage III (n=16)	44	6	-	25	-	25
	Stage IV (n=28)	50	7	4	4	7	29
	Carers (n=19)	26	16	-	11	21	26
Limitations in doing either your work or other daily activities	Stage I (n=17)	18	18	6	18	18	24
	Stage II (n=10)	60	-	-	10	-	30
	Stage III (n=16)	19	6	12	31	25	6
	Stage IV (n=28)	18	7	7	25	32	11
	Carers (n=19)	11	-	26	21	26	16
Limitations in pursuing your hobbies or other leisure time activities	Stage I (n=17)	6	12	12	18	35	18
	Stage II (n=10)	50	10	10	10	-	20
	Stage III (n=15)	13	7	-	53	7	20
	Stage IV (n=27)	15	7	7	33	26	11
	Carers (n=19)	5	5	11	32	26	21
Short of breath	Stage I (n=17)	18	29	-	6	12	35
	Stage II (n=10)	40	10	10	-	-	40
	Stage III (n=16)	31	6	12	12	19	19
	Stage IV (n=28)	43	4	-	29	7	18
	Carers (n=19)	22	11	11	-	22	33
Pain	Stage I (n=16)	12	12	12	19	6	38
	Stage II (n=10)	50	10	-	10	-	30
	Stage III (n=16)	25	6	6	38	19	6
	Stage IV (n=28)	36	4	11	21	11	18
	Carers (n=19)	17	-	11	11	50	11

Table 6 - Examples of questions in current HRQoL questionnaires EORTC-QLQ-C30 and FACT-M compared to responses from our study population about what patients find important in HRQoL

Questionnaire	Question	Relevance to patient population	Difference in wording	Example of patient response
EORTC QLQ-MEL38	Have you felt able to carry on with things as normal?	Relevant	Wording similar	'Wish to continue life as before.' 'Ability to life my life as normal as possible.'
EORTC QLQ-MEL38	Have you felt confident that a psychological support service would be available if you needed it?	Relevant	Wording similar	'Care and mental support (professionals and personal network).' 'Piece of mind that help is just at the end of a phone.'
EORTC QLQ-MEL38	Have you received realistic and reliable information about the extent (spread) of your disease?	Relevant	Wording may differ	'More facts and less fantasy. I could need statistics and knowledge.' 'Not being treated like a passive idiotic patient but being informed according to my intellectual and emotional needs.'
EORTC QLQ-MEL38	Have you had problem with pain at or near your melanoma site?	Relevant	Wording may differ	'Worry and fears about future pain and mortality.' 'Being able to live without pains.'
EORTC QLQ-MEL38	Have you been given enough time to think about the treatment options available to you?	Less relevant	Wording may differ	'Having treatment options explained and discussed with me.' 'Up to date knowledge of available treatments.'
EORTC QLQ-MEL38	Have you had swelling near your melanoma site?	Less relevant	NA	NA
EORTC QLQ-MEL38	Have you felt able to accept that melanoma is a serious condition?	Less relevant	Wording may differ	'Understanding how hard it is to live with cancer (friends, relatives and work).' 'Doctors who don't take your worries seriously.'
FACT-M	I get emotional support from my family	Relevant	Wording similar	'Being surrounded by people who support you through every step of the treatment.' 'Family and friends support.'

Table 6 - continued

Questionnaire	Question	Relevance to patient population	Difference in wording	Example of patient response
FACT-M	I worry that my condition will get worse	Relevant	Wording similar	'Worry every time it I have to go for my liver scan.' 'To be free from the constant worry and stress about mets.'
FACT-M	I have a lack of energy	Relevant	Wording may differ	'Have the energy to play with my children not be impatient because of fatigue.' 'Being able to exercise fully.'
FACT-M	I am bothered by side effects of treatment	Relevant	Wording may differ	'I'm very anxious about potential side-effects from treatment.' 'Being able to control drug side-effects.'
FACT-M	I have good range of movement in my arm or leg	Less relevant	NA	NA

NA, Not Applicable (e.g. respondents did not discuss anything regarding this question).

(e.g. women or patients with early stages of melanoma)(28;29). The geographic spread of patients reached through social media is considerable, ranging in this study from the United States of America (U.S.A.), to Norway, Serbia and Romania. Employing social media to collect such data would require less resources than multi-centre trials or point-of-care studies and has been confirmed in previous research on cervical cancer(21). Similarly, the use of social media may be less resource-intensive and more efficient than similar data collection through the establishment and conduct of patient/citizen panels at various stages of HTA decision making. Moreover, data collection through social media allows patients the option to provide HRQoL information at their own pace and within a trusted environment of their own choice. In contrast, patients participating in patient/citizen panels may hesitate to share particular personal insights related to their HRQoL due to the formal setting in which panels are conducted or due to the dynamics of ongoing discussions. Such advantages of social media use may help increase the impact of HRQoL on REA of drugs by: increasing availability of HRQoL data for HTA, widening the scope of information from a broader patient group and increasing candidness of responses collected.

Findings from this study illustrated a difference between what patients and carers may regard as important aspects for HRQoL. Similar findings have been reported in previous research exploring responses of patients and carers to validated HRQoL questionnaires(30-32). Despite the efforts invested by stakeholders such as clinicians and scientists to develop HRQoL questionnaires, it can thus be argued that questionnaires may not be equally implementable across patients and carers. Moreover, differences on important aspects of HRQoL extended to patients' disease stage. Comparable findings in previous research have enticed discussions for the development of individualised HRQoL questionnaires(13;33;34). This raises the question of which form of HRQoL questionnaires HTA agencies should resort to within REA's. Moreover, it raises doubts as to whether current questionnaires are sufficient to distinguish between HRQoL of patients with different stages of melanoma. In fact, the incremental value of cancer- or melanoma-specific questionnaires for REA's may be questionable when compared to more general tools such as the EQ-5D, considering the fact that even they may be unable to distinguish between the HRQoL of patients with different disease stages. Provided that innovative, expensive drugs are targeted at stage III/IV patients (i.e. metastatic melanoma), as well as the marginal relative incremental gains in overall survival amongst innovative drugs and toxicity profiles associated with their use, it may therefore become necessary to develop separate stage-specific HRQoL questionnaires for patients and carers to better delineate HRQoL gains with new treatments in the future (35;36).

Meanwhile, findings on the varying relevance of questions posed in available cancer-specific or melanoma-specific questionnaires to patients may provide insights as to why completion rates for HRQoL questionnaires remain low, whether in the setting of RCTs or routine practice (5;37). Controversy regarding the relevance of questions posed in HRQoL in comparison to patient needs has been repeatedly cited in literature on several disease areas(13;14;38). A possible explanation for this phenomenon may be that HRQoL

questionnaires are conventionally developed with a physician- or scientific focus whereby the emphasis is set on aspects such as reliability, validity, and cross-cultural relevance, rather than a patient-centred approach which elicits thorough patient input at all stages of development(13;14;39). The subsequent irrelevance of certain questions, in combination with factors such as disease burden and practical difficulties associated with completing paper-based questionnaires, may result in patients feeling less inclined to provide responses. Consequently, a paucity of HRQoL data for purposes such as REA ensues. If developers of new HRQoL questionnaires would address abovementioned limitations of current questionnaires, it may thus be worthwhile to use insights provided by patients and carers through social media to ensure that the newly-developed questionnaires are deemed relevant to their personal perspectives, thereby encouraging them to complete such questionnaires.

In this study, social media was used as a medium for spreading a survey to collect patient perspectives on HRQoL. Although this approach allows for the collection of hypothesis-generated, structured data, some may argue that channels other than social media may be used to the same effect (e.g. distribution of questionnaires via mailing lists). Meanwhile, another potential approach for using social media pertains to the mining of data already available on different channels, such as patient forums, (micro-)blogs, or social networking services for data on HRQoL(16;17;40). In the latter approach, little intervention is applied by the researcher (e.g. through directed questions), implying that insights gained on HRQoL may better represent the "real world". On the other hand, it involves the mining of highly unstructured data which was initially not designed to answer specific research questions hence raising fundamental methodological challenges in relation to the collection, analysis and interpretation of such data for HTA. At the moment, little research exists on whether results generated by the two approaches for using social media correlate or differ, the reasons behind potential discrepancies in such results and thus the interpretation of such findings. Therefore, it remains unclear whether both approaches complement or contradict one another and subsequently how they may potentially lend themselves to HTA decision-making. This is a topic of ongoing research by the authors. Additionally, ongoing projects of IMI including Real-World Outcomes Across the AD Spectrum for Better Care (IMI-ROADMAP), and Healthcare Alliance for Resourceful Medicines Offensive Against Neoplasms in Hematology (IMI-HARMONY) are currently exploring the potential of using social media to gain insights on big data use for Alzheimer's disease and haematological malignancies, respectively (41;42).

Strengths

Three different social media channels were used to distribute the survey, representing two different forms of social media: Twitter (micro-blogs), Facebook and LinkedIn (social networking sites).

Open-ended questions regarding HRQoL were posed in the survey, allowing respondents the opportunity to express which aspects of HRQoL were important to them in

their own terms and length. This ensured that responses compiled were likely to represent the views of their writers accurately and comprehensively.

Inductive content analysis on free text in compiled responses was conducted independently by 2 researchers. This approach avoids limitations associated with computerised approaches such as missing misspelled words or misinterpreting slang and sarcasm. Moreover, all discrepancies related to the analysis were resolved by consensus amongst both researchers to ensure validity.

Limitations

The low number of survey respondents precluded quantitative analysis of differences between patients vs. carers and patients of different stages. However, provided that the aim of this study was to test the feasibility of employing social media to assess patient perspectives on HRQoL, the composition and constitution of respondents allowed for the identification of such differences in perspectives on HRQoL and generated hypotheses for future research.

The survey developed to assess perspectives on HRQoL was written in English. This was not the native language of a considerable number of the respondents in this study, which may have impacted their ability to adequately represent their thoughts on the issues raised. It may have also led to selection bias since English-speaking respondents were more likely to participate in the study. Initially, the authors aimed to translate the survey into languages spoken by all MPNE members yet this proved unfeasible within the project timelines given that members span at least 13 European countries.

Data collection was performed at a single time-point using one set of answers per respondent, thereby providing a cross-sectional analysis of melanoma patient perspectives on HRQoL. Although this information is valuable in the context of this feasibility study, HTA decision-making on the effectiveness of melanoma drugs in practice conventionally requires longitudinal data collection on HRQoL. Therefore, the current study does not shed light on potential attrition rates in questionnaire completion or the robustness of findings from longitudinal data collection through social media. Therefore, further studies may be required to assess the feasibility of long-term, longitudinal data collection through social media to inform decision-making.

Comparison of patient and carers´ perspectives on HRQoL was performed against three validated questionnaires: the cancer-specific EORTC-QLQ-C30 and melanoma-specific EORTC-MEL38 and FACT-M. Selection of the three instruments was agreed upon by consensus amongst the research team. Other generic HRQoL instruments exist which were not included, such as the SF-36 and EQ-5D questionnaires. Provided the relevance of such generic measures for REA of drugs, this may impact the relevance of findings for HTA. On the other hand, it may be argued that the relevance of such generic measures for the comparison made would have been predictably lower than for the selected disease-specific instruments.

CONCLUSIONS

Social media may provide a valuable tool to assess patient and carer perspectives on HRQoL, thus potentially increasing the availability and impact of HRQoL data in REA of drugs. However insights gleaned through social media are not easily generalizable to the broader melanoma patient population. Differences emerge between what patients of varying melanoma stages and carers consider important for HRQoL. Cancer- and melanoma- specific HRQoL questionnaires currently available do not seem to correlate fully with what patients view as important in HRQoL, particularly in relation to wording of issues. This raises the question of how information generated from current cancer- and melanoma-specific HRQoL questionnaires could be used for HTA decision-making and whether new, patient-centred, stage-specific instruments should be developed that better reflect patient perspectives on HRQoL.

Furthermore, current knowledge on the potential approaches for using social media to inform HTA decision-making is sparse. Although this study sheds light on the potential use of social media as a medium for gathering cross-sectional data on melanoma patient perspectives on HRQoL through questionnaires, future research should also aim to address the wide array of other potential uses, such as: the use of social media to collect longitudinal data on HRQoL, the use of data-mining approaches to glean insights on HRQoL from other channels (e.g. patient forums) and the methods for combining the potential value of the two different approaches for the use of social media (i.e. as a medium vs. data mining) for HTA decision-making. Additionally, since this was a feasibility study, a similar study on larger scale would allow for robust quantitative analysis of aspects that are important to the HRQoL of melanoma patients.

REFERENCE LIST

1. Health Technology Assessment International. HTAi Glossary. 2017 [cited Aug 2017]. Available from: http://htaglossary.net/tiki-index.php?page=List+all+terms.
2. HLPF. High Level Pharmaceutical Forum 2005-2008. Final Conclusions and Recommendations of the High Level Pharmaceutical Forum. 2017 [cited Mar 2015].
3. Matikas A, Mavroudis D. Beyond CTLA-4: novel immunotherapy strategies for metastatic melanoma. Future Oncology 2015;11(6):997-1009.
4. Valpione S, Campana LG. Immunotherapy for advanced melanoma: future directions. Immunotherapy 2016;8(2):199-209.
5. Kleijnen S, Lipska I, Alves TL, Meijboom K, Elsada A, Vervölgyi V, et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Ann Oncol 2016;mdw233.
6. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 17-11-2015. Zorginstituut Nederland.
7. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. National Institute for Health and Care Excellence. 3-7-2015.
8. EuroQoL. About EQ-5D. 2017 [cited Dec 2016]. Available from: http://www.euroqol.org/about-eq-5d.html.
9. EORTC. EORTC-QLQ-C30. 2017 [cited Dec 2016]. Available from: http://groups.eortc.be/qol/eortc-qlq-c30.
10. Cella D. FACT-M: Functional Assessment of Cancer Therapy - Melanoma. 2017 [cited Dec 2016]. Available from: http://www.facit.org/FACITOrg/Questionnaires.
11. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;324(7351):1417.
12. Makady A, Kalf R, Goettsch W., Lees M. D1.5 - Case Study on Metastatic Melanoma. 2017 [cited Dec 2016]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20WP1%20deliverables/Deliverable%201.5%20Report%20-%20metastatic%20melanoma%20-%20webversion.pdf.
13. Carr AJ, Higginson IJ. Are quality of life measures patient centred? Br Med J 2001;322(7298):1357.
14. Haywood KL, Wilson R, Staniszewska S, Salek S. Using PROMs in Healthcare: Who Should Be in the Driving Seat: Policy Makers, Health Professionals, Methodologists or Patients? 2016. Springer.
15. IMI-GetReal. IMI-GetReal: Overall objectives. 2017 [cited Dec 2016]. Available from: https://www.imi-getreal.eu/About-GetReal/Overall-objectives.
16. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014;37(5):343-50.
17. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013;22(3):256-62.
18. Pages A, Bondon-Guitton E, Montastruc JL, Bagheri H. Undesirable effects related to oral antineoplastic drugs: comparison between patientsGÇÖ internet narratives and a national pharmacovigilance database. Drug Saf 2014;37(8):629-37.
19. Heijden L, Piner SR, Sande MAJ. Pigmented villonodular synovitis: a crowdsourcing study of two hundred and seventy two patients. Int Orthop 2016;1-10.
20. McCarrier KP, Bull S, Fleming S, Simacek K, Wicks P, Cella D, et al. Concept Elicitation Within Patient-Powered Research Networks: A Feasibility Study in Chronic Lymphocytic Leukemia. Value Health 2016;19(1):42-52.
21. Zaid T, Burzawa J, Basen-Engquist K, Bodurka DC, Ramondetta LM, Brown J, et al. Use of social media to conduct a cross-sectional epidemiologic and quality of life

survey of patients with neuroendocrine carcinoma of the cervix: a feasibility study. Gynecol Oncol 2014;132(1):149-53.

22. MPNE. Melanoma Patient Network Europe. 2016 May 28 [cited Dec 2016]. Available from: http://www.melanomapatientnetworkeu.org/home.html.
23. Eriksson H, Lyth J, Mansson-Brahme E, Frohm-Nilsson M, Ingvar C, Lindholm C, et al. Low level of education is associated with later stage at diagnosis and reduced survival in cutaneous malignant melanoma: A nationwide population-based study in Sweden. Eur J Cancer 2013;49(12):2705-16.
24. Scott N, Fayers P, Aaronson N, Bottomley A, de Graeff A, Groenvold M, et al. EORTC QLQ-C30 Reference Values. EORTC 2008 July EORTC ;
25. Bay C, Kejs A, Storm H, Engholm G. Incidence and survival in patients with cutaneous melanoma by morphology, anatomical site and TNM stage: A Danish population-based register study 1989-2011. Cancer Epidemiology 2015;39(1):1-7.
26. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-88.
27. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015 R Foundation for Statistical Computing ; Available from: https://www.R-project.org/.
28. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLOS Clin Trial 2006;1(1):e9.
29. Holterheus C. Burden of Melanoma. Erasmus University Rotterdam; 2011.
30. Shearer J, Green C, Ritchie CW, Zajicek JP. Health state values for use in the economic evaluation of treatments for AlzheimerGÇÖs disease. Drugs & aging 2012;29(1):31-43.
31. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. J Clin Epidemiol 2002;55(11):1130-43.
32. Dirven L, Reijneveld JC, Aaronson NK, Bottomley A, Uitdehaag BM, Taphoorn MJ. Health-related quality of life in patients with brain tumors: limitations and additional outcome measures. Current neurology and neuroscience reports 2013;13(7):359.
33. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Qual Life Res 1999;8(1):79-91.
34. Waldron D, O'Boyle CA, Kearney M, Moriarty M, Carney D. Quality-of-life measurement in advanced cancer: assessing the individual. J Clin Oncol 1999;17(11):3603-11.
35. Fojo T, Mailankody S, Lo A. Unintended consequences of expensive cancer therapeuticsGÇöthe pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: the John Conley Lecture. JAMA OtolaryngologyGÇôHead & Neck Surgery 2014;140(12):1225-36.
36. Light DW, Lexchin J. Why do cancer drugs get such an easy ride. BMJ 2015;350:h2068.
37. Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trialsGÇöguidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. Eur J Cancer 1997;33(1):20-8.
38. Neudert C, Wasner M, Borasio GD. Patients' assessment of quality of life instruments: a randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. J Neurol Sci 2001;191(1):103-9.
39. Sprangers MAG, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer approach to quality of life assessment: guidelines for developing questionnaire modules. Qual Life Res 1993;2(4):287-95.
40. Beusterien K, Tsay S, Gholizadeh S, Su Y. Real-world experience with colorectal cancer chemotherapies: patient web forum analysis. Ecancermedicalscience 2013;7:361.
41. The Innovative Medicines Initiative. Big data, medicines safety and respiratory disease focus of €93 million IMI Call for proposals. 2017 [cited Dec 2016]. Available

from: http://www.imi.europa.eu/content/imi-2-call-6-launch.

42. ROADMAP. Real world Outcomes across the AD spectrum for better care: Multi-modal data Access Platform. 2017 [cited Dec 2016]. Available from: http://www.roadmap-alzheimer.org/index.html.

SECTION VI

General Discussion

CHAPTER 10

Real-World Evidence for Health Technology Assessment of Pharmaceuticals: Perspectives on the status quo and the future

RE-STATEMENT OF RESEARCH QUESTION

In this thesis, we set out to address the current gap of knowledge on the use of real-world evidence (RWE) in health technology assessment (HTA) and decision making for drugs. This gap pertained to several aspects, including: whether stakeholders have a unanimous understanding of what RWE is, whether RWE is used by HTA agencies, in which contexts RWE is used and the practical and cultural obstacles associated with the use of RWE in HTA and decision making. In the final chapter of this thesis, we will summarize and briefly discuss the implications of the main findings of the research conducted. Subsequently, we will highlight new initiatives in the field of RWE use in decision making and their relevance to the main findings of this thesis. Finally, we will discuss potential steps towards more optimal use of RWE in HTA and decision making.

MAIN FINDINGS & IMPLICATIONS

Diverging understanding of RWE and differing policies for RWE use among HTA agencies

In the first section, we explored the definitions available in current literature (both academic and non-academic) for the concepts of Real-World Data (RWD), as well as the definitions different stakeholders in the field of healthcare assign to this concept during interviews.

The results demonstrated a general lack of consensus of what RWD precisely is. Qualitatively, the definitions for RWD identified in literature or cited in interviews could be classified into different categories, each category imposing different requirements on what does or does not qualify as RWD. In a sub-analysis of HTA agencies, regulatory agencies and pharmaceutical industry, two-thirds of interviewees could not provide an official definition for RWD which is adopted on an institutional level. Therefore, it became clear that there is no consensus between stakeholders (including HTA agencies) on what RWD precisely is.

The implications of such findings are far-reaching, since they comprise the fundamental understanding of stakeholders regarding RWD, the sources of data that qualify as RWD and their use in decision making. As an example, we consider Pragmatic Clinical Trials (PCTs). In PCTs, patients are included based on less stringent inclusion/exclusion criteria than randomized controlled trials (RCTs) and are initially randomized to treatment arms. However, at a later point in time (after a pre-determined interim analysis) randomization and blinding of treatment allocation may be ended (1). Such PCTs involve elements whereby researchers intervene with inclusion/exclusion criteria and treatment allocation procedures yet not to the same extent as RCTs. For some stakeholders, PCTs would thus qualify as RWD (i.e. data generated in a non-RCT setting). Yet for others stakeholders, who believe that RWD is collected in a strictly non-experimental setting, PCTs would not. Therefore, a dialogue amongst stakeholders on the value of RWD for decision making is hindered by these underlying differences in the concept of RWD and its associated evidence sources.

Although we recognize that agreement by all stakeholders on a single definition of RWD may seem far-fetched, we strongly encourage consensus-seeking on this topic.

In recent years, several definitions have been developed by different initiatives like the International Society for Pharmacoeconomic and Outcomes Resarch (ISPOR) in 2007 (2) up to the Innovative Medicines Initiative (IMI)-GetReal consortium in 2016 (1). We believe that such definitions provide a good starting point for further dialogue amongst different stakeholders on this topic.

In the second section, we explored the current policies of 6 European HTA agencies regarding the RWE they accept in HTA submissions, as well their appraisal of RWE in HTA submissions.

The results demonstrated that the policies for the use and appraisal of RWE by HTA agencies in decision making hinges on two main factors; the context of use and subsequently the parameters for which RWE is used. In general, three distinct contexts could be sketched whereby RWE use is considered, namely: Initial Reimbursement Discussions (IRD), Pharmacoeconomic Analysis (PEA) and Conditional Reimbursement Schemes (CRS).

In all three contexts, RWE use for parameters of (relative) effectiveness of drugs was associated with some degree of controversy. Moreover, agencies differed in how acceptable they deemed the use of RWE for treatment effect estimates to compensate for the lack of evidence for RCTs (e.g. for rare diseases). In contrast to this, RWE use for parameters relating to the prevalence and incidence of the disease, treatment-related costs, resource use and adherence to treatment was largely accepted.

Inter-context variation in how RWE is used in HTA decision making may seem logical at first sight; ultimately, the research question at hand determines whether the evidence is relevant for the envisioned purpose. However, as is the case in a number of HTA agencies, separate assessors are usually assigned to assess evidence on relative effectiveness and cost-effectiveness. As a result, RWE submitted as part of an HTA submission may be assessed and appraised differently by the separate assessors. Another factor to this problem presents itself in agencies which offer the opportunity for CRSs, since the same assessors would be obliged to assess and appraise the RWE submitted differently depending on whether the dossier is submitted in the context of IRD or CRS. Meanwhile, inter-agency differences in policies on RWE accepted and RWE appraisal may present Marketing Authorization Holders (MAHs) with an array of challenging questions when developing strategies for evidence generation across the product life cycle (3-5).

In light of the aforementioned points, it may be useful for HTA agencies in Europe to align policies on RWE and provide guidance on practical aspects of RWE generation and analysis. This is especially important in light of the increasing trend of new (oncology or orphan) drugs granted conditional marketing authorization on the basis of limited phase II data or surrogate outcomes rather than confirmatory phase III RCT data (6-8). A harmonized set of policies on RWD use by HTA agencies would provide MAHs with the ability to plan evidence generation pathways that, next to RCTs, provide more data from real-world studies; the latter theoretically yielding outcomes more generalizable for HTA purposes (9-11). The European Network of HTA (EUnetHTA) may provide a platform for discussions

on aligning RWD policies. We will return to the efforts undertaken by EUnetHTA later in this chapter.

The fragmented reality of RWE use in HTA practice: standard HTA and Managed Entry Agreements (MEA)

In the third section, we explored how and in which contexts RWE is currently used in HTA practice. **In Chapter 4**, we examined the use of RWE in IRD (i.e. standard HTA) of innovative drugs for metastatic melanoma by 5 European HTA agencies.

In general, RWD was more often included in cost-effectiveness assessments (CEAs) than in relative effectiveness assessments (REAs) of HTA reports. The main reason for inclusion in REA was estimating prevalence and/or incidence of melanoma and in the setting of CEA for extrapolating long-term effectiveness of new drugs. If RWD was included in reports, clear statements regarding its appraisal were often not identified. When identified, critical appraisal outcome was mostly unknown or negative. Inclusion of RWD in REAs differed substantially between the 5 agencies; some citing RWD only for prevalence and/or incidence and others for drug effectiveness and safety.

These results raise the question why RWE plays a relatively minor role in HTA, especially for parameters relating to drug effectiveness. A possible reason could be the lack of RWE available at the time of initial HTA assessments. Since these assessments take place soon after regulatory approval of a drug, there might be insufficient time for MAHs to collect RWE through registries or observational studies. Another factor could be the lack of guidance on systematic approaches for the inclusion, analysis and interpretation of RWE for HTA purposes. In this context, HTA agencies have only recently begun collaborating on strengthening understanding of appropriate study designs for generating RWD and developing further analytic methods for synthesis of RWE from different sources through initiatives such as IMI-GetReal and EUnetHTA (12;13). Further dialogue amongst HTA agencies is necessary to ensure that the product of these ongoing collaborations will be deemed useful by decision-makers. We will return to this later in this chapter.

In Chapter 5, we explored the experiences gained in the implementation of Conditional Financing (CF) of expensive hospital drugs in the Netherlands, an example of Managed Entry Agreements (MEA) scheme.

The study highlighted a number of shortcomings related to the procedure for CF. An important example is the fact that the elapsed time between the initial HTA report (T=0) and final reassessment report (T=4) extended beyond 4 years for almost all drugs. This may be due to a myriad of factors, including the time needed to set up registries required for data collection, and subsequently assess the evidence generated (14;15). Moreover, one may wonder whether a 4-year period is applicable to all indications for which the CF drugs were approved. The use of tailored approaches for determining required time-frames to answer the questions raised at T=0, rather than a fixed 4-year window, may provide a more appropriate design for future MEAs.

The study also highlighted shortcomings in CF from a methodological standpoint. For example, it was apparent that despite the necessary feedback rounds from HTA committees on the proposed outcomes research, the RWE generated between T=0 and T=4 through outcomes research was deemed of insufficient quality to answer a third of questions. The often poor quality of RWE generated by outcomes research meant that much uncertainty remained unresolved regarding key reimbursement criteria, including cost-effectiveness. Literature alludes to numerous reasons for these findings, such as challenges with analyzing and interpreting RWE generated for decision making (16;17). Safeguards proposed in ZIN guidelines could have prevented such shortcomings in hindsight. Firstly, the conduct of value of information (VoI) analyses at T=0 to highlight the feasibility and intrinsic value of data collection for specific parameters within the timelines projected. Secondly the mid-term reporting of outcomes research progress and interim results between T=0 and T=4 could have led to more timely decisions regarding adjustment or termination of the CF procedure for drugs, thereby avoiding waste of valuable time and money for all stakeholders involved. Both design aspects may be essential for future design of MEAs, as has been iterated in previous literature (18).

In Chapter 6, we explored the opinions of different stakeholders on CF. The stakeholders provided their insights on aspects such as the functioning of CF, possible improvement points and the future need for CF as a policy tool.

Many of the insights provided by stakeholders confirmed those from the dossiers analyzed in Chapter 5. Notable new insights gained included the fact that the ever-changing reality of clinical practice often affected the validity and relevance of outcomes research planned at T=0. In combination with the low quality of RWE generated, the majority of stakeholders were skeptical of the value of RWE for eventual decision making at the time. Therefore, none of the stakeholders deemed CF successful in achieving all of its aims and only half indicated that its goals were partially met. Moreover, some considered it a "back alleyway" to allow the reimbursement of controversial drugs through the national reimbursement package.

On the other hand, all stakeholders alluded to positive aspects of CF. For example, it served as an important alternative to the binary decision system on reimbursement and generated valuable lessons for inter-stakeholder collaboration. Furthermore, CF boosted societal awareness of cost-effectiveness and budget impact of drugs in healthcare. Therefore, almost all of stakeholders unanimously agreed that CF should be reintroduced, albeit with the necessary improvements relating to procedure, governance, methodology and implementation of the scheme in practice.

The emergence of innovative, yet expensive medications is occurring rapidly. Moreover, novel oncology treatments increasingly gain marketing authorization based on less conclusive evidence on safety or efficacy (e.g. phase I/II studies) within the context of accelerated/conditional approval pathways (6). Consequently, HTA agencies increasingly encounter submissions with more uncertainties on aspects such as long-term health outcomes and effectiveness in clinical practice leading to increasing reliance on MEAs

by agencies worldwide (19;20). One may argue that without systematic evaluations of established MEAs, novel schemes are likely to suffer similar caveats as previous ones. To counter this potential risk, knowledge regarding the successes, failures, strengths and weaknesses of established MEAs in other jurisdictions should be the focus of future research, in order to avoid repeating historical mistakes when setting up new schemes within the Netherlands and elsewhere.

Accessibility of RWE and the negative consequences on changing longstanding paradigms

In the fourth section, we examined the obstacles associated with accessing and using individual patient data (IPD) from RWE repositories in comparative effectiveness research (CER; an equivalent term for REA).

IMI-GetReal's experiences in accessing IPD from RWE repositories were disparate. For half of the case studies, IPD was accessed from registries and observational studies. However, for the remaining case studies, access to IPD was denied. Reasons for inaccessibility mostly related to datasets not being research-ready within project timelines or unwillingness to share data. The consortium's experiences with inaccessibility of IPD RWE for research purposes were echoed by many external stakeholders, implying that lack of access to IPD RWE remains a wide-spread issue. Bearing in mind that RWE is generated by patients in clinical practice, this barrier to data access diminishes the potential benefit of using RWE to provide critical insights on the effectiveness of treatments for all patients in healthcare practice.

Furthermore, stakeholders' views varied on the acceptability of RWE use in CER and subsequent decision making. The reasons behind such controversy generally hinged on two inter-related aspects: a lack of trust in the robustness of findings based on RWE compared with RCT data, as well as a lack of experience with using RWE in currently available methods to address (comparative) drug effectiveness. Arguably, the inaccessibility to IPD RWE contributes both to the lack of concrete examples demonstrating the potential added value of RWE use in CER and the wide lack of trust among decision-makers regarding the robustness of findings based on RWE.

An array of potential solutions lend themselves to overcoming this persistent inaccessibility to RWE and maximizing societal gain from its use in CER. For example, the EU Clinical Trial Directive was recently adopted whereby sponsors of RCTs conducted for marketing authorization applications agreed to provide access to all patient-level clinical reports of trial subjects online (21). Presently, no equivalent initiative exists for the publication of similar IPD for RWE generated through observational studies and may be a worthwhile endeavour for the future. Another example relates to the US FDA-Sentinel initiative, whereby external researchers can send standardized data queries to multiple nodes of a decentralized network of participating databases (22). In this model, databases can opt-in or out of the sentinel initiative without having to relinquish complete access to their IPD, yet still run external research queries. The main advantage of such a model

is its ability to circumvent the sensitive aspects relating to full-fledged access to IPD while delivering the required information for furthering scientific pursuits. Similar frameworks have been implemented in other fields of research, such as DataSHIELD to conduct international research as part of the Healthy Obese Project and the Biobank Standardisation and Harmonisation for Research Excellence Project (BioSHaRE-EU) (23).

Social media demonstrates potential for RWE collection but within confines

In the fifth section, we explored the possibilities for RWE generation through a novel source, namely social media. **In Chapter 8**, we conducted a scoping review of academic and grey literature to examine whether social media was being used to inform (relative) effectiveness research.

The results indicated that Social media may be a potential source of RWD for REA, particularly on aspects such as adverse events (AEs), occurrence of disease-specific symptoms, adherence behaviour, and health-related quality of life (HRQoL). This potential has not yet been fully acknowledged due to methodological limitations that accompany social media-generated health data, such as information bias and selection bias, as well as the limited acceptability of such data.

One caveat of using social media to collect health data that requires special attention is the lack of clear methodological guidance. Standardized approaches regarding data collection, analysis and interpretation are necessary to ensure comparability and reproducibility between studies. For example, although social media posts may be extracted by automated processes, interpretation of the posts requires manual work since automated processes cannot successfully interpret aspects such as sarcasm or the use of slang (24-26). Additionally, there is currently a lack of guidance on managing the drawbacks of social media highlighted above regarding validation of authenticity, selection bias and information bias. The development of clear and uniform methodological guidance is thus critical to improve the extraction, interpretation and subsequent use of social media in decision making.

In Chapter 9, we conducted a feasibility study on the use of social media to collect patients' and carers' perspectives on HRQoL in melanoma.

Findings from this study indicated that social media may provide a valuable tool to quickly and efficiently assess patient and carer perspectives on HRQoL, potentially increasing the availability and impact of HRQoL data in REA of drugs. Responses can be gathered within a short period of time and from audiences not usually included in RCTs (e.g. women or patients with early stages of melanoma) (27;28). Employing social media to collect such data would require less resources than multi-centre trials or point-of-care studies and has been confirmed in previous research on cervical cancer (29).

On a different note, the use of social media may be less resource-intensive and more efficient than data collection through patient/citizen panels at various stages of HTA decision making. Patients participating in patient/citizen panels may hesitate to share particular personal insights related to their HRQoL due to the formal setting in which panels

are conducted or due to the dynamics of ongoing discussions. On the other hand, through social media, patients can provide candid answers at their own pace and within a trusted environment of their choice.

In this study, social media was used as a medium for spreading a survey to collect patient perspectives on HRQoL. Meanwhile, another potential approach for using social media pertains to the mining of data already available on different channels, such as patient forums, (micro-)blogs, or social networking services for data on HRQoL (24;30;31). In the latter approach, little intervention is applied by the researcher (e.g. through directed questions), implying that insights gained on HRQoL may better represent the "real world". On the other hand, it involves the mining of unstructured data which was initially not designed to answer specific research questions hence raising fundamental methodological challenges in relation to the collection, analysis and interpretation of such data for HTA. At the moment, little research exists on whether results generated by the two approaches for using social media correlate or differ, the reasons behind potential discrepancies in such results and thus the interpretation of such findings. This is a topic of ongoing research by the authors. Additionally, initiatives such as IMI-WEB-RADR, and the IMI-Big Data for Better Outcomes (IMI-BD4BO) are currently exploring the potential of using social media to gain insights on adverse effects of drugs or to focus on big data use for cardiovascular disease, Alzheimer's disease and haematological malignancies, respectively (32;33).

ONGOING INITIATIVES ON RWE & RELEVANCE TO MAIN FINDINGS

In parallel to the research conducted for this thesis between 2014 and 2018, several reputable initiatives have been instigated in relation to the collection and use of RWE in HTA and decision making. A number of these initiatives have provided (or will soon provide) complimentary insights to those presented in this thesis. Below, we highlight a number of themes these initiatives have (or aim to) address and the relevance thereof for the findings of this thesis.

Good practices for the conduct of RWE studies, analyzing RWE and reporting findings: Addressing the cultural barrier to RWE use in decision making

The ISPOR and the International Society for Pharmacoepidemiology (ISPE) have recently established a Special Task Force (ISPOR-ISPE STF) on RWE. This task force published two reports that addressed two topics, namely: good procedural practices for the conduct of RWE studies that enhance the confidence of health care decision makers to factor these studies into their decision making (34); and good practices to structurally improve reporting so that healthcare database studies become reproducible, hereby instilling confidence in RWE (35).

As part of the first report by Berger et al., the authors explicitly call for the pre-registration of RWE study protocols (including data analysis protocols) on a publicly-accessible website, particularly for studies aiming to confirm pre-established hypotheses on (relative)

treatment effects. In doing so, decision-makers' trust in the face validity of results from such studies would be enhanced and concerns regarding data-mining and their impact on study results would be diminished. In concert with these recommendations, the second report by Wang et al. establishes a comprehensive list of analytical choices and data parameters that research teams conducting RWE studies based on large observational databases should aim to report. It is worth noting that this list builds on work from previous reporting guidelines (e.g. STROBE (36), RECORD (37), PCORI Methodology Report (38) and EnCePP (39)). As a result of such reporting, different research teams should be able to reproduce the findings generated using the same datasets and decision-makers should be able to assess the validity of analytic choices made throughout the study.

Another initiative worth noting is Reproducible Evidence: Practices to Enhance and Achieve Transparency (REPEAT) which aims to improve the transparency, reproducibility and validity of longitudinal healthcare database research. In particular, the initiative will quantitatively assess the current state of replicability of current RWE from studies on large healthcare databases and subsequently quantitatively evaluate the robustness of such RWE (40). In doing so, the initiative hopes to diminish hesitancy amongst drug regulators, patients, clinicians, and payers to trust evidence from databases due to high profile controversies with overturned and conflicting results.

Bearing in mind the cultural reluctance amongst decision makers regarding RWE use demonstrated in chapters 3 to 7, we hope that efforts such as the ISPOR-ISPE STF and REPEAT will contribute to increased confidence in RWE, particularly in relation to (relative) effectiveness of drugs.

Harmonizing RWE requirements and registry standards for HTA decision making in Europe: the EUnetHTA collaborative

The European network for HTA (EUnetHTA) is a longstanding collaboration aiming to create an effective and sustainable network for HTA across Europe (41). Established in 2005, the collaboration has since striven to: facilitate efficient use of resources available for HTA, create a sustainable system of HTA knowledge sharing and promote good practice in HTA methods and processes. One of the Work Packages (WP) within Joint Action 3 (JA3, 2016-2020) of EUnetHTA, specifically WP5, is dedicated to the life-cycle approach to improve evidence generation of drugs and medical devices (13). Within WP5, two activity strands are defined, Strand A on early dialogues (initial evidence generation) and Strand B on post-launch evidence generation & registries.

With regards to Strand B (we will return to Strand A later in this chapter), the work group aims to generate a checklist tool for registries to assess their fit for informing HTA (13). This tool is based upon work previously performed by the PAtient REgistry INiTiative Joint Action (PARENT JA) (42) which provided European Union Member States with a set of guidelines, recommendations and tools to support establishment of interoperable patient registries. Moreover, the work group plans to conduct pilot projects whereby registries will be set up for RWE collection based on evidence gaps identified in HTA reports. This

entails producing a common research question and minimum data set for registries, and, if possible, a definition of a core common protocol (13).

Recalling the negative experiences stakeholders have cited with regards to the implementation of CF in the Netherlands in Chapter 6, whereby simple registries with limited sets of parameters exponentially expanded as more research questions were raised in time and the negative impact of such expansion on the quality of data generated, there seems to be a need for guidance on core data sets which truly matter to decision making in HTA. Not only would work by WP5 Strand B thus inform better implementation of registries within the Netherlands but across Europe, provided there is agreement among HTA agencies on core common protocols. The implementation of robust, standardized, inter-operable RWE generation across Europe could deliver a vital tool for the conduct of successful MEAs on a national and international level. This is much needed, provided HTA agencies' increasing reliance on MEAs to strike the balance between expedited patient access to innovative drugs versus adequate evidence generation (19;20). Furthermore, agreement on core common protocols amongst HTA agencies suggests an implicit harmonization of RWE requirements. Streamlining on this front would significantly minimize obstacles for stakeholders including MAHs for generating valuable evidence for HTA decision making (3-5).

Adaptive Pathways and the Product Life-Cycle Approach: Improving synergy between regulatory agencies and HTA agencies regarding RWE generation

Coinciding with efforts amongst HTA agencies to explore the potential of RWE use in decision making, the European Medicines Agency (EMA; pan-European regulatory authority) began work on Adaptive Pathways (AP), a scientific concept for drug development and data generation which allows for early and progressive patient access to a medicine (43). The AP concept rests on three fundamental principles: iterative drug development (e.g. approval within a restricted patient population with subsequent expansion based on evidence generation); gathering evidence on real-life use of the drug to supplement RCT data; and early involvement of all stakeholders including patients, healthcare providers, HTA agencies and payers in scientific discussions on the drug development pathway. Between March 2014 and August 2016, the EMA also completed a pilot project during which the AP concept was implemented in practice (43).

The second principle of AP indicates a central role for RWE studies to supplement RCT's. This concept has inspired work within the NEW Drug Development ParadIGmS (NEWDIGS) program at the Massachusetts Institution of Technology (MIT), resulting in proposals for "efficacy-to-effectiveness" trials in which an effectiveness trial (RWE) would commence seamlessly upon completion of the efficacy trial (RCT) (44). Moreover, the EMA has recently published scientific guidance on post-authorization effectiveness studies to deliver RWE (45).

The third principle of AP on early dialogue between regulatory agencies, HTA agencies and patients is the focus of ongoing projects. Building on progress made within the Shaping

Early European Dialogues (SEED) consortium (46), the EUnetHTA WP5 Strand A workgroup represents an European effort to establish a sustainable system for early dialogues with the EMA on initial (pre-authorization) evidence generation for new drugs (13). Naturally, discussions within the parallel scientific advice between the EMA and HTA agencies include deliberation on RCT evidence to be developed for new products. However, the generation of RWE for new products (e.g. in the post-marketing setting to supplement phase III RCT data) may also be discussed during early dialogues.

Therefore, as the drug development landscape shifts from the conventional approach of phases I-IV development stages to an iterative, product life-cycle approach, there will be an increasing need to harmonize RWE requirements between HTA agencies and regulatory agencies. Initiatives such as AP at the EMA and EUnetHTA WP5A&B thus complement one another well and may pave the road to better synergy between HTA agencies and regulatory agencies on RWE use in decision making. This collaboration has been formalized in the EMA-EUnetHTA work plan (47). It is also our hope that such initiatives will inform the development of more coherent systems for MEAs, whereby the outputs of the regulatory pathway include a clear RWE generation pathway that meet the needs of MEAs.

Alternative (pragmatic) trial designs and approaches to deliver RWE for HTA decision making

A wealth of literature is present on pragmatic clinical trials and their relevance for decision making, beginning with the seminal paper by Shwartz et al. in 1967 (9;48). Moreover, throughout research conducted for Chapters 2-3 and 5-6 of this thesis, several interviewees and authors have indicated the potential benefit of pragmatic trials for generating relevant evidence for HTA.

Clustered Randomized Controlled Trials (CRCT) are often used to examine the effects of healthcare interventions in routine clinical practice (49). One variant of such trials are the Stepped Wedge (SW-CRCT) design. In this approach, clusters (e.g. healthcare institutions based on geographical area) are identified to take part in a CRCT. Before the trial begins, each cluster is randomly allocated to time at which the intervention (e.g. a new drug) is applied. Essentially, the SW-CRCT thus resembles a one-way cross-over design, where all clusters receive the treatment eventually but at different time intervals (50). Consequently, SW-CRCTs circumvent ethical considerations relating to clinical equipoise since all participating clusters eventually receive treatment. This may be particularly important when developing HTA evidence on orphan therapies, whereby conduct of RCTs may be unethical. Additionally, the design maximizes statistical power since the effect of the intervention is measured both between-clusters and within-clusters comparisons (51). This may reduce the temporal and financial burden for evidence generation for HTA decision making when compared to RCTs. However, we are also aware that SW-CRCTs are associated with several drawbacks which should be considered before recommendations are made for their conduct (49;52).

With regards to such drawbacks, the IMI-GetReal initiative has conducted research on the operational and ethical challenges associated with the implementation of PCTs to provide RWE in early drug development. One of the outputs of IMI-GetReal on this front has been the publication of a series of scientific articles discussing the complex interplay between: pragmatic design options, study feasibility, stakeholder acceptability of trial outputs, validity, precision and generalizability of outcomes (53). In addition to this, a tool has been developed (PragMagic) to guide teams through the potential decisions in trial design, the associated challenges and practical solutions to overcome such challenges (54).

Another promising approach to expedited and less cumbersome generation of high-quality RWE lies in embedding RCTs in patient registries or Electronic Health Records (EHRs). The Clinical Trials Transformation Initiative (CTTI) is an ongoing initiative seeking to develop and drive adoption of practices that will increase the quality and efficiency of clinical trials (55). The Registry Trial Project of the CTTI specifically aims to increase the practice of leveraging registries to facilitate high-quality clinical trials at lower costs and has recently published guidance to this effect (56). To date, numerous examples of registry-based RCTs have been successfully completed (56). On a similar note, van Staa et al provide examples whereby RCTs have been embedded within EHRs in the United Kingdom (57). Notably, literature from both initiatives points to challenges related to governance of individual patient data (IPD), whether within registries or EHRs (56;57). Such challenges, along with recommendations the authors make to address them, resonate well with the findings presented in Section D of this thesis, once again highlighting the need for multi-stakeholder collaboration on the governance of data access and use.

LIMITATIONS OF RESEARCH

The focus of research conducted alternated between the European and Dutch perspectives in most chapters. For example, the policies for RWE use amongst HTA agencies were reviewed for only 6 agencies in Europe, representing only 6 of the 29 jurisdictions. Furthermore, the experiences of implementing RWE collection and use for MEAs were only analyzed for the Netherlands. The reasons behind such strategic choices are discussed in detail in the corresponding chapters and include practical considerations and the placement of the research committee within Europe. However, the authors recognize that the issue of RWE use in decision making extends beyond the boundaries of a single nation or continent. The authors are also aware that considerable research efforts on similar topics are occurring worldwide and have referred to these in the discussion section above.

The approach chosen to assess the use of RWE in HTA practice was fragmented. For example, we only analyzed RWE use in standard HTA (i.e. REA and CEA) and in the context of MEAs. Therefore, all the research conducted examines the use of RWE in decision making within the post-marketing setting. Ideally, the authors would have preferred to investigate issues pertaining to defining evidence gaps and RWE requirements, collection methods, analysis and interpretation throughout the entire drug lifecycle (i.e. from early scientific

advice prior to regulatory approval until the finalization of MEA commitments). This may be possible in the future should researchers have the ability to follow the full pathway of new drugs in the context of AP.

Little attention was devoted to potential synergies amongst regulatory agencies (e.g. EMA and the Food and Drug Administration (FDA; USA)) and HTA agencies on the subject of RWE generation and use for decision making within this thesis. As cited above, this is an active field of research within a number of initiatives such as EUnetHTA (13), the AP project (43), NEWDIGS (44) and CTTI (55). Previous research conducted as part of IMI-GetReal and discussions as part of the IMI-ADAPTSMART consortium (58) allude to the importance of such synergies amongst regulatory agencies and HTA agencies in order to reduce the burden of RWE collection and to increase the impact of RWE generated for both stakeholders (5). Therefore, the outcomes of further research on this particular theme may have tangible effects on the RWE landscape in the future.

As a discipline, HTA encompasses a broad variety of health technologies such as medical devices, surgical procedures or even health applications on smartphones and wearable devices (59). The regulation and assessment of drugs is embedded in a highly structured framework; drug development begins with phase I trials, proceeds through to phase III pivotal RCTs in human patients and subsequently to phase IV in the post-marketing setting (60). Each phase is also associated with well-documented evidential requirements, often published in institutional guidelines (60). On the other hand, the regulation and assessment of medical devices is conventionally embedded in a more adaptive framework, whereby RCT evidence is not always mandatory (or feasible) for marketing authorization purposes (61). As such, it may be argued that the role of non-RCT evidence (i.e. RWE) is prominent for medical devices and should have warranted attention in this thesis. In fact, the FDA has recently issued draft guidance on the use of RWE for regulatory applications for medical devices which cited numerous historic examples for RWE use for different medical devices (62). On a similar note, NICE seeks to publish guidance for HTA of health applications on smartphones and wearable devices in the near future (63). Again, the assessment of such health technologies is conventionally not based on the same evidence requirements for drugs, implying a more prominent role for RWE within decision making processes.

WHICH STEPS CAN BE TAKEN TOWARDS MORE OPTIMAL USE OF RWE IN HTA DECISION MAKING? RECOMMENDATIONS BASED ON FINDINGS

In order to advance the use of RWE in HTA and decision making, we propose recommendations related to policy as well as methodological research. Ultimately, both aspects should influence and inform one another; policies should be founded on and harness recent advances in scientific methodology and methodological research should aim to inform the needs of decision-makers.

To begin with, we list the main policy recommendations and briefly elaborate on each point below:

- HTA agencies should aim to harmonize policies regarding RWE generation and use in decision making. To this effect, we recall the efforts of EUnetHTA cited above. However, it is equally important to seek harmonization beyond the European context. We believe that recent guideline developments citing RWE by the Canadian HTA agency (CADTH) provide an example for further collaboration (64).
- In light of increasing attention for the product life-cycle approach, HTA agencies should seek to harmonize with regulatory agencies on RWE requirements both during pre-clinical development and in the post-marketing setting. To this effect, we recall the efforts of the EMA on AP (43), of EUnetHTA and EMA on early dialogues and parallel consultations (13;65), as well as NEWDIGS on AP in the USA (44).
- As reliance on MEAs increases worldwide, HTA agencies should invest in research on the implementation of MEAs to distill the main learnings from previous schemes. This research should cover different forms of MEAs, such as price-volume, coverage with evidence development and performance-based outcomes schemes. Importantly, the learnings discovered should be taken into consideration when designing and implementing future MEAs.
- Bearing in mind that RWE generation may be the common responsibility of several stakeholders (e.g. HTA and regulatory agencies, pharmaceutical industry and healthcare providers), stakeholders should seek to achieve transparency on the distribution of roles, funding of data collection and regulation of access to data for RWE repositories. Moreover, alternative governance structures for (international) data access may be explored to minimize access barriers. To the latter, we recall initiatives such as FDA sentinel in the USA (22), BioSHaRE-EU in Europe (23) and the FAIR guiding principles for scientific data management and stewardship (66).

Finally, we propose the following with regards to recommendations for future methodological research:

- Investigate which sources of RWE could be collected, using which methods and for which policy questions. For example, IMI-GetReal has touched upon the use of RWE for (relative) effectiveness research in network meta-analyses (67;68), to implement propensity scoring techniques (69) and to extrapolate long-term outcomes beyond RCTs (70). However, this research has raised further methodological questions that remain unanswered. Moreover, current guidance on the implementation of such methodologies in decision making remains sparse.
- In order for decision makers to develop trust in RWE, it is imperative to understand the impact of strategic choices regarding study parameters and analysis methods on results from studies using RWE. Future research should seek to quantify such impact and its effect on the interpretation of the results from RWE. To this effect, we recall the efforts of the REPEAT initiative (40) and others (34;35) cited above.

- Despite its potential as a data source, guidance on methods for RWE generation and use from social media for (relative) effectiveness research is lacking. Topics where scientific research would be of impact include: validation of responder authenticity, selection bias and information bias in data collected from social media and cross-sectional versus longitudinal data collection from social media.

CONCLUSION

Real-World Evidence bears promise for HTA and decision making on reimbursement of drugs. However, a number of challenges complicate the potential implementation of RWE in HTA and subsequently for decision making on resource allocation in healthcare systems. The challenges pertain to a number of issues, including: unresolved differences amongst stakeholders on the definition of RWE, dissimilarities in policies for RWE use amongst HTA agencies, a fragmented reality of RWE use in standard HTA and MEAs, inaccessibility to RWE and the need for alternate governance structures.

On the other hand, numerous initiatives are ongoing to address the challenges cited above. Many of these will provide critical insights that may help improve the methodology behind RWE collection and analysis, create awareness for alternative approaches in RWE generation throughout the product lifecycle and stimulate joint action amongst all relevant stakeholders for developing appropriate governance structures for RWE repositories. Alongside policy and research recommendations made above, the outcomes of such initiatives may lead to progressively optimal use of RWE for HTA and decision making. It is our aspiration that in doing so, HTA agencies and decision makers will inch ever closer to realizing the full potential healthcare systems can bear for all citizens.

REFERENCE LIST

1. Makady A, Goettsch W, Hummel N, Jonsson P, Nordon C, Zuidgeest C. GetReal. Updated glossary of definitions of common terms. 2016 June 14 Available from: https://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary%20-%20Update.pdf.
2. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real World Data for Coverage and Payment Decisions: The ISPOR Real World Data Task Force Report. Value Health 2007;10(5):326-35.
3. Mestre-Ferrandiz J, Deverka P, Pistallato M, Rosenberg E. The current drug development paradigm: responding to US and European demands for evidence of comparative effectiveness and relative effectiveness. Office of Health Economics 2014.
4. The Academy of Medical Sciences. Real world evidence: Summary of a joint meeting held on 17 September 2015 by the Academy of Medical Sciences and the Association of the British Pharmaceutical Industry. 2015 September 17 [cited Oct 2016]. Available from: https://www.acmedsci.ac.uk/viewFile/56cab22108cf9.pdf.
5. Makady A, Goettsch W. Review of Policies and Perspectives on Real-World Data. 2015 [cited Apr 2017].
6. Lipska I, Hoekman J, McAuslane N, Leufkens HGM, H+Âvels AM. Does conditional approval for new oncology drugs in Europe lead to differences in health technology assessment decisions? Clin Pharmacol Ther 2015 Nov 1;98(5):489-91.
7. Hoekman J, Boon WPC, Bouvy JC, Ebbers HC, de Jong JP, De Bruin ML. Use of the conditional marketing authorization pathway for oncology medicines in Europe. Clin Pharmacol Ther 2015 Nov 1;98(5):534-41.
8. Kaaniche A, Troubat A, Sherwood A. Does Conditional Marketing Authorisation Influence Market Access in France, England, and Germany? Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research 2015;18(7):A559.
9. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. J Clin Epidemiol 2010;63(10):1053-8.
10. ABPI. Demonstrating Value with Real World Data: A practical guide. 2015 [cited Jul 2015]. Available from: http://www.abpi.org.uk/our-work/library/guidelines/Pages/real-world-data.aspx.
11. Berger ML, Lipset C, Gutteridge A, Axelsen K, Subedi P, Madigan D. Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines. Value Health 2015;18(1):127-30.
12. IMI-GetReal. IMI-GetReal: Overall objectives. 2017 [cited Dec 2016]. Available from: https://www.imi-getreal.eu/About-GetReal/Overall-objectives.
13. EUnetHTA. Work Package 5 - Life cycle approach to improve Evidence Generation. 2017 Available from: http://eunethta.eu/activities/eunethta-joint-action-3-2016-20/work-package-5-life-cycle-approach-improve-evidence-gener.
14. Alemayehu D, Riaz Ali MPP, Alvir JM, Mardekian J, Mullins CD, Perfetto EM, et al. Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data". J Manag Care Pharm 2011;17(9 suppl.):S3-S37.
15. Alemayehu D, Mardekian J. Infrastructure Requirements for Secondary Data Sources in Comparative Effectiveness Research. J Manag Care Pharm 2011;17(9 Supp A):S16-S21.
16. Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: lessons learned from the PHAROS registry. Pharmacoecon 2015;33(6):551-60.
17. Mohseninejad L, van Gils C, Uyl-de Groot CA, Buskens E, Feenstra T. Evaluation of patient registries supporting reimbursement decisions: the case of oxaliplatin for treatment of stage III colon cancer. Value Health 2015;18(1):84-90.

18. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. Value Health 2012;15(3):570-9.

19. Faulkner SD, Lee M, Qin D, Morrell L, Xoxi E, Sammarco A, et al. Pricing and reimbursement experiences and insights in the European Union and the United States: Lessons learned to approach adaptive payer pathways. Clin Pharmacol Ther 2016;100(6):730-42.

20. Garrison Jr LP, Carlson JJ, Bajaj PS, Towse A, Neumann PJ, Sullivan SD, et al. Private sector risk-sharing agreements in the United States: trends, barriers, and prospects. The American journal of managed care 2015;21(9):632-40.

21. European Commission. Clinical trials - Regulation EU No 536/2014. 2017 [cited Mar 2017]. Available from: https://ec.europa.eu/health/human-use/clinical-trials/regulation_en#ct4.

22. McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. Pharmacoepidemiol Drug Saf 2012;21(S1):18-22.

23. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014;43(6):1929-44.

24. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014;37(5):343-50.

25. Sarker A, Ginn R, Nikfarjam A, OGÇÖConnor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. Journal of biomedical informatics 2015;54:202-12.

26. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: a review of the opportunities and challenges. Br J Clin Pharmacol 2015;80(4):910-20.

27. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLOS Clin Trial 2006;1(1):e9.

28. Holterheus C. Burden of Melanoma. Erasmus University Rotterdam; 2011.

29. Zaid T, Burzawa J, Basen-Engquist K, Bodurka DC, Ramondetta LM, Brown J, et al. Use of social media to conduct a cross-sectional epidemiologic and quality of life survey of patients with neuroendocrine carcinoma of the cervix: a feasibility study. Gynecol Oncol 2014;132(1):149-53.

30. Beusterien K, Tsay S, Gholizadeh S, Su Y. Real-world experience with colorectal cancer chemotherapies: patient web forum analysis. Ecancermedicalscience 2013;7:361.

31. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013;22(3):256-62.

32. The Innovative Medicines Initiative. Big data, medicines safety and respiratory disease focus of €93 million IMI Call for proposals. 2017 [cited Dec 2016]. Available from: http://www.imi.europa.eu/content/imi-2-call-6-launch.

33. ROADMAP. Real world Outcomes across the AD spectrum for better care: Multi-modal data Access Platform. 2017 [cited Dec 2016]. Available from: http://www.roadmap-alzheimer.org/index.html.

34. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler H, Goettsch W, et al. 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. Pharmacoepidemiol Drug Saf 2017;26(9):1033-9.

35. Wang SV, Schneeweiss S, Berger ML, Brown J, Vries F, Douglas I, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1. 0. Pharmacoepidemiol Drug Saf 2017;26(9):1018-32.

36. Von Elm E, Altman DG, Egger M, Pocock SJ, G+©tzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE) Statement: guidelines for reporting observational studies. International Journal of Surgery 2014;12(12):1495-9.

37. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS medicine 2015;12(10):e1001885.
38. Committee PP-CORIM. The PCORI Methodology Report. 2013.
39. Takac MJ-M. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. 2011.
40. REPEAT. The REPEAT Initiative. 2017 [cited Oct 2017]. Available from: http://www.repeatinitiative.org/about.html.
41. EUnetHTA. About us: EUnetHTA. 2017 [cited Oct 2017]. Available from: http://www.eunethta.eu/about-us.
42. PARENT JA. The Cross-Border Patient Registries Initiative. 2016 [cited Mar 2017]. Available from: http://patientregistries.eu/.
43. European Medicines Agency. Adaptive Pathways. 2017 [cited Oct 2017]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp.
44. NEWDIGS. About us: MIT NEWDIGS. 2017 [cited Oct 2017]. Available from: https://newdigs.mit.edu/about.
45. European Medicines Agency. Scientific guidance on post-authorisation efficacy studies. 12-10-2016.
46. SEED. Shaping European Early Dialogues for health technologies. 2017 [cited Oct 2017]. Available from: http://www.eunethta.eu/seed.
47. EUnetHTA, European Medicines Agency. EMA-EUnetHTA three-year work plan: 2017-2020. 2017 November 13 [cited Dec 2017]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2017/11/news_detail_002850.jsp&mid=WC0b01ac058004d5c1#.
48. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20(8):637-48.
49. Mdege ND, Man MS, Taylor CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J Clin Epidemiol 2011;64(9):936-48.
50. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology 2006;6(1):54.
51. Pearson D, Torgerson D, McDougall C, Bowles R. Parable of two agencies, one of which randomizes. The ANNALS of the American Academy of Political and Social Science 2010;628(1):11-29.
52. Heim N, van Stel HF, Ettema RG, van der Mast RC, Inouye SK, Schuurmans MJ. HELP! Problems in executing a pragmatic, randomized, stepped wedge trial on the Hospital Elder Life Program to prevent delirium in older patients. Trials 2017;18(1):220.
53. Zuidgeest MG, Goetz I, Groenwold RH, Irving E, van Thiel GJ, Grobbee DE. Series: Pragmatic trials and real world evidence: Paper 1. Introduction. J Clin Epidemiol 2017;88:7-13.
54. IMI-GetReal. PragMagic. 2017 [cited Oct 2017]. Available from: http://www.imi-getreal.eu/Tools/PragMagic.
55. CTTI. Clinical Trials Transformation Initiative. 2017 [cited Oct 2017]. Available from: https://www.ctti-clinicaltrials.org/.
56. CTTI. Registry Trials Project Multi-Stakeholder Expert Meeting: Summary of the meeting held March 30th, 2016. 2016.
57. van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. BMJ 2012;344:e55.
58. IMI-ADAPT SMART. Accelerated Development of Appropriate Patient Therapies: a Sustainable, Multi-stakeholder Approach from Research to Treatment-outcomes. 2017 December 1 [cited Dec 2017]. Available from: http://adaptsmart.eu/.

59. Banta HD, Behney CJ, Andrulid DP. Assessing the efficacy and safety of medical technologies. Washington: Office of Technology Assessment 1978.

60. Food and Drug Administration. The Drug Development Process. 2015 June 24 [cited Oct 2017]. Available from: https://www.fda.gov/ForPatients/Approvals/Drugs/default.htm.

61. Tarricone R, Boscolo PR, Armeni P. What type of clinical evidence is needed to assess medical devices? European Respiratory Review 2016 Sep 1;25(141):259.

62. Food & Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. 27-7-2016.

63. Mulryne J, Clemence B. NICE to Introduce HTA-style App Assessment Process from March 2017. 2017 February 28 [cited Oct 2017]. Available from: https://www.digitalhealthdownload.com/2017/02/nice-introduce-hta-style-app-assessment-process-march-2017/.

64. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada - 4th Edition. 2017 March [cited Oct 2017]. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition.

65. EUnetHTA, European Medicines Agency. Guidance for parallel consultation EMA-EUnetHTA. 2017 June 30 [cited Dec 2017]. Available from: http://www.eunethta.eu/outputs/ema-eunethta-parallel-consultation.

66. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016 Mar 15;3:160018.

67. Ericson A, Gsteiger S. Deliverable D1.6 WP1 Case Study Review: Rheumatoid Arthritis. 2016 [cited Mar 2017]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.6%20Report%20-%20Rheumatoid%20Arthritis_webversion.pdf.

68. Hummel N, Debray TPA, Didden EM, Efthimiou O, Egger M, Fletcher C, et al. Methodological guidance, recommendations and illustrative case studies for (network) meta-analysis and modelling to predict real-world effectiveness using individual participant and/or aggregate data. 2017 [cited May 2017]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/2017-03-30%20-%20WP4%20-%20Methodological%20guidance%2C%20recommendations%20and%20illustrative%20case%20studies.pdf.

69. NICE, Novartis, University of Leicester. Deliverable 1.6 (Multiple Sclerosis Case Study). 2016 Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.6%20Report%20-%20multiple%20sclerosis%20case%20study_webversion.pdf.

70. Makady A, Kalf R, Goettsch W., Lees M. Deliverable D1.6 WP1 Case Study Review: Metastatic Melanoma. 2016 [cited Mar 2017]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20WP1%20deliverables/Deliverable%201.6%20Report%20-%20Rheumatoid%20Arthritis_webversion.pdf.

CHAPTER 11

Appendices

CHAPTER 2 - APPENDIX

((((Perspective[tiab] OR "guideline"[tiab] OR "regulation"[tiab] OR approach[tiab] OR policy*[tiab]) AND ("HTA agency" OR "Regulatory agency" OR industry[tiab] OR "healthcare provider"[tiab] OR "healthcare payer"[tiab] OR stakeholder*[tiab]) AND ("real world data" OR "real world evidence" OR "real world outcome" OR "clinical effectiveness data" OR "hospital data" OR "electronic health records" OR "patient registry" OR "effectiveness"[tiab] OR "alternative study design") AND ("Pragmatic clinical trial" OR "observational design" OR "post-marketing study" OR comparative OR observ*[tiab] OR design*[tiab]) AND ("comparative effectiveness research" OR "outcomes research" OR "relative effectiveness assessment" OR "evidence"[tiab] OR "decision making"[tiab] OR "comparative effectiveness"[tiab]))) AND ("2005/01/01"[PDat] : "2016/12/31"[PDat]))*

Figure I - PubMed search strategy

Table I - Websites searched to locate grey literature across 8 stakeholder groups.

Stakeholder Group	Stakeholder
Health Technology Assessment (HTA) Agencies	National Institute for Health and Care Excellence (NICE)
	Zorginstituut Nederland
	Haute Autorite de Sante
	Institute for Quality and Efficiency in Health (IQWiG)
	Agencia Italaina de Farmaco
	Canadian Agency for Drugs and Technologies in Health (CADTH)
	Centre for Practice and Technology Assessment (USA)
Pharmaceutical Industry	GlaxoSmithKline
	Pfizer
	Merk, Sharp & Dohme (MSD)
	Novartis
	Genzyme
Regulatory Agencies	European Medicines Agency (EMA)
	Food and Drug Administration (FDA)
Healthcare Providers	The Federal Join Committee (G-BA)
	European Hospital & Healthcare Federation (HOPE)
	The Standing Committee of European Doctors (CPME)
Healthcare Payers/ Insurers	European Social Insurance Platform (ESIP)
	Zorgverzekeraars Nederland
	Caisse nationale de l'Assurance Maladie des travailleurs salaries (CNAMTS)
	Association of Standing Health Insurance Funds (GKV Spitzerband)
Patient Organisations	International Alliance of Patient Organisations (IAPO)
	European Patients' Forum (EPF)
Initiatives	Patient-Centered Outcomes Research Institute (PCORI)
	International Society for Pharmacoeconomics and Outcomes Research (ISPOR)
	Clinical Practice Research Datalink (CPRD)
	New Drug Development Paradigms (NEWDIGS)
	Institute of Medicine (IOM)
	Health Technology Assessment International (HTAi)
	National Institute for Health Research (NIHR) – HTA Program
	Quintiles
	McKinzie
	PriceWaterhouseCooper
	National Pharmaceutical Council (USA)

Table I - continued

Stakeholder Group	Stakeholder
	RAND Corporation
	Ernst & Young
	PatientsLikeMe
	Centre for Medical Technology Policy (CMTP)
	Centre for Medicare and Medicaid Services (CMS)
	Eye for Pharma
	Computer Sciences Corporation
	Association of British Pharmaceutical Industry
	European Alliance for Personalised Medicine (EAPM)
	Paraxel
	The Galen Institute

Table II - Inclusion and exclusion criteria for selection of documents from PubMed and grey literature searches.

Inclusion criteria	Exclusion criteria
Document is published in English	Document does not focus on Real World Data (RWD) use in the context of pharmaceutical drug development, drug regulation and drug assessment.
Document is a scientific article, opinion article, editorial, report or guideline.	Document only focuses on methodology of RWD analysis, best practices of evidence synthesis, or evidence synthesis.
In the case of a scientific article, opinion article or editorial, it must be published in a peer-reviewed journal. In the case of a guideline or report, the document must be published on the official website of a recognised institute/organisation.	Document only comprises a summary or abstract (i.e. no access to full document).

RWD

1. What is your understanding of the term real-world data (RWD)?
   a. Could you provide a specific definition, in your opinion, of RWD?
2. Do you collect RWD for all your licensed products? Why or why not?
   a. Does it vary depending on the type of product?
   b. If you do not collect RWD for all your products, could you specify for which types of products you collect RWD?
   c. What is the type of RWD collected in such cases?
   d. Is real-life data also collected for comparators of your products or more generally, e.g. for a disease area?
3. What is timing of collection of RWD in relation to the lifecycle of your products?
   a. Does your company only collect RWD after marketing authorization or also premarketing authorization? Could you specify the timing?
4. Is the collection of RWD mostly connected to mandatory obligations from EMA (e.g. risk management) or part of national reimbursement requirements (coverage with evidence or conditional reimbursement)?
   a. Are there other reasons for your company to collect RWD, for example, for relative effectiveness assessments?
5. Are results from studies with RWD made public, for instance by publication in peer-reviewed journals?
   a. If not, under what conditions, and in what form, would it be likely for RWD data to be made public?

Perceived usefulness

Extent to which a person believes RWD can positively contribute to drug development licensing and market access

1. What are, according to your perceptions, the added benefits of using RWD in drug development, in comparison to, for example, RCT data?
2. What are, according to your perceptions, the limitations of collecting and using RWD for drug development. And what are the possible solutions to such limitations?
   a. How do you value the quality of RWD that are collected during studies?
   b. Do you have any suggestions to improve the quality of RWD?
3. To what extent do you use RWD data in relative effectiveness modelling performed by your institution?
4. Can RWD be used as evidence in pre-licensing studies, market applications and/or to forecast clinical effectiveness?
   a. Is RWD presently included in submission files to regulators and reimbursement agencies?
5. Are you familiar with evidence synthesis strategies, such as meta-analysis, mixed treatment comparisons or network meta-analysis, and how do you value the quality of information resulting from such analyses?
6. What is your opinion regarding the quality of the methodology and/or software used to synthesise the evidence for relative effectiveness assessments? Do you have any suggestions for improvements?
7. To what extent is the methodology available for evidence synthesis of relative effectiveness applicable in a real-world setting?
   a. Can the available methodology directly be implemented?

b. If not, which types of required data are typically not at hand? (i.e. can the available methodology directly be implemented, or is some of the required data typically not at hand?

8. Would you be willing to consider/ perform an assessment of relative effectiveness that is predicted from the available RWD data sources? If so, what types of structural uncertainty regarding, for example, assumptions made or parameter definitions, should primarily be addressed?
9. What is your opinion regarding uncertainty arising from synthesising evidence for relative effectiveness assessment that are due to, for example, assumptions made or parameter definitions?
    a. Are sufficient sensitivity analyses performed relative to key assumptions being made?
    b. Which data sources may enhance the credibility of predictions regarding relative effectiveness?
10. Are you satisfied with text-based reports of RWD evidence used as an input for evidence synthesis/ predictive modelling, or would you prefer these reports to be supported by the underlying structured data sets and/or statistical models (in electronic format)?
11. What software do you currently use (if any) for evidence synthesis and/or predictive modelling?
    a. What is your opinion of such software? Are there any important gaps in functionality or usability of such software?
12. What (if any) should be the role of structured decision aids such as multiple criteria decision analysis (MCDA) in decisions on relative effectiveness?

Perceived ease of use

Degree to which effort is needed to collect and use RWD

1. What are the current obstacles faced in the collection of RWD as well as the implementation of policies for the use of RWD in the decision-making process of drug development?
    a. Do you have any suggestions for improvements?
2. How challenging is the implementation (or assessment) of statistical/mathematical models for data synthesis in relative effectiveness assessment?
    a. Is this a routine in-house task or do you frequently need external expertise?
3. What is the role of software in enabling efficient use of RWD (for example, efficient analysis of data, efficient communication of results)?
    a. Is there key software that you use or that you feel is needed but currently missing?

Figure II - Interview questionnaire sent to stakeholders of the pharmaceutical industry group

RWD

1. What is your understanding of the term real-world data (RWD)?
   a. Could you provide a specific definition, in your opinion, of RWD?
2. To which extent is the collection of RWD officially linked to official regulatory requirements of your institution?
   a. Could you please specify?
3. Do you request the use of RWD as supportive evidence in marketing authorisation applications?
   a. What sort of RWD is ideally preferred and requested for clinical efficacy assessments?
   b. What sort of RWD is currently available, in comparison to ideal requirements?
   c. Specific types of products/ disease areas?
   d. Is this particularly relevant for orphan diseases?
   e. Relevant examples?
4. What are the policies of your organisation governing the collection of RWD data from post-marketing studies?
   a. Did you publish any guidelines on this subject?

Perceived usefulness

Extent to which a person believes RWD can positively contribute to drug development licensing and market access

1. What are, according to your perceptions, the added benefits of using RWD for marketing authorization submissions in comparison to, for example, RCT data?
2. What are, according to your perceptions, the limitations of collecting and using RWD for drug development. And what are the possible solutions to such limitations?
   a. How do you value the quality of RWD that are collected during studies?
   b. Do you have any suggestions to improve the quality of RWD?
3. To what extent do you use RWD data in relative effectiveness modelling performed by your institution?
4. Can RWD currently generated in a post-marketing setting (e.g. PASS, PAES or other observational approaches) be used to predict real-world efficiency of drugs?
5. Are you familiar with evidence synthesis strategies, such as meta-analysis, mixed treatment comparisons or network meta-analysis, and how do you value the quality of information resulting from such analyses?
6. What is your opinion regarding the quality of the methodology and/or software used to synthesise the evidence for relative effectiveness assessments? Do you have any suggestions for improvements?
7. To what extent is the methodology available for evidence synthesis of relative effectiveness applicable in a real-world setting?
   a. Can the available methodology directly be implemented?
   b. If not, which types of required data are typically not at hand? (i.e. can the available methodology directly be implemented, or is some of the required data typically not at hand?
8. What software do you currently use (if any) for evidence synthesis and/or predictive modelling?
   a. What is your opinion of such software? Are there any important gaps in functionality or usability of such software?
9. What (if any) should be the role of structured decision aids such as multiple criteria decision analysis (MCDA) in decisions on relative effectiveness?

Perceived ease of use

Degree to which effort is needed to collect and use RWD

1. What are the current obstacles faced in the collection of RWD as well as the implementation of policies for the use of RWD in the decision-making process of drug assessment at your institution?
    a. Do you have any suggestions for improvements?
2. How challenging is the implementation (or assessment) of statistical/mathematical models for data synthesis in relative effectiveness assessment?
    a. Is this a routine in-house task or do you frequently need external expertise?
3. What is the role of software in enabling efficient use of RWD (for example, efficient analysis of data, efficient communication of results)?
    a. Is there key software that you use or that you feel is needed but currently missing?

Figure III - Interview questionnaire sent to stakeholders of the regulatory agencies group

RWD

1. What is your understanding of the term real-world data (RWD)?
   a. Could you provide a specific definition, in your opinion, of RWD?
2. Do you request the use of RWD in HTA submissions for the purposes of decision- making for reimbursement?
   a. What sort of RWD is ideally preferred and requested for HTA assessments?
   b. What sort of RWD is currently available, in comparison to ideal requirements?
   c. Is this related to Coverage with Evidence Development (CED) or conditional reimbursement after market authorization?
   d. Specific types of products/ disease areas?
      - Is this particularly relevant for orphan diseases?
   e. Relevant examples?
3. What are the policies governing the use of RWD data in HTA submissions at your organization?
   a. Did you publish any guidelines regarding the use of RWD for reimbursement decision-making?
4. Are you satisfied with text-based reports of the submitted evidence, or would you prefer these reports to be supported by the underlying structured data sets and/or statistical models (in electronic format)?

Perceived usefulness

Extent to which a person believes RWD can positively contribute to drug development licensing and market access

1. What are, according to your perceptions, the added benefits of using RWD for HTA submissions, in comparison to, for example, RCT data?
2. What are, according to your perceptions, the limitations of submitting RWD for HTA submissions. And what are the possible solutions to such limitations?
   a. How do you value the quality of RWD that are submitted to you?
   b. Do you have any suggestions to improve the quality of RWD?
3. To what extent do you use RWD data in relative effectiveness modelling performed by your institution?
4. Can RWD be used as evidence in pre-licensing studies, market applications and/or to forecast clinical effectiveness?
   a. Is this expected in reimbursement files from manufacturers?
   b. If yes, how is this assessed by your organization?
5. Are you familiar with evidence synthesis strategies, such as meta-analysis, mixed treatment comparisons or network meta-analysis, and how do you value the quality of information resulting from such analyses?
6. What is your opinion regarding the quality of the methodology and/or software used to synthesise the evidence for relative effectiveness assessments? Do you have any suggestions for improvements?
7. To what extent is the methodology available for evidence synthesis of relative effectiveness applicable in a real-world setting?
   a. Can the available methodology directly be implemented?
   b. If not, which types of required data are typically not at hand? (i.e. can the available methodology directly be implemented, or is some of the required data typically not at hand?

8. Would you be willing to consider/ perform an assessment of relative effectiveness that is predicted from the available RWD data sources? If so, what types of structural uncertainty regarding, for example, assumptions made or parameter definitions, should primarily be addressed?
9. What is your opinion regarding uncertainty arising from synthesising evidence for relative effectiveness assessment that are due to, for example, assumptions made or parameter definitions?
    a. Are sufficient sensitivity analyses performed relative to key assumptions being made?
    b. Which data sources may enhance the credibility of predictions regarding relative effectiveness?
10. What software do you currently use (if any) for evidence synthesis and/or predictive modeling?
    a. What is your opinion of such software? Are there any important gaps in functionality or usability of such software?
11. What (if any) should be the role of structured decision aids such as multiple criteria decision analysis (MCDA) in decisions on relative effectiveness?

Perceived ease of use

Degree to which effort is needed to collect and use RWD

1. What are the current obstacles faced in the collection of RWD as well as the implementation of policies for the use of RWD in the decision-making process of your institution?
    a. Do you have any suggestions for improvements?
2. How challenging is the implementation (or assessment) of statistical/mathematical models for data synthesis in relative effectiveness assessment?
    a. Is this a routine in-house task or do you frequently need external expertise?
3. What is the role of software in enabling efficient use of RWD (for example, efficient analysis of data, efficient communication of results)?
    a. Is there key software that you use or that you feel is needed but currently missing?

Figure IV - Interview questionnaire sent to stakeholders of the HTA agencies group

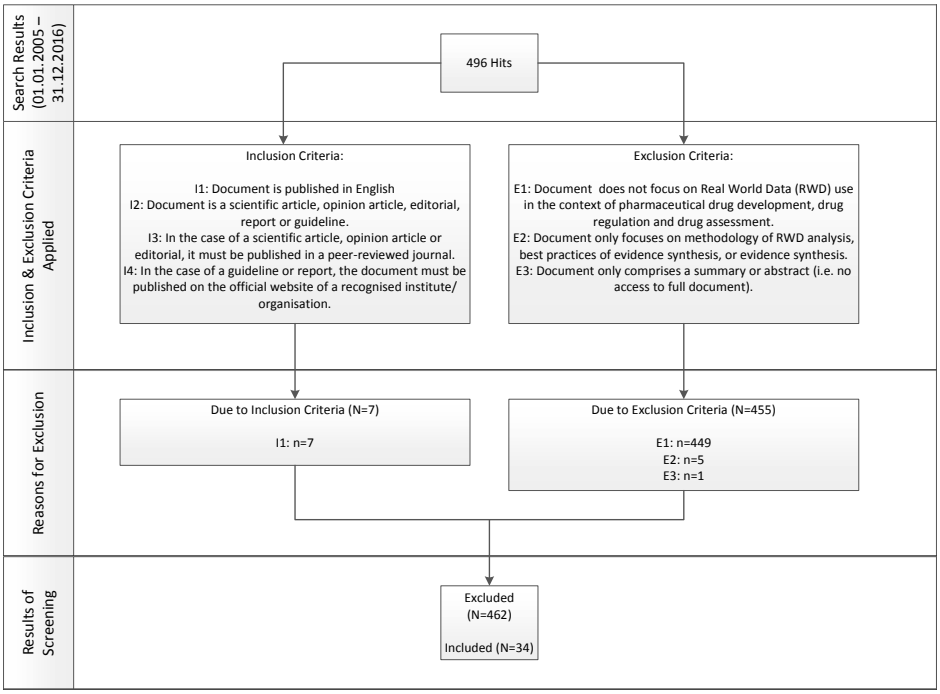


Figure V - Inclusion and exclusion of documents retrieved through PubMed search.

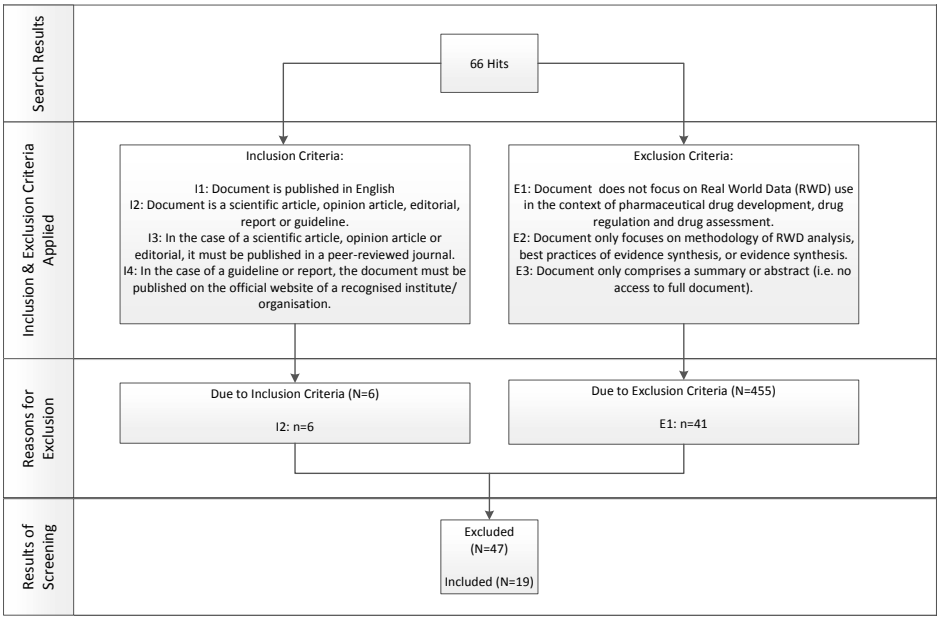


Figure VI - Inclusion and exclusion of documents retrieved from grey literature.

Table III - List of documents selected from academic and grey literature search for analysis.

Primary Author	Date of Publication	Document Title
Alemayehu, D.	2011	Examination of Data, Analytical Issues and Proposed Methods for Conducting Comparative Effectiveness Research Using "Real-World Data"
Annemans, L.	2007	Real-Life Data: A growing Need.
Association of British Pharmaceutical Industry	2011	Demonstrating Value with Real World Data: A practical guide.
Barker, R.	2010	A flexible blueprint for the future of drug development.
Berger, M.	2010	Comparative Effectiveness Research
Berger, M.	2014	Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines
Carpenter, W.	2012	A framework for understanding cancer comparative effectiveness research data needs
Doležal,T	2008	Real-world data in Czech Republic 2008
Dubois, R.	2012	Looking at CER from the Pharmaceutical Industry Perspective
Eichler, H. G.	2012	Adaptive Licensing: taking the next step in the evolution of drug approval
Eichler, H.G.	2011	Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response
Epstein, M.	2007	Guidelines for good pharmacoepidemiology practices (GPP)
European Alliance for Personalised Medicine	2014	MEP's Briefing Paper 2014-2019 Legislature.
European Commission	2010	Directive 2010/84/EU of the European Parliamant and of the Council
European Medicines Agency	2010	The ENCePP Code of Conduct for Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies.
European Union	2012	eHealth Task Force Report: Redesigning health in Europe for 2020
Eye for Pharma	2014	Real World Data Report, 2013-2014: How Real World data are being used to change the pharmaceutical business model.
Foltz, D.	2013	Real-World Data Research: A case for action.
Food and Drug Administration	2013	Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data.
Food and Drug Administration	2011	Postmarketing Studies and Clinical Trials - Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act.
Freemantle, N.	2010	Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials
Fung, V.	2011	Using Medicare Data for Comparative Effectiveness Research: Opportunities and Challenges
Garrison, L.	2007	Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report

Table III - continued

Primary Author	Date of Publication	Document Title
Healthcare Leadership Council	2014	Accesss to Federal Health Data± A key imperative for improving health and health care.
Heranowski, T.	2008	Real-world data and transferability of economic evaluations in Poland
Holve	2012	A tall order on a tight timeframe: stakeholder perspectives on comparative effectiveness research using electronic clinical data
HOPE	2013	Towards patient-focused financing for healthcare provision
IQWiG	2009	Working Paper: Modelling
IQWiG	2013	General Methods
Kaló,Z.	2008	Real World Data for Pharmacoeconomic Evaluation in Hungary
Keohane, P.	2011	The Reality of Real World Data and its Use in Health Care Decisions in Europe
Knottnerus, J.	2010	Real world research
Leyens L.	2016	Use of big data for drug development and for public and personal health and care
Luce, B.	2008	Can managed care organizations partner with manufacturers for comparative effectiveness research
Merck	2013	Merck and Israel's Maccabi Healthcare to Leverage Unique Real-World Database to Inform Novel Health Approaches
Messner, D.	2015	The future of comparative effectiveness and relative efficacy of drugs± an international perspective.
Mohr, P.	2012	Looking at CER from Medicare's Perspective
Neely, J.G.	2013	Practical Guide to Understanding Comparative Effectiveness Research (CER)
NICE	2013	Guide to the methods of technology appraisal 2013
Novartis	2014	Leaders in Clinical Trial Data Transparency
Olson, N.	2013	Introduction to the use of Observational Data
Palozzo, A.	2012	New drugs: How much are they worth? The Italian registries: a model to evaluate appropriateness and effectiveness
Paraxel	2012	Unlocking the Value of Observational Research.
Pleil, A. M.	2013	Using Real World Data in Pharmacoeconomic Evaluations: Challenges, Opportunities and Approaches
Rawlins, M.	2008	De testimonio: on the evidence for decisions about the use of therapeutic interventions.
Romio, S.	2013	Real-world data from the health decision maker perspective. What are we talking about?
Sanofi	2013	Main Sanofi positions on CSR topics
Tesar, T.	2008	Using real-world data for pricing and reimbursement decision within the Slovak republic
Turner, G. M.	2014	Real World Dara and its promise for medicine and research
Umscheid, C.A.	2010	Maximizing the Clinical Utility of Comparative Effectiveness Research

Table III - continued

Primary Author	Date of Publication	Document Title
van Nooten, F.	2013	Use of relative effectiveness information in reimbursement and pricing decisions in Europe
van Staa, T. P.	2013	Background Paper 8.4 Real-life data and learning from practice to advance innovation
Weissman, J.	2015	Translating comparative effectiveness research into Medicaid payment policy: views from medical and pharmacy directors.

Table IV - Overview of interviews conducted per stakeholder group.

Stakeholder Group	Stakeholder	Number of interviewees
Health Technology Assessment (HTA) Agencies	HTA Agency A	2
	HTA Agency B	3
	HTA Agency C	2
	HTA Agency D	2
	HTA Agency E	1
Pharmaceutical Industry	Industry A	2
	Industry B	2
	Industry C	3
	Industry D	1
Regulatory Agencies	Regulatory Agency A	2
	Regulatory Agency B	1
Academia	Academia A	1
	Academia B	1
	Academia C	1
Healthcare Payers/ Insurers	Payer/ Insurer A	1
Healthcare Provider	Provider A	1
Patient Organisations	Patient Organisation A	1
	Patient Organisation B	1
Initiatives	Initiative A	1
	Initiative B	1

CHAPTER 3 - APPENDIX

(((((Zorginstituut Nederland[Affiliation]) OR Agenzia Italiana del Farmaco[Affiliation]) OR Haute Autorite de Sante[Affiliation]) OR Institut fur Qualitat und Wirtschaftlichkeit im Gesundheitswesen[Affiliation]) OR (National Institute for Health[Affiliation] AND Care Excellence[Affiliation])) OR Tandvards-lakemedel sformansverket[Affiliation] AND (("2006/01/01"[PDat] : "2016/06/21"[PDat]))

Search strategy used for searching PubMed.

Interview Questionnaire

RWD

1. What is your understanding of the term real-world data (RWD)?
   a. Could you provide a specific definition, in your opinion, of RWD?
2. Do you request the use of RWD in HTA submissions for the purposes of decision- making for reimbursement?
   a. What sort of RWD is ideally preferred and requested for HTA assessments?
   b. What sort of RWD is currently available, in comparison to ideal requirements?
   c. Is this related to Coverage with Evidence Development (CED) or conditional reimbursement after market authorization?
   d. Specific types of products/ disease areas?
      - Is this particularly relevant for orphan diseases?
   e. Relevant examples?
3. What are the policies governing the use of RWD data in HTA submissions at your organization?
   a. Did you publish any guidelines regarding the use of RWD for reimbursement decision-making?
4. Are you satisfied with text-based reports of the submitted evidence, or would you prefer these reports to be supported by the underlying structured data sets and/or statistical models (in electronic format)?

Perceived usefulness

Extent to which a person believes RWD can positively contribute to drug development licensing and market access

1. What are, according to your perceptions, the added benefits of using RWD for HTA submissions, in comparison to, for example, RCT data?
2. What are, according to your perceptions, the limitations of submitting RWD for HTA submissions. And what are the possible solutions to such limitations?
   a. How do you value the quality of RWD that are submitted to you?
   b. Do you have any suggestions to improve the quality of RWD?
3. To what extent do you use RWD data in relative effectiveness modelling performed by your institution?
4. Can RWD be used as evidence in pre-licensing studies, market applications and/or to forecast clinical effectiveness?
   a. Is this expected in reimbursement files from manufacturers?
   b. If yes, how is this assessed by your organization?

5. Are you familiar with evidence synthesis strategies, such as meta-analysis, mixed treatment comparisons or network meta-analysis, and how do you value the quality of information resulting from such analyses?
6. What is your opinion regarding the quality of the methodology and/or software used to synthesise the evidence for relative effectiveness assessments? Do you have any suggestions for improvements?
7. To what extent is the methodology available for evidence synthesis of relative effectiveness applicable in a real-world setting?
    a. Can the available methodology directly be implemented?
    b. If not, which types of required data are typically not at hand? (i.e. can the available methodology directly be implemented, or is some of the required data typically not at hand?
8. Would you be willing to consider/ perform an assessment of relative effectiveness that is predicted from the available RWD data sources? If so, what types of structural uncertainty regarding, for example, assumptions made or parameter definitions, should primarily be addressed?
9. What is your opinion regarding uncertainty arising from synthesising evidence for relative effectiveness assessment that are due to, for example, assumptions made or parameter definitions?
    a. Are sufficient sensitivity analyses performed relative to key assumptions being made?
    b. Which data sources may enhance the credibility of predictions regarding relative effectiveness?
10. What software do you currently use (if any) for evidence synthesis and/or predictive modeling?
    a. What is your opinion of such software? Are there any important gaps in functionality or usability of such software?
11. What (if any) should be the role of structured decision aids such as multiple criteria decision analysis (MCDA) in decisions on relative effectiveness?

Perceived ease of use

Degree to which effort is needed to collect and use RWD

1. What are the current obstacles faced in the collection of RWD as well as the implementation of policies for the use of RWD in the decision-making process of your institution?
    a. Do you have any suggestions for improvements?
2. How challenging is the implementation (or assessment) of statistical/mathematical models for data synthesis in relative effectiveness assessment?
    a. Is this a routine in-house task or do you frequently need external expertise?
3. What is the role of software in enabling efficient use of RWD (for example, efficient analysis of data, efficient communication of results)?
    a. Is there key software that you use or that you feel is needed but currently missing?

Figure I - Interview questionnaire used to guide interviews.

CHAPTER 4 - APPENDIX

Appendix 1: Data-extraction form (DEF)

General information

1. Topic (drugs)
2. Country
3. URL of report (?)
4. Date/data of extraction
5. Date of HTA recommendation
6. Which indication was used under assessment?
7. Goal of treatment:
   - Extend life (improves morbidity or mortality)
   - Improve symptoms or QoL
   - Other (prophylaxis)

Assessment

Effectiveness

1. Are real-world data (RWD) included in the effectiveness-assessment?
2. *If so, what was the reason for inclusion?*
3. What are the characteristics of the real-world data? (please describe per study)
   - Study-design (prospective/retrospective, controlled, randomised, blinded, etc.)
   - PICO (patient, intervention, comparison and outcome)
   - Number of patients

Cost-effectiveness

1. Are real-world data (RWD) included in the cost-effectiveness-assessment?
2. *If so, what was the reason for inclusion?*
3. Is a model used to estimate the ICER?
4. *If so, for which parameters was RWD used? (e.g. prevalence, effectiveness, cost-effectiveness and/or costs)*
5. *Which type of sources were used for the different parameters (observational, registries, PCTs, etc.)?*

Appraisal

Effectiveness

1. Is information originating from RWD mentioned in the appraisal?
2. What was the impact of the RWD for decision-making?
   - Positive impact, statement in the recommendations section identifying a positive opinion regarding the role of data derived from RWD.
   - Negative impact, statement in the recommendations section identifying a negative opinion regarding the role of data derived from RWD.

- Neutral impact, statement in the recommendations section identifying a neutral opinion regarding the role of data derived from RWD.
- Impact unknown, statement in the recommendations section that cannot clearly be identified as positive, negative or neutral.
- Not identified, no statement in the recommendations section, regarding the RWD.
- Not included, RWD was not included in the assessment.

Cost-effectiveness

1. Is information originating from RWD mentioned in the appraisal?
2. What was the appraisal of the RWD for decision-making?
   - Positive, statement in the recommendations section identifying a positive opinion regarding the role of data derived from RWD.
   - Negative, statement in the recommendations section identifying a negative opinion regarding the role of data derived from RWD.
   - Neutral, statement in the recommendations section identifying a neutral opinion regarding the role of data derived from RWD.
   - Impact, statement in the recommendations section that cannot clearly be identified as positive, negative or neutral.
   - Not identified, no statement in the recommendations section, regarding the RWD.
   - Not included, RWD was not included in the assessment.

Final

1. What was the final recommendation of the dossier for effectiveness?
   - Positive or added benefit
   - Equal benefit or added benefit not proven
   - Negative or lesser benefit
2. What was the final recommendation of the dossier for cost-effectiveness?
   - Positive cost-effectiveness
   - Negative cost-effectiveness
3. In case of a negative recommendation, what was the primary reason for the negative recommendation?
   - Clinical
   - Cost/cost-effectiveness
   - Both
   - Other

Appendix 2: Inter-Rater Reliability

The inter-rater reliability (IRR) was calculated twice in 2 different rounds. In each round, authors independently extracted data from 4 randomly-selected reports (see list below for reports per round). Authors' extraction for closed questions were compared using the Fleiss' kappa method, whereby a score of 0 indicates poor agreement and a score of 1 indicates perfect agreement[11]. Authors' extraction for open-ended questions was compared by a third, independent researcher. Once IRR was established, the remaining 44 reports were equally divided amongst both authors.

List of reports used to calculate IRR

1st round

Agency	*Drug*	*Date of publication*
NICE	Vemurafenib	12-12-2012
SMC	Dabrafenib	09-03-2015
IQWiG	Pembrolizumab	12-11-2015
HAS	Ipilimumab	14-12-2011

2nd round

Agency	*Drug*	*Date of publication*
NICE	Ipilimumab	23-07-2014
SMC	Pembrolizumab	12-12-2016
IQWiG	Nivolumab	11-12-2015
HAS	Vemurafenib	03-10-2012

Appendix 3: Health Technology Assessment (HTA) panel members

List of panel-members

Hedi Schelleman	Zorginstituut Nederland (ZIN)
Pall Jonsson	National Institute for Health and Care Excellence (NICE)
Owen Moseley	Scottish Medicines Consortium(SMC), Healthcare Improvement Scotland (HIS)
Beate Wieseler	Institut für Qualität und Wirtschaflichkeit im Gesundheitswesen (IQWiG)
Anne d'Andon	Haute Autorité de Santé (HAS)

Appendix 4: Health Technology Assessment (HTA) reports included

List of included reports	Date of publication
ZIN	
Ipilimumab	10-02-2012
Vemurafenib	24-02-2014
NICE	
Ipilimumab	12-12-2012
Ipilimumab	23-07-2014
Vemurafenib	12-12-2012
Dabrafenib	22-10-2014
Trametinib (combined with dabrafenib)	22-06-2016
Cobimetinib (combined with vemurafenib)	26-10-2016
Nivolumab	18-02-2016
Nivolumab (combined with ipilimumab)	27-07-2016
Pembrolizumab (treated)	01-10-2015
Pembrolizumab (untreated)	25-11-2015
SMC	
Ipilimumab	14-05-2012
Ipilimumab	08-04-2013
Ipilimumab	10-11-2014
Vemurafenib	10-09-2012
Vemurafenib	09-12-2013
Dabrafenib	09-03-2015
Trametinib (combined with dabrafenib)	12-09-2016
Nivolumab	07-03-2016
Nivolumab	08-08-2016
Nivolumab (combined with ipilimumab)	07-11-2016
Pembrolizumab (treated)	09-11-2015
Pembrolizumab (untreated)	09-11-2015
Pembrolizumab (treated)	12-12-2016
IQWiG	
Ipilimumab	27-04-2012
Ipilimumab	13-03-2014
Ipilimumab (First Addendum)	26-03-2014
Ipilimumab (Second Addendum)	16-05-2014
Vemurafenib	13-06-2012
Vemurafenib	11-12-2013

List of included reports	Date of publication
Dabrafenib	23-12-2013
Dabrafenib (Addendum)	14-03-2014
Dabrafenib (Combined with trametinib)	28-12-2015
Dabrafenib (Combined with trametinib – Addendum)	24-02-2016
Trametinib (Alone or combined with dabrafenib)	28-12-2015
Cobimetinib	11-03-2016
Cobimetinib (combined with vemurafenib)	12-05-2016
Nivolumab	13-10-2015
Nivolumab (Addendum)	11-12-2016
Nivolumab	12-09-2016
Nivolumab (Addendum)	21-11-2016
Pembrolizumab	12-11-2015
Pembrolizumab (Addendum)	14-01-2016
HAS	
Ipilimumab	14-12-2011
Ipilimumab	19-11-2014
Vemurafenib	03-10-2012
Dabrafenib	07-05-2014
Trametinib (Alone or combined with dabrafenib)	20-01-2016
Cobimetinib (Combined with vemurafenib)	16-03-2016
Nivolumab	13-01-2016
Pembrolizumab	16-03-2016

Appendix 5: Supplementary Figures S1 & S2 and Table S1

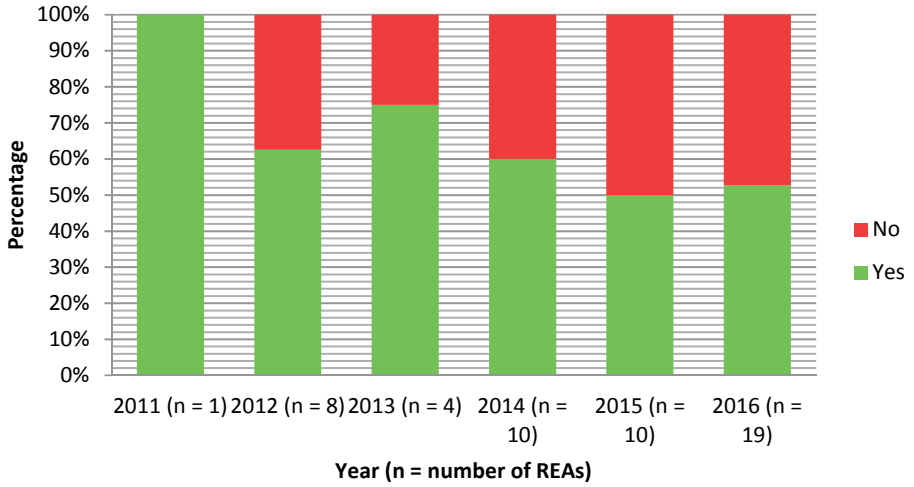


Figure S1 - Inclusion of Real-World Data (RWD) in relative effectiveness assessments (REAs) over time.

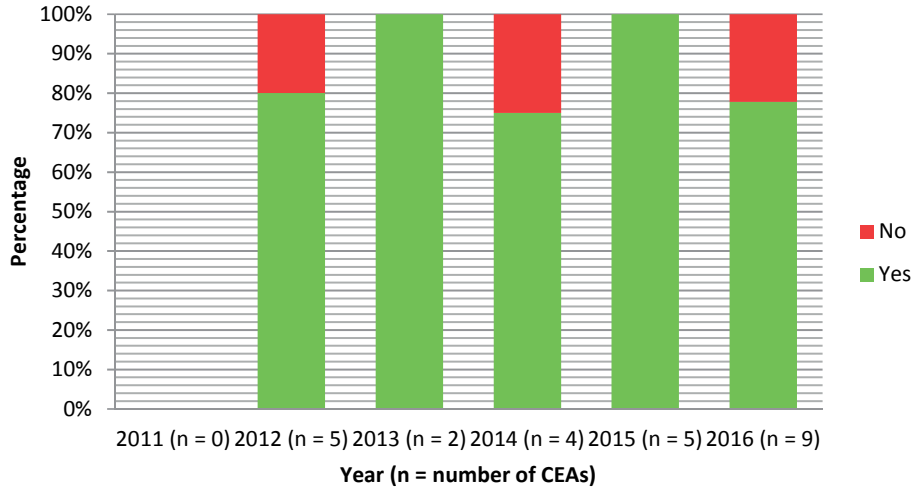


Figure S2 – Inclusion of Real-World Data (RWD) in cost-effectiveness assessments (CEAs) over time.

Table S1 – Studies used to provide real-world data (RWD) on effectiveness and safety of new drugs in relative effectiveness assessments (REAs).

	Summary of RWD study characteristics		
Study Name	**Aim**	**Patient Population**	**Intervention(s)**
CA184-089	Early Access Programme (Italian subset) to find optimal dosing	Italian subset of European patient population from routine care setting.	(1) ipilimumab 10mg/kg every 3 weeks for 4 doses (induction), followed by additional doses of ipilimumab every 12 weeks (maintenance) (2) ipilimumab 3mg/kg every 3 weeks for 4 doses (induction), followed by re-induction if necessary.
CA184-332	The primary objectives were to describe 1) patient demographic and disease characteristics 2) patterns of care 3) survival outcomes 4) AEs of patients with treatment-naïve AM who were treated with ipilimu-mab 3 mg/kg monotherapy in a community practice setting	Eligible patients were required to have AM, be ≥18 years old, have started ipilimumab monotherapy at 3 mg/kg in the first-line setting during April 1, 2011 to September 30, 2012 and received care at an MSH/USON Comprehensive Strategic Alliance Network site that used full iKM EHR capabilities over the entire study period. Ex-clusion criteria included prior systemic treatment for AM, current or pending enrollment in a clinical trial, and treatment for other cancers.	ipilimumab monotherapy

Comparator(s)	Outcome(s) measured	Study Design	Reference
2 different doses of ipilimumab	Overall survival	Prospective observational cohort; non-randomised study.	https://www.nice.org.uk/guidance/ta268/documents/melanoma-stage-iii-or-iv-ipilimumab-updated-analysis-with-revised-patient-access-scheme-from-bristol-myerssquibb2
N/A	1) Variables to describe patient demographic and disease characteristics included gender, age, race, primary site at diagnosis, time since initial melanoma diagnosis, disease stage at start of ipilimumab treatment, location of metastases, ECOG performance status (PS), presence of brain metastases, serum lactate dehydrogenase (LDH) level, and BRAF mutation status. 2) Variables to evaluate patterns of care included ipilimumab dosing, treatment delays and discontinuations, reasons for treatment termination, and concomitant and subsequent supportive and anticancer therapy. 3) OS was the primary measure of treatment effectiveness and was defined as the time from start of ipilimumab until death from any cause. Safety was assessed by reported AEs. 4) Because AE grade/intensity information, causality (e.g. treatment-related or not), and date of occurrence were not available in the iKM source data, AE data are reported irrespective of grade or causality. However, any AE reported that occurred during ipilimumab treatment was captured. Any AEs documented as serious AEs (SAEs) in patients' charts or leading to a hospital/emergency room visit or death were defined as SAEs.	Retrospective electronic health record (EHR) observational cohort study of patients treated with first-line ipilimumab monotherapy in the clinical practices of the McKesson Specialty Health/US Oncology Network (MSH/USON).	http://file.scirp.org/pdf/JCT_2014102311333330.pdf

Table S1 – continued

	Summary of RWD study characteristics		
Study Name	**Aim**	**Patient Population**	**Intervention(s)**
CA184-338	The primary objectives were to describe 1) patient demographic and disease characteristics 2) patterns of care 3) survival outcomes 4) AEs of patients with treatment-naïve AM who were treated with ipilimu-mab 3 mg/kg monotherapy in a community practice setting	Eligible patients were required to have AM, be ≥18 years old, have started ipilimumab monotherapy at 3 mg/kg in the first-line setting during April 1, 2011 to September 30, 2012 and received care at an MSH/USON Comprehensive Strategic Alliance Network site that used full iKM EHR capabilities over the entire study period. Ex-clusion criteria included prior systemic treatment for AM, current or pending enrollment in a clinical trial, and treatment for other cancers.	ipilimumab monotherapy
Periodic Safety Update Report (PSUR data)	Periodic update of the safety of ipilimumab use in clincal practice	As present in clinical practice across Europe	ipilimumab monotherapy

Comparator(s)	Outcome(s) measured	Study Design	Reference
N/A	1) Variables to describe patient demographic and disease characteristics included gender, age, race, primary site at diagnosis, time since initial melanoma diagnosis, disease stage at start of ipilimumab treatment, location of metastases, ECOG performance status (PS), presence of brain metastases, serum lactate dehydrogenase (LDH) level, and BRAF mutation status. 2) Variables to evaluate patterns of care included ipilimumab dosing, treatment delays and discontinuations, reasons for treatment termination, and concomitant and subsequent supportive and anticancer therapy. 3) OS was the primary measure of treatment effectiveness and was defined as the time from start of ipilimumab until death from any cause. Safety was assessed by reported AEs. 4) Because AE grade/intensity information, causality (e.g. treatment-related or not), and date of occurrence were not available in the iKM source data, AE data are reported irrespective of grade or causality. However, any AE reported that occurred during ipilimumab treatment was captured. Any AEs documented as serious AEs (SAEs) in patients' charts or leading to a hospital/ emergency room visit or death were defined as SAEs.	Retrospective electronic health record (EHR) observational cohort study of patients treated with first-line ipilimumab monotherapy in the clinical practices of specialized centers (Cytokine Working Group [CWG] or CWG-affiliated)	Margolin, K., Wong, S. L., Penrod, J. R., Song, J., Chang, I. F., Johnson, D. B., ... & Mcdermott, D. (2013). Effectiveness and safety of first-line ipilimumab for advanced melanoma-A Us multisite retrospective study. Pigment Cell & Melanoma Research, 26(6), 986-987.
N/A	Drug-relatd adverse effects	N/A; non-randomised study	http://www.has-sante.fr/portail/upload/docs/evamed/CT-12741_YERVOY_PIC_REEV_Avis3_CT12741.pdf

Table S1 – continued

	Summary of RWD study characteristics		
Study Name	**Aim**	**Patient Population**	**Intervention(s)**
BRIM1	Phase 1 dose ranging study	Previously treated BRAF mutation positive patients with solid tumours (advanced melanoma and metastatic colorectal). Inclusion criteria were: Male or female > 18 Solid tumours Histologically refractory to standard care or no care available ECOG status 0-1 Life expectancy of > 3 months Absence of known progressing or unstable brain metastases Adequate haematological, renal and hepatic function	vemurafenib 960 mg twice daily
BRIM2	Long-term follow-up of vemurafenib-treated patients with BRAF-V600 positive mutation	Previously treated BRAF mutation-positive melanoma patients. Inclusion criteria were: Male or female > 18 Histologically proven stage IV melanoma BRAF V600-positive mutation Progressive disease following at least one prior systemic treatment ECOG status 0-1 Brain metastasis controlled for at least 3 months No other invasive cancer in last 5 years Adequate haematological, renal and hepatic function	vemurafenib 960 mg orally twice daily
BREAK-2	Assess the safety and clinical activity of dabrafenib in BRAF(V600E/K) mutation-positive metastatic melanoma	Patients with histologically confirmed BRAFV600E/K metastatic melanoma (stage IV) were enrolled. Additional eligibility criteria included measurable disease according to Response Evaluation Criteria in Solid Tumors version 1.1 and an Eastern Cooperative Oncology Group performance status of 0 or 1. Patients were eligible whether or not they had received prior systemic therapy for metastatic disease, exclusive of other BRAF/MEK inhibitors. History or evidence of brain metastases was exclusionary. Adequate bone marrow, liver, renal, and cardiac function and normal clotting parameters were also required.	dabrafenib 150 mg twice daily

Comparator(s)	Outcome(s) measured	Study Design	Reference
N/A	1) Safety/Adverse events 2) Pharmacodynamic activity in tumour tissue	Phase 1 dose-ranging; non-randomised study	Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. New England Journal of Medicine. 2010; 363:809-19.
N/A	Primary: Overall response rate (ORR) Secondary: Overall survival (OS)	Phase II open label – single arm	Sosman J, Kim K, Schuchter L, Gonzalez R, Pavlick A, Weber J, et al. Long-term follow-up of BRAF V600 mutated metastatic melanoma patients treated with vermrafenib reveals prolonged survival. New England Journal of Medicine. 2012; 366:707-14.
N/A	Primary: investigator-assessed overall response rate Secondary: progression-free survival (PFS) and overall survival (OS)	Phase II, single-arm, multicentre study	Ascierto, P. A., Minor, D., Ribas, A., Lebbe, C., O'Hagan, A., Arya, N., ... & Hamid, O. (2013). Phase II trial (BREAK-2) of the BRAF inhibitor dabrafenib (GSK2118436) in patients with metastatic melanoma. Journal of clinical oncology, 31(26), 3205-3211.

Table S1 – continued

	Summary of RWD study characteristics		
Study Name	**Aim**	**Patient Population**	**Intervention(s)**
BREAK-MB	Assess overall intracranial response rate (OIRR) to dabrafenib in patients with BRAF V600E/k mutation-positive melanoma with brain metastases.	Stage IV pts with ≥ 1 intracranial met (0.5 cm–4 cm assessed by MRI) without prior brain therapy (Cohort A) or with progression following prior brain therapy (Cohort B) were eligible with V600E/K mutation.	dabrafenib 150 mg twice daily
Checkmate-003	Phase I dose-escalation cohort expansion trial evaluating safety and clinical activity of nivolumab in patients with advanced NSCLC, melanoma, and kidney, colorectal, and castration-resistant prostate cancer.	Patients eligibility criteria were as follows: pathologically confirmed advanced NSCLC, age ≥ 18 years, Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 (before implementation of amendment 4 in October 2010, ECOG performance status of 0 to 2 was allowed), adequate organ function, and one to five prior systemic treatment regimens for advanced NSCLC. Patients also had to have experienced progression through at least one platinum- or taxane-based regimen and have at least one measurable lesion by RECIST (version 1.0).14 Patients with treated brain metastases stable for at least 8 weeks were eligible. Exclusion criteria included autoimmune disease, prior therapy with T cell–modulating antibodies (eg, anti–CTLA-4, anti–PD-1, and anti–PD-L1), conditions requiring immunosuppressive medications, history of infection by HIV, and active infection by hepatitis B or C viruses.	nivolumab 1-, 3-, or 10-mg/kg

Comparator(s)	Outcome(s) measured	Study Design	Reference
N/A	Primary: investigator-assessed overall intracranial response rate (OIRR)	Phase II, single-arm study	Kirkwood, J. M., Long, G. V., Trefzer, U., Davies, M. A., Ascierto, P. A., Chapman, P. B., ... & Goodman, V. L. (2012). BREAK-MB: A phase II study assessing overall intracranial response rate (OIRR) to dabrafenib (GSK2118436) in patients (pts) with BRAF V600E/k mutation-positive melanoma with brain metastases (mets).
N/A	Primary: drug-related adverse events (AEs), objective response rate (ORR; percentage of patients with confirmed complete or partial responses among all treated patients) Secondary: Overall survival	Phase I dose-escalation cohort; non-randomised study	Gettinger, S. N., Horn, L., Gandhi, L., Spigel, D. R., Antonia, S. J., Rizvi, N. A., ... & Sequist, L. V. (2015). Overall survival and long-term safety of nivolumab (anti–programmed death 1 antibody, BMS-936558, ONO-4538) in patients with previously treated advanced non–small-cell lung cancer. Journal of clinical oncology, 33(18), 2004-2012.

Table S1 – continued

	Summary of RWD study characteristics		
Study Name	**Aim**	**Patient Population**	**Intervention(s)**
Keynote-001 (Part B1)	To evaluate the safety profile of pembrolizumab (formerly called lambrolizumab) assess tumour response every 12 weeks	Patients with measurable metastatic or locally advanced unresectable melanoma, both those who had received prior therapy with ipilimumab and those who had not.	Pembrolizumab 2 mg/kg every 3 weeks

Comparator(s)	Outcome(s) measured	Study Design	Reference
Pembrolizumab 10 mg/kg every 2 weeks Pembrolizumab 10 mg/kg every 3 weeks	Efficacy end-points: overall responses derived from investigator-reported data, with assessment according to immune-related response criteria; and overall responses derived from independent, central, blinded radiologic review, with assessment according to the Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1 Toxic effects: graded with the use of the National Cancer Institute Common Terminology Criteria for Adverse Events, version 4.0.	Expansion cohort of phase 1 dose-escalation study; non-randomised study.	Hamid O et al (2013) Safety and tumor responses with lambrolizumab (Anti-PD-1) in melanoma NEJM 369:2 134-144.

CHAPTER 5 - APPENDIX

Table A - Report sections for which critical commentary was collected for the outcomes research proposal (T=0) and the total number of comments per section.

Outcomes research proposal (T=0) report section	**Number of critical comments made (percentage of total comments; %)**
Research question	2 (3%)
Indication	2 (3%)
Patient population	8 (12%)
Comparator treatment	5 (7%)
Outcomes (effects)	13 (19%)
Outcomes (costs)	7 (10%)
Time horizon	2 (3%)
Data collection method	8 (12%)
Model	14 (21%)
Study feasibility	0 (0%)
Anticipated bottleneck(s)	7 (10%)
Total number of comments	68

Table B - Report sections for which critical commentary was collected for appropriate use assessment (T=4) and the total number of comments per section.

Appropriate use assessment (T=4) report section	**Number of critical comments made (percentage of total comments; %)**
Research question	0 (0%)
Indication	0 (0%)
Patient population	9 (16%)
Comparator treatment	3 (5%)
Outcomes (effects)	7 (12%)
Data collection method	6 (10%)
Representativeness of included patients (generalisability)	4 (7%)
Appropriate use in clinical practice	7 (12%)
Clinical effectiveness	10 (17%)
Quality of Life	12 (21%)
Total number of comments	58

Table C - Report sections for which critical commentary was collected for cost-effectiveness assessment (T=4) and the total number of comments per section.

Cost-effectiveness assessment (T=4) report section	**Number of critical comments made (percentage of total comments; %)**
Patient population	11 (9%)
Comparator treatment	2 (2%)
Outcomes (effects)	14 (11%)
Outcomes (costs)	20 (16%)
Model structure	15 (12%)
Input parameters	11 (9%)
Analysis technique	0 (0%)
Study perspective	2 (2%)
Time Horizon	1 (0%)
Discounting	0 (0%)
Assumptions	0 (0%)
Sensitivity analyses (planned)	10 (8%)
Incremental & total costs	10 (8%)
Incremental & total effects	5 (4%)
ICER	2 (2%)
Sensitivity analyses (results)	12 (10%)
Other	8 (7%)
Total number of comments	123

Table D - Reports available per component of T=0 and T=4 reports published for finalized drugs.

	T=0 Dossier				
Finalized drug	**Pharmacotherapeutic assessment**	**Cost-effectiveness assessment**	**Budget Impact Analysis**	**Outcomes research proposal for appropriate use**	**Outcomes research proposal for cost-effectiveness**
alglucosidase alpha	Yes	Yes[b]	Yes	Yes	Yes
agalsidase alpha	Yes	Yes[b]	Yes	Yes	Yes
agalsidase beta	Yes	Yes[b]	Yes	Yes	Yes
eculizumab	Yes	Yes[b]	Yes	Yes	Yes
rituximab	Yes	Yes	No	Yes	Yes
natalizumab	Yes	Yes	Yes	Yes	Yes
trastuzumab	Yes	Yes	No	No	No
omalizumab	Yes	Yes	Yes	Yes	Yes
voriconazol	Yes	Yes	Yes	Yes	Yes
ranibizumab	Yes	Yes	Yes	Yes	Yes
methyl aminolevulinate	Yes	Yes	Yes	Yes	Yes
pemetrexed	Yes	Yes	Yes	Yes	Yes

[a] As outlined in the guideline for outcomes research, published by ZIN in 2008.

[b] MAH did not submit a full cost-effectiveness assessment. Instead, plans for a cost-effectiveness analysis were proposed. These plans were subsequently assessed.

	T=4 Dossier			
Value of Information analysis[a]	**Pharmacotherapeutic assessment**	**Report on outcomes research for appropriate use**	**Cost-effectiveness assessment**	**Budget Impact Analysis**
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	No	Yes	Yes

Table E - The committee that conducted the appraisal of evidence at T=4 in relation to the 4 package criteria for finalized drugs.

Finalized drug	Committee conducting appraisal of evidence
alglucosidase alpha	Appraisal Committee
agalsidase alpha	Appraisal Committee
agalsidase beta	Appraisal Committee
eculizumab	Appraisal Committee
rituximab	Assessment Committee
natalizumab	Assessment Committee
trastuzumab	Assessment Committee
omalizumab	Appraisal Committee
voriconazol	Assessment Committee
ranibizumab	Appraisal Committee
methyl aminolevulinate	Appraisal Committee
pemetrexed	Assessment Committee

Table F - Assessment Committee's conclusions on the scientific adequacy of evidence submitted at T=4 to answer quest

Finalized drug	Conclusions on quality of evidence for appropriate use (AU) submitted at T=4	Conclusions on appropriate use (AU) in clinical practice
alglucosidase alpha	Sufficient	AU confirmed
agalsidase alpha	Sufficient	AU confirmed
agalsidase beta	Sufficient	AU confirmed
eculizumab	Sufficient	AU not confirmed; implemented in broader population than intended
rituximab	Sufficient	AU confirmed
natalizumab	Sufficient	AU confirmed
trastuzumab	Sufficient	AU confirmed
omalizumab	Not sufficient	AU not confirmed; no conclusions could be reached based on evidence submitted
voriconazol	Sufficient	AU confirmed
ranibizumab	Not sufficient	AU not confirmed; no conclusions could be reached based on evidence submitted
methyl aminolevulinate	Sufficient	AU confirmed
pemetrexed	Not sufficient	AU not confirmed; no conclusions could be reached based on evidence submitted

Abbreviations: ICER: Incremental Cost-Effectiveness Ratio; QALY: Quality-Adjusted Life-Year.

tions raised at T=0, appropriate use (AU) at T=4 and cost-effectiveness (CE) at T=4 per finalized drug.

	Conclusions on quality of evidence for cost-effectiveness (CE) submitted at T=4	Conclusions on cost-effectiveness (CE) in clinical practice
	Sufficient	ICERs presented substantiated by evidence submitted and are above threshold value of €80,000/QALY.
	Sufficient	ICERs presented substantiated by evidence submitted and are above threshold value of €80,000/QALY.
	Sufficient	ICERs presented substantiated by evidence submitted and are above threshold value of €80,000/QALY.
	Not sufficient	ICERs presented not substantiated by evidence submitted and are above threshold value of €80,000/QALY.
	Not sufficient	Due to absence of added therapeutic value and comparable costs with the comparator, there is little risk for incurring high ICER´s. Therefore, despite insufficient evidence, no additional data should be collected.
	Not sufficient	ICERs presented not substantiated by evidence submitted. No conclusion can be reached regarding CE.
	Sufficient	The calculated ICER of €15,535/QALY are acceptable and well substantiated.
	Not sufficient	ICERs presented not substantiated by evidence submitted. No conclusion can be reached regarding CE.
	Sufficient	The cost-effectiveness of Vfend compared to L-Amb (comparator) is favourable.
	Not sufficient	ICERs presented not substantiated by evidence submitted. No conclusion can be reached regarding CE.
	Sufficient	ICERs presented substantiated by evidence submitted and are above threshold value of €80,000/QALY.
	Sufficient	At a reference ICER value of €80,000/QALY, the probability for Pemetrexed being cost-effective is between 10% to 40%. Should it go out of patent soon, cost-effectiveness will be improved due ot the introduction of generic products.

Table G - ZIN advice on reimbursement of finalized drugs based on advice of its Committees.

Finalized drug	ZIN advice
alglucosidase alpha	Keep drug in basic healthcare package based on certain conditions.
agalsidase alpha	Keep drug in basic healthcare package based on certain conditions.
agalsidase beta	Keep drug in basic healthcare package based on certain conditions.
eculizumab	Remove drug from basic healthcare package.
rituximab	Keep drug in basic healthcare package
natalizumab	Keep drug in basic healthcare package based on certain conditions.
trastuzumab	Keep drug in basic healthcare package
omalizumab	Keep drug in basic healthcare package based on certain conditions.
voriconazol	Keep drug in basic healthcare package
ranibizumab	Remove drug from basic healthcare package.

Extra conditions

- Temporarily continue reimbursement of the drug from the basic healthcare package.
- Develop a separate financial framework for drugs for orphan diseases.
- Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases.
- Negotiate price negotiations with the marketing authorisation holder (MAH).
- Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification).
- Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria.
- Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.
- Temporarily continue reimbursement of the drug from the basic healthcare package.
- Develop a separate financial framework for drugs for orphan diseases.
- Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases.
- Negotiate price negotiations with the marketing authorisation holder (MAH).
- Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification).
- Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria.
- Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.
- Temporarily continue reimbursement of the drug from the basic healthcare package.
- Develop a separate financial framework for drugs for orphan diseases.
- Transfer the reimbursement of the drug to the new framework specific to drugs for orphan diseases.
- Negotiate price negotiations with the marketing authorisation holder (MAH).
- Discuss with clinicians if, and how, costs per QALY can be reduced (e.g. through dose modification).
- Demand the necessary parties to set up a (European) study to investigate predictive factors for clinical effectiveness, develop start- & stop-criteria and develop a more transparent system for the implementation of start- and stop-criteria.
- Consider establishing an independent committee to advise clinicians in practice on start- and stop-decisions for treatment with this drug.

N/A

N/A

ZIN postpones its final decision for removal of this drug from the basic healthcare package until results from the [separate] Round Table on Multiple Sclerosis are presented.

N/A

To guarantee continued reimbursement, the marketing authorisation holder (MAH) should sign Pay-for-Performance (PfP) agreements with all hospitals whereby the drug will be prescribed. In the case of defaults on PfP agreements (e.g. due to lack of cooperation from individual hospitals or no refunds to hospitals based on outcomes), ZIN will advise for the removal of this drug from the basic healthcare package.

N/A

N/A

Table G - continued

Finalized drug	ZIN advice
methyl aminolevulinate	Keep drug in basic healthcare package based on certain conditions.
pemetrexed	Keep drug in basic healthcare package

Abbreviations: QALY: Quality-Adjusted Life-Year; ZIN: Zorginstituut Nederland.

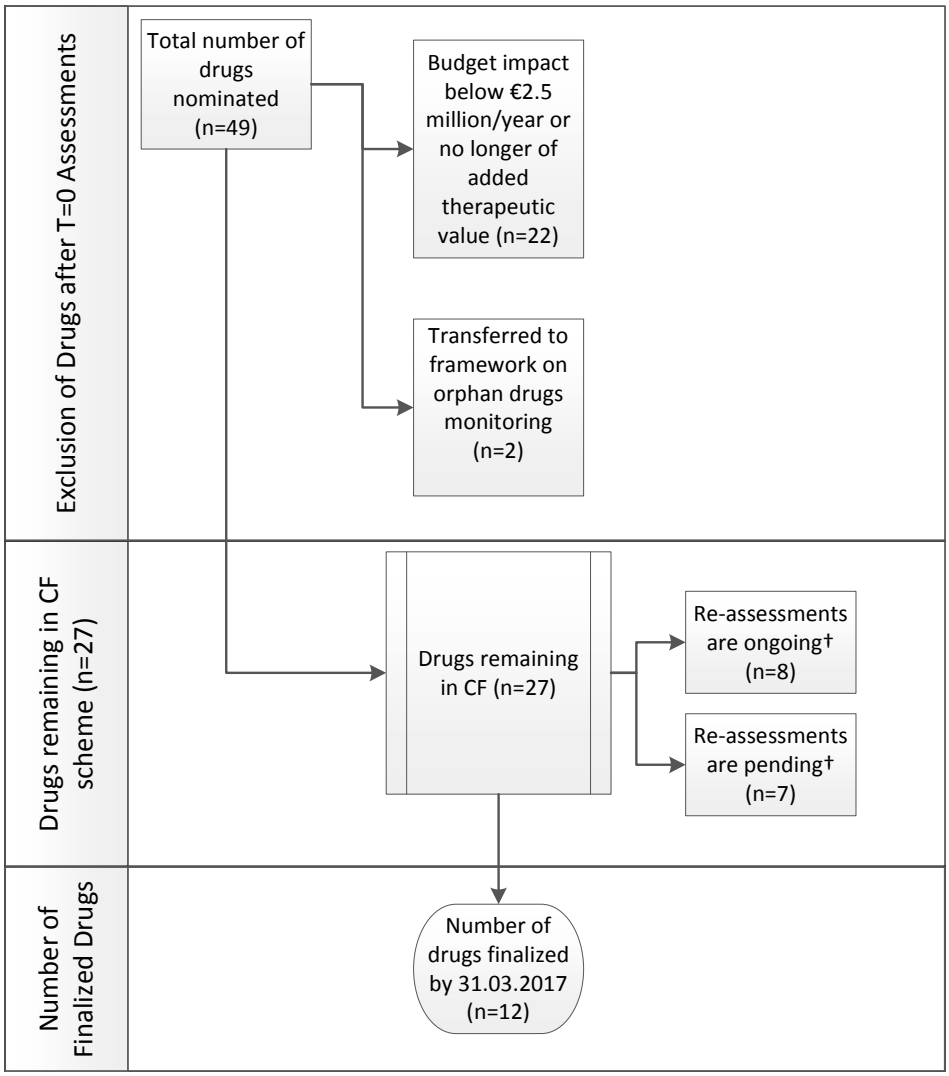


Figure A - Overview of in- and exclusion and status of drugs for the Conditional Financing (CF) scheme. † No information regarding the status of these drugs was available in public documents on the ZIN website. The authors were thus obliged to retrieve the relevant information from assessors within ZIN.

Extra conditions
ZIN requests the clinicians' societies to update the clinical guideline, in order to clarify and specify the criteria for treatment with methyl aminolevulinate thus ensuring that the implementation of such criteria becomes feasible in practice.
N/A

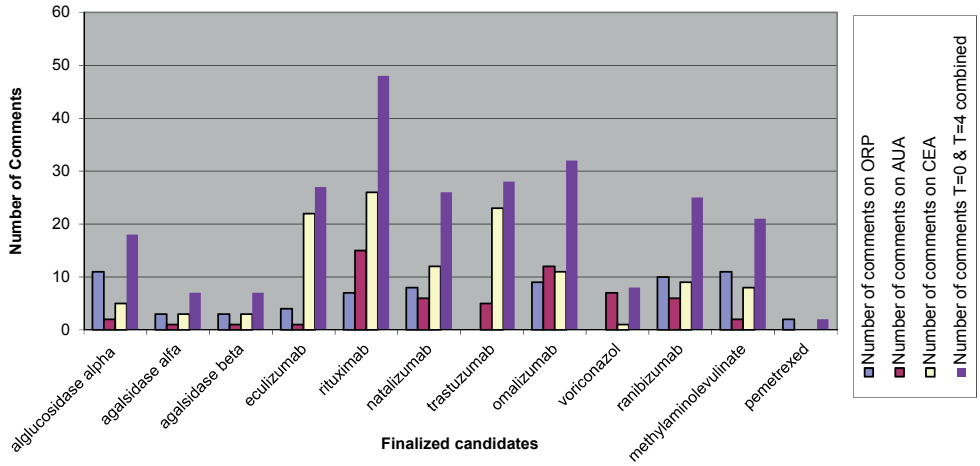


Figure B - The number of critical comments provided by the Assessment Committee on the outcomes research proposal (ORP; T=0), appropriate use assessment (AUA; T=4), cost-effectiveness assessment (CEA; T=4), and total number of comments for the finalized drugs.

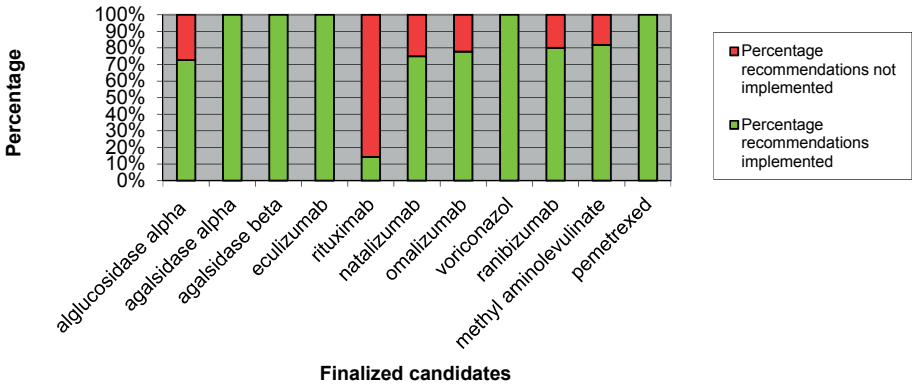


Figure C - Percentage of recommendations made by the Assessment Committee regarding the outcomes research proposal (T=0) implemented in the outcomes research study per finalized drug.

CHAPTER 6 - APPENDIX

Table 1 – Roles of different stakeholders in the conditional financing (CF) scheme.

Stakeholders	Role at T=0	Role during outcomes research phase
External Public Bodies (i.e. non-ZIN)	• The Netherlands Organization for Health Research and Development (ZonMW) provided methodological feedback on the outcomes research proposals for some drugs.	• ZonMW financed a limited number of outcomes research studies (n=2) implemented as part of the CF scheme.
Zorginstituut Nederland (ZIN; previously CVZ), including: Assessment Committee, Appraisal Committee and Assessors	• Assess and appraise the scientific evidence submitted by the MAH on therapeutic value of the drug, budget impact analysis and preliminary cost-effectiveness analysis. • Provide methodological feedback on the outcomes research proposals for drugs.	N/A
Pharmaceutical Industry	• Prepare and submit evidence on therapeutic value of the drug, budget impact analysis and preliminary cost-effectiveness analysis. • Prepare and submit an outcomes research proposal to provide evidence on appropriate use and cost-effectiveness of the drug.	• Collaborate with academic/private hospitals, medical specialists and medical societies to implement the outcomes research study. • Finance the implementation of outcomes research studies.
Healthcare Insurers	N/A	N/A
Medical Specialists Societies	• Provide input during the establishment of outcomes research proposals. • Provide feedback on preliminary versions of ZIN reports at T=0.	• Implement outcomes research studies in practice (i.e. data collection).
Academic/ Private Hospitals	N/A	• Collaborate with pharmaceutical industry, medical specialists and medical society to facilitate outcomes research studies.
Patient Organizations	• Provide feedback on preliminary versions of ZIN reports at T=0.	N/A

Abbreviations: CVZ: College voor Zorgverzekeringen; MAH: Marketing Authorization Holder; NZa: Nederlandse Zorgautoriteit; ZIN: Zorginsitituut Nederland; ZonMW: De Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie.

Role at T=4	Other
N/A	• The Dutch Healthcare Authority (NZa) established the underlying policy framework for CF. • The Ministry of Health (VWS) was responsible for the ultimate decision to keep, or remove, a drug from the reimbursement package after T=4 based on ZIN's advice. • ZonMW provided feedback on the proposed design for the CF scheme in 2006.
• Re-assess and re-appraise the scientific evidence submitted by the MAH on therapeutic value of the drug, budget impact analysis, **appropriate use and cost-effectiveness analysis**. • Issue final advice to VWS to keep, or remove, a drug from the reimbursement package.	• Was responsible for designing and implementing the CF scheme in practice. • Members of the Assessment and Appraisal Committees provided feedback on the proposed design for the CF scheme in 2006.
• Prepare and submit evidence for the re-assessment of therapeutic value of the drug, budget impact analysis, **appropriate use and cost-effectiveness analysis**.	• Provided feedback on the proposed design for the CF scheme in 2006.
N/A	• Provided feedback on the proposed design for the CF scheme in 2006. • Reimbursement of CF drugs from T=0 onwards based on hospital claims.
• Provide feedback on preliminary versions of ZIN reports at T=4.	• Some medical specialists provided feedback on the proposed design for the CF framework in 2006.
N/A	N/A
• Provide feedback on preliminary versions of ZIN reports at T=4.	N/A

Table 2 – The completed consolidated criteria for reporting qualitative studies (COREQ) 32-item checklist.
Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

No. Item	Guide questions/description	Response
Domain 1: Research team and reflexivity		
Personal Characteristics		
1. Inter viewer/facilitator	Which author/s conducted the interview or focus group?	AM and HN conducted the interviews for the public stakeholders. AM and SA conducted the interviews for the private stakeholders.
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	AM: Pharm D, MSc. HN: BSc. SA: BSc.
3. Occupation	What was their occupation at the time of the study?	AM: Policy Advisor and PhD candidate HN: Master's student SA: Master's student
4. Gender	Was the researcher male or female?	AM: Male HN: Male SA: Female
5. Experience and training	What experience or training did the researcher have?	AM: Conducted several qualitative research projects published in scientific literature. HN: Followed courses on qualitative research methods within Master's program. SA: Followed courses on qualitative research methods within the Master's program.

Table 2. - continued

No. Item	Guide questions/description	Response
Relationship with participants		
6. Relationship established	Was a relationship established prior to study commencement?	AM: Through his employment at ZIN, AM had professional relationships with some interviewees from ZIN. AM had no professional relationships with any of the remaining stakeholders. HN: No relationships with any of the interviewed stakeholders prior to study commencement. SA: No relationships with any of the interviewed stakeholders prior to study commencement.
7. Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	AM: Participants were informed of the affiliations of the interviewer and the research aims. HN: Participants were informed of the affiliations of the interviewer and the research aims. SA: Participants were informed of the affiliations of the interviewer and the research aims.
8. Interviewer characteristics	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	By explicitly informing interviewees of the affiliations of the interviewers and aims of the research project beforehand, interviewees were aware of any biases/ conflicts of interest of the interviewers.
Domain 2: study design		
Theoretical framework		
9. Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	Directed content analysis was the methodological orientation of this study as detailed by Hsieh et al. (2005).

Table 2. - continued

No. Item	Guide questions/description	Response
Participant selection		
10. Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	A combination of purposive sampling and snowballing was used to select participants. In first instance, the authors established a list of stakeholders based on their knowledge of the Conditional Financing framework (purposive). Additionally, potential interviewees contacted were provided the chance to recommend further interviewees; whether colleagues from the same institution or other stakeholders (snowballing).
11. Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Participants were approached by email using a standardized email invitation.
12. Sample size	How many participants were in the study?	In total, there were 35 interviewees involved, representing 30 stakeholders.
13. Non-participation	How many people refused to participate or dropped out? Reasons?	None of the interviewees approached refused to participate or dropped out.
Setting		
14. Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	The data was either collected through face-to-face interviews at the workplace of the interviewees or through the phone at the interviewers' workplace.
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	No, only the researchers and participants were present during interviews.
16. Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	The sample is fully described in the tables accompanying the manuscript. Importantly, the sample consisted of interviewees with senior functions. All interviewees had academic degrees equivalent to a master of science (MSc) degree or higher. The sample consisted of 17 (49%) females and 18 (51%) males. Of the 30 interviews conducted, 24 (80%) included one interviewee and 6 (20%) included two or more.

Table 2. - continued

No. Item	Guide questions/description	Response
Data collection		
17. Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	An interview guide was developed in 2016 for interviewing phase 1 stakeholders. The guide was pilot tested in the first 3 interviews. Subsequently, the guide was finalized and used for the remaining interviews. The same interview guide was used in 2017 for interviewing phase 2 stakeholders. The guide addressed the following domains: perceived aims of CF, perceived functioning of CF, impact of CF and conclusions on CF and future perspectives.
18. Repeat interviews	Were repeat inter views carried out? If yes, how many?	No repeat interviews were conducted.
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Audio recording was used for 29 (97%) interviews. In only 1 (3%) interview did the stakeholder refuse to have the interview recorded.
20. Field notes	Were field notes made during and/or after the interview or focus group?	Field notes were made during and after the interviews.
21. Duration	What was the duration of the inter views or focus group?	Each interview lasted between 60 and 90 minutes.
22. Data saturation	Was data saturation discussed?	Data saturation was discussed amongst authors. This discussion is what motivated the decision not to conduct more interviews.
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	Extensive summaries that were made based upon audio recordings were sent to the interviewees for comment and correction.
Domain 3: analysis and findings		
Data analysis		
24. Number of data coders	How many data coders coded the data?	AM and HN coded the data for phase 1 stakeholders. AM and SA coded the data for phase 2 stakeholders.

Table 2. - continued

No. Item	Guide questions/description	Response
25. Description of the coding tree	Did authors provide a description of the coding tree?	The basic framework is based upon the domains and questions stated in the interview guide. Please see Figure 3 in the Appendix for the coding tree developed.
26. Derivation of themes	Were themes identified in advance or derived from the data?	Themes were derived from the data.
27. Software	What software, if applicable, was used to manage the data?	MaxQDA version 11 (release 11.1.2, build 151125) was used.
28. Participant checking	Did participants provide feedback on the findings?	Interviewees have been provided the chance to provide feedback on the manuscript.
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Quotations have been presented to illustrate themes and findings in the manuscript.
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Consistency between the data and findings was reviewed by the authors of the manuscript.
31. Clarity of major themes	Were major themes clearly presented in the findings?	Clarity of the presentation of the major themes was reviewed by the authors of the manuscript.
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Description of diverse cases and discussion of minor themes was reviewed by the authors of the manuscript and by interviewees during the feedback period.

Table 3 – Summary of the results of directed content analysis. The themes identified for all topics are cited in order of decreasing mention for all stakeholders, phase 1 stakeholders and phase 2 stakeholders.

Topic	All Stakeholders (n=30; %)	Phase 1 Stakeholders (n=14; %)	Phase 2 Stakeholders (n=16; %)
Functioning of CF – Procedural aspects	4-year time period (90%) Ambiguity regarding roles (50%) Distribution of roles in relation to conflicting of interests (43%) Lack of monitoring mechanisms (27%) Financing of outcomes research studies (27%)	The 4-year time period. (86%) Disincentives due to conflicts of interest and procedural ambiguity. (71%) Ambiguous procedure and distribution of roles across stakeholders. (43%) Monitoring of progress throughout 4-year period. (36%) Financing of outcomes research. (36%)	The 4-year time period. (94%) Ambiguous procedure and distribution of roles across stakeholders. (56%)
Functioning of CF – Methodological aspects	No clear methodological requirements for outcomes studies (e.g. study design and cut-off points) (40%) Poor methodological quality of outcomes research. (33%) Rapid changes in clinical practice influence value of outcomes research. (33%) Excessive collection of data (e.g. additional, non-critical parameters) (27%)	No clear methodological requirements for outcomes studies (e.g. study design and cut-off points). (64%) Poor methodological quality of outcomes research. (57%) Rapid changes in clinical practice influence value of outcomes research. (36%) Low patient recruitment rates for outcomes research. (36%) Excessive collection of data (e.g. additional, non-critical parameters). (29%) Lack of implementation of advice on study design at T=0. (29%)	Rapid changes in clinical practice influence value of outcomes research. (31%) Excessive collection of data (e.g. additional, non-critical parameters). (25%) Difficult to set up patient registries. (25%)

Table 3 - continued

Topic	All Stakeholders (n=30; %)	Phase 1 Stakeholders (n=14; %)	Phase 2 Stakeholders (n=16; %)
Functioning of CF – Decision-making aspects	External (political) factors influenced final decisions at T=4. (50%) Minimal contribution of results from outcomes research to final decisions at T=4. (43%) Impossible to remove drugs from the national reimbursement package beyond T=4. (40%) Conclusions on cost-effectiveness and/or appropriate use were apparent from the start of the process. (33%)	Impossible to remove drugs from the national reimbursement package beyond T=4. (64%) Minimal contribution of results from outcomes research to final decisions at T=4. (64%) External (political) factors influenced final decisions at T=4. (57%) Conclusions on cost-effectiveness and/or appropriate use were apparent from the start of the process. (43%) Influence of legal criteria on effectiveness versus non-binding criteria on cost-effectiveness. (29%)	External (political) factors influenced final decisions at T=4. (44%) Minimal contribution of results from outcomes research to final decisions at T=4. (25%) Conclusions on cost-effectiveness and/or appropriate use were apparent from the start of the process. (25%)
Impact of CF – Positive effects	Cost-effectiveness and displacement of healthcare became important topics of public debate. (53%) More (real-world) evidence generated on the use of drugs in routine clinical practice. (40%) (Policy) Experiences gained in the implementation of managed entry agreements. (33%) Increased awareness of need for multi-stakeholder collaboration. (27%)	Cost-effectiveness and displacement of healthcare became important topics of public debate. (79%) (Policy) Experiences gained in the implementation of managed entry agreements. (36%) More (real-world) evidence generated on the use of drugs in routine clinical practice. (36%) Increased awareness of need for multi-stakeholder collaboration. (29%)	More (real-world) evidence generated on the use of drugs in routine clinical practice. (44%) (Policy) Experiences gained in the implementation of managed entry agreements. (31%) Cost-effectiveness and displacement of healthcare became important topics of public debate. (31%) Increased awareness of need for multi-stakeholder collaboration. (25%) More appropriate use of drugs in clinical practice. (25%)

Table 3 - continued

Topic	All Stakeholders (n=30; %)	Phase 1 Stakeholders (n=14; %)	Phase 2 Stakeholders (n=16; %)
Impact of CF – Negative effects	No major or minor themes identified.	CF perceived as a "back-door" into the national reimbursement package and an excuse to postpone difficult decisions. (36%)	No major or minor themes identified.
Conclusions and future perspectives - Improvement of CF	Agreement amongst SH's on aims of policy framework and better SH collaboration. (37%)	Agreement amongst SH's on aims of policy framework and better SH collaboration. (50%)	Better governance and distribution of roles throughout the CF procedure. (38%)
	Re-consideration of underlying incentives in policy framework in relation to diverging SH interests. (33%)	Re-consideration of underlying incentives in policy framework in relation to diverging SH interests. (43%)	Make (definitive) conclusions on cost-effectiveness at T=0. (31%)
	Better governance and distribution of roles throughout the CF procedure. (30%)	Need for extra steps for monitoring of progress per drug throughout CF process. (43%)	Agreement amongst SH's on aims of policy framework and better SH collaboration. (25%)
	Need for extra steps for monitoring of progress per drug throughout CF process. (30%)	Make (definitive) conclusions on cost-effectiveness at T=0. (29%)	Abandon rigid 4-year timeframe. (25%)
	Make (definitive) conclusions on cost-effectiveness at T=0. (30%)		Re-consideration of underlying incentives in policy framework in relation to diverging SH interests. (25%)

Abbreviations: CF: Conditional Financing.

Table 4 – Illustrative quotes for themes identified in interviews.

Topic	Theme	Quotes from phase 1 interviews
Functioning of CF – Procedural aspects	4-year time period	"This [4-year period, ed.] was a too rigid and short timeframe. One needs tailored approaches to be able to determine how much time would be needed to generate sufficient evidence on appropriate use or cost-effectiveness for drugs in different disease indications." – ZW1
Functioning of CF – Procedural aspects	Ambiguity regarding roles	"Even worse, it was not clear to us which stakeholder was directly responsible for delivering the data from outcomes research to ZIN (i.e. the pharmaceutical company or healthcare providers) making it difficult to hold a specific stakeholder accountable." - ZA1 "It was never clear to me which stakeholder eventually made the decisions in this framework. For example: who governs the framework? What were the roles of the appraisal committee (ACP), ZonMW and VWS in the decision-making process in this framework?" – ZA3
Functioning of CF – Procedural aspects	Distribution of roles in relation to conflicting of interests	"The actual implementation of data collection was in the hands of pharmaceutical industry and medical specialists who had little incentive to conduct robust research on cost-effectiveness. Should a negative advice on reimbursement be issued by ZIN at T=4 due to the drugs being not cost-effective (which was to be expected for many of these drugs), the initial decision for reimbursement would be reversed. This would mean that the drug would no longer be reimbursed, causing problems for both stakeholders." – ZW2
Functioning of CF – Procedural aspects	Lack of monitoring mechanisms	"Throughout my involvement, there was no mention of monitoring of outcomes research studies [by ZIN, ed.] There wasn't even mention of a mid-term evaluation of progress of the studies. If one wants ZIN to monitor this, then ZIN should also be given the authority to end outcomes research studies that seem to falter or that are of inadequate quality." – ZS4 "It wasn't the case that better evidence was to be expected if the outcomes research studies would have proceeded longer. On the other hand, a more direct involvement by ZIN in evaluating the progress of the studies may have delivered better results; for example through yearly progress meetings." – FG1

Quotes from phase 2 interviews

"We have repeatedly stated that the 4-year period is impractical. It may have been motivated by the parliamentary period of 4 years. However, for the context of drugs, this choice is illogical because for some drugs one knows quickly if they work, while for others it could take a long time before that happens. For example, extra overall survival of 7 months can be adequately captured in a registry within 4 years. However, for overall survival period of up to 9 to 10 years in metastatic breast cancer, a 4-year registry would not be informative." – MS3

"The public perception of CF was that it was a tug-of-war between ZIN and pharmaceutical companies. That meant that little attention was paid to the responsibilities of other parties, such as healthcare providers and medical specialists, who were also responsible for the appropriate use of drugs in clinical practice." – PI1

"Although it was often clear at T=0 what should have been done, there was much confusion regarding T=4. In that final year, it was unclear what each stakeholder's responsibilities were and why. The uncertainty was thus increasingly obvious towards the end of the process." – MS2

"Data collection was the responsibility of clinicians and medical specialists. However, they were often less willing to cooperate with industry on doing so. Without the support of the clinicians and medical specialists, data collection cost excessive time and the quality of data collected was affected." – PI3

"Monitoring mechanisms should have been designed and implemented at the start of the CF process. It becomes useless to inspect results at T=4, otherwise. By then, it is difficult to correct for any errors that have happened along the way." – HI2

"Another aspect of CF that did not function well was the monitoring and auditing of research progress throughout the whole procedure." – PO1

Table 4 - continued

Topic	Theme	Quotes from phase 1 interviews
Functioning of CF – Procedural aspects	Financing of outcomes research studies	"In the first round of CF, several outcomes research studies were financed by ZonMW. However, some stakeholders had the feeling that this financing mechanism was forced upon them. In subsequent rounds of CF, ZonMW funds were only used for specific outcomes studies, with the remaining studies being financed by pharmaceutical industry. Within the study protocols, some research questions were made mandatory and other non-mandatory questions were raised by medical specialists. Many financing structures emerged, with numerous research questions and numerous resources that research teams attempted to address." – PE3 "It is quite disappointing to see how the financing structure of outcomes research studies removed all authority from the hands of public bodies. It was not wise to let pharmaceutical industry finance outcomes research studies on their own products." – ZA3 "Another problem that emerged was the lack of independence (i.e. independent research, independent financing of research). Two outcomes research studies performed by ZonMW [public body, ed.] were assessed with a positive outcome. In these cases there was not financing through industry thus no influence on the outcomes of the research." – FG1
Functioning of CF – Methodological aspects	No clear methodological requirements for outcomes studies	"Throughout the CF framework, there was no consensus on the design of the outcomes research studies. Some stakeholders expected RCT's to be conducted, while others did not. Furthermore, there was no consensus on the aims and expectations of the studies. This resulted in huge databases with a predominantly clinical perspective, rather than a cost-effectiveness perspective. The infrastructure required to collect the data was also not readily available." – ZW2 ZS3: "There was not enough time to reflect on which outcomes parameters ZIN deemed relevant. The pharmaceutical industry also raised critical questions: they wanted to know when their product would be deemed "adequate" for reimbursement based on such outcome parameters. No answer could be provided to such questions since ZIN also had little experience with these drugs/ indication areas at the time." – ZS3
Functioning of CF – Methodological aspects	Low quality of outcomes research	"The quality of information submitted at T=4 and the quality of outcomes research studies at T=4 was quite poor. Elements such as low patient recruitment, fragmented data collection in clinical practice and the change in standard clinical practice throughout the period of 4 years really affected the relevance of the evidence." – ZS2 "A main problem encountered with outcomes research studies was the absence of a control group or that the intervention and control groups were not suitable for comparisons. In the latter case, patients who did not wish to be treated with the new drug became the control group. This undoubtedly led to selection bias, making any comparisons of little value." – FG1

Quotes from phase 2 interviews

"The governance and financing of patient registries differed significantly per registry." – PI3

"If you bring a large number of stakeholders with different interests together, you get a huge set of variables for which information is desired. In contrast, we should be looking to core datasets that really matter within smaller patient populations. We should also be looking to better incorporation of patient perspectives in outcomes research studies." – PI1

"It was not the case that data collection did not happen but rather a lack of knowledge on what you exactly want to collect and in which patient groups. Therefore, it is a fundamental design error. In some cases, extra variables were requested by ZIN while the studies were running. These were subsequently added to the protocol. From a methodological viewpoint, this cannot be allowed in prospective research." – MS3

"It was not the case that data collection did not happen but rather a lack of knowledge on what you exactly want to collect and in which patient groups. Therefore, it is a fundamental design error." – MS3

Table 4 - continued

Topic	Theme	Quotes from phase 1 interviews
Functioning of CF – Methodological aspects	Rapid change in clinical practice	"It was often the case that a second-line drug would be included into a registry, only to find out that it has been registered for first-line treatment a year later." – PE3 "Drug development is happening at a rapid pace, particularly in oncology. As a result, the standard of care changed several times throughout the CF procedure, complicating the use of evidence generated by the outcomes research studies in pharmacoeconomic models." – ZS2
Functioning of CF – Decision-making aspects	Effect of external (political factors) on ZIN's advice	"This became very apparent when the Appraisal Committee (ACP) had to issue advice for the first T=4 dossiers. The ACP could have stayed its ground on negative advice on reimbursement but could also not punish the citizen for the non-robust evidence generated by the pharmaceutical industry." – ZA1 "External political factors certainly had an influence on the advice ZIN eventually issued at T=4. In the summer of 2012, the CVZ/ZIN was daily mentioned in the evening news in association with drugs for Pompe's and Fabry's disease. Patients also deliver pressure by (rightly) saying that they should not be denied hope. Emotional responses and accusations on that front definitely impact the decisions being made." – ZS1
Functioning of CF – Decision-making aspects	Outcomes research contributed little to decision making T=4	"In my opinion, it still remains the question whether the outcomes research studies actually delivered more useful information at T=4 in comparison to the information available at T=0. Even after conduct of the studies, the uncertainties relating to the cost-effectiveness of drugs remained large at T=4." – PE2 "A fundamental problem resides in the question: what can we do with observational, real-world data? There are many methodological challenges associated with the analysis of observational data, decreasing trust in the outcomes at T=4. People said too easily at the beginning that RCTs would not be conducted within CF and that we will rely on observational data for our decisions. This just doesn't work." – ZA3
Functioning of CF – Decision-making aspects	Impossible to remove drugs from reimbursement package at T=4	"Unfortunately, the initial decision to remove a set number of drugs from the national reimbursement package had to be reversed. The legal implications of the decision were judged to be too large." – PE1 "It is very difficult to remove something that is provided temporarily but experienced as a permanent solution by the patient." – ZS1
Functioning of CF – Decision-making aspects	Conclusions at T=4 already predictable at T=0	"In many cases we warned that the collection of new information on these drugs was useless. We had enough evidence at T=0 to already conclude that the drug would never be cost-effective. These warnings were ignored. So in some cases, you could say that the conclusions at T=4 were already obvious at T=0." – ZW2 "One should have been wondering: do we really need these outcomes research studies? Did we not already know the answers before, or during, the studies?" – ZS4

Quotes from phase 2 interviews

"One disadvantage of CF is that it takes too long, while the healthcare environment is very dynamic and where many new drugs reach the market. Research designed now can therefore not mean much 2 years later." – HI3

"In the field of oncology, many drugs rapidly reach the market because of which treatments can move through different treatment lines or be administered in different combinations. Because of that, the drugs we choose to assess at T=0 are no longer the relevant ones by T=4. This is the case in hemato-oncology, metastatic melanoma and prostate cancer… Therefore, the data on initial use patterns are no longer relevant." – PO2

"With these types of frameworks, it becomes extremely difficult to remove a product from the reimbursement package. This is usually due to external political pressure by stakeholders. In my opinion, ZIN should dare to say "no" [to reimbursement, ed.]. more often." – HI1

"External factors certainly had an influence, given the examples of Pompe's and Fabry's diseases. The public outcry around these two drugs left a lasting impact on how ZIN proceeded with the remaining drugs in CF." – PO2

"The re-assessment of drugs by ZIN was quite technocratic; ZIN only wanted to tick off all the items on the submissions checklist. As a result, one lost sight of the bigger picture." – MS1

"ZIN repeatedly came to the conclusion that the data at generated through outcomes research studies by T=4 was of little use because of patient populations being too small to generate significant evidence." – MS3

"Once you have treated a patient with a drug while it is in the national reimbursement package, it becomes difficult to remove it from the package. Your negotiation power and argumentation to do so is immediately diminished." – HI3

"If patients have already been treated using this drug for 10 years already, we should also take our responsibility as decision makers and not suddenly decide to cut the drug out of the reimbursement package." – HO2

"Outcomes research on omalizumab provided new insights on the drug but no new conclusions." – PI2

"We were excited to being with our outcomes research. Not because we didn't know what the conclusion would be: that we already knew. We knew the drug cost €200,000 per patient and that we theoretically cannot generate more than 1 QALY per patient. There were many other unsolved, interesting questions though." – HO2

Table 4 - continued

Topic	Theme	Quotes from phase 1 interviews
Impact of CF scheme – Positive effects	Societal debate on cost-effectiveness and displacement of healthcare	"One of the positive effects of CF is the debate on costs and cost-effectiveness. Due to the agreement on growth in health costs not exceeding 1% to 1.5%, the coming of these new drugs displaced other drugs in the package which could be more cost-effective and deliver more to society. The societal awareness of this displacement and its consequences was boosted by CF." – ZA2
Impact of CF scheme – Positive effects	Valuable real-world evidence generated	"Evidence generated through outcomes research studies allowed some of the involved medical societies to develop start- and stop- criteria for treatment administration." – FG1
Impact of CF scheme – Positive effects	Policy experience gained	"If anything, we may have probably learned how we shouldn't do managed entry agreements in the future. Because of CF, a new line of thought has been set regarding the operationalisation of cost-effectiveness in decision making. Examples relate to the discussions on drugs that came after the CF scheme (e.g. with eculizumab and pertuzumab). These were the modern T=0's, where the lessons learnt from CF have been applied. " – ZS3
Impact of CF scheme – Positive effects	Increased awareness for collaboration	"Since the re-assessment of drugs for Pompe's and Fabry's disease, the HTA process has changed. It has become a more collaborative process, including collaboration even in the early stages of HTA (e.g. through scoping meetings). In that way, one knows more of the factors really impacting the indication field in question and you arrive at the best strategy with all stakeholders." – FG2
Conclusions & Future Perspectives - Improvement of CF	Consensus on scheme aims and importance of collaboration	"Partly due to learnings from CF, ZIN's working methods have shifted from one-sided HTA to more collaboration with other stakeholders on issues such as appropriate care and quality of healthcare. Looking to the future, medical societies, ZIN and healthcare insurers should discuss conditions for conditional reimbursement schemes together. In my opinion, healthcare insurers should also be the ones to impose sanctions (where needed) based on these conditions. As the payers, they're best positioned to do so. We should all take up our (new) responsibilities in future schemes." – ZS2
Conclusions & Future Perspectives - Improvement of CF	Underlying incentives and accountability	"One should not implement schemes such as CF without the correct sets of checks-and-balances. There should be a system for sanctions; without that, schemes such as these will not yield benefits. Moreover, the sanctions should be proportional to the effort stakeholders (including industry) should invest to meet the demands of the scheme." – ZS3 "Most importantly, do not finance the CF drugs out of the national reimbursement package. A better alternative would be a subsidiary, temporary funds structure. That way, stakeholders including medical specialists and patients would also be aware that re-assessment could affect access at a later stage." – FG1 "Consider making patient inclusion in outcomes research is obligatory in return for access to the drug. Otherwise, the current incentives of stakeholders can once again lead to underpowered, low-quality studies." – FG2

Quotes from phase 2 interviews

"Even doctors on the work floor began to think more about the costs associated with the treatments they prescribe and appropriate care, due to their heightened awareness of these topics." – MS2

"It's good to have seen that with the right research questions and parameters list, valuable information could have been generated about the treatment in practice. This should be done for all drugs, not only those in CF." – PI1

"The advantage of having gone through the CF experience is to have the knowledge to design future policy instruments, such as the Conditional Coverage scheme [other managed entry agreement currently implemented in the Netherlands*, ed.]." – HI1

"All stakeholders in the healthcare arena started collaborating more as a result of the discussions around CF. That is an important gain, even for ZIN. If you look at the facts, the evidence generated on drugs through CF may have been mediocre but the increased collaboration is remarkable." – MS3

"A lesson learned is to do this together, rather than sitting on opposite sides of the table. We should also have dialogues at an earlier stage of the process." – PI2

"I would like to see patient organizations being incorporated into all steps of the procedure. We could see with CF that important aspects were unknown to us, such as the establishment of start- and stop criteria for treatment." – PO3

"The scope, organization and financing of outcomes research studies and registries would best be conducted on a European level. With regards to financing both, specifically, industry should play an important role. In return, they would receive valuable information on their products to enable them to improve these products." – HO2

Table 4 - continued

Topic	Theme	Quotes from phase 1 interviews
Conclusions & Future Perspectives - Improvement of CF	Monitoring procedures and mid-term reviews	"In my opinion, there should be continuous monitoring of the outcomes research studies. The question is: how and who should do that? Additionally, if ZIN would have the mandate to do so, then it should also have the authority to stop the studies whereby monitoring results indicate little progress." – ZS3
		"There should be a pre-defined time point, or set of time points, when ZIN would monitor the progress made in the outcomes research studies." – FG1
Conclusions & Future Perspectives - Improvement of CF	Definitive conclusions on cost-effectiveness at T=0	"We can say that at initial assessment (T=0), much is already known regarding the cost-effectiveness of the new drug. Based upon that, agreements can be made on other aspects such as: appropriate use, financial-based agreements or restriction of reimbursement. We otherwise doubt the usefulness of re-assessment of cost-effectiveness at T=4." – PE2
		"Although the current form of CF served as a good basis for future efforts, we believe that we can say much more on the effectiveness and cost-effectiveness of drugs at T=0." – PE3

Abbreviations: CF: Conditional Financing; HTA: Health Technology Assessment; VWS: Ministerie voor Volksgezondheid, Welzijn en Sport; ZIN: Zo

Quotes from phase 2 interviews

"There should be harder rules with regards to acceptable and non-acceptable cost-effectiveness. As usually done in the United Kingdom by the NHS, we should define clear threshold values and say no if new drugs exceed these values. Other aspects (e.g. appropriate use) come later, after this system is established." – MS1

"The prices for these new drugs are too high. We should be quite strict regarding cost-effectiveness at T=0. What we're trying to do now with CF is lessen the rate of drainage, whereas we could better shut the tap at the beginning." – MS2

Table 5 – Unique themes identified through comparative analysis of phase 1 and phase 2 studies.

Topic	Theme (%)
Methodological aspects	4/14 (29%) of phase 1 stakeholders iterated that feedback provided on the outcome research study proposals by public bodies at T=0 was often not incorporated into the final studies implemented[ZS2,ZS4]; an important example relates to recommendations to include Health-Related Quality of Life (HRQoL) outcomes collection using the EuroQol 5-dimensional scale (EQ-5D)[FG2].
Decision-making aspects	4/14 (29%) of phase 1 stakeholders emphasized how national legislation in the Netherlands may have diminished the impact of ZIN advices at T=4. The health insurance law ("Zorgverzekeringswet") states that drugs with a scientifically proven added therapeutic value ("stand van de wetenschap en praktijk") should be part of the reimbursement package[17]. However, there are no equivalent clauses in current legislation on requirements pertaining to the cost-effectiveness of drugs in the reimbursement package. In the absence of such legislation, it became impossible to remove cost-ineffective drugs from the reimbursement package after re-assessment at T=4 from a legal perspective.
Positive effects of CF	4/16 (25%) of phase 2 stakeholders stated that the conduct of outcomes research studies in itself stimulated appropriate use of drugs in clinical practice. For example, in the case of treatments for metastatic melanoma, CF led to centralization of healthcare delivery to expertise centers and increased awareness on drug use in practice[MS2]. In the case of omalizumab for the treatment of severe asthma, the pay-for-performance scheme enticed adherence to strict start- and stop-treatment criteria in clinical practice[PO2].
Negative effects of CF	5/14 (36%) of phase 1 stakeholders experienced the scheme as a "back-door" to the reimbursement package and an excuse to postpone difficult decisions. In their opinion, drugs with relatively higher costs and relatively higher uncertainties on appropriate use and cost-effectiveness were admitted at T=0 under conditions of less scrutiny than for conventional drugs. In doing so, difficult decisions were shifted to T=4. By that time, it became apparent that the drugs would not be removed. In the words of one stakeholder "Once they [drugs] were in, there was no turning back"[ZW2].

Abbreviations: CF: Conditional Financing; ZIN: Zorginstituut Nederland.

Geachte [naam interviewee],

Gezien alle recente nationale en internationale discussies over flexibele vergoedingssystemen is vanuit Zorginstituut Nederland het plan opgevat om het gebruik van voorwaardelijke financiering, dat een onderdeel was van de beleidsregels dure en weesgeneesmiddelen (2006-2012), te evalueren. Het is de bedoeling dat de uitkomsten van deze evaluatie zullen bijdragen aan de inrichting van toekomstige vormen van voorwaardelijke vergoeding/financiering. De resultaten van dit onderzoek zullen ook worden gebruikt voor een reeds lopend promotietraject aan de Universiteit Utrecht omtrent het gebruik van observationele gegevens (real-world data (RWD)) voor de beoordeling van de effectiviteit van geneesmiddelen.

Wij, Amr Makady (farmacoeconomische beoordelaar) en Hugo Neijmeijer (masterstuden Radboud MC Nijmegen)/Sandrine van Acker (masterstudent VU Amsterdam), werken momenteel aan deze evaluatie onder begeleiding van Wim Goettsch. U ontvangt deze mail omdat u op een manier betrokken bent geweest bij de uitvoering van deze beleidsregels. Om deze reden willen wij u graag wat vragen stellen over uw ervaringen met voorwaardelijke financiering in het kader van deze beleidsregels. In de bijlage kunt u een tabel vinden met medicijnen die uiteindelijk de volledige procedure van voorwaardelijke financiering zijn ondergaan.

In hoofdlijnen willen we de volgende zaken bespreken:

- Uw rol in voorwaardelijke financiering als onderdeel van de beleidsregel dure en weesgeneesmiddelen
- Uw perceptie van de totstandkoming van de voorwaardelijke financiering also onderdeel van de beleidsregels
- Uw ervaringen met voorwaardelijke financiering als onderdeel van de beleidsregels
- Enkele toekomst scenario's voor voorwaardelijke financiering

Uw bijdrage aan dit interview is van grote toegevoegde waarde. Wij willen spoedig een datum en plaats vaststellen voor het interview en vernemen daarom graag of u hieraan wilt deelnemen.

Met vriendelijke groet,
[naam onderzoeker]

Figure 1 – Standardized e-mail invitation for stakeholder representatives.

Inleiding:

Gezien alle recente nationale en internationale discussies over flexibele vergoedingssystemen is vanuit Zorginstituut Nederland het plan opgevat om het gebruik van voorwaardelijke financiering, dat een onderdeel was van de beleidsregels dure en weesgeneesmiddelen (2006-2012), te evalueren. Dit kader kan misschien beter bekend zijn als de t=0,t=4 dynamiek.

Het is de bedoeling dat de uitkomsten van deze evaluatie zullen bijdragen aan de inrichting van toekomstige vormen van voorwaardelijke vergoeding/financiering. De resultaten van dit onderzoek zullen ook worden gebruikt voor een reeds lopend promotietraject aan de Universiteit Utrecht omtrent het gebruik van observationele gegevens (real-world data (RWD)) voor de beoordeling van de effectiviteit van geneesmiddelen.

Rol en belang:

- Wat is /was uw rol in het kader van voorwaardelijke financiering?
- Wat is/was de rol van uw organisatie in het kader van voorwaardelijke financiering?
- Wat is het belang van u of uw organisatie in het kader van voorwaardelijke financiering?

Tot stand komen van voorwaardelijke financiering (aim):

- Wat was, vanuit uw perspectief of het perspectief van uw organisatie, het oorspronkelijke doel van voorwaardelijke financiering?
- In hoeverre zijn er eerdere pogingen gedaan om het zelfde doel te verwezenlijken als dit kader van voorwaardelijke financiering?
- In welke mate bent u/ is uw organisatie betrokken geweest bij het tot stand komen van voorwaardelijke financiering?
- In hoeverre zijn er concrete criteria gesteld waaraan voorwaardelijke financiering moest voldoen? Dit om later het kader te kunnen evalueren.

De huidige situatie van voorwaardelijke financiering (functioning):

- Hoe denkt u over het functioneren van voorwaardelijke financiering? (evt. toelichten dat maar 12 van 47 de volledig procedure zijn doorgekomen en mening daarover vragen)

Procedurele aspecten:

- Hoe denkt u over de afgesproken tijdsperiode van 4 jaar voor kandidaten en het overschrijden van die periode?

Methodologische aspecten:

- In welke mate is er op T=0 sprake van goed geplande uitkomstenonderzoeken met duidelijke doelstellingen voor gepast gebruik en kosteneffectiviteit geweest?
- In welke mate is er op T=4 sprake van uitgevoerde uitkomstenonderzoeken die goed aan doelstellingen voor gepast gebruik en kosteneffectiviteit aansluiten/aansloten?
- Zou u uw mening kunnen geven over de kwaliteit van de uitgevoerde studies?
- In welke mate is er op T=4 sprake van opgeleverd data middels uitkomstenonderzoek die relevante antwoorden op primaire doelstelling?
- Zou u uw mening kunnen geven over de kwaliteit van opgeleverde data en betrouwbaarheid van de resultaten?

Het beoordelen van data en besluitvorming op T=4:

- In hoeverre denkt u/uw organisatie dat resultaten uit uitkomstenonderzoeken invloed hebben gehad op het uiteindelijke advies op T=4? Kunt u uw antwoord toelichten?
- In hoeverre denkt u/uw organisatie dat externe factoren (los van evidentie op T=4) invloed hebben gehad op het uiteindelijke advies op T=4? Kunt u uw antwoord toelichten?

Uw opvatting over het kader (advantages and disadvantages):

- In welke mate denkt u/uw organisatie dat voorwaardelijke financiering zijn gewenste doel heeft bereikt? Kunt u uw antwoord toelichten?
- Wat zijn volgens u/uw organisatie enkele voordelen of positieve effecten van voorwaardelijke financiering? Kunt u uw antwoord toelichten?
- Wat zijn volgens u/uw organisatie de nadelen of matig functionerende aspecten van voorwaardelijke financiering? Kunt u uw antwoord toelichten?

Toekomst voorwaardelijke financiering (future perspectives):

- Wat zijn volgens u de belangrijkste punten voor verbetering van voorwaardelijke financiering?
- Wat zijn volgens u concrete punten die meegenomen moeten worden voor toekomstig beleid?

Aantal scenario's voor VF in de toekomst

- Het VF kader stoppen.
- Het VF kader continueren, zoals het nu is.
- Het VF kader aanpassen naar aanleiding van geïdentificeerde leerpunten.
- Nieuw kader oprichten met uitgangspunten van "adaptivepathways"[1]. (Buiten vergoedingssysteem, afgebakend fonds, participatie plicht patiënten, 1 a 2 centra, duidelijke doelstelling mbt effectiviteit/kosteneffectiviteit/gepast gebruik.
- Ziet u een duidelijke rol voor voorwaardelijke financiering in de toekomst? Kunt u uw antwoord toelichten?

Figure 2 – Interview guide.

[1]Bron:http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp

Adaptive pathways is based on three principles:

- 1. iterative development, which either means:
  - approval in stages, beginning with a restricted patient population then expanding to wider patient populations;
  - confirming the benefit-risk balance of a product, following a conditional approval based on early data (using surrogate endpoints) considered predictive of important clinical outcomes;
- gathering evidence through real-life use to supplement clinical trial data;
- early involvement of patients and health-technology-assessment bodies in discussions on a medicine's development.

Figure 3a – The full coding tree developed for content analysis (part 1 of 4)

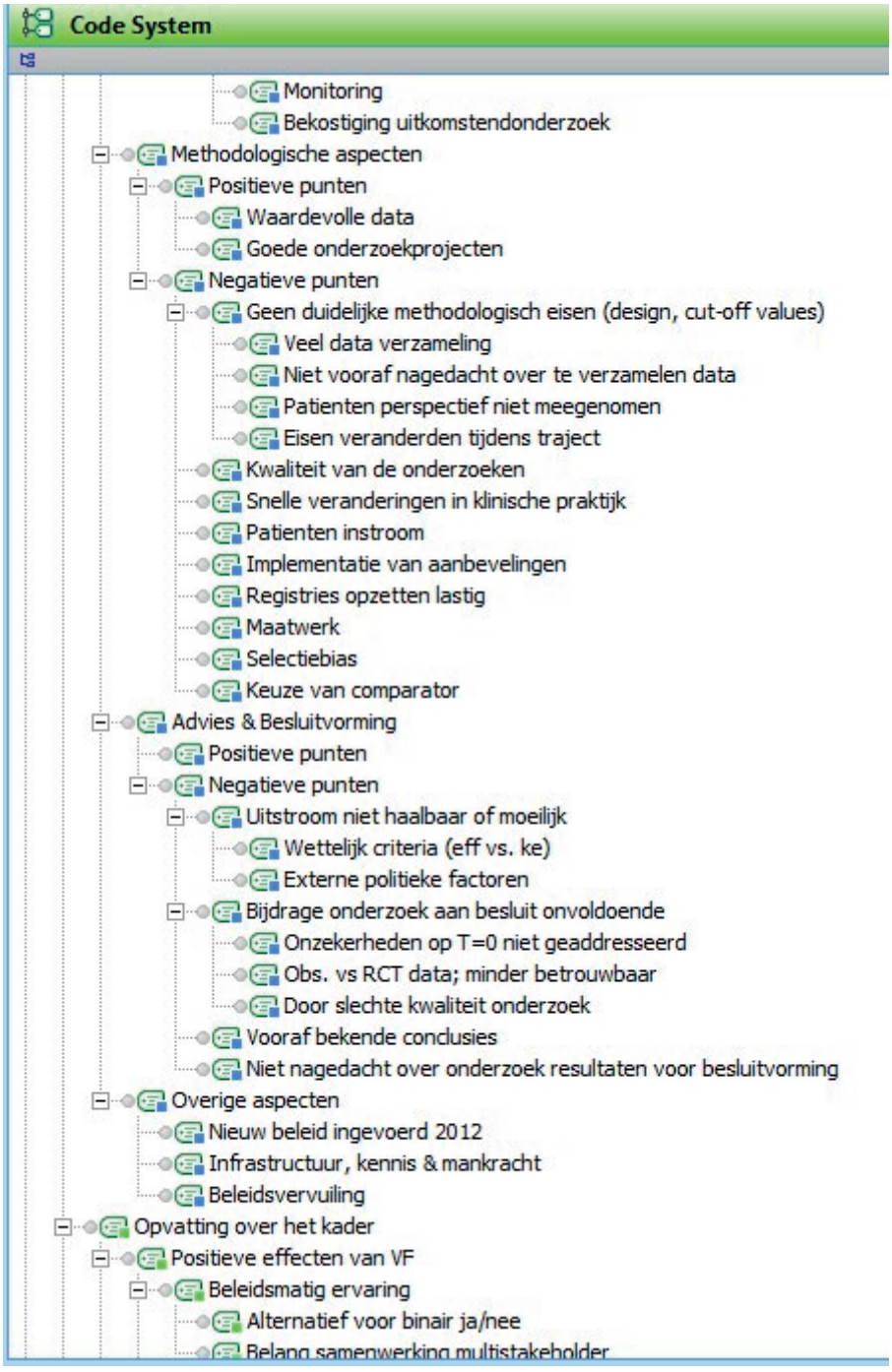


Figure 3b – The full coding tree developed for content analysis (part 2 of 4)

Figure 3c – The full coding tree developed for content analysis (part 3 of 4)

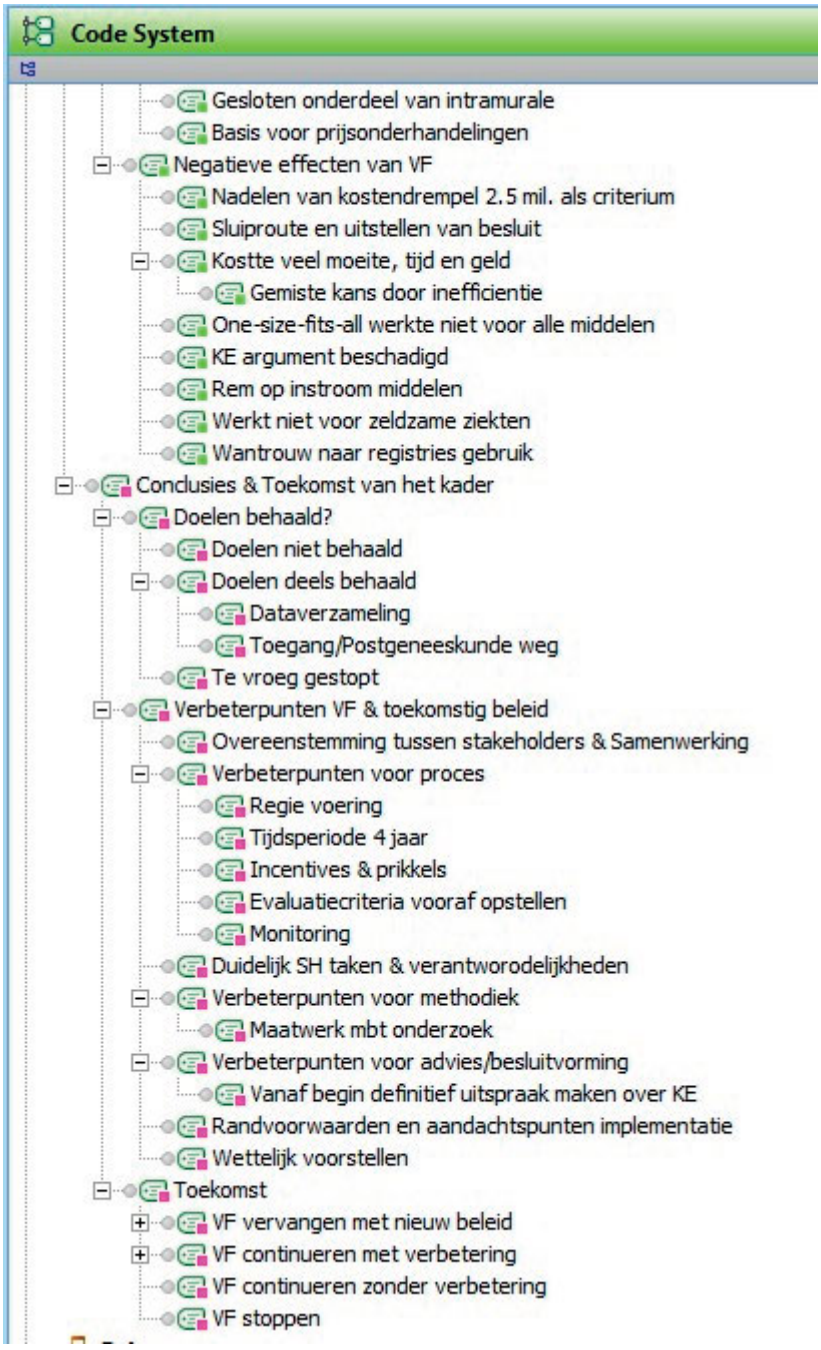
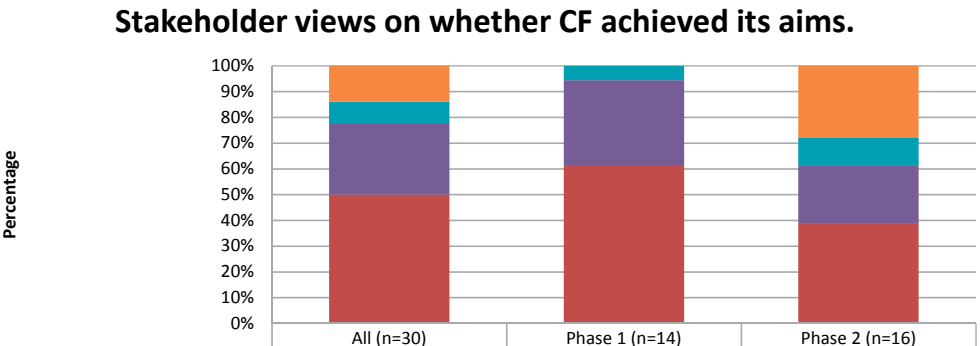
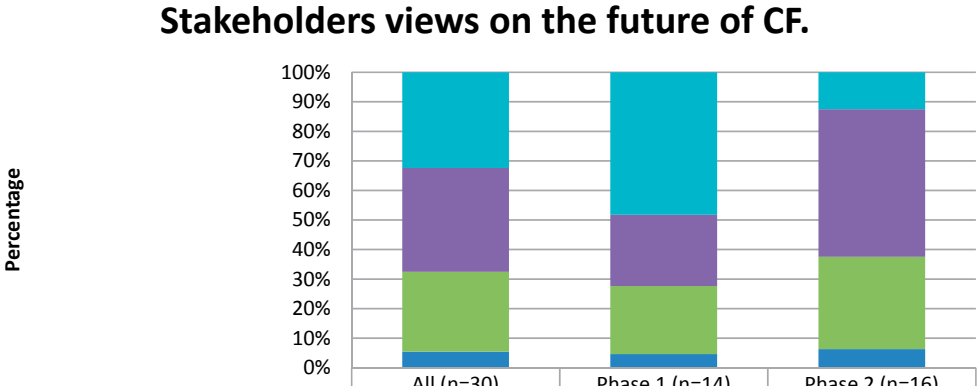


Figure 3d – The full coding tree developed for content analysis (part 4 of 4)



	All (n=30)	Phase 1 (n=14)	Phase 2 (n=16)
Partially (other)	14,71%	0,00%	31,25%
Partially (evidence generation)	8,82%	5,56%	12,50%
Partially (early access)	29,41%	33,33%	25,00%
No	52,94%	61,11%	43,75%
Yes	0,00%	0,00%	0,00%

Figure 4 – Comparative analysis of stakeholder views on achievement of conditional financing (CF) aims.



	All (n=30)	Phase 1 (n=14)	Phase 2 (n=16)
Replace CF with new policy (other)	32,43%	47,62%	12,50%
Replace CF with new policy (AP)	35,14%	23,81%	50,00%
Improve & reintroduce CF	27,03%	22,73%	31,25%
Continue CF in current form	0,00%	0,00%	0,00%
Stop CF/ MEAs	5,41%	4,55%	6,25%

Figure 5 – Comparative analysis of stakeholder views on future perspectives.
Abbreviations: AP: Adaptive Pathways; CF: Conditional Financing; MEA: Managed Entry Agreement.

CHAPTER 9 - APPENDIX

Appendix 1 - Relevance of questions from EORTC QLQ-C30 questionnaire in our study population

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Trouble doing strenuous activities	Stage I (n=17)	18	24	6	12	12	29
	Stage II (n=10)	40	10	20	-	-	30
	Stage III (n=16)	19	12	-	25	31	12
	Stage IV (n=28)	18	7	18	29	18	11
	Carers (n=19)	11	5	21	11	32	21
Trouble taking a long walk	Stage I (n=17)	12	24	6	6	24	29
	Stage II (n=10)	40	30	-	-	-	30
	Stage III (n=16)	19	6	12	6	38	19
	Stage IV (n=28)	18	7	18	29	18	11
	Carers (n=19)	5	5	21	21	21	26
Trouble taking a short walk outside the house	Stage I (n=18)	28	22	-	-	17	33
	Stage II (n=10)	40	20	-	-	10	30
	Stage III (n=16)	25	19	6	19	12	19
	Stage IV (n=28)	18	25	14	14	7	21
	Carers (n=19)	17	17	-	17	22	28
Need to stay in bed or a chair during the day	Stage I (n=17)	18	24	12	6	-	41
	Stage II (n=10)	50	10	-	10	-	30
	Stage III (n=16)	25	12	6	31	6	19
	Stage IV (n=28)	25	21	11	14	11	18
	Carers (n=19)	11	21	11	11	21	26
Need help with eating, dressing, washing yourself or using the toilet	Stage I (n=17)	18	18	12	-	6	47
	Stage II (n=10)	70	-	-	-	-	30
	Stage III (n=16)	44	6	-	25	-	25
	Stage IV (n=28)	50	7	4	4	7	29
	Carers (n=19)	26	16	-	11	21	26
Limitations in doing either your work or other daily activities	Stage I (n=17)	18	18	6	18	18	24
	Stage II (n=10)	60	-	-	10	-	30
	Stage III (n=16)	19	6	12	31	25	6
	Stage IV (n=28)	18	7	7	25	32	11
	Carers (n=19)	11	-	26	21	26	16
Limitations in pursuing your hobbies or other leisure time activities	Stage I (n=17)	6	12	12	18	35	18
	Stage II (n=10)	50	10	10	10	-	20
	Stage III (n=15)	13	7	-	53	7	20
	Stage IV (n=27)	15	7	7	33	26	11
	Carers (n=19)	5	5	11	32	26	21

Appendix 1 - continued

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Short of breath	Stage I (n=17)	18	29	-	6	12	35
	Stage II (n=10)	40	10	10	-	-	40
	Stage III (n=16)	31	6	12	12	19	19
	Stage IV (n=28)	43	4	-	29	7	18
	Carers (n=19)	22	11	11	-	22	33
Pain	Stage I (n=16)	12	12	12	19	6	38
	Stage II (n=10)	50	10	-	10	-	30
	Stage III (n=16)	25	6	6	38	19	6
	Stage IV (n=28)	36	4	11	21	11	18
	Carers (n=19)	17	-	11	11	50	11
Needed more time to rest	Stage I (n=16)	6	19	12	12	25	25
	Stage II (n=10)	20	10	10	20	10	30
	Stage III (n=15)	7	7	7	53	20	7
	Stage IV (n=28)	21	4	7	29	32	7
	Carers (n=19)	16	16	5	26	21	16
Trouble sleeping	Stage I (n=17)	6	6	24	18	18	29
	Stage II (n=10)	20	10	-	30	10	30
	Stage III (n=16)	12	12	25	38	12	-
	Stage IV (n=28)	21	-	7	25	36	11
	Carers (n=19)	-	5	11	32	32	21
Feeling weak	Stage I (n=16)	6	19	25	19	-	31
	Stage II (n=10)	40	10	10	10	-	30
	Stage III (n=16)	12	6	19	44	12	6
	Stage IV (n=27)	37	7	7	15	22	11
	Carers (n=19)	11	5	5	26	37	16
Lack of appetite	Stage I (n=17)	18	29	18	-	-	35
	Stage II (n=10)	50	-	20	-	-	30
	Stage III (n=16)	25	19	6	19	12	19
	Stage IV (n=27)	33	7	15	19	4	22
	Carers (n=19)	26	5	11	11	21	26
Nausea/ Feeling sick	Stage I (n=17)	24	18	6	6	-	47
	Stage II (n=10)	60	-	-	10	-	30
	Stage III (n=16)	25	19	6	6	25	19
	Stage IV (n=28)	32	11	7	11	11	29
	Carers (n=19)	21	26	5	11	16	21

Appendix 1 - continued

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Have you vomited?	Stage I (n=17)	29	18	-	-	-	53
	Stage II (n=10)	70	-	-	-	-	30
	Stage III (n=16)	50	-	-	19	6	25
	Stage IV (n=28)	43	7	11	7	7	25
	Carers (n=19)	26	21	11	5	-	37
Were you constipated?	Stage I (n=16)	19	19	6	6	-	50
	Stage II (n=10)	60	10	-	-	-	30
	Stage III (n=16)	38	6	6	25	6	19
	Stage IV (n=28)	29	4	11	18	11	29
	Carers (n=19)	11	5	26	16	-	42
Diarrhea	Stage I (n=16)	19	19	12	-	-	50
	Stage II (n=10)	70	-	-	-	-	30
	Stage III (n=16)	38	6	6	25	6	19
	Stage IV (n=28)	32	11	11	18	14	14
	Carers (n=19)	16	26	5	21	5	26
Tiredness	Stage I (n=16)	12	6	-	50	12	19
	Stage II (n=10)	20	-	-	50	10	20
	Stage III (n=16)	12	-	12	38	25	12
	Stage IV (n=28)	18	-	7	21	54	-
	Carers (n=19)	11	-	-	32	42	16
Did pain interfere with your daily activities?	Stage I (n=17)	18	12	-	29	6	35
	Stage II (n=10)	30	20	-	20	-	30
	Stage III (n=16)	25	6	12	25	12	19
	Stage IV (n=28)	36	7	11	7	21	18
	Carers (n=19)	11	11	11	16	21	32
Difficulty in concentrating on things	Stage I (n=17)	18	6	29	18	12	29
	Stage II (n=9)	22	-	11	22	22	22
	Stage III (n=15)	33	-	13	33	20	13
	Stage IV (n=28)	11	7	18	11	32	7
	Carers (n=19)	5	26	5	26	11	26
Feeling tense	Stage I (n=17)	6	6	12	35	29	12
	Stage II (n=10)	10	10	-	30	30	20
	Stage III (n=16)	12	-	25	25	25	12
	Stage IV (n=28)	11	4	18	29	36	4
	Carers (n=19)	-	11	5	26	47	11

Appendix 1 - continued

Question in EORTC QLQ-C30:		Relevance					
		Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Worrying	Stage I (n=17)	-	-	12	35	41	12
	Stage II (n=10)	-	10	-	30	40	20
	Stage III (n=16)	-	6	-	31	44	19
	Stage IV (n=28)	7	11	11	32	32	7
	Carers (n=19)	5	16	-	16	47	16
Feeling irritable	Stage I (n=16)	-	6	19	38	25	12
	Stage II (n=10)	10	10	20	10	20	30
	Stage III (n=14)	-	21	29	29	14	7
	Stage IV (n=28)	11	7	11	50	18	4
	Carers (n=19)	-	16	11	37	21	16
Feeling depressed	Stage I (n=17)	-	6	12	35	18	29
	Stage II (n=10)	10	-	20	20	30	20
	Stage III (n=16)	-	12	25	19	38	6
	Stage IV (n=28)	4	18	21	18	21	18
	Carers (n=19)	5	16	5	12	42	11
Difficulty remembering things	Stage I (n=15)	7	-	27	33	7	27
	Stage II (n=10)	10	20	-	10	30	30
	Stage III (n=16)	19	19	19	19	12	12
	Stage IV (n=27)	19	7	19	19	22	15
	Carers (n=19)	11	21	11	21	16	21
Physical condition or medical treatment interfered with your family life	Stage I (n=16)	6	6	6	19	25	38
	Stage II (n=10)	20	-	20	20	20	20
	Stage III (n=16)	12	-	25	25	25	12
	Stage IV (n=28)	7	-	21	32	32	7
	Carers (n=19)	5	5	5	32	32	21
Physical condition or medical treatment interfered with your social activities	Stage I (n=16)	6	6	6	25	31	25
	Stage II (n=10)	20	-	-	30	20	30
	Stage III (n=16)	12	-	12	31	31	12
	Stage IV (n=28)	4	-	18	29	36	14
	Carers (n=19)	5	5	5	32	32	21
Physical condition or medical treatment caused you financial difficulties	Stage I (n=18)	12	12	-	24	18	35
	Stage II (n=10)	30	-	10	20	20	20
	Stage III (n=16)	19	12	12	19	19	19
	Stage IV (n=28)	11	7	25	25	18	14
	Carers (n=19)	11	16	-	11	42	21

Appendix 2. The Melanoma Quality of Life Survey: a 25-item web-based survey

Dear Melanoma patient or carer,

What is Quality of Life in Melanoma for YOU?

This study is part of the GetReal project and is conducted as collaboration between MPNE, the Melanoma Patient Network Europe, and ZIN, the Dutch National Healthcare Institute. So far, few studies have looked at what Melanoma patients themselves consider important for their own Quality of Life. The aim of this study is therefore to find out what truly matters to the Melanoma patients reached through our network. Quality of Life data is also increasingly used for the approval and reimbursement of new therapies – so please take the time to share your thoughts! We would like to understand the influence of the Melanoma stage, the time of diagnosis, the country you live in and Melanoma therapies on the Quality of Life of Melanoma patients. We also want to see if social media could be used to collect such information on patient perspectives. More information about this collaboration can be found on our website.

This survey should take 20 minutes to complete. Your answers are confidential and we will only publish anonymous results. Insights and reports will obviously be shared via the Melanoma Patient Network Europe channels!

Thank you for you time and effort.

MPNE and ZIN

Melanoma Patient Network Europe and National Healthcare Institute

We value your opinion

1. Quality of Life in Melanoma – which aspects come to your mind?

1. [Open Field]
2. [Open Field]
3. [Open Field]
4. [Open Field]
5. [Open Field]
6. [Open Field]
7. [Open Field]
8. [Open Field]
9. [Open Field]
10. [Open Field]

2. What is Quality of Life in Melanoma for you?

[Open Field]

3. **On a scale from 1 to 7, please rate your/the** patient's Quality of Life today.

	1 – poor	2	3	4	5	6	7 – Excellent
Quality of Life	O	O	O	O	O	O	O

4. **The 3 things that today make your/ the** Melanoma patient's Quality of Life good

1. [Open Field]
2. [Open Field]
3. [Open Field]

5. **The 3 things that today make your/ the** Melanoma patient's Quality of Life good

1. [Open Field]
2. [Open Field]
3. [Open Field]

6. **The single thing that would improve your/ the** Melanoma patient's Quality of Life right now?

[Open Field]

We value your opinion – 2

7. How important are for you

	Not important at all	Not important	Neutral	Important	Very important
Physical well-being (e.g. energy level, nauseau, pain)	O	O	O	O	O
Social/ Family well-being (e.g. support from family and friends, sex life)	O	O	O	O	O

Emotional well-being (e.g. feeling sad or nervous, worries related to Melanoma or treatments)	O	O	O	O	O
Functional well-being (e.g. ability to work, sleep and enjoy life)	O	O	O	O	O
Other (please specify below)	O	O	O	O	O
Other (please specify)	[Open Field]				

8. Please comment on question 7

[Open Field]

9. **How relevant are the** following aspects for you

	Not relevant at all	Not relevant	Neutral	Relevant	Very relevant	Does not apply to me
Trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase	O	O	O	O	O	O
Trouble taking a long walk	O	O	O	O	O	O
Trouble taking a short walk outside of the house	O	O	O	O	O	O
Need to stay in bed or a chair during the day	O	O	O	O	O	O
Need help with eating, dressing, washing yourself or using the toilet	O	O	O	O	O	O
Limitations in doing either your work or other daily activities	O	O	O	O	O	O
Short of breath	O	O	O	O	O	O
Pain	O	O	O	O	O	O
Needed more time to rest	O	O	O	O	O	O
Trouble sleeping	O	O	O	O	O	O
Feeling weak	O	O	O	O	O	O
Lack of appetite	O	O	O	O	O	O
Nausea/ Feeling sick	O	O	O	O	O	O
Have you vomited?	O	O	O	O	O	O
Were you constipated?	O	O	O	O	O	O
Diarrhoea	O	O	O	O	O	O
Tiredness	O	O	O	O	O	O
Did pain interfere with your daily activities?	O	O	O	O	O	O
Difficulty in concentrating on things, like reading a newspaper or watching television	O	O	O	O	O	O
Feeling tense	O	O	O	O	O	O

Worrying	O	O	O	O	O	O
Feeling irritable	O	O	O	O	O	O
Feeling depressed	O	O	O	O	O	O
Difficulty remembering things	O	O	O	O	O	O
Physical condition or medical treatment interfered with your family life	O	O	O	O	O	O
Physical condition or medical treatment interfered with your social activities	O	O	O	O	O	O
Physcial condition or medical treatment caused you financial difficulties	O	O	O	O	O	O
Other (please specify)	[Open Field]					

10. Please comment on question 8

[Open Field]

Tell us about yourself

11. I am

O Female
O Male

12. What is your Country of Residence?

13. What is your age?

14. **What is the** highest level of education you have completed?

15. Where did you find this survey?

16. Your relationship to Melanoma

	Stage I	Stage II	Stage III	Stage IV	N/A
I am a Melanoma patient	O	O	O	O	O
I am the carer or a Melanoma patient whose disease is in	O	O	O	O	O
Other (please specify)	[Open Field]				

17. **The** Melanoma diagnosis was

18. **What type of Melanoma do you or the** patient have?

Other (please specify) [Open Field]

19. **Which mutations does your/ the** patient's Melanoma have?

O BRAF mutant
O BRAF wild-type
O NRAS mutant
O c-kit mutant
O GNAQ/ GNA11
O I don't know
O Other (please specify)
[Open Field]

Melanoma therapies and treatments

20. Did you have surgery for your Melanoma?

O No
O Yes

If yes, what type of surgery?

[Open Field]

21. Did/ do you have radiotherapy for your Melanoma?

O No
O Yes

If yes, what type of radiotherapy?

[Open Field]

22. Did/ do you have chemotherapy for your Melanoma?

No
Yes

If yes, what type of chemotherapy?

[Open Field]

23. Did/ do you have immune therapies for your Melanoma? (please tick all that apply)

O No
O Ipilimumab/ YERVOY® - BMS
O Pembrolizumab/ KEYTRUDA® - MSD
O Nivolumab/ OPDIVO® - BMS
O T-Veck/ Talimogene Laherparepvec/ IMLYGIC® - Amgen
O Pidilizumab (CT011 anti-PD1) – Curetech
O Atezolizumab (anti-PD-L1) – BMS
O BMS936559 (anti-PDL1) – BMS
O Dendritic Cell Vaccine – academic
O Adaptive Cell Therapies like TILs (T-infiltrating Lymphocytes) – academic
O Other (please specify)
[Open Field]

24. Did/ do you have targeted therapies for your Melanoma? (please tick all that apply)

O No
O Vemurafenib/ ZELBORAF® - Roche
O Dabrafenib/ TAFINLAR® - Ex-GSK, now Novartis
O Trametinib/ MEKINIST® - Ex-GSK, now Novartis
O Cobimetinib/ COTELLIC® - Roche
O Encorafenib/ LGX8181 – Ex-Novartis, now Array
O Binimetinib MEK 162 – Ex-Novartis, now Array
O Other (please specify)
[Open Field]

Thank you

25. Anything else you would like to let us know?

[Open Field]

Thank you for helping us understand what Quality of Life means to Melanoma patients.

The results of this survey will be shared in any anonymous form with the MPNE network and the general public. To make sure you don't miss updates, please sign up to the MPNE newsletter.

MPNE and ZIN

CHAPTER 12

Summary
Samenvatting in het Nederlands
Acknowledgements
List of publications
About the author

SUMMARY

Achieving good health has been recognized as a human right in the constitution of the World Health Organization (WHO). In most countries, governmental and non-governmental parties participate in healthcare systems which aim to provide the general population with access to good healthcare. However, healthcare budgets are not infinite. Meanwhile, healthcare expenditures continue to rise worldwide. Consequently, decision makers are faced with challenging questions on how to allocate resources to achieve the greatest health gains for their citizens.

One framework facilitating transparent and accountable decision making on this front pertains to health technology assessment (HTA). In general, HTA is conducted by (public or private) HTA agencies and comprises two stages; the assessment of scientific evidence available for the question at hand and the appraisal of evidence to reach decisions. In this thesis, we limit our scope to HTA of pharmaceutical drugs, as opposed to other health technologies and interventions.

When conducting health technology assessments, HTA agencies often resort to scientific evidence of the highest degree available; thus conventionally randomized controlled clinical trials (RCTs). Several limitations pertaining to RCTs (e.g. highly selected patient populations and controlled follow-up of patients) make them less suitable for answering all questions of relevance to HTA agencies, which aim to assess the value of drugs for the entire patient population under conditions of routine clinical practice. A potential alternative data source for scientific evidence which may complement RCTs in this regard is so-called real-world data (RWD). At the moment, little is known with regards to RWD use in HTA of pharmaceutical drugs and subsequent decision making. This thesis aims to address this gap in knowledge.

To begin with, it is important to define what real-world data (RWD) is. Chapter 2 presents a study of definitions available in literature and stakeholder interviews. The results indicated a different understanding amongst different stakeholders of what RWD precisely is. The definitions identified could be categorized into 4 categories, each imposing different criteria on what does, or does not, qualify as RWD. Moreover, not all stakeholders could refer to an established definition of the concept. Discrepancies in definitions of RWD amongst different stakeholders could undoubtedly lead to confusions throughout discussions on the value of RWD for HTA; stakeholders could disagree on whether particular data sources really do represent the "real-world" and thus qualify as RWD. On the other hand, several definitions for RWD have been developed which may form a good starting point for future consensus-seeking. For the remainder of this thesis, RWD was defined as health data generated outside the context of RCTs. Meanwhile, real-world evidence (RWE) refers to evidence that is derived from the analysis and/or synthesis of RWD.

Consensus on definitions notwithstanding, it is important to explore available policies of HTA agencies for the use of RWE. Chapter 3 presents a study exploring the respective policies of 6 European HTA agencies. The results indicate that policies depend on two factors:

the context of use and the parameters for which RWE is used. Three contexts emerged whereby RWE is used: relative effectiveness assessment (REA), pharmacoeconomic analysis (PEA) and conditional reimbursement schemes (CRS). In general, RWE use for (relative) effectiveness estimates in all contexts was not encouraged by agencies. On the other hand, it was directly requested for other parameters specific to PEAs and CRSs. Differences in policies also emerged between agencies, for instance with regards to RWE use for disease areas where conducting RCTs is challenging (e.g. orphan diseases).

Having created an overview of HTA agencies' policies on RWE, the following step would be to assess the actual use of RWE within the different contexts identified above. Chapter 4 presents a study on the use of RWE in REA and PEA (i.e. cost-effectiveness assessments; CEA) of drugs for metastatic melanoma by 5 European HTA agencies. In general, RWE inclusion was higher in CEAs than REAs. It was mostly used to estimate melanoma prevalence in REAs or to predict long-term effectiveness in CEAs. Moreover, several differences emerged between agencies' use of RWE in their assessments.

In the Netherlands, the use of RWE is HTA and decision making has been implemented in a CRS, namely the conditional financing (CF) of hospital drugs. Chapters 5 and 6 present in-depth analyses of experiences gained with the implementation of CF in the Netherlands. Analysis of HTA dossiers published for CF drugs in chapter 5 indicated shortcomings related to procedural, methodological and decision-making aspects of the scheme. For example, the CF process extended beyond the pre-defined period of 4 years for nearly all drugs. Focusing on RWE-related issues, the reports indicated that the scientific robustness of RWE collected in the context of CF was often of inadequate scientific quality. As a result, the RWE submitted for two-thirds of the HTA questions posed did not provide the envisioned answers. Stakeholder interviews conducted in chapter 6 demonstrated differences amongst the stakeholders on the perceived aim of CF. Conversely, there was some agreement amongst stakeholders on the positive impact of CF on the Dutch healthcare setting and improvement points for CF. For example, stakeholders referred to the valuable insights RWE can provide with regards to the appropriate use of drugs in practice. However, stakeholders also emphasized that RWE could not be regarded as a sole evidence source for HTA decision making on effectiveness and cost-effectiveness. In this regard, stakeholders referred to several factors that impacted the relevance of RWE to HTA; one example being the inability to account for rapid changes in clinical practice when implementing RWE study protocols. Stakeholders also referred to the difficulty of collecting RWE in practice, which required extensive inter-stakeholder collaboration, substantial funding and raised issues related to data ownership and governance. Despite the belief that CF only partially met its aims, there was agreement on the need for new policy to address the same aims of CF in the future. However, stakeholders diverged on whether CF should be improved based on the learnings identified and re-introduced into practice, or replaced with new policy schemes.

Currently, few examples exist in literature whereby RWE is used for HTA purposes. In principle, increased adoption of RWE use for HTA could thus be facilitated by the conduct of

robust scientific research whereby the advantages and limitations of its use could be brought to light. The IMI-GetReal initiative attempted to conduct such research in the context of numerous case studies between 2015 and 2017. Chapter 7 presents a summary of IMI-GetReal's experiences in accessing and using RWE. In general, only a third of all requests for access to individual patient-level data (IPD) from RWE repositories submitted across all case studies were successful. Reasons for inaccessibility mostly related to datasets not being research-ready within project timelines or unwillingness to share data. As an alternative to accessing IPD, case study teams explored options for using aggregate data (AD) from registries and observational studies. However, findings indicated that although AD can be easily obtained from literature, it is often of limited usefulness, mostly lending itself to descriptive statistical analyses rather than to analysis of treatment effects across different settings and populations.

Looking beyond the conventional sources of RWE, such as registries and observational studies, could RWE be generated from novel sources that have not yet been exploited for HTA purposes? Chapters 8 and 9 aim to address this question by exploring the potential use of social media to gather evidence for REAs. Chapter 8 presents a study whereby a literature review was conducted on this topic. The results demonstrated that social media may provide a potential source of RWE for REA, particularly on aspects such as adverse events of drugs, symptom occurrence, quality of life, and drug adherence behaviour. Meanwhile, chapter 9 presents a study whereby social media was used to gather melanoma patient perspectives on health-related quality of life (HRQoL) through a survey. The results imply that social media may provide a quick and time-efficient manner to assemble valuable data on patient perspectives on HRQoL. Employing social media to collect such data would require less resources than multi-centre trials or point-of-care studies. Similarly, the use of social media may be less resource-intensive and more efficient than similar data collection through the establishment and conduct of patient/citizen panels at various stages of HTA decision making. However, many limitations are also associated with social media use for HTA, including a lack of methodological guidance and standard practices to do so.

Provided the points discussed above, how should the HTA community move towards more optimal use of RWE in decision making? This is the topic of Chapter 10. Firstly, a summary is provided of chapters 2-9, followed by an overview of new and ongoing RWE initiatives that overlap with questions addressed in this thesis. Finally, eight recommendations are provided addressing both policy measures and methodological research. In our opinion, these recommendations stipulate important areas of focus for future developments in this rapidly-evolving field. It is our aspiration that the recommendations will help HTA agencies and decision makers inch ever closer to realizing the full potential RWE can bear, thus better healthcare systems for all citizens.

SAMENVATTING IN HET NEDERLANDS

In de constitutie van de Wereldgezondheidsorganisatie (WHO) staat dat goede gezondheid een mensenrecht is. Overheidsinstanties en privépartijen werken vaak samen in het kader van gezondheidssystemen met als doel goede gezondheid voor burgers te realiseren en gezondheidszorg toegankelijk te maken. Echter, enerzijds zijn de beschikbare budgetten voor gezondheidszorg beperkt en anderzijds, stijgen de afgelopen jaren de uitgaven aan gezondheidszorg opmerkelijk snel. Als gevolg hiervan moeten er vaak moeilijke beslissingen genomen worden om met de beperkte middelen zo veel mogelijk gezondheidswinst te behalen voor alle burgers.

Health Technology Assessment (HTA) biedt een transparant en verantwoord kader om besluitvorming in de gezondheidszorg te ondersteunen. In het algemeen wordt HTA uitgevoerd door (publieke of privé) HTA instanties en kent twee fasen: de wetenschappelijke beoordeling van de beschikbare evidentie (oftewel "assessment") en vervolgens de bespreking van de evidentie in de context van overige (bijvoorbeeld: maatschappelijke) overwegingen om tot een besluit te komen (oftewel "appraisal"). Deze thesis richt zich met name op HTA van geneesmiddelen, niet van hulpmiddelen en overige gezondheidsinterventies.

Bij het uitvoeren van HTA maken HTA instanties het liefst gebruik van wetenschappelijke evidentie van de hoogste kwaliteit en betrouwbaarheid. In de meeste gevallen betekent dit gerandomiseerde klinische studies (RCTs). Een aantal eigenschappen van RCTs (bijvoorbeeld geselecteerde, homogene patiëntpopulaties en gecontroleerde protocollen voor het monitoren van patiënten) leiden er toe dat ze niet altijd geschikt zijn voor het beantwoorden van alle HTA-gerelateerde vragen die zich vaak richten op de waarde van geneesmiddelen voor de hele patiëntpopulatie in de context van de dagelijkse klinische praktijk. Een alternatieve bron voor wetenschappelijke evidentie die aanvullende informatie zou kunnen opleveren ten opzichte van RCTs is "real-world data" (RWD). Op dit moment is er nog weinig bekend omtrent het gebruik van RWD in HTA voor geneesmiddelen en besluitvorming. Deze thesis heeft als doel hierover kennis te genereren.

Allereerst moet het begrip RWD gedefinieerd worden. Hoofdstuk 2 bevat een studie van beschikbare definities uit de literatuur en interviews met stakeholders. De resultaten laten zien dat stakeholders verschillende definities hanteren voor RWD. De definities kunnen opgesplitst worden in 4 categorieën, met verschillende omschrijvingen voor wat wél of géén RWD is. Een aantal stakeholders waren niet in staat een bestaande definitie voor RWD citeren. Verschillen in RWD definities kunnen leiden tot misverstanden bij het bespreken en bediscussiëren van RWD voor HTA; stakeholders kunnen het oneens zijn of specifieke databronnen de "real-world" weerspiegelen en dus RWD zijn. Anderzijds, zijn meerdere definities voor RWD ontwikkeld die een goed startpunt zouden kunnen vormen voor toekomstige consensus over dit begrip. Voor deze thesis is RWD gedefinieerd als gezondheidsdata die buiten de context van RCTs wordt verzameld. Verder is "real-world evidence" (RWE) gedefinieerd als evidentie die gebaseerd is op het analyseren van RWD (alleen óf in combinatie met andere bronnen).

Naast consensus over definities, is het belangrijk het beleid van HTA instanties omtrent het gebruik van RWE in kaart te brengen. Hoofdstuk 3 bevat een studie waarbij het RWE beleid van 6 Europese HTA instanties wordt onderzocht. De resultaten geven aan dat RWE beleid afhankelijk is van twee factoren: het perspectief voor het gebruik van RWE en de parameters waarvoor RWE gebruikt wordt. Onderscheid kan gemaakt worden tussen drie perspectieven: relatieve effectiviteit ("relative effectiveness assessment"; REA), kosteneffectiviteit ("pharmacoeconomic analysis"; PEA) en voorwaardelijke vergoedingskaders ("conditional reimbursement schemes"; CRS). In het algemeen wordt het gebruik van RWE voor inschattingen van (relatieve) effectiviteit afgeraden door HTA instanties. Daarentegen is er een directe vraag naar RWE voor andere parameters voor PEAs en CRSs. Verder bestaan er beleidsverschillen tussen instanties, bijvoorbeeld, in beleid voor het gebruik van RWE bij ziekten waarbij RCTs moeilijk uit te voeren zijn, zoals bij weesziekten.

Nadat een overzicht van RWE beleid gemaakt is, is de volgende stap om het daadwerkelijk gebruik van RWE in de praktijk te analyseren. Hoofdstuk 4 bevat een studie waarbij het gebruik van RWE voor REAs en PEAs (oftewel "cost-effectiveness assessments"; CEA) bij 5 Europese instanties onderzocht wordt voor de beoordeling van geneesmiddelen voor de behandeling van melanomen. Over het algemeen wordt RWE vaker gebruikt in CEAs dan REAs. RWE wordt grotendeels gebruikt om de prevalentie en incidentie voor REAs in te schatten en om de effectiviteit van middelen op de lange termijn in te schatten voor CEAs. Verder zien wij verschillen in de manier waarop instanties RWE meenemen in hun beoordelingen.

In Nederland is het gebruik van RWE in besluitvorming toegepast in het kader van de voorwaardelijke vergoeding, namelijk voorwaardelijke financiering (VF), van dure intramurale geneesmiddelen. Hoofdstuk 5 bevat een analyse van HTA rapporten voor VF geneesmiddelen. De resultaten laten zien dat er tekortkomingen waren omtrent procedurele, methodologische en besluitvormingsaspecten van het VF kader. De procedure voor VF duurde voor nagenoeg alle geneesmiddelen langer dan de afgesproken periode van 4 jaar. Wanneer men focust op RWE-gerelateerde punten, laten de rapporten zien dat de wetenschappelijke kwaliteit en betrouwbaarheid van de verzamelde RWE matig was. Derhalve heeft RWE geen antwoorden kunnen opleveren voor twee-derde van alle beleidsvragen. Hoofdstuk 6 bevat een studie waarin stakeholder interviews zijn uitgevoerd omtrent hun ervaringen met VF. De resultaten laten verschillen zien in hoe stakeholders zich het doel van VF voorstellen. Anderzijds, waren er overeenkomsten met betrekking tot de voordelen van het VF kader. Stakeholders gaven bijvoorbeeld aan dat RWE belangrijke inzichten gaf omtrent het gepast gebruik van middelen in de praktijk. Echter, gaven ze ook aan dat RWE niet als enige bron beschouwd kon worden als bewijs voor (kosten) effectiviteit. Stakeholders noemden ook factoren die de relevantie van RWE verminderde, bijvoorbeeld de snelle veranderingen in de praktijk en de impact daarvan op vooraf gestelde studieprotocollen. Verder gaven stakeholders aan dat de verzameling van RWE in de praktijk moeizaam was omdat het intensieve samenwerking eiste tussen stakeholders,

aanzienlijke financiering vereiste en ingewikkelde vragen omtrent de eigendom en beheer van de data opriep. Echter waren de meeste stakeholders het met elkaar eens dat nieuw beleid nodig is in de toekomst om de doelen van VF te verwezenlijken.

Op dit moment zijn er weinig voorbeelden waarbij RWE gebruikt wordt voor HTA. Het uitvoeren van een wetenschappelijk robuust onderzoek waarin de voor- en nadelen van RWE goed toelicht worden, kan het opnemen van RWE in besluitvorming faciliteren. Het IMI-GetReal project heeft geprobeerd hierover onderzoek uit te voeren in het kader van meerdere case studies tussen 2015 en 2017. Hoofdstuk 7 bevat een samenvatting van ervaringen van IMI-GetReal met het krijgen van toegang tot RWE en het gebruik hiervan. In het algemeen slaagde men maar in een derde van de pogingen om toegang te krijgen tot individuele patiënt data (IPD) uit RWE bronnen. De belangrijkste redenen, dat er geen toegang tot de data was, waren dat datasets niet bruikbaar waren voor onderzoek óf er geen bereidheid was om de data te delen. Als alternatief voor IPD gebruik, probeerden onderzoekers geaggregeerde data (AD) uit registraties en observationele studies te gebruiken. Ondanks dat AD makkelijker te verkrijgen is, geven de resultaten aan dat AD niet bruikbaar is voor robuust onderzoek omtrent de effectiviteit van geneesmiddelen in de klinische praktijk.

Als onderdeel van de thesis is ook onderzocht in hoeverre RWE gegenereerd wordt uit nieuwe bronnen, in plaats van bekende bronnen zoals patiëntregistraties en observationele studies. Hoofdstukken 8 en 9 richten zich op deze vraag. Hoofdstuk 8 bevat een literatuur review met als onderzoeksvraag of sociale media een potentiele bron van gegevens kan zijn voor REA. De resultaten laten zien dat sociale media wordt gebruikt voor gegevensverzameling omtrent bijwerkingen van geneesmiddelen, frequentie van symptomen, kwaliteit van leven en therapietrouw. Hoofdstuk 9 bevat een studie waarbij sociale media gebruikt werden om patiënt perspectieven te verzamelen omtrent kwaliteit van leven (HRQoL). De resultaten suggereren dat sociale media een effectievere en efficiëntere manier zouden kunnen bieden om deze perspectieven te verzamelen ten opzichte van burgerpanels, (internationale) klinische studies of ad-hoc registraties in de klinische praktijk. Anderzijds zijn er belangrijke beperkingen met betrekking tot het gebruik van sociale media voor HTA die verder onderzoek vragen, waaronder de afwezigheid van methodologische richtlijnen en gestandaardiseerde methodes.

De focus van hoofdstuk 10 is het beschrijven van een aantal suggesties hoe de HTA gemeenschap kan streven naar optimaal gebruik van RWE in de besluitvorming op basis van de eerder beschreven resultaten. Ten eerste worden hoofdstukken 2-9 samengevat. Ten tweede wordt een overzicht gegeven van lopende en toekomstige RWE initiatieven die overeenkomsten laten zien met de vraagstelling van de huidige thesis. Ten slotte, worden acht concrete aanbevelingen naar voren gebracht voor zowel beleidsmaatregelen als verder methodologisch onderzoek. Naar onze mening wijzen deze aanbevelingen belangrijke aandachtspunten aan voor toekomstige ontwikkelingen op dit gebied. Het is onze hoop dat deze aanbevelingen HTA instanties en besluitnemers verder zullen helpen om het meeste uit RWE te halen en zal meehelpen tot de realisatie van betere gezondheidssystemen voor iedereen.

ACKNOWLEDGEMENTS

It is an irrevocable truth that the people we become and the milestones we reach in our lives are the outcome of more than merely our personal qualities and efforts. I am quite humbled and thankful to be writing this piece at the end of a wonderful chapter. However, I also feel indebted to a vast network of colleagues, friends and family; each of whom has had a valuable and lasting impact on my modest and often tumultuous journey through life. Here's a humble man's attempt to thank all of you here.

Dear Ton, Hans, Olaf and Wim… These words often marked the beginning of 4 years of e-mails. Gentlemen, thank you for your patience and guidance throughout this period. Seldom does a student have the pleasure and privilege of working with such renowned scientists. Better still, those whom, despite their considerable achievements and positions, remain as accessible, encouraging and accepting as you have been. I sincerely wish that this thesis marks the beginning of our continued collaboration, rather than farewell.

Any claim on quality and impact this thesis may be perceived to make is also due to the contributions of many co-authors. To them, I wish to extend a word of thanks. Your many affiliations and diverse interests have enriched the chapters written and ensured their relevance to many stakeholders, instead of one. To Rachel Kalf, Anne Willemsen, Renske ten Ham, Sandrine van Acker, Kim Meijboom, Ard van Veelen and Hugo Nijmeijer: thank you for helping me through the grit of research and the challenges of exploring new territory. I hope you enjoyed our time together as much as I have and I wish you success with all your future endeavours.

When literally translated, our department at Zorginstituut Nederland could either mean the "Health Sector" or the "Worry Sector". With that in mind, to my colleagues at ZIN, thank you for never making me worry about combining my PhD with a full-time job. If it weren't for the welcoming (and inspiring) group that you are, this would have all been a lot harder! I continue to learn a lot from you and enjoy our many discussions together. As for my colleagues at Utrecht University, though we only recently met, thank you for making me feel as welcome and involved as you have this past year.

Academic journeys are riddled with obstacles, frustrations and set-backs. In the midst of these grim circumstances, the gentle touch of a hobby reminds one of the niceties in life. Weathertown Big Band, thank you for the joy of jazz! The animated notes, subtle dynamics and lovely group that you are were often a remedy to a tired mind. Furthermore, as the saying goes: "Big brains are important, but big biceps are importanter." To my buddies at the gym, thank you for the muscle aches… Who knew they worked so well against headaches?

Gianni, in Egypt we have a saying: "ibn balad". You don't have to look far to find its true meaning. We once stood alongside each other as defensive mid-fielders on the pitch. Today, we are defenders in a very different arena. Here's to a lifetime of friendship and a promise to always return the favour.

Finally, one's greatest wealth resides in that one elusive, yet eternal place we call home and the family we share it with. Throughout these years, they have been my rock and my strength.

To my dearest Mother and Father, your unconditional generosity, devotion and commitment to the lives of your children have been my biggest lesson and an honour to experience. Thank you for the gift of a wonderful childhood, for instilling in us the passion for knowledge and for opening up our eyes to a world of joy and compassion.

To my dearest Femke, we did it! As we close this chapter at each other's side, we start our next: except, that new one has no end. Oh, how I look forward to it. Always together, come what may!

Yours Faithfully,
Amr

LIST OF PUBLICATIONS

Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What is real-world data? A review of definitions based on literature and stakeholder interviews. Value in Health. 2017 Jul 1;20(7):858-65.

Makady A, ten Ham R, de Boer A, Hillege H, Klungel O, Goettsch W. Policies for use of real-world data in health technology assessment (HTA): a comparative study of six HTA agencies. Value in Health. 2017 Apr 1;20(4):520-32.

Makady A, van Veelen A, Jonsson P, Moseley O, D'Andon A, de Boer A, Hillege H, Klungel O, Goettsch W. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. PharmacoEconomics. 2017 Dec 6:1-0.

Makady A, van Veelen A, de Boer A, Hillege H, Klungel O, Goettsch W. Implementing Managed Entry Agreements in Practice: The Dutch reality check. Health Policy. 2018 Jan *(under review).

Makady A, van Acker S, Nijmeijer H, de Boer A, Hillege H, Klungel O, Goettsch W. Conditional Financing of Drugs in the Netherlands: Past, Present and Future. Results from Stakeholder Interviews. Value in Health. 2018 Jan *(article prepared for submission).

Makady A, Stegenga H, Ciaglia A, Debray TP, Lees M, Happich M, Ryll B, Abrams K, Thwaites R, Garner S, Jonsson P. Practical implications of using real-world evidence (RWE) in comparative effectiveness research: learnings from IMI-GetReal. Journal of Comparative Effectiveness Research. 2017 Aug 6:6.

Kalf R, Makady A, ten Ham R, Meijboom K, Goettsch W. Use of Social Media in the Assessment of Relative Effectiveness: An explorative review with examples from oncology. Hopeful or hopeless? Journal of Medical Internet Research. 2018 Mar *(Accepted for publication).

Makady A, Kalf R, Ryll B, Spurrier G, de Boer A, Hillege H, Klungel O, Goettsch W. Social Media as a Tool for Assessing Patient Perspectives on Quality of Life in Metastatic Melanoma: a feasibility study. Health and Quality of Life Outcomes. 2018 Jan *(under review).

Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. Real-World Evidence for Health Technology Assessment of Pharmaceuticals: Perspectives on the status quo and the future. Journal of Comparative Effectiveness Research. 2018 Jan *(article prepared for submission).

Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, Madigan D, Makady A, Schneeweiss S, Tarricone R, Wang SV. 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. 2017 Sep 1;20(8):1003-8.

Makady A, Goettsch W. WP1: Deliverable D1.2 Review of current policies/perspectives [Internet]. Utrecht (NL): GetReal; 2014. Report No.: 115546. [cited 2017 Dec 7]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/GetReal%20D1.2%20Current%20Policies%20and%20Perspectives%20FINAL_webversion.pdf

Goettsch W, Makady A. WP1: Deliverable D1.3. Glossary of definitions of common terms (including comments & replies from consultation rounds) [Internet]. Utrecht (NL): GetReal; 2017. Report No.: 115546. [cited 2018 Jan 18]. Available from: https://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/D1.3%20-%20GetReal%20Glossary%20of%20Definitions%20of%20Common%20Terms%20%28Including%20Comments%20%26%20Replies%20from%20Consultation%20Rounds%29.pdf

Makady A, Kalf R, Lees M, Goettsch W. WP1: Deliverable D1.5 Report – Metastatic Melanoma [Internet]. Utrecht (NL): GetReal; 2016. [cited 2017 Dec 7]. Available from: http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%201.5%20Report%20-%20metastatic%20melanoma%20-%20webversion.pdf

Makady A, Kalf R, Lees M, Goettsch W. WP1: Deliverable D1.6 Report – Metastatic Melanoma [Internet]. Utrecht (NL): GetReal; 2016. [cited 2017 Dec 7]. Available from: http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%20D1.6%20-%20Metastatic%20Melanomav2_website%20version.pdf

Jonsson P, Thwaites R, Makady A, Happich M, Granados A, Goettsch W, Garner S, Cerrata F, Barefoot B. Advancing Evidence Generation for New Drugs: IMI-GetReal's Recommendations on Real-World Evidence. Manchester (UK): GetReal; 2017. [cited 2017 Dec 7]. Available from: http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/2017-03-29%20-%20WP1%20-%20Advancing%20Evidence%20Generation%20for%20New%20Drugs.pdf

ABOUT THE AUTHOR

Amr Makady was born in Adan, Kuwait. After completing his secondary education at The New English School of Kuwait, he then moved to Cairo, Egypt. There, Amr completed his Bachelor's studies for Pharmacy and Biotechnology at the German University in Cairo.

Having secured a Huygen's scholarship, Amr moved to Utrecht, the Netherlands where he completed his Master's studies for Drug Innovation at Utrecht University. Throughout his Master's studies, Amr has conducted research in various fields, including: counterfeit in drugs, corruption in healthcare, risk-assessment modelling of generic drug quality and standardization of pharmacoeconomic modelling methods.

In 2014, Amr began his career at the Dutch National Healthcare Institute (ZIN) as a project manager for the IMI-GetReal project. Currently, he works as a pharmacoeconomic assessor and policy advisor at ZIN. Within this function, Amr participates in numerous international projects related to real-world evidence generation and use in decision-making, including: IMI-ROADMAP, EUnetHTA JA3 and REPEAT. Moreover, he participates in numerous teaching activities, including the ISPOR HTA Training Program.