



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com



Original Research

Real-world outcomes associated with new cancer medicines approved by the Food and Drug Administration and European Medicines Agency: A retrospective cohort study



Jemma M. Boyle ^{a,1}, Gemma Hegarty ^{b,1}, Christopher Frampton ^c,
Elizabeth Harvey-Jones ^d, Joanna Dodkins ^d, Katharina Beyer ^e,
Gincy George ^e, Richard Sullivan ^{d,f}, Christopher Booth ^{g,2},
Ajay Aggarwal ^{a,d,f,*}

^a Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom

^b Royal Marsden Hospital, London, United Kingdom

^c Department of Medicine, University of Otago, Christchurch, New Zealand

^d Department of Clinical Oncology, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

^e Translational and Oncology Research (TOUR), King's College London, United Kingdom

^f Institute of Cancer Policy, King's College London, United Kingdom

^g Department of Oncology, Queen's University, Kingston, Canada

Received 30 May 2021; received in revised form 1 July 2021; accepted 6 July 2021

Available online 6 August 2021

KEYWORDS

Real-world data;
Drug approval;
FDA/EMA;
Study quality;
Survival;
Novel cancer drugs

Abstract Purpose: Real-World Data (RWD) studies are increasingly used to support regulatory approvals, reimbursement decisions, and changes in clinical practice for novel cancer drugs. However, few studies have systematically appraised their quality or compared outcomes to pivotal trials.

Methods: All RWD studies (2010–2019) for drugs approved by the Food and Drug Administration (FDA) and European Medicines Agency (EMA) from 2010 to 2015 for solid organ tumours in the non-curative setting were identified. Quality assessment was undertaken using the Newcastle Ottawa Scale. Survival differences between each RWD study and the pivotal trial were determined using a related sample Wilcoxon signed-rank test.

Results: 293 RWD studies for 45 of the 57 drug indications approved by the FDA/EMA were identified. The most common tumour types were prostate cancer (29%, $n = 86$) and melanoma (15%, $n = 43$). A quarter of the studies had industry funding. No high-quality studies were identified, and 78% were low quality. Comparative survival analysis between RWD and

* Corresponding author: Department of Health Services Research and Policy, Tavistock Place, London, WC1H 9SH, United Kingdom.

E-mail address: ajay.aggarwal@lshtm.ac.uk (A. Aggarwal).

¹ JB and GH are joint first authors. ² AA and CB are joint final authors.

pivotal trials was possible for 224 studies (37 drug indications). Differences in median survival between the RWD studies and their corresponding trial ranged from –32 months to 21 months (IQR –4.2 months to 1.6 months). Low-quality studies were more likely to report superior survival outcomes (23%) compared to higher quality studies (8%) ($p = 0.02$).

Conclusion: RWD study quality for novel cancer drugs is low and of insufficient rigour to inform reimbursement decisions and clinical practice. RWD studies seeking publication should provide a completed quality assessment tool on submission. Greater investment in properly designed RWD studies is required.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Real-World Data (RWD) is population- or institution-level data collected either prospectively or retrospectively from non-randomised observational sources such as electronic health records [1], billing claims, insurer databases, and disease registries [2,3]. RWD has rapidly expanded to influence and inform a wide range of activities from regulatory approvals through to health technology assessment (HTA) and clinical guidelines [4].

RWD studies allow an assessment of treatment effectiveness in diverse non-selected populations and are an important adjunct to Randomised Controlled Trials (RCTs), particularly to obtain data on late toxicities, rare events, and long-term outcomes [1,5,6]. They also offer critical insights into the quality and outcomes achieved in routine practice, particularly for the vast majority of patients who would not meet eligibility criteria for RCTs. In contrast, RCTs have been considered the gold standard for measuring the efficacy and short-term toxicity of an intervention. The simple act of randomisation ensures high internal validity (ie ability to capture ‘true’ treatment effect); however, strict eligibility criteria can limit the external validity (ie generalisability) of results [7].

Regulatory agencies such as the Food and Drug Administration (FDA) have demonstrated an increased willingness to use RWD data to support marketing authorisations [4]. HTAs are also increasingly requesting RWD to assess whether interventions provide clinically meaningful benefits, particularly in light of the lowering of thresholds for regulatory approval by both the FDA and European Medicines Agency (EMA) [8,9]. In circumstances where uncertainties exist in the post-marketing setting, RWD is being used to define conditional reimbursement schemes or managed entry agreements. For example, coverage with evidence development (CED), risk-sharing agreements, or payments for outcomes [10].

However, several potential flaws with the use of RWD for these processes have been highlighted, particularly around the use of historical controls and surrogate endpoints, as well as issues regarding data quality and its validity [4,11].

The role of RWD to support drug regulatory decisions, HTA, and clinical practice guidelines, therefore, continues to be the subject of ongoing consultation and debate [12,13]. However, these discussions are being undertaken in the absence of a systematic assessment of RWD to understand the quality of the studies presently undertaken and the outcomes typically delivered in ‘real-world’ populations.

In this study, we sought to empirically evaluate all contemporary RWD studies reporting the effectiveness of cancer drugs for the treatment of advanced/metastatic solid organ malignancies published over the last 10 years. Our specific objectives were to describe their quality and the extent to which survival outcomes for patients in the real world were comparable to those observed in their pivotal RCTs.

2. Methods

This retrospective cohort study included all published RWD studies reporting the effectiveness of new cancer therapies that had been approved by the FDA and EMA between 1st January 2010 and 31st December 2015. We limited the cohort analysis to all antineoplastic and immunomodulating agents for solid tumours used in the non-curative setting (which accounts for approximately 97% of all indications). For all drug indications, we identified the corresponding pivotal trial through a review of the EMA and FDA approval documents (RCT or Phase 2 trial).

To identify relevant observational studies, we searched PubMed for each approved drug indication. Our search strategy included the drug name, approved indication, and search terms for ascertaining real-world studies (‘real world’, ‘population based’, ‘cohort’, ‘observational’, ‘registry’, ‘access scheme’). This was repeated for each drug indication and the list of relevant studies to enable comparison. Our latest search was on 31st July 2019, which allowed at least 4.5 years for the completion and publication of RWD studies.

Fully published reports were eligible if they reported survival outcomes in the same indication, combination, and line of therapy as the FDA/EMA approval. Observational studies with mixed populations ie, had a proportion of patients receiving the drug in an earlier

line than approved or with a different chemotherapy backbone (other standard chemotherapy agents given with the drug of choice) were also included. Studies were excluded if the drug was used solely in an earlier line of therapy compared to the pivotal trial, a different chemotherapy backbone was used, and/or the study did not seek to assess the survival of patients in the study cohort.

Of the studies meeting the inclusion criteria, survival analysis was restricted to those RWD studies for which both the RWD study and pivotal trial reported the median overall survival. RWD studies reporting the median survival for only a subset of the whole study population, different treatment backbone, or for an earlier line of therapy were excluded.

Data extracted included: location of study, study type (multicentre/single centre), number of patients receiving an intervention, prospective or retrospective evaluation, line of therapy, and chemotherapy backbone used, where applicable. The age, gender, and performance status of participants were included, as well as the median overall survival observed where available. The source of study funding and whether the article was open access was also recorded.

2.1. Quality assessment

To assess the quality of the included studies, we used the Newcastle Ottawa Scale (NOS) that scores cohort studies 0–9 according to eight criteria across three domains: (1) selection of the study groups; (2) comparability of groups; and (3) ascertainment of exposure and outcome. However, most studies identified in our cohort were single-centre case series evaluating the intervention alone or sought to compare the intervention drug with another comparator drug not used in the pivotal study. To evaluate the intervention arm, we used the previously validated modified Newcastle Ottawa scoring system for the case series, which scores the intervention arm on a scale of 0–6 [14]. This scoring system is derived from the original NOS but excludes an assessment of the comparability of the intervention and non-intervention arm. In line with the previous use of the NOS [15] appraisal tool, studies with scores of 0–3 have a high risk of bias and are considered of low quality. Studies with scores of 4–6 are of moderate quality, and studies with scores of 7–9 are at low risk of bias and considered of high quality. Quality assessment was performed independently by GH, JB, EHJ, and JD. AA performed a random duplicate assessment of 20% of the papers.

2.2. Survival analysis

For each drug-approved indication, we collated RWD studies meeting our inclusion criteria for the survival analysis and plotted the difference in median overall survival between each observational study and their

corresponding pivotal trial. A frequency table was then produced to demonstrate the proportion of observational studies across all drug indications for which the median overall survival outcomes were greater or less than the index study. A Related Sample Wilcoxon signed-rank test was performed to assess whether the overall difference across all studies was statistically significant. We also assessed the proportion of RWD studies that reported median overall survival, which was better, comparable or worse than the index study using both a 10% and then a 20% survival threshold.

3. Results

From 2010 to 2015, the EMA and FDA approved the use of 50 drugs for 57 solid organ cancer indications in the advanced or metastatic setting ([Supplementary Table 1](#)). These approvals were supported by 60 clinical trials, 55 of which were phase III RCTs. Of 4672 RWD studies identified by the search strategies, 323 full-text articles were deemed potentially eligible and reviewed in full. We subsequently excluded 30 studies ([Supplementary Fig. 1](#)) because they did not assess overall survival; a composite survival was reported across different drug combinations; or the study was a duplicate, which presented data from a cohort of patients already included as part of another study. The final study cohort included 293 RWD studies for 45 of the 57 indications.

3.1. Characteristics of RWD studies

Characteristics of all RWD ($n = 293$) studies identified are shown in [Table 1](#); 98% ($n = 288$) of RWD studies were case series. Five studies were comparative cohort studies, which replicated the comparison undertaken in the pivotal trial. The most common tumour types were prostate (29%, $n = 86$), melanoma (15%, $n = 43$), colorectal (12%, $n = 34$), lung (11%, $n = 33$), and renal (11%, $n = 32$). The mechanism of action of agents in the studies of RWD were small molecule inhibitors (37%, $n = 108$), cytotoxic agents (22%, $n = 65$), hormonal agents (17%, $n = 50$), and monoclonal antibodies (16%, $n = 48$).

Three-quarters of studies (76%, $n = 223$) originated from eleven countries across Europe (Italy, France, Spain, UK, the Netherlands, and Poland), South East Asia (Japan, China, and South Korea), and North America (the USA and Canada). Of these countries, Italy published the largest number of studies (18%, $n = 52$) followed by Japan (10%, $n = 30$), and China (10%, $n = 28$) ([Table 1](#)).

Only 2% of studies ($n = 6$) used data from national cancer registries, and 38% ($n = 112$) were from single-

Table 1

Characteristics of the RWD studies (n = 293 for 45 drug indications) identified for FDA and EMA approved indications, including RWD studies included in the survival analyses (n = 224 for 37 drug indications).

| | Observational RWD Studies | | Observational RWD studies included in survival analyses | |
|-------------------------------------|---------------------------|------|---|------|
| | No. | % | No. | % |
| Indication | | | | |
| Prostate | 86 | 29.4 | 72 | 32.1 |
| Melanoma | 43 | 14.7 | 33 | 14.7 |
| Colorectal | 34 | 11.6 | 34 | 15.2 |
| Lung | 33 | 11.3 | 16 | 7.1 |
| Renal | 32 | 10.9 | 24 | 10.7 |
| Breast | 23 | 7.8 | 16 | 7.1 |
| Gastric | 16 | 5.5 | 14 | 6.3 |
| Sarcoma | 13 | 4.4 | 7 | 3.1 |
| Thyroid | 7 | 2.4 | 4 | 1.8 |
| Pancreatic | 3 | 1.0 | 3 | 1.3 |
| Ovarian | 3 | 1.0 | 1 | 0.4 |
| Drug Type | | | | |
| Small molecule inhibitor | 108 | 36.9 | 73 | 32.6 |
| Cytotoxic | 65 | 22.2 | 57 | 25.4 |
| Hormonal | 50 | 17.1 | 40 | 17.9 |
| Monoclonal antibody | 48 | 16.4 | 40 | 17.9 |
| Radionuclide | 14 | 4.8 | 11 | 4.9 |
| Immunotherapy | 7 | 2.4 | 3 | 1.3 |
| Vaccine | 1 | 0.3 | 0 | 0.0 |
| Country | | | | |
| Italy | 52 | 17.7 | 44 | 19.6 |
| Japan | 30 | 10.2 | 22 | 9.8 |
| China | 28 | 9.6 | 22 | 9.8 |
| USA | 26 | 8.9 | 16 | 7.1 |
| France | 22 | 7.5 | 18 | 8.0 |
| Spain | 15 | 5.1 | 12 | 5.4 |
| Canada | 14 | 4.8 | 9 | 4.0 |
| South Korea | 13 | 4.4 | 12 | 5.4 |
| UK | 8 | 2.7 | 6 | 2.7 |
| Netherlands | 8 | 2.7 | 8 | 3.6 |
| Poland | 7 | 2.4 | 5 | 2.2 |
| Multi-country | 14 | 4.8 | 9 | 4.0 |
| Other individual country | 56 | 19.1 | 41 | 18.3 |
| Study Type | | | | |
| Prospective | 50 | 17.1 | 39 | 17.4 |
| Retrospective | 242 | 82.6 | 184 | 82.1 |
| Unknown | 1 | 0.3 | 1 | 0.4 |
| Multicentre | | | | |
| Yes | 180 | 61.4 | 135 | 60.3 |
| No | 112 | 38.2 | 88 | 39.3 |
| Unknown | 1 | 0.3 | 1 | 0.4 |
| Registry | | | | |
| Yes | 6 | 2.0 | 3 | 1.3 |
| No | 287 | 98.0 | 221 | 98.7 |
| Study size (<50 patients) | | | | |
| Yes | 103 | 35.2 | 88 | 39.3 |
| No | 190 | 64.8 | 136 | 60.7 |
| Funding | | | | |
| Industry | 79 | 27.0 | 63 | 28.1 |
| Other (non-industry) | 39 | 13.3 | 25 | 11.2 |
| No | 163 | 55.6 | 126 | 56.3 |
| Unknown | 12 | 4.1 | 10 | 4.5 |

centre evaluations. Almost a third of studies were funded by the pharmaceutical industry (27%, n = 79).

The overall median age for the pivotal RCT studies was 61 years (IQR 56–62 years) compared to 64 years (IQR 58–69 years) for the RWD studies. Twenty per cent (n = 60) of RWD studies included patients with a median age >5 years older than the index trial. Where reported, few studies included patients with an ECOG performance status of 3, but the lack of a reported performance status breakdown in most studies meant comparison of the proportion of patients of ECOG performance status 0–2 was not feasible. Less than 20% of RWD studies reported comorbidity.

3.2. Study quality

All 293 studies were scored out of nine according to the Newcastle Ottawa Scale (NOS). The distribution of scores across the studies is presented in Fig. 1. Seventy-eight per cent of studies (n = 230) were classified as low quality (score 0–3), 22% (n = 63) moderate quality (score 4–6), and no studies were classified as high quality (score 7–9).

The proportion of studies meeting each individual scoring criteria in the NOS is outlined in Table 2a for case series (n = 288) and Table 2b for cohort studies (n = 5). For the case series, we found that only 22% of studies adequately evaluated survival (eg, method of ascertaining patient deaths), and only 33% of studies had sufficient follow-up to estimate survival. In addition, only 38% met the criteria for an adequate description of the selection of participants.

Study quality was evaluated according to funding status. For studies that were funded, there was a trend towards better quality compared to those without funding, although this was not statistically significant (p = 0.082). However, when studies with industry (pharmaceutical) funding were compared to those without, there was a statistically significant difference with 35% (n = 28) of studies classified as medium quality, compared to 17% (n = 34) of studies without industry funding (p = 0.001).

Supplementary Figs. 2–4 report quality trends according to study country, tumour type, and individual drug indication. The Netherlands had the highest proportion of studies scoring 4–6 with 63% (n = 5 out of 8), followed by multicentre international studies with 57% (n = 8 out of 14), Italy with 36% (n = 15 out of 42), and Spain with 33% (n = 5 out of 15) (Supplementary Fig. 2). According to tumour type, sarcoma (46%, n = 6 out of 13), gastric cancer (31%, n = 5 out of 16), and breast cancer (30%, n = 7 out of 23) had the highest proportion of studies scoring 4–6 (Supplementary Fig. 3). When assessing quality according to individual drug indications, almost half (44%, n = 20) had no RWD studies that scored 4–6 (Supplementary Fig. 4).

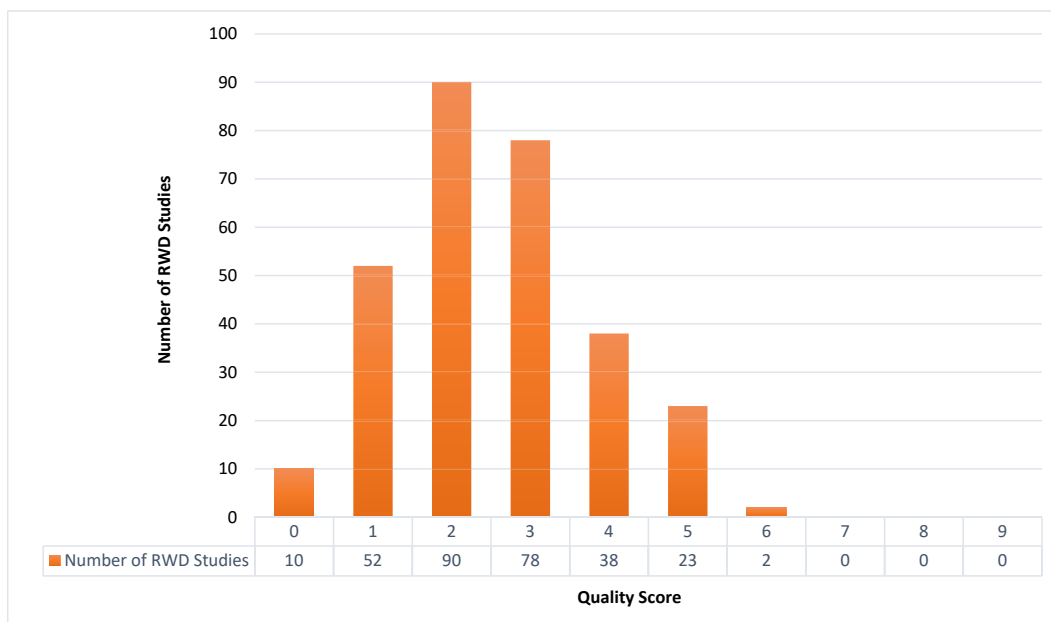


Fig. 1. Histogram of the distribution of total scores for RWD studies appraised using the Newcastle Ottawa Scale.

Table 2a

Breakdown of scores for case series (n = 288) using the modified Newcastle Ottawa Score.

| Validation Question | Proportion meeting criteria for YES score |
|---------------------------|---|
| 1; Selection | 108 (38%) |
| 2; Exposure | 188 (65%) |
| 3; Outcome | 64 (22%) |
| 4; Confounding | 173 (60%) |
| 5; Follow-up | 94 (33%) |
| 6; Replication/inferences | 102 (35%) |

Table 2b

Breakdown of scores for cohort studies (n = 5) using the Newcastle Ottawa Score.

| Validation Question | Proportion meeting criteria for YES score |
|---------------------------------|---|
| 1; Representativeness | 0 (0%) |
| 2; Selection of exposed cohort | 3 (60%) |
| 3; Exposure | 1 (20%) |
| 4; Outcome not present to start | 5 (100%) |
| 5 ^a ; Comparability | 1 (20%) |
| 6; Outcome | 1 (20%) |
| 7; Follow-up | 1 (20%) |
| 8; Adequacy of follow-up | 3 (60%) |

^a Worth two points.

3.3. Survival outcomes of systemic agents in the real world

The survival analysis included 224 of the initial 293 RWD studies identified (characteristics are described in Table 1) for 37 of the 45 drug indications with RWD studies. We summarised as a frequency chart the difference in median overall survival reported for all 224 studies compared to the index trial for each of the 37 drugs

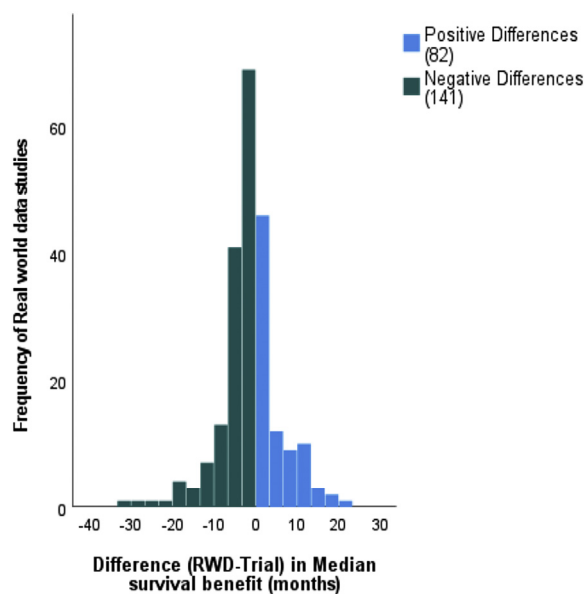


Fig. 2. Frequency chart reporting survival differences between RWD^a studies (n = 224 for 37 drug indications) and their corresponding pivotal trial for that drug indication. ^aReal-world data

approved indications (Fig. 2). Survival differences between the RWD studies and their corresponding trial ranged from −32 months to +21 months, with an interquartile range (IQR) (25th to 75th centile) from −4.2 months to +1.6 months. Thirty-seven per cent (82/224) of studies had superior survival outcomes compared to the pivotal trial, compared to 63% (141/224) of studies that had inferior survival outcomes. Across all RWD studies, the median survival difference was statistically inferior to the pivotal trial by −1.2 months (95% CI −1.7 to −0.6, p < 0.001).

We also undertook an analysis to assess the proportion of RWD studies that had survival outcomes that were superior, inferior, or comparable with the pivotal trial using $\pm 10\%$ (better/worse) and $\pm 20\%$ thresholds. The median overall survival across all 224 RWD studies was 13 months. At the 10% threshold, 26% of RWD studies ($n = 58$) had superior survival, 53% ($n = 119$) inferior survival, and 22% ($n = 49$) comparable survival to the pivotal trials. At the 20% threshold, 15% ($n = 35$) of RWD studies had superior survival, 40% ($n = 90$) inferior survival, and 45% ($n = 101$) had comparable survival.

When assessing the correlation between better survival and study quality, we found that lower quality RWD studies were more likely to report superior survival outcomes compared to the pivotal RCT. At the 10% level, 30% of low-quality studies reported better survival compared to 16% of moderate quality studies ($p = 0.129$). At the 20% level, 23% of low-quality studies reported better survival, compared to 8% of moderate quality studies ($p = 0.020$).

We observed numerous examples of RWD studies for the same indication that presented consistently superior

(eg panitumumab for colorectal cancer) or inferior (eg sorafenib for locally advanced thyroid cancer) median overall results compared to the pivotal study (Fig. 3). In addition, for some indications, we found contradictory survival benefits reported (superior and inferior) compared to the pivotal trial, with a wide range in the survival outcomes reported (eg eribulin for 2nd to 5th line metastatic breast cancer (RWD survival range -7.0 to $+14.8$ months)).

This analysis was stratified according to study quality. We found that compared to the pivotal trial, moderate quality studies reported less variation in the median OS (IQR -3.9 to 0.2) compared to low-quality studies (IQR -4.3 to 2.4). In addition, there was markedly less variation in outcomes reported across RWD studies of moderate quality for the same drug indication compared to low-quality studies for the same drug indication (Supplementary Figs. 5 and 6). A similar trend was noted when comparing studies according to funding status, with industry-funded studies demonstrating less variation in survival outcomes compared to those without industry funding (Supplementary Figs. 7 and 8).

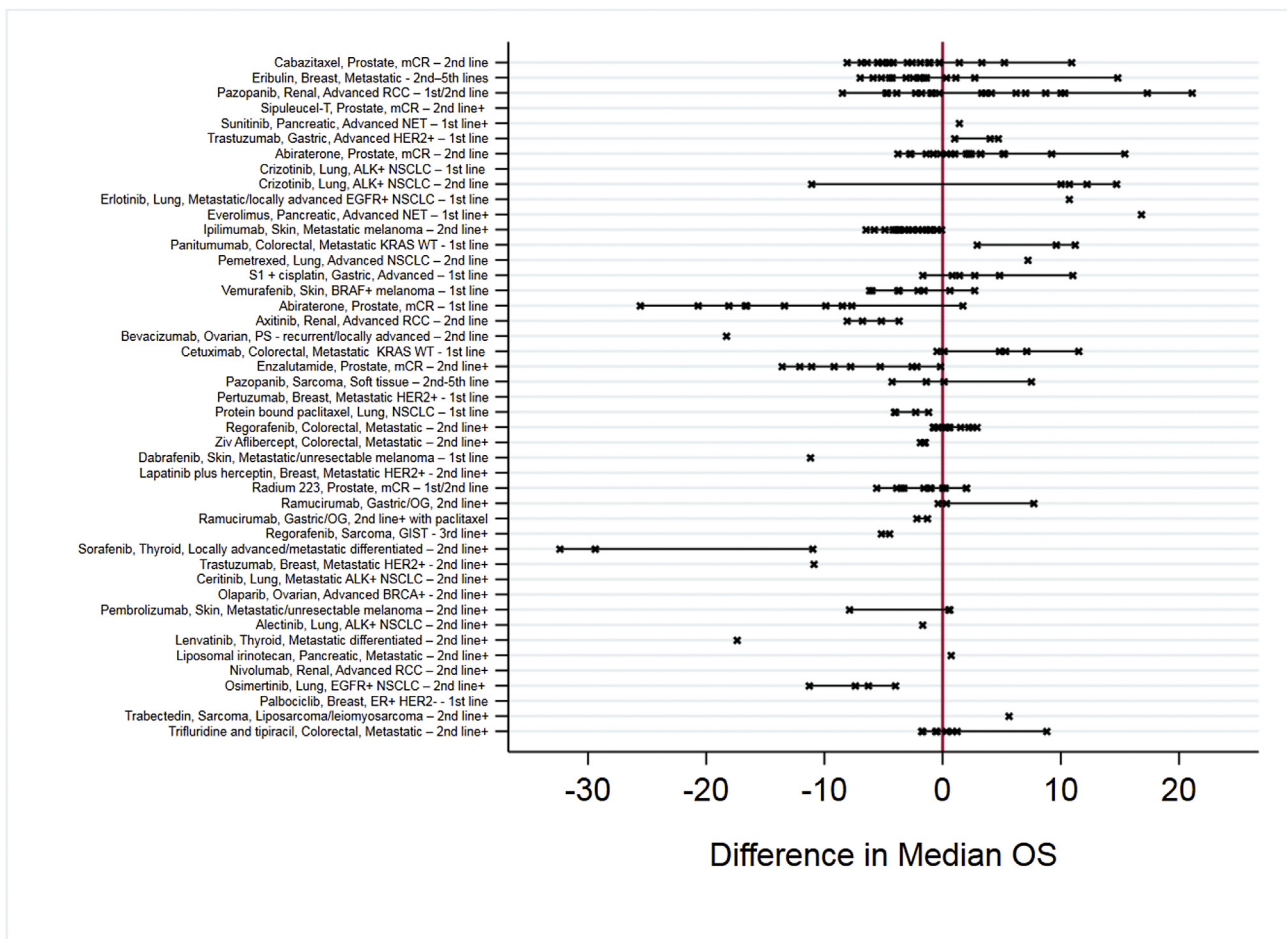


Fig. 3. Range of differences in the median overall survival (OS) (months) reported in individual RWD studies ($n = 224$ for 37 drug indications) for each drug indication relative to the median OS reported in the pivotal trial.

4. Discussion

This retrospective cohort study provides the first systematic evaluation of RWD studies reporting the effectiveness of cancer drugs for the treatment of solid organ malignancies approved by the EMA and FDA between 2010 and 2015. Overall, most EMA/FDA drugs now have RWD studies available. However, their methodologic quality is generally poor, with no high-quality studies identified and approximately 80% of the 293 studies evaluated scoring 0–3 (out of 9) using the Newcastle Ottawa Scale.

Patient selection, assessment and control of confounders, and evaluation of the study endpoint were identified as the main limitations of these studies. These studies would therefore not be considered of sufficient methodological rigour to inform practice or policy, with small, single-centre retrospective case series predominating. Only five of the 293 studies undertook a comparative assessment of the intervention drug with a comparator that was the same as that evaluated in the pivotal trial.

The range of differences in median overall survival outcomes in RWD studies compared to the corresponding pivotal trial was large (over two years in some examples) and conflicting, with some demonstrating superior and others inferior outcomes for the same drug indication. Importantly, the variation observed was considerably more for low-quality studies compared to those of high quality. Furthermore, low-quality studies were more likely to report superior outcomes for the intervention drug than the pivotal RCT.

Our findings are important as the narrative around real-world evidence does not highlight the broad range of studies of highly variable design and quality that come under this umbrella term, including retrospective case series data. While guidance on the design and conduct of RWD studies are available, the reality is that the term is used ubiquitously to cover all non-randomised studies that use routine health records [12]. A major issue is that data from such studies can be used as evidence of effect in routinely managed populations without any explicit reference to their methodological quality. For example, even for single-arm observational studies – in particular those evaluations within the context of compassionate access schemes for drugs awaiting reimbursement – selection bias remains a concern, as patients may receive a particular intervention over and above a comparator because physicians deem them to be fitter or have a greater likelihood of tolerating treatment [11].

One immediate policy change that could be implemented is for publishers to require that authors routinely complete a methodological critical appraisal checklist prior to submission to ensure transparency or that peer reviewers are expected to complete this as part of their assessment [16,17].

From a wider structural perspective, the study highlights the importance of ensuring adequate funding is available to develop high-quality outcomes research programs to inform policy and practice. We found that 54% of RWD studies included in our evaluation did not receive funding [18]. However, industry-funded studies were more likely to be of better quality than other studies. The actual cost of investing in RWD studies is very small compared to drug development and the amount spent on systemic therapies.

Even for well-designed observational studies, data may be missing, incomplete or not coded according to an established protocol. As such, extensive methodological work is necessary to curate and develop specific indicators (eg skeletal-related events) to enable meaningful evaluation of interventions using routinely collected data [19,20].

Worryingly, in our analysis, we found that only a very small proportion of studies used cancer registry data (2%). A major advantage of registry data relative to data from single or selected centres is the very large sample size and coverage of eligible patients, especially in single-payer systems.

However, the reality is that few countries have such large-scale linked hospital registration systems in place. This is due to the absence of a centralised data infrastructure to collate these measures, heavily fragmented public and private systems, and a lack of incentive amongst physicians and providers [21]. In addition, epidemiological research is still not a strategic priority amongst both public and philanthropic funders who have orientated almost exclusively around novel pharmaceutical and basic cancer research [22].

The efficacy to effectiveness gap has been used to describe differences in the outcomes observed in RCTs for new interventions and their subsequent impact under routine prescribing practice [23]. The efficacy to effectiveness gap observed in our analysis is likely to be due to differences in the characteristics of the population treated or the way in which care is delivered in routine practice [24]. With regards to the former, 40% of the RWD studies included in our analysis had significantly older populations compared to their corresponding pivotal trial, and some studies included specific sub-populations, which are frequently excluded from RCTs eg men and women with brain metastases. However, data on performance status and comorbidity was either not included or not presented with sufficient granularity to enable direct comparison. Our study findings of an efficacy to effectiveness gap complement a recent study, which has focused on Medicare patients in the United States treated with FDA approved cancer drugs between 2018 and 2020 [25], and a further study assessing the correlation between hazard ratios of observational studies undertaken using US population-based registries and their matched RCT across different cancer interventions [26]. Of note, neither study appraised the methods of the RWD studies as we have done.

There are a number of limitations in the present study. We identified the corresponding pivotal trial from a review of the regulatory documents at the time of approval but acknowledge this is not always a one-to-one match. We have used a single database (PubMed) for identifying relevant articles up to August 2019, and it is possible that additional studies meeting our inclusion and exclusion criteria may have been missed. However, given the number and breadth of drugs identified over a 10-year period ($n = 293$) in this evaluation – which uses an established search database – we would not expect additional studies to significantly change our findings with respect to quality and outcomes. Similar methods have been used in other studies that sought to compare outcomes reported in observational studies with a defined cohort of pivotal trials [29].

The study was limited in scope to an assessment of those drugs reporting survival and did not include RWD studies reporting QOL and alternative outcomes such as progression-free survival if they did not also include overall survival. The study was not designed to give a precise estimate of effectiveness for each drug approved indication but a broad overview of the variation in results as reported. Given the overall poor quality of studies, the majority of which were case series, it is not possible to undertake a formal meta-analysis of the study results, and survival analyses reported are exploratory in nature. The strengths of the analysis are the inclusion of all relevant studies pertaining to consecutive drugs approved over a six-year period. It also provides an evaluation of study quality using an established critical appraisal framework.

Our study is timely given that RWD studies are increasingly being utilised to generate pharmaco-economic data [27], and gradually more countries are looking to RWD studies to generate outcome-based reimbursement data [28,29]. The trend for using RWD has also extended to multiple Health Technology Authorities such as the National Institute for Health and Care Excellence (NICE) in England and Wales and the Haute Autorite de Sante (HAS) in France [30].

Given the limitations of published RWD studies that this study has identified, significant improvements in the reporting of study methods and the design of studies are necessary before RWD data is used to support changes in clinical practice or reimbursement policy. Going forward, we recommend researchers consider the methodological frameworks developed by several organisations to support the design and conduct of RWD studies to improve their quality and their ability to assess variation in access to and outcomes of care [3,5,31,32]. In addition, we recommend the routine inclusion of a critical appraisal checklist as part of the submission process of RWD studies, as the entrenchment of poor-quality research can result in misinformation regarding the clinical effectiveness of cancer drugs.

5. Conclusion

Our study provides the first global systematic evaluation of RWD studies assessing the effectiveness of FDA and EMA approved drugs. We find that most new FDA and EMA approved drugs for solid organ cancers have RWD studies; however, the overall quality is very low and would presently be of insufficient rigour to support regulatory approvals and reimbursement. We also find that the majority of RWD studies report survival outcomes that are inferior to RCTs suggesting that the benefits observed in trials are not translated into the real world. Of concern is that low-quality studies are more likely to overstate the benefits of new cancer drugs. The standard of RWD studies of cancer drugs needs to improve and be more consistent prior to its routine use in support of clinical practice and policy change.

Author contribution

Jemma Boyle: Methodology, Investigation, Formal Analysis, Writing – original draft, Writing – review and editing; Gemma Hegarty: Investigation, Writing – review and editing; Christopher Frampton: Conceptualisation, Formal Analysis, Writing – review and editing, Visualisation; Elizabeth Harvey-Jones: Investigation, Writing – review and editing; Joanna Dodkins: Investigation, Writing – review and editing; Katharina Beyer: Investigation, Writing – review and editing; Gincy George: Investigation, Writing – review and editing; Richard Sullivan: Conceptualisation, Methodology, Writing – original draft, Writing – review and editing, Supervision; Christopher Booth: Conceptualisation, Methodology, Writing – original draft, Writing – review and editing, Supervision; Ajay Aggarwal: Conceptualisation, Validation, Methodology, Investigation, Writing – original draft, Writing – review and editing, Supervision.

Funding

This work was supported by a National Institute for Health Research (NIHR) Advanced Fellowship [NIHR300599] to [AA] and by the UK Research and Innovation Economic and Social Research Council [ES/P010962/1] to [RS]. These grant sources had no involvement in the study design, collection, analysis and interpretation of data, in the writing of the report, or the decision to submit the article for publication.

Conflict of interest

The authors have declared no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejca.2021.07.001>.

References

- [1] Skovlund E, Leufkens H, Smyth J. The use of real-world data in cancer drug development. *Eur J Canc* 2018;101:69–76.
- [2] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – what is it and what can it tell us? *N Engl J Med* 2016; 375(23):2293–7.
- [3] Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol* 2019; 16(5):312–25.
- [4] Raphael MJ, Gyawali B, Booth CM. Real-world evidence and regulatory drug approval. *Nat Rev Clin Oncol* 2020:1–2.
- [5] Garrison Jr 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.
- [6] Vandembroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004;363(9422):1728–31.
- [7] Lyman GH. Comparative effectiveness research in oncology. *The oncologist* 2013;18(6):752.
- [8] Bolislis WR, Fay M, Kühler TC. Use of real-world data for new drug applications and line extensions. *Clin Therapeut* 2020;42(5):926–38.
- [9] Hall PS. Real-world data for efficient health technology assessment. *Eur J Canc* 2017;79:235–7.
- [10] Lewis J, Kerridge I, Lipworth W. Coverage with evidence development and managed entry in the funding of personalized medicine: practical and ethical challenges for oncology. 2015. American Society of Clinical Oncology; 2015.
- [11] Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med* 2020; 382(7):674–8.
- [12] ESMO. What role for real-world evidence in cancer research?. 2021.
- [13] GOV.UK. Consultation document: MHRA draft guidance on randomised controlled trials generating real-world evidence to support regulatory decisions. In: Agency MHpR; 2020.
- [14] Murad MH, Sultan S, Haffar S, Bazerbachi F. Methodological quality and synthesis of case series and case reports. *BMJ Evid Based Med* 2018;23(2):60–3.
- [15] Wallis CJD, Saskin R, Choo R, et al. Surgery versus radiotherapy for clinically-localized prostate cancer: a systematic review and meta-analysis. *Eur Urol* 2016;70(1):21–30.
- [16] Wells GA, Shea B, Da O'Connell, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000. Oxford.
- [17] Adelaide) JBIUo. Critical appraisal tools. 2021.
- [18] Aggarwal A, Nossiter J, Parry M, et al. Public reporting of outcomes in radiation oncology: the national prostate cancer audit. *Lancet Oncol* 2021;22(5):e207–15.
- [19] Parry MG, Cowling TE, Sujenthiran A, et al. Identifying skeletal-related events for prostate cancer patients in routinely collected hospital data. *Canc Epidemiol* 2019;63:101628.
- [20] Sujenthiran A, Nossiter J, Charman SC, et al. National population-based study comparing treatment-related toxicity in men who received Intensity-Modulated versus 3D-Conformal Radical Radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2017;99(5):1253–60.
- [21] Palta JR, Efstathiou JA, Bekelman JE, et al. Developing a national radiation oncology registry: from acorns to oaks. *Pract Radiat Oncol* 2012;2(1):10–7.
- [22] Loucaides EM, Fitchett EJA, Sullivan R, Atun R. Global public and philanthropic investment in childhood cancer research: systematic analysis of research funding, 2008–16. *Lancet Oncol* 2019;20(12). e672–e84.
- [23] Nordon C, Karcher H, Groenwold RH, et al. The "Efficacy-Effectiveness gap": historical background and current conceptualization. *Value Health* 2016;19(1):75–81.
- [24] Templeton AJ, Booth CM, Tannock IF. Informing patients about expected outcomes: the efficacy-effectiveness gap. *J Clin Oncol* 2020;38(15):1651–4. 02035. JCO.19.
- [25] Green AK, Curry M, Trivedi N, Bach PB, Mailankody S. Assessment of outcomes associated with the use of newly approved oncology drugs in Medicare beneficiaries. *JAMA Network Open* 2021;4(2). e210030-e.
- [26] Soni PD, Hartman HE, Dess RT, et al. Comparison of population-based observational studies with randomized trials in oncology. *J Clin Oncol* 2019;37(14):1209–16.
- [27] Suvarna VR. Real world evidence (RWE) – are we (RWE) ready? *Perspect Clin Res* 2018;9(2):61–3.
- [28] Jørgensen J, Hanna E, Kefalas P. Outcomes-based reimbursement for gene therapies in practice: the experience of recently launched CAR-T cell therapies in major European countries. *J Mark Access Health Pol* 2020;8(1):1715536.
- [29] Justo N, Espinoza MA, Ratto B, et al. Real-world evidence in healthcare decision making: global trends and case studies from Latin America. *Value Health* 2019;22(6):739–49.
- [30] Makady A, van Veelen A, Jonsson P, et al. Using real-world data in health technology assessment (HTA) practice: a comparative study of five HTA agencies. *Pharmacoeconomics* 2018;36(3): 359–68.
- [31] Mahendraratnam N, Silcox C, Mercon K, Kroetsch A, Romine M, Harrison N. Determining real-world data's fitness for use and the role of reliability. 2019.
- [32] Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther* 2019;105(4): 867–77.