


Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies

Amr Makady^{1,2}  · Ard van Veelen² · Páll Jonsson³ · Owen Moseley⁴ · Anne D'Andon⁵ · Anthonius de Boer² · Hans Hillege⁶ · Olaf Klungel² · Wim Goettsch^{1,2}

Published online: 6 December 2017

© The Author(s) 2017. This article is an open access publication

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, as well as differences in RWD use between agencies and across time.

Methods HTA reports published between 1 January 2011 and 31 December 2016 were retrieved from websites of agencies representing five jurisdictions: England [National Institute for Health and Care Excellence (NICE)], Scotland [Scottish Medicines Consortium (SMC)], France [Haute Autorité de santé (HAS)], Germany [Institute for Quality and Efficacy in Healthcare (IQWiG)] and The Netherlands [Zorginstituut Nederland (ZIN)]. A standardized data extraction form was used to extract information on RWD inclusion for both REAs and CEAs.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40273-017-0596-z>) contains supplementary material, which is available to authorized users.

✉ Amr Makady
amakady@zinl.nl

Ard van Veelen
ardvanveelen@live.nl

Páll Jonsson
pall.jonsson@nice.org.uk

Owen Moseley
o.moseley@nhs.net

Anne D'Andon
a.dandon@has-sante.fr

Anthonius de Boer
a.deboer@uu.nl

Hans Hillege
h.hillege@umcg.nl

Olaf Klungel
o.h.klungel@uu.nl

Wim Goettsch
wgoettsch@zinl.nl

- 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, UK
- ⁴ The Scottish Medicines Consortium (SMC), Healthcare Improvement Scotland (HIS), Delta House (8th floor), 50 West Nile Street, Glasgow G1 2NP, Scotland, UK
- ⁵ La Haute Autorité de Santé (HAS), 5 Avenue du Stade de France, Saint-Denis La Plaine Cedex, 93218 Paris, France
- ⁶ Department of Epidemiology, University Medical Centre Groningen, Broerstraat 5, 9712 CP Groningen, The Netherlands

¹ The National Healthcare Institute (ZIN), Eekholt 4, 1112 XH Diemen, The Netherlands

Results Overall, 52 reports were retrieved, all of which contained REAs; CEAs were present in 25 of the reports. RWD was included in 28 of the 52 REAs (54%), mainly to estimate melanoma prevalence, and in 22 of the 25 (88%) CEAs, mainly to extrapolate long-term effectiveness and/or identify drug-related costs. Differences emerged between agencies regarding RWD use in REAs; the ZIN and IQWiG cited RWD for evidence on prevalence, whereas the 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, and 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.

Key Points for Decision Makers

Real-world data (RWD) may provide useful evidence on relative effectiveness (REAs) and cost effectiveness assessments (CEA) for reimbursement decisions.

This study showed that RWD is more often included in CEAs than REAs. In REAs and CEAs, RWD is often used to describe the effectiveness/safety of a new drug in clinical practice and to predict the long-term effectiveness of the new drug, respectively.

Differences emerged between agencies in how they use RWD for reimbursement decisions.

1 Introduction

Melanoma is the most serious and fatal form of skin cancer [1], and its incidence has been increasing, largely caused by increased exposure to ultraviolet 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 the literature, prior to 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 four novel mechanisms of action, thereby substantially increasing treatment options [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 with one or more alternative interventions for achieving the desired results and when provided under the routine setting of healthcare 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 controlled 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 to 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 with 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 the natural history of diseases, delineate clinical treatment pathways, determine costs and resource use associated with treatments, and to examine 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 five HTA agencies in Europe.

2 Methods

Methods used were comparable with 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 six European jurisdictions were selected for inclusion, since they make full reports publicly available: National Institute for Health and Care Excellence (NICE), England; Scottish Medicines Consortium (SMC), Scotland; Haute Autorité de santé (HAS), France; Institute for Quality and Efficacy in Healthcare (IQWiG), Germany; Agency for Health Technology Assessment and Tariff System (AOTMiT), Poland; and Zorginstituut Nederland (ZIN), The Netherlands. However, due to the authors' inability to read Polish reports, the study proceeded with five agencies.

HTA reports on seven 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 1 January 2011 and 31 December 2016, and the availability of at least three reports, published by three 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 ESM Appendix 1). The inclusion of RWD in REAs and CEAs was examined separately. When RWD was included, two 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 and AvV, the inter-rater reliability (IRR) was calculated twice in two different rounds. In each round, authors independently extracted data from four randomly selected reports (see ESM 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 among 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 five senior assessors representing the five respective agencies was consulted (see ESM 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, who 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.

2.1 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. The source(s) of RWD used per parameter and the frequency thereof were then 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 among the five 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.

3 Results

Sixty-five reports were identified for the seven drugs on the agencies' websites, of which 52 were indicated for melanoma; all 52 were published between 1 January 2011 and 31 December 2016. NICE, HAS, and IQWiG published at least one report for all seven drugs, allowing for the inclusion of all 52 reports (see ESM Appendix 4 for the full list). The distribution of reports across 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, the IQWiG and HAS reports did not include CEAs. In total, 25 CEAs were located in the reports from NICE, SMC and ZIN. It is important to note that ZIN reports entailed initial assessments as part of conditional reimbursement schemes (CRSs), and, as such, 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 and 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 ESM 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-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.

Table 1 Parameters for which real-world data are included, and real-world data 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 (<i>n</i> = 22)	Long-term effectiveness	21	Registry (<i>n</i> = 21)
		National statistics database (<i>n</i> = 9)			National statistics database (<i>n</i> = 12)
		Observational study (<i>n</i> = 6)	Costs	12	Claims database (<i>n</i> = 10)
		Claims database (<i>n</i> = 2)			Observational study (<i>n</i> = 4)
Effectiveness	7	Observational study (<i>n</i> = 6)	Resource use	8	Cost-of-illness study (<i>n</i> = 1)
		Non-randomized phase I/II trial (<i>n</i> = 6)			Observational study (<i>n</i> = 7)
		Registry (<i>n</i> = 1)			Claims database (<i>n</i> = 4)
Safety	6	Non-randomized phase I/II trial (<i>n</i> = 4)	Quality-of-life data	4	Registry (<i>n</i> = 1)
		Observational study (<i>n</i> = 3)			Quality-of-life study (<i>n</i> = 3)

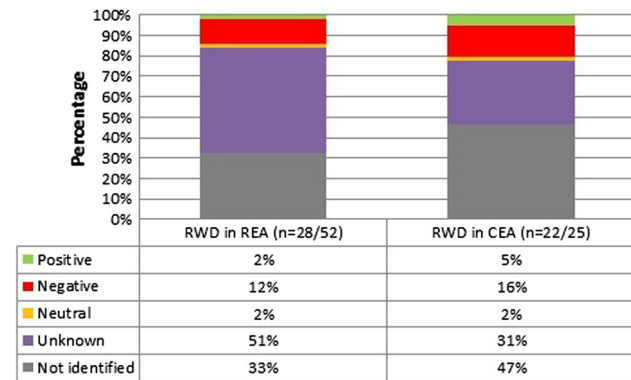


Fig. 1 Appraisal of the validity of RWD use and sources chosen when included in REAs and CEAs

The inclusion of RWD in REAs differed between the five agencies. For example, NICE reports cited RWD in 10/10 (100%) REAs, while SMC reports cited RWD in

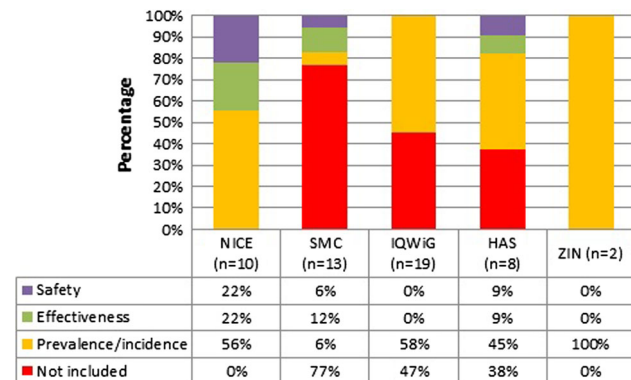


Fig. 2 Inclusion of RWD in REAs and the reasons for inclusion per agency

3/10 (33%) (Fig. 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 > 75% for all three agencies (Fig. 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 (Fig. 4), and is shown separately in Figs. S1 and S2 in ESM Appendix 5. No trend was visible for RWD inclusion in REAs; however, the inclusion of RWD in CEAs exceeded 75% in all years (2011–2016), displaying no visible variation in trend.

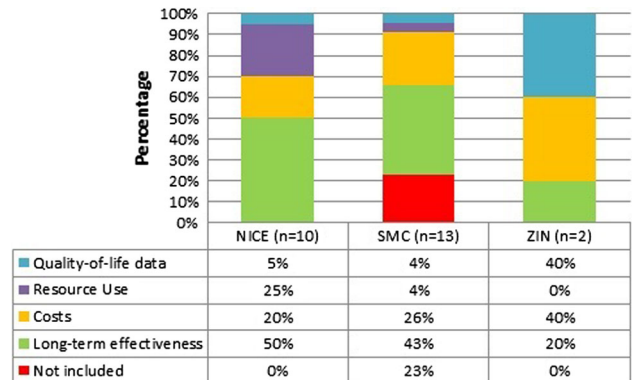


Fig. 3 Inclusion of RWD in CEAs across the 3 agencies and reasons for inclusion per agency

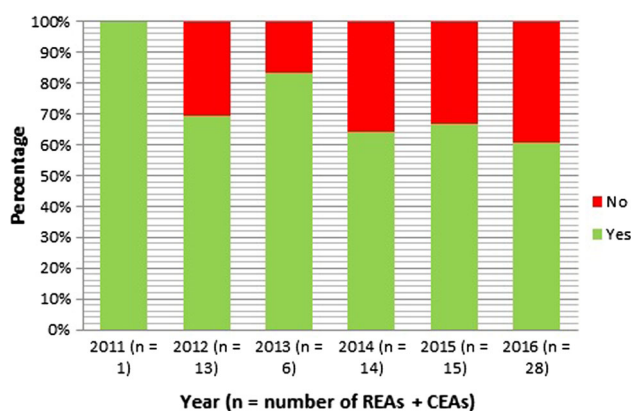


Fig. 4 Inclusion of RWD in REAs and CEAs (combined) over time

In the current study, only 2 of the 52 reports were initial assessment reports within conditional reimbursement schemes (CRSs), namely those published by ZIN; however, the respective reassessment reports have not yet been published. We will return to the implications of this in the Sect. 5 below.

4 Discussion

This study examined the extent with which RWD was included and its appraisal in HTA reports of seven melanoma drugs from five 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.

The 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 among six HTA agencies (four 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 CRSs, 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 RCTs 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 RCTs remain the preferred source for relative effectiveness estimates in CEAs.

In the past 10 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-Centered 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 regard 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 as to 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 among 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, i.e. 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 crossover between arms based on interim analyses [14, 28]. The advantages of PCT 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 prespecified 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 regard 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, CRSs provide an ideal context for incorporating RWD in HTA. The value of RWD generated in CRSs would play a critical role in the reassessment of drugs (e.g. to confirm previous efficacy estimates, cost-effectiveness ICER estimates or budget impact). According to previous research, policies for CRSs implemented by three agencies indicated that RWD is largely accepted within this context, provided data collection and analysis abide by predefined conditions [23]. In the current study, only 2 of the 52 reports were initial assessment reports within CRSs, namely those published by ZIN; however, the respective reassessment reports have not yet been published. Moreover, HAS reports examined were not part of CRSs implemented in France. As such, the potential role of RWD in melanoma reports within CRSs could not be assessed. To our knowledge, work is ongoing within ZIN and HAS to reassess melanoma drugs using RWD. Therefore, provided no similar study on RWD inclusion and appraisal within CRSs across HTA agencies has been performed, this should be the focus of future research once reassessment reports are published.

4.1 Strengths

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

The IRR between the two 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 five senior assessors representing all five agencies included, to verify whether the results accurately represent practice within their agency, thus improving their plausibility.

4.2 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. Nonetheless, the inclusion of the 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 with those presented in this current study. 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 (Canadian Agency for Drugs and Technologies in Health (CADTH)) and Australia (Pharmaceutical Benefits Advisory Committee (PBAC))].

The comparison of RWD inclusion and RWD appraisal between the five agencies and over time was complicated by the varying number of reports published per agency, per year, and the 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 in this study 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, the authors 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 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 the generalizability of results to broader HTA practice.

5 Conclusions

In general, RWD was more often included in CEAs than in REAs of HTA reports. The main reason for inclusion in REAs was the prevalence and/or incidence of melanoma, and in CEAs the main reason for inclusion was 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 five agencies, with 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 CRSs implemented by different HTA agencies, which provide an ideal context for RWD use in HTA practice, and across multiple disease indications.

Acknowledgements The authors would like to thank Ms. Rachel Kalf (ZIN) for her help in assessing agreement in data extracted by the authors for open-ended questions, as part of determining the IRR.

Author contributions AM was involved in establishing the study aim, designing the study protocol, implementing the study protocol, writing the manuscript and corresponding with the journal throughout the review process. AvV was involved in designing the study protocol, implementing the study protocol and writing the initial manuscript, and providing feedback on subsequent versions of the manuscript throughout the review process. PJ, OM and AA were involved in the HTA panel consulted to verify the results of the manuscript, and providing feedback on the initial manuscript written and on subsequent versions throughout the review process. AdB, HH, OK and WG were involved in establishing the study aim, designing the study protocol, reviewing findings generated by the study, and providing feedback on the initial manuscript written and on subsequent versions throughout the review process.

Compliance with Ethical Standards

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

Availability of data and materials The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflict of interest Amr Makady, Ard van Veelen, Páll Jonsson, Owen Moseley, Anne D'Andon, Anthonius de Boer, Hans Hillege and Wim Goettsch declare that they have no conflicts of interest. Olaf Klungel declares that grants have been received by the Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, where he is employed, but no conflicts of interest exist with regard to the subject matter of this article. He also declares to have received a small fee for an educational lecture on unmeasured confounding for Roche B.V., but, again, no conflicts of interest exist with regard to the subject matter of this article.

Funding None received.

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.

References

- Liu Y, Sheikh MS. Melanoma: molecular pathogenesis and therapeutic management. *Mol Cell Pharmacol*. 2014;6:228.
- Kumar P, Clark ML. *Kumar and Clark's Clinical Medicine*, 9th edn. 2012. Elsevier, Canada
- Azoury C, Lange R. Epidemiology, risk factors, prevention, and early detection of melanoma. *Surg Clin N Am*. 2014;94:945–62.
- Makady A, Kalf R, Goettsch W, Lees M. Deliverable D1.6 WPI case study review: metastatic melanoma. 2016. http://www.imigetreal.eu/Portals/1/Documents/01%20deliverables/Deliverable%20D1.6%20-%20Metastatic%20Melanomav2_website%20version.pdf. Accessed 17 Nov 2017.
- Agarwala SS. Current systemic therapy for metastatic melanoma. *Expert Rev Anticancer Ther*. 2009;9:587–95.
- Julia F, Thomas L, Dumontet C, Dalle S. Targeted therapies in metastatic melanoma: toward a clinical breakthrough? *Anticancer Agents Med Chem*. 2010;10:661–5.
- Rubio-Rodríguez D, Blanco SDD, Pérez M, Rubio-Terrés C. Cost-effectiveness of drug treatments for advanced melanoma: a systematic literature review. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-017-0517-1>. (epub 27 May 2017).
- 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;27:1768–75.
- Directorate-General for Enterprise and Industry (European Commission), Directorate-General for Health and Consumers (European Commission). High level pharmaceutical forum 2005-2008 conclusions and recommendations. 2008. <https://publications.europa.eu/en/publication-detail/-/publication/4fddf639-47cc-4f90-9964-142757d2515a>. Accessed 17 Nov 2017.
- Kleijnen S, George E, Goulden S, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. *Value health*. 2012;15:954–60.
- Drummond MF, Sculpher MJ, Claxton K, et al. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2015.
- 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.
- 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.
- Makady A, Goettsch W, Hummel N, et al. GetReal. D1.3 - GetReal Glossary of Definitions of Common Terms. 2016. http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/D1.3%20-%20Revised%20GetReal%20glossary%20-%20FINAL%20updated%20version_25Oct16_webversion.pdf. Accessed 2 Nov 2017.
- 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.
- Alemayehu D, Mardekian J. Infrastructure requirements for secondary data sources in comparative effectiveness research. *J Manag Care Pharm*. 2011;17:S16–21.
- 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.
- 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:S3–37.
- 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.
- Schmidt AF, Klungel OH, Groenwold RH, Consortium G. Adjusting for confounding in early postlaunch settings: going beyond logistic regression models. *Epidemiology*. 2016;27:133–42.
- Myrtveit I, Stensrud E, Olsson UH. Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods. *IEEE Trans Softw Eng*. 2001;27:999–1013.
- 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. <http://www.imi-getreal.eu/Portals/1/Documents/01%20deliverables/2017-03-30%20-%20WP4%20-%20Methodological%20guidance%2C%20recommendations%20and%20illustrative%20case%20studies.pdf>. Accessed 17 Nov 2017.
- 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;20:520–32.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 159–174.
- Makady A, Goettsch W. Review of policies and perspectives on real-world data. IMI-GetReal deliverable. *Value Health*. 2015;18(7):A567.
- Overall Objectives of IMI-GetReal. 2017. <http://www.imi-getreal.eu/About-GetReal/Overall-objectives>. Accessed 17 Nov 2017.
- EUnetHTA. Work package 5—life cycle approach to improve evidence generation. 2017. <http://www.eunetha.eu/activities/eunetha-joint-action-3-2016-20/work-package-5-life-cycleapproach-improve-evidence-gener>. Accessed 17 Nov 2017.
- Zuidgeest M, Goetz I, Groenwold R, et al. Pragmatic trials and real world evidence: paper 1. Introduction. *J Clin Epidemiol*. 2017;88:7–13.
- 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? A review of definitions based on literature and stakeholder interviews. *Value Health*. 2017;20(7):858–65.
31. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. *Handb Health Econ Eval*. 2006.
32. Nederland Z. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland; 2015.
33. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: 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;35:1111–2.
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. *J Health Policy Outcomes Res*. 2015;2:23–30.