

REVIEW

Open Access



How is overall survival assessed in randomised clinical trials in cancer and are subsequent treatment lines considered? A systematic review

Kara-Louise Royle^{1*} , David Meads², Jennifer K. Visser-Rogers³, Ian R. White⁴ and David A. Cairns¹

Abstract

Background Overall survival is the “gold standard” endpoint in cancer clinical trials. It plays a key role in determining the clinical- and cost-effectiveness of a new intervention and whether it is recommended for use in standard of care. The assessment of overall survival usually requires trial participants to be followed up for a long period of time. In this time, they may stop receiving the trial intervention and receive subsequent anti-cancer treatments, which also aim to extend survival, during trial follow-up. This can potentially change the interpretation of overall survival in the context of the clinical trial. This review aimed to determine how overall survival has been assessed in cancer clinical trials and whether subsequent anti-cancer treatments are considered.

Methods Two searches were conducted using MEDLINE within OVID® on the 9th of November 2021. The first sought to identify papers publishing overall survival results from randomised controlled trials in eight reputable journals and the second to identify papers mentioning or considering subsequent treatments. Papers published since 2010 were included if presenting or discussing overall survival in the context of treating cancer.

Results One hundred and thirty-four papers were included. The majority of these were presenting clinical trial results (98, 73%). Of these, 45 (46%) reported overall survival as a (co-) primary endpoint. A lower proportion of papers including overall survival as a (co-) primary endpoint compared to a secondary endpoint were published in recent years. The primary analysis of overall survival varied across the papers. Fifty-nine (60%) mentioned subsequent treatments. Seven papers performed additional analysis, primarily when patients in the control arm received the experimental treatment during trial follow-up (treatment switching).

Discussion Overall survival has steadily moved from being the primary to a secondary endpoint. However, it is still of interest with papers presenting overall survival results with the caveat of subsequent treatments, but little or no investigation into their effect. This review shows that there is a methodological gap for what researchers should do when trial participants receive anti-cancer treatment during trial follow-up. Future research will identify the stakeholder opinions, on how this methodological gap should be addressed.

Keywords Systematic review, Overall survival, Randomised controlled trial, Cancer, Subsequent treatment lines

*Correspondence:

Kara-Louise Royle
k.l.royle@leeds.ac.uk

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Overall survival (OS), disease-free survival, progression-free survival and cause-specific survival are all used to assess the clinical effectiveness of experimental treatments in cancer clinical trials. However, due to subjectivity and variability in the definitions of response, progression, and cause of death, OS is the only endpoint that has a definite event which cannot be questioned. As a result of this, OS is often referred to as the “gold standard” endpoint for assessing the clinical effectiveness of experimental interventions for the treatment of cancer. In England, OS is used in conjunction with other evidence in appraisals by NICE (National Institute for Health and Care Excellence) to determine whether interventions should be recommended for use in the NHS. A recent review of Health Technology Assessment (HTA) systems in the European Union, UK, Canada and Australia, found that the model in England which uses clinical and cost-effectiveness data in their review alongside other evidence is similar to eight other HTA bodies including France and Sweden. In addition, there were 46 other HTA bodies across multiple countries including Italy, The Netherlands and Spain which also use both clinical and cost-effectiveness data in their appraisals [1].

To ensure consistency between clinical trials, guidance exists on their analysis and interpretation. The International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) topic E9 is entitled “Statistical Principles for Clinical Trials” and includes, for example, advice on how to analyse a trial when stratification factors are used during randomisation [2]. In addition, for time-to-event endpoints like OS, a stakeholder consultation on presenting results using Kaplan-Meier plots was published in 2019 [3]. Finally, OS itself has the published definitive definition as time from randomisation to death from any cause [4].

The assessment of OS requires trial participants to be followed up for what can be a considerable length of time. This poses a problem when it comes to interpreting the trial results, as participants may discontinue their trial intervention prior to death. For example, participants may discontinue at disease progression or earlier due to withdrawal from trial treatment (as a result of toxicity or patient choice), or an intervention stopping rule. Trial participants may then go on to receive multiple subsequent lines of anti-cancer treatment during trial follow-up. These subsequent treatments will also aim to control a patient’s cancer and therefore may improve OS. This improvement may reduce the number of OS events observed in the trial and in turn reduce the power available to detect a difference in OS between the experimental and control interventions. Consequentially, using

England as an example, this could lead to a beneficial intervention not being recommended for use in the NHS, resulting in sub-optimal patient outcomes.

This issue with interpretation is well known and is referenced in clinical trial recommendations [4]. Methodological solutions exist to account for scenarios where participants in the control arm switch to receive the experimental intervention, by design or in a later line of therapy (treatment switching). These methods were summarised as part of a technical support document in 2014 and include the rank preserving structural failure time model (RPSFTM) and the Inverse Probability Censoring Weights (IPCW) method [5].

It is currently unknown how widely the published guidance on the analysis of OS is followed and how researchers at present perceive subsequent treatment lines as a whole compared to the specific scenario of treatment switching. Therefore, the aim of this systematic review was to determine how OS is currently analysed in practice and whether subsequent lines of treatment are mentioned in the main manuscript or supplementary material of cancer clinical trials publications.

Methods

Search strategy

The electronic database MEDLINE using OVID® was searched on 9 November 2021. To ensure that the search was not biased towards trials that did or did not report the limitations surrounding the analysis of OS in terms of subsequent treatments, two approaches were taken (Table 1). The first approach aimed to gain a broad view of the analysis of OS in cancer clinical trials. Due to the breadth of publications available, the search was restricted to several major journals focusing on general medicine and cancer: *Lancet*, *Lancet Oncology*, *JCO*, *PLOS One*, *PLOS Medicine*, *BMJ*, *JAMA* and *New England Journal of Medicine*. The second approach aimed to target papers which were specifically discussing the issue around subsequent treatments in conjunction with the analysis of OS in cancer clinical trials. Both searches were restricted to the English language and to papers published since 1 January 2010. The results of the two searches were combined with the Boolean operator OR and all abstracts were extracted into Microsoft Excel (Microsoft, Redmond, WA, USA) and Endnote™ (Clarivate, Philadelphia, PA, USA).

Initial screening process

The results were screened based on their abstracts. The screening process was conducted by K-LR, reviewed by DAC and any uncertainty discussed. The screening process categorised the papers into red (not relevant), amber (potentially relevant), and green (relevant) (Fig. 1). Any

Table 1 Literature search terms

Approach 1 — Overall Survival				
Concept	"Overall Survival".tw.	Randomi?ed Controlled Trial.tw.	Cancer.tw.	
Synonyms	OS mortality	RCT	Carcinoma Neoplasm Sarcoma	
MESH	"Survival Analysis"	Randomized Controlled Trials as Topic	Neoplasms	
Approach 2 — Subsequent Treatment Lines				
Concept	"Overall Survival".tw.	Randomi?ed Controlled Trial.tw.	"Subsequent Treatments"	Cancer.tw.
Synonyms	OS mortality	RCT	(Line* adj3 therap*) (Line* adj3 treatment*) (Subsequent adj3 therap*) (Subsequent adj3 treatment*) (Subsequent adj3 medication*)	Carcinoma Neoplasm Sarcoma
MESH	"Survival Analysis"	Randomized Controlled Trials as Topic		Neoplasms

Overview of the literature search. Each row for each approach (concept, synonyms, MESH term) was combined with the Boolean operator OR and each column was combined with the Boolean operator AND. Terms which were required to be in the title or abstract were suffixed with .tw., full phrases were specified with quotation marks, terms which could have multiple endings were suffixed with a * and terms which could have English or American spellings were noted with a? So both were included. Any terms where phrases could be given in several ways were specified using the adj3 term which looks for the terms within 3 words of each other. MESH subjects were considered as focussed terms only

OS Overall survival, RCT randomised controlled trial

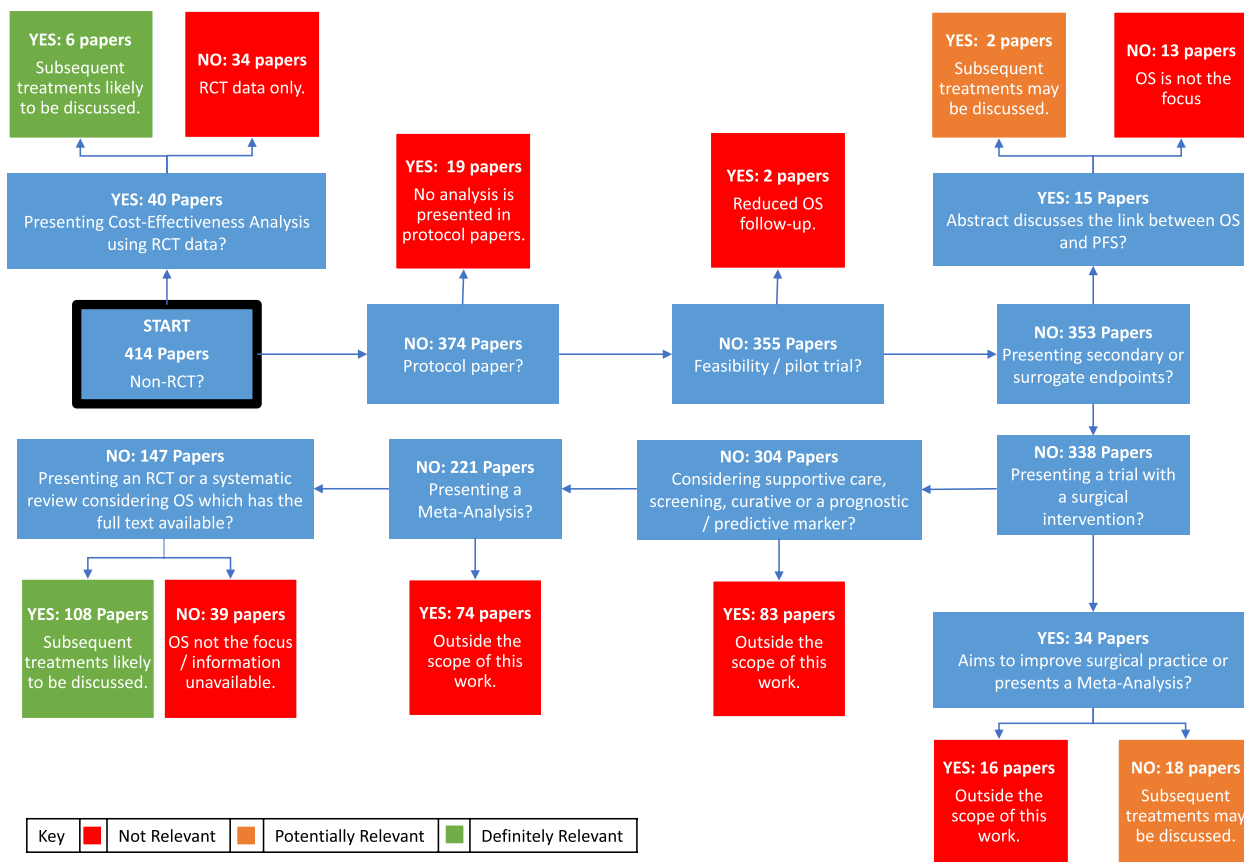


Fig. 1 Screening process for abstracts. Flow diagram showing the process for abstracts to be included in the review

paper which did not mention OS in the abstract was classed as red. In addition, if the abstract or full publication was not available it was also classed as red.

Reading process

All amber and green papers which remained were read in full by K-LR. and categorised as presenting clinical effectiveness results of a clinical trial, presenting a cost-effectiveness analysis, or a methodology or review paper discussing multiple treatment lines. Category-specific information was extracted from each publication (Table 2) and summarised using descriptive statistics. Note that for the review, crossover by design was included in the definition a subsequent treatment line. Any uncertainty in the categorisation or interpretation of the information being extracted was discussed and agreed with DAC. The papers presenting cost-effectiveness were also read by DM given his experience as a Health Economist performing and reviewing cost-effective analysis and any disagreements discussed. The Kappa coefficient of agreement was also calculated.

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards were followed for reporting (Additional file 1, Supplementary Tables 1 and 2) [6].

Results

Overview and sample size

A total of 415 results were identified from the literature search. Following the screening process 281 (68%) were classed as red, 26 (6%) amber and 108 (26%) green (Fig. 1). No changes were made to the screening process following the DAC review. Therefore, a total of 134 papers were included in the review. Following the reading process these were identified to be 98 (73%) clinical effectiveness papers, 27 (20%) methodology or review papers and 9 (7%) cost-effectiveness papers.

Clinical effectiveness papers

Overall survival

The 98 categorised clinical effectiveness papers along with citations are listed in Additional file 1,

Table 2 Information extracted from each article type

Article type	Information extracted
Clinical effectiveness paper	<p><i>Overall survival (OS)</i></p> <ul style="list-style-type: none"> • Endpoint level (primary, secondary, exploratory, unclear) • Definition • The analysis used for the interpretation of the primary endpoint • Whether a Kaplan-Meier Plot for OS was included <ul style="list-style-type: none"> ○ Whether an at-risk table was included with the plot ○ Whether the at-risk table included the number censored/number of events ○ Whether confidence intervals were included on the plot • Whether a logrank test for OS was conducted <ul style="list-style-type: none"> ○ Whether the test was adjusted or unadjusted for other factors. • Whether a Cox proportional hazards model for OS was included <ul style="list-style-type: none"> ○ Whether the results were adjusted or unadjusted for other factors. <p><i>Subsequent treatment^a</i></p> <ul style="list-style-type: none"> • Whether the number or percent of trial participants receiving treatment in follow-up is included in the results or appendix. • Whether a summary of the number of treatment lines and/or types of treatments trial participants received. • Whether subsequent treatments are discussed in the discussion. <ul style="list-style-type: none"> ○ If yes — what are their thoughts? • Whether they did some form of analysis to account for subsequent treatments. <ul style="list-style-type: none"> ○ If yes — what analysis?
Cost-effectiveness paper	<p><i>Cost-effectiveness</i></p> <ul style="list-style-type: none"> • What OS results fed into the cost-effectiveness analysis and how were they determined. <p><i>Subsequent treatment</i></p> <ul style="list-style-type: none"> • Whether subsequent treatment costs were incorporated into the cost-effectiveness analysis • Whether subsequent treatments are discussed in the discussion. <ul style="list-style-type: none"> ○ If yes — what are their thoughts? • Whether they did some form of analysis to account for subsequent treatments. <ul style="list-style-type: none"> ○ If yes — what analysis?
Methodology/review paper	<p><i>General information</i></p> <ul style="list-style-type: none"> • Treatment overview or methodology paper? <p><i>Subsequent treatment</i></p> <ul style="list-style-type: none"> • Whether subsequent treatments are discussed. <ul style="list-style-type: none"> ○ If yes — what are their thoughts?

^a Note that subsequent treatments included cross-over by design

Supplementary Table 3. In total 45 (46%) of the clinical effectiveness papers reported OS to be the primary endpoint of the trial. Forty-eight papers (49%) reported OS as a secondary endpoint. Five papers (5%) stated that OS was an exploratory endpoint only.

OS was defined to be time from randomisation to death from any cause in 87 (89%) papers. Two papers (2%) defined the starting point to be diagnosis, three papers (3%) intervention initiation, one paper enrolment (1%) and one paper pre-inclusion (1%), but it was unclear what these meant in relation to the point of randomisation. Four papers (4%) did not explicitly define OS within the text, two of which were not presenting OS results.

Most papers (86/98, 88%) presented a Kaplan-Meier curve for OS in the main manuscript. Of these, the majority included the numbers at risk (79/86, 92%). However, only a small proportion included the number of events or number censored in their at-risk table (13/79, 16%). None of the papers provided confidence intervals on their Kaplan-Meier curve. Eight papers, with mature OS data, did not present a Kaplan-Meier curve for OS in their main manuscript. Four papers included them in the appendix [7–10] and four omitted a Kaplan-Meier curve entirely [11–14]. A similar proportion of papers presented a logrank test (73/98, 74%) and/or a Cox proportional hazards model (71/98, 72%). However, the level of adjustment (also referred to as stratification) of the tests varied across papers (Fig. 2). Of the 73 papers which included a logrank test, 31 (42%) were not clear as to whether they reported adjusted or unadjusted results (Fig. 2a). In contrast, around half of the papers which presented results of the Cox model reported results which clearly had been adjusted for some or all of the trial's stratification factors (Fig. 2b, 35/71, 49%).

A summary of the primary analysis methods used for OS is shown by endpoint level in Table 3, by year in Additional file 1, Supplementary Table 4 and by cancer type in Additional file 1, Supplementary Table 5. The most popular primary analysis method was the combination of a p -value from a logrank test and the hazard ratio from a Cox proportional hazards model. In total 36 of the 98 papers (37%) papers interpreted OS results using the p -value from a stratified or unstratified logrank test and the hazard ratio from an adjusted or unadjusted Cox model. Of the 98 papers, 15 (15%) used a hazard ratio and

p -value from an adjusted Cox model to interpret OS. Of these 15, 12 (80%), were when OS was a primary or co-primary endpoint. When considered by year (Additional file 1, Supplementary Table 4) of publication the proportion of papers reporting OS as a (co-)primary endpoint fluctuates over the years. Since 2018 OS was more often a secondary rather than (co-)primary endpoint (Additional file 1, Supplementary Table 4). Considering cancer type (Additional file 1, Supplementary Table 5), upper gastrointestinal/hepato-biliary cancers had the highest proportion of papers presenting OS as a primary endpoint.

Subsequent treatment lines

In total, 59 of the 98 (60%) papers mentioned subsequent treatment lines in the main manuscript or supplementary material. Of those that mentioned subsequent treatment lines, the majority (50/59, 85%) provided the number or percentage of participants who received a later treatment line during trial follow-up. Four of the 59 (7%) papers provided a breakdown of the number of later lines a participant received. However, 45 of the 59 (76%) provided a summary of the types of treatments participants received during follow-up. Finally, seven of the 59 (12%) papers performed additional analysis addressing subsequent treatments. The most common reason for the additional analysis was due to participants switching to receive the experimental intervention [10, 19–21]. Miles et al. [19] included time-dependent covariates in the Cox proportional hazards model to adjust for trial participants crossing over from placebo to active treatment. Dimopoulos et al. [22] conducted a post hoc landmark analysis which measured OS from time of progression. Hodi et al. [23] included a sensitivity analysis for OS which censored participants at the point of subsequent treatment. Middleton et al. [20] conducted a post hoc landmark analysis which compared the survival of patients on the active arm who received standard chemotherapy after progression with those who were still alive on the standard chemotherapy arm at this timepoint. Goldhirsch et al. [10] conducted an analysis which cut off follow-up at specific timepoints to report how the effect changed over time after patients switched from observation to active in follow-up. Penichoux et al. [12] included time-dependent covariates in various Cox proportional hazards models to account for progression, toxicity and introduction of third-line

(See figure on next page.)

Fig. 2 a Was a logrank test presented in the main text? Stacked bar chart showing whether a logrank test was presented in the main text of the effectiveness papers (quantified by the percentages at the top of the bar) and whether it was adjusted or unadjusted for additional factors (quantified by the percentages within the bar). **b** Was a Cox-PH model presented in the main text? Stacked bar chart showing whether a Cox-proportional hazards model was presented in the main text of the effectiveness papers (quantified by the percentages at the top of the bar) and whether it was adjusted or unadjusted for additional factors (quantified by the percentages within the bar)

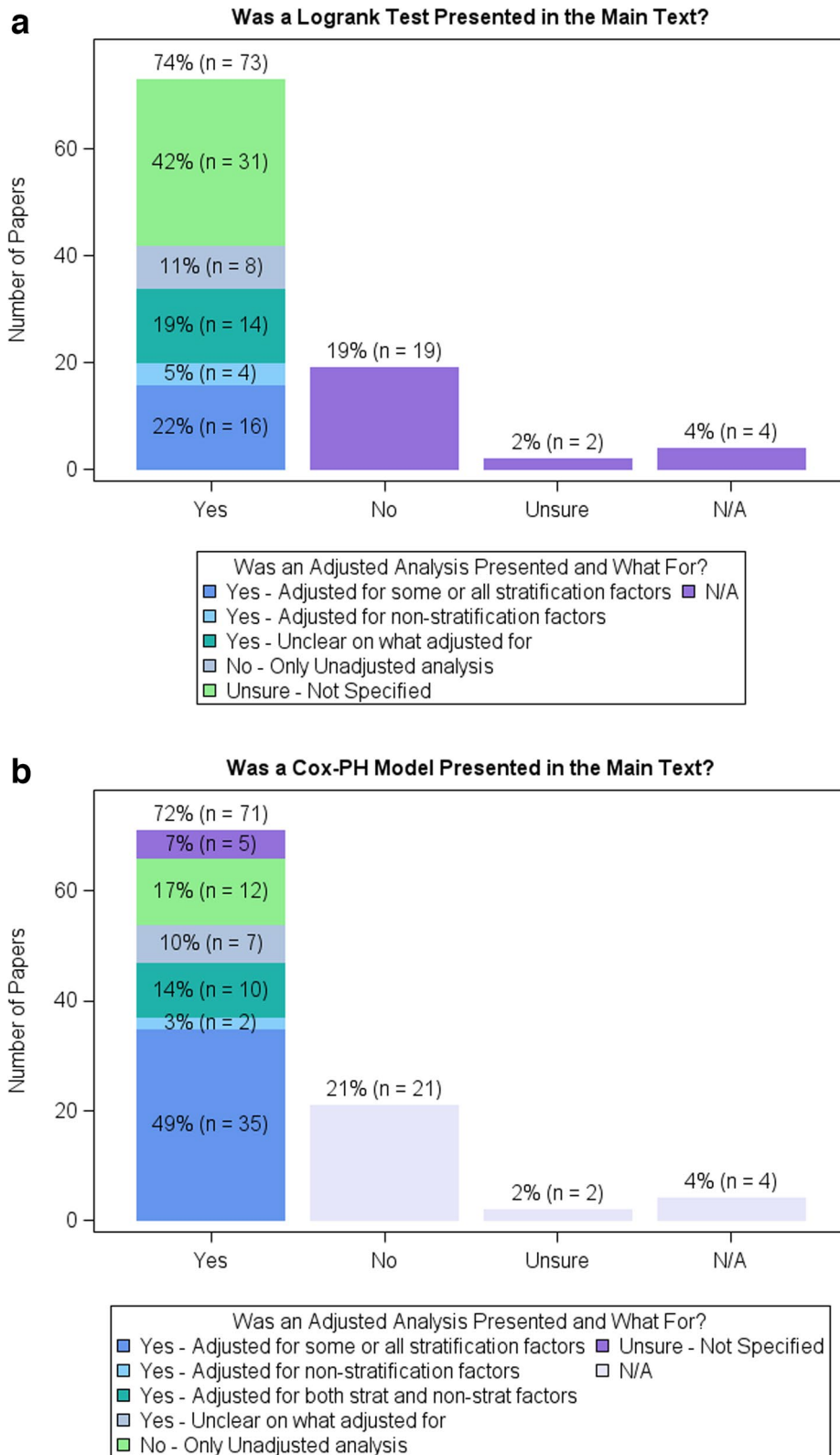


Fig. 2 (See legend on previous page.)

Table 3 Summary of the clinical effectiveness papers by endpoint type

	Primary or co-primary (n=45)	Secondary (n=48)	Exploratory (n=5)	Total (n=98)
Was a Kaplan-Meier curve for OS presented in the main text?				
Yes	40 (88.9%)	41 (85.4%)	5 (100.0%)	86 (87.8%)
No	2 (4.4%)	6 (12.5%)	0 (0.0%)	8 (8.2%)
N/A ^a	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
Were the numbers at risk included in the Kaplan-Meier curve?				
Yes	36 (80.0%)	38 (79.2%)	5 (100.0%)	79 (80.6%)
No	4 (8.9%)	3 (6.3%)	0 (0.0%)	7 (7.1%)
N/A	5 (11.1%)	7 (14.6%)	0 (0.0%)	12 (12.2%)
Were the number of patients censored or number of events included in the at risk table				
Yes	9 (20.0%)	3 (6.3%)	1 (20.0%)	13 (13.3%)
No	27 (60.0%)	35 (72.9%)	4 (80.0%)	66 (67.3%)
N/A	9 (20.0%)	10 (20.8%)	0 (0.0%)	19 (19.4%)
Were confidence intervals included in the Kaplan-Meier curve?				
No	40 (88.9%)	41 (85.4%)	5 (100.0%)	86 (87.8%)
N/A	5 (11.1%)	7 (14.6%)	0 (0.0%)	12 (12.2%)
Primary analysis				
Hazard ratio and 95% CI from an unadjusted Cox model	1 (2.2%)	0 (0.0%)	0 (0.0%)	1 (1.0%)
Hazard ratio and 95% CI from an adjusted Cox model	1 (2.2%)	1 (2.1%)	0 (0.0%)	2 (2.0%)
Hazard ratio and <i>P</i> -value from a logrank test	8 (17.8%)	11 (22.9%)	0 (0.0%)	19 (19.4%)
Hazard ratio and <i>P</i> -value from a stratified logrank test.	3 (6.7%)	3 (6.3%)	0 (0.0%)	6 (6.1%)
Hazard ratio and <i>p</i> -value from an adjusted Cox model	12 (26.7%)	2 (4.2%)	1 (20.0%)	15 (15.3%)
Hazard ratio and <i>p</i> -value from an unadjusted Cox model	2 (4.4%)	3 (6.3%)	0 (0.0%)	5 (5.1%)
Hazard ratio from a Cox model and <i>p</i> -value from a logrank test	3 (6.7%)	7 (14.6%)	0 (0.0%)	10 (10.2%)
Hazard ratio from an adjusted Cox model and <i>p</i> -value from a logrank test.	2 (4.4%)	4 (8.3%)	0 (0.0%)	6 (6.1%)
Hazard ratio from an adjusted Cox model and <i>p</i> -value from a stratified logrank test.	5 (11.1%)	5 (10.4%)	3 (60.0%)	13 (13.3%)
Hazard ratio from an unadjusted Cox model and <i>p</i> -value from a logrank test.	1 (2.2%)	4 (8.3%)	0 (0.0%)	5 (5.1%)
Hazard ratio from an unadjusted Cox model and <i>p</i> -value from a stratified logrank test.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
N/A ^a	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
Other	4 (8.9%)	5 (10.4%)	1 (20.0%)	10 (10.2%)
Were subsequent treatments mentioned in the paper?				
Yes	33 (73.3%)	23 (47.9%)	3 (60.0%)	59 (60.2%)
No	12 (26.7%)	25 (52.1%)	2 (40.0%)	39 (39.8%)
Was the percentage or number of participants who received later lines included in the main text or appendix?				
Yes	31 (68.9%)	18 (37.5%)	1 (20.0%)	50 (51.0%)
No	1 (2.2%)	5 (10.4%)	2 (40.0%)	8 (8.2%)
N/A	13 (28.9%)	25 (52.1%)	2 (40.0%)	40 (40.8%)
Was a breakdown or summary of the number of subsequent treatments (two, three, four) included in the main text or appendix?				
Yes	3 (6.7%)	1 (2.1%)	0 (0.0%)	4 (4.1%)
No	28 (62.2%)	22 (45.8%)	3 (60.0%)	53 (54.1%)
N/A ^b	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)
Was a breakdown or summary of the type of subsequent treatments (treatment 1, treatment 2, treatment 3) included in the main text or appendix?				
Yes	26 (57.8%)	18 (37.5%)	1 (20.0%)	45 (45.9%)
No	5 (11.1%)	5 (10.4%)	2 (40.0%)	12 (12.2%)
N/A ^b	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)

Table 3 (continued)

	Primary or co-primary (n=45)	Secondary (n=48)	Exploratory (n=5)	Total (n=98)
Was additional analysis conducted to account for subsequent treatment lines?				
Yes	2 (4.4%)	3 (6.3%)	2 (40.0%)	7 (7.1%)
No	30 (66.7%)	20 (41.7%)	1 (20.0%)	51 (52.0%)
N/A ^b	14 (31.1%)	25 (52.1%)	2 (40.0%)	41 (41.8%)
Were subsequent treatment lines mentioned in the discussion?				
Yes	22 (48.9%)	15 (31.3%)	2 (40.0%)	39 (39.8%)
No	11 (24.4%)	8 (16.7%)	1 (20.0%)	20 (20.4%)
N/A	12 (26.7%)	25 (52.1%)	2 (40.0%)	39 (39.8%)
How were subsequent treatments mentioned in the discussion?				
The uptake of subsequent treatments is given as a reason for the discrepancy between OS and surrogate endpoints.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
The uptake of subsequent treatments is given as a reason for better OS results or a reduced event rate than expected.	4 (8.9%)	1 (2.1%)	1 (20.0%)	6 (6.1%)
The uptake of subsequent treatments is given as a reason for reduced OS effect/stated as may have affected the results/used to caveat the results.	6 (13.3%)	7 (14.6%)	1 (20.0%)	14 (14.3%)
The lack of uptake of subsequent treatments is given as a reason for no OS effect/may have negatively affected OS.	4 (8.9%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
Subsequent treatments are stated to not have affected the OS results as an OS benefit was observed or OS was similar between those who did and did not receive a subsequent treatment.	2 (4.4%)	1 (2.1%)	0 (0.0%)	3 (3.1%)
Randomisation/balance of subsequent treatment lines is given as a reason as to why subsequent treatment lines will not have affected the OS results.	2 (4.4%)	2 (4.2%)	0 (0.0%)	4 (4.1%)
None or limited options of subsequent treatment lines for patients are given as a reason as to why subsequent treatment lines will not have affected the OS results.	0 (0.0%)	2 (4.2%)	0 (0.0%)	2 (2.0%)
Other	4 (8.9%)	0 (0.0%)	0 (0.0%)	4 (4.1%)
N/A	23 (51.1%)	33 (68.8%)	3 (60.0%)	59 (60.2%)

N/A Not applicable, OS Overall survival

^a Not applicable here refers to papers where OS is a pre-specified endpoint, but the data is not mature at the time of the publication [15–18]

^b Note one paper stated no patients received subsequent treatments so is classed as N/A for these summaries

treatment, they also modelled progression and toxicity events as repeating events using shared gamma frailty models where death was classed as the definitive event. Finally, Gianni et al. [21] conducted an analysis which censored patients at the point of cross-over from observation to active treatment. In the discussion section, 39 of the 59 (66%) papers mentioned subsequent treatment lines (Table 3). The most common discussion point for 14 of the 39 (36%) papers was “The uptake of subsequent treatments is given as a reason for reduced OS effect / stated as may have affected the results / used to caveat the results.” However, four papers also stated that “The lack of uptake of subsequent treatments is given as a reason for no OS effect / may have negatively affected OS.”

Cost-effectiveness papers

In total nine papers whose primary purpose was to report cost-effectiveness analysis were identified during

the literature review (Additional file 1, Supplementary Table 6). There were no disagreements between K-LR and DM following review. Therefore, the Kappa coefficient was one for each summary.

The majority of the papers (6/9, 67%) used a partitioned survival analysis model where the OS estimate was extrapolated from a Kaplan-Meier plot (5/9, 56%). Five of the nine (56%) papers included subsequent costs and four of the nine (44%) papers performed additional analysis to account for subsequent lines of therapy. One paper used the rank preserving structural failure time model (RPS-FTM) [24], another used the two-stage method [25] and the other two adjusted for the cross-over in the trial, but the method used was not specified [26, 27].

Methodology papers

In total, 27 methodology/review papers were identified (Additional file 1, Supplementary Table 7). Of the

papers which mentioned subsequent treatments (14/27, 52%), the majority (9/14, 64%) classed them as a limitation for the interpretation of OS in clinical trials [28–36]. One paper (1/14, 7%) stated that they were not a limitation for interpretation if they reflected the care patients would receive after the trial [37] and four papers (4/14, 29%) did not mention subsequent treatments in this context [38–41].

Discussion

This review identified 134 papers which had reported or discussed OS in the last decade with the aim to determine how it was currently analysed and if subsequent treatment lines were being accounted for and/or discussed.

The main finding in terms of the analysis of OS is that whilst it is generally defined consistently as an endpoint, and is commonly visualised with the Kaplan-Meier curve, the primary analysis is not as consistent as might have been expected. The authors expected that the majority of papers would report the results of a Cox proportional hazards model adjusted for at least the trial's stratification factors, where both the hazard ratio and the p -value from the adjusted model would be used to interpret OS. This would be in line with guidance in ICH E9 which states, "If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis" [2]. However, this was not the case. The majority of authors used a combination of a Cox model and a logrank test to interpret their primary analysis and make their conclusions. Whilst this may be statistically sound, it could be considered inconsistent as the confidence interval around the hazard ratio for the Cox model does not directly relate to the p -value from the logrank test as it would under the laws of duality if the p -value from the Cox model was used. It was reassuring to see that where a Cox model was included in the text, as a secondary analysis for OS, the majority of papers did report that it was adjusted for some or all of the stratification factors. This was less clear when a logrank test was presented. This is unexpected given that ICH E9 was first published in 1998, before the exclusion publication date for this literature review. In order to encourage consistency and transparency in reporting researchers in the future should consider planning and describing their analysis using the estimand framework proposed within ICH E9 (R1) [42]. These results are consistent with the overview of reporting survival analysis endpoints by Batson et al. [43] which was identified as a methodology/review paper in the search (ORN 209). In the paper, Batson and colleagues concluded that the majority of papers identified in their review reported the results of a Cox model and a logrank test with varying levels of adjustment. However, it is important to note that they were focusing on

time-to-event endpoints broadly rather than just OS. It is also interesting that whilst a recent survey found that respondents consider adding a confidence interval to a Kaplan-Meier plot useful [3] no papers included in this review included confidence intervals.

The findings around subsequent treatment lines were less surprising. As whilst the majority which mentioned subsequent treatment lines summarised the types of treatments being received. A finding which complements a recent review by Oliver et al. [44] which was conducted over a shorter time period. Only seven papers, five of which were primary or interim results papers [10, 20–23], implemented some form of analysis to account for anti-cancer treatment received during trial follow-up. The majority of these only considered when participants switched to the alternative trial intervention and only two used methods akin to that described by Latimer et al. [5] (censoring at point of switch [21], sequencing models [12]). However, both were published prior to these recommendations. The work conducted by Penichoux et al. [12] using a shared frailty model to model progression and toxicity events as repeating events seems novel and does not seem to have received a lot of attention with only three papers citing this work. One is reporting the results of a single-arm study so is not relevant to this review [45]. One of the remaining citations by Malka et al. [46] is an editorial piece which references Penichoux's work as an example of how to use random effects to address patients receiving the same intervention at different treatment lines. The final paper to cite Penichoux's work is relevant to this research. Petracci et al. [47] used recurrent event analysis to assess the overall effectiveness of an intervention, in terms of progression-free survival, when it could have been used in first or second line by trial design. In the paper, the authors propose OS as an area for future work. There is a question as to whether this is restricted to the context of trials comparing specific sequences of treatment or whether it could be applied to account for subsequent lines of therapy that are not defined by trial design. As whilst Latimer et al. [5] suggest that sequencing models are a naïve approach given the need for data on each treatment line is hard to come by, there is a question as to whether this is still the case given the utilisation of routine data and patients awareness of how their data can be used through strategies such as DATA-CAN. In terms of the papers which discussed subsequent treatment lines, the most common discussion point was that uptake of subsequent treatments may have reduced the OS effect, affected the results or used to caveat the results. Interestingly the lack of subsequent treatment lines was also given as a reason for no OS effect. Whilst we agree with Oliver et al. that each paper reporting OS should summarise subsequent

anti-cancer treatment [44]. We believe that descriptive statistics alone are not sufficient to support the interpretation of OS when patients receive anti-cancer therapy during trial follow-up. Instead, methodological work is required to develop estimation techniques to disentangle, if present, the effects of the trial experimental treatment and subsequent treatment lines.

Conversely, whilst only nine cost-effectiveness papers were identified they did seem to have a consistent approach to their primary analysis. In addition, where it was clear, the majority of papers did include subsequent treatment costs with a good proportion going further and including additional analysis which adjusted the OS estimate for subsequent treatment lines. Given the contrast to the clinical effectiveness papers, there is a question as to whether adjusting for subsequent lines is being left to be considered by those that evaluate the intervention for HTA review, which often requires cost-effectiveness to be fully analysed, rather than at the publication of the primary clinical effectiveness results.

The majority of the identified methodology/review papers were discussing treatment as a whole and did not consider subsequent lines in the context being considered here. As expected, a number of the methodology papers mentioned subsequent lines as a reason for assessing surrogate endpoints in their field [28, 30, 35] as this is in line with existing guidance [4]. This may also explain the shift in OS from a primary to secondary endpoint as seen in the analysis of the clinical effectiveness papers. In addition, one of the papers published in 2011 stated that subsequent lines are only an issue for the interpretation of clinical trial results if they fall outside of standard of care (SoC) [37]. Whilst this may be true, cancer research has accelerated in the last decade with more varied treatment options being available to patients. For example, in England, there are calls for guidance to be published quickly, and appraisals to be conducted efficiently, so that patients can get access to new treatments faster as part of the NHS, where, once approved, treatments should be available to patients within three months [48]. Currently there are 377 trials in cancer listed as “ongoing” on the ISCRTN database [49]. In addition, between 1 Jan 2021 and 1 Jan 2022, there were 35 technology appraisal guidelines published by NICE to advise on the treatment of cancer [50]. This is more than four times the eight appraisals which were published in between 1 Jan 2010 and 1 Jan 2011 [51]. This rapidly evolving clinical landscape means that the SoC is changing during the trial timeline and will continue to change after the trial closes and the treatment is incorporated into standard practice. Therefore, if OS is analysed without consideration of the subsequent lines, it is possible that the results will soon be invalid

after publication. Methodological research is needed now to better account for subsequent lines so that the true effect of the new experimental intervention in trials is understood and considered when interventions are reviewed for use in SoC. This would work towards ensuring that interventions are properly evaluated so that beneficial interventions are not rejected due to subsequent lines diminishing any effect, and the time and resources it takes for a phase III confirmatory trial to report OS results are not wasted.

This review provides evidence that whilst OS is less frequently the primary endpoint, it is still of interest, and even when treatment switching is not the case, subsequent treatments are being considered as a limitation. This is the same across the different perspectives from the clinical effectiveness papers to the methodological and cost-effectiveness papers. At present, the approach to the assessment of both OS and subsequent treatment lines is not consistent. Whilst there will never be a one size fits all to the analysis of clinical trials, there is a need for an approach to subsequent treatment lines to aid interpretation of clinical trial results as more interventions are developed and recommended for use in standard or care.

There are some limitations to this systematic review. As there was only one researcher conducting the review, the more general search to gauge practice was restricted to only include top UK and international journals. It may be that the exploratory analysis which we are interested in here is reported in other, less-prominent journals. However, this was accounted for in the two approaches where the first search was not restricted by journal. In addition, articles prior to 2010 were excluded from the search. However, this is considered to be less of an issue as the technology support document advocating accounting for subsequent lines in the form of treatment switching was published in 2014 [5]. Having an additional researcher conduct and read all the results may have also helped to overcome some of the inconsistencies in interpretation of the terminology. As whilst all uncertainties were discussed with DAC, in some cases it was unclear whether stratification of a Cox model referred to applying different baseline hazards for different strata or including the variable in the model as an independent variable (the latter more commonly referred to as adjustment). However, the logrank tests were also reported to be adjusted when they can only be stratified. Finally, only papers which were written in English and available through free-text or institutional access were included. Papers written in other languages or behind a paywall may also have considered subsequent treatment lines and how they impact on the analysis of OS.

Conclusion

Over the last decade, OS has changed from being the primary endpoint of cancer clinical trials to more commonly a secondary endpoint. Despite this, subsequent treatment lines in relation to OS are being consistently mentioned in articles, confirming that whilst OS may not be the primary endpoint, the analysis and results are still of interest. The majority of articles concluded that subsequent lines were a limitation of the interpretation of their OS analysis. However, analysis accounting for subsequent lines was generally only conducted in the event of treatment switching. The implemented methods were often naïve, and researchers did not utilise the more sophisticated and less well-known methods of IPCW or the RPSFTM. In fact, RPSFTM was only observed in one paper aiming to determine cost-effectiveness rather than clinical effectiveness. There is a question of whether researchers are aware of these more advanced methods (a possible educational gap) or whether they are just not being published (a reporting gap). This will be explored through the next stage of this research, in which stakeholders (patients and the public, statisticians and other data analysts, healthcare professionals, payers and industry partners) will be invited to complete a questionnaire on their views of OS and its analysis. Within the questionnaire, statisticians and other data analysts (e.g. health economists) will be asked to answer whether they have heard of and whether they have applied the various methods to analyse OS in cancer clinical trials. The overarching aim of the questionnaire will be to agree a consensus across the stakeholder groups in terms of the direction of this research and how statistical methodology should be extended or developed to account for subsequent treatment lines in the analysis of cancer clinical trials.

Abbreviations

BMJ	British Medical Journal
HTA	Health Technology Assessment
ICH	International Council for Harmonisation
IPCW	Inverse Probability Censoring Weights
ISCRTN	International Standard Randomised Controlled Trial Number
JAMA	Journal of the American Medical Association
JCO	Journal of Clinical Oncology
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OS	Overall Survival
PLOS	Public Library of Science
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RPSFTM	Rank preserving structural failure time model
SoC	Standard of care
UK	United Kingdom
USA	United States of America

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13063-023-07730-1>.

Additional file 1: Supplementary Table 1. PRISMA Abstract Checklist. **Supplementary Table 2.** PRISMA Reporting Checklist. **Supplementary Table 3.** Summarised clinical effectiveness papers. **Supplementary Table 4.** Summary of the clinical effectiveness papers by year of publication. **Supplementary Table 5.** Summary of the Effectiveness papers by cancer. **Supplementary Table 6.** Summarised cost-effectiveness papers. **Supplementary Table 7.** Summarised Methodology / Review papers.

Acknowledgements

Not applicable.

Authors' contributions

Under supervision from DM, JKV-R, IRW and DAC, K-LR conceived the idea for the review, planned the project, performed the screening, extracted the data, and drafted the manuscript. All authors read and approved the final manuscript.

Funding

This project is funded by the NIHR Doctoral Fellowship Programme [NIHR301510]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by Core Clinical Trials Unit Infrastructure from Cancer Research UK [C7852/A25447]. IRW was supported by the Medical Research Council [Programme MC_UU_00004/07].

Availability of data and materials

The material from this literature review is available upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Leeds Cancer Research UK Clinical Trials Unit, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK. ²Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK. ³PHASTAR, London, UK. ⁴MRC Clinical Trials Unit at UCL, London, UK.

Received: 29 June 2023 Accepted: 13 October 2023

Published online: 06 November 2023

References

1. Fontrier AM, Visintin E, Kanavos P. Similarities and differences in health technology assessment systems and implications for coverage decisions: evidence from 32 countries. *Pharmacoecoon Open*. 2022;6(315–328):20211129. <https://doi.org/10.1007/s41669-021-00311-5>.
2. EMA. ICH topic E9 statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (1998, Accessed 6 Sept 2022).
3. Morris TP, Jarvis CI, Cragg W, et al. Proposals on Kaplan-Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open*. 2019;9:e030215. <https://doi.org/10.1136/bmjopen-2019-030215>.
4. FDA. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry, <https://www.fda.gov/media/71195/download> (2018, Accessed 6 Sept 2022).

5. Latimer NR, Abrams KR. NICE DSU technical support document 16: adjusting survival time estimates in the presence of treatment switching. 2014.
6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10:1–11.
7. Chopra S, Gupta S, Kannan S, et al. Late toxicity after adjuvant conventional radiation versus image-guided intensity-modulated radiotherapy for cervical cancer (PARCER): a randomized controlled trial. *J Clin Oncol*. 2021;39:3682–92. <https://doi.org/10.1200/JCO.20.02530>.
8. Carrie C, Hasbini A, de Laroche G, et al. Salvage radiotherapy with or without short-term hormone therapy for rising prostate-specific antigen concentration after radical prostatectomy (GETUG-AFU 16): a randomised, multicentre, open-label phase 3 trial. *Lancet Oncol*. 2016;17:747–56. [https://doi.org/10.1016/S1470-2045\(16\)00111-Xr](https://doi.org/10.1016/S1470-2045(16)00111-Xr). Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial.
9. Kunkler IH, Williams LJ, Jack WJ, et al. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. *Lancet Oncol*. 2015;16:266–73. [https://doi.org/10.1016/S1470-2045\(14\)71221-5](https://doi.org/10.1016/S1470-2045(14)71221-5). Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't.
10. Goldhirsch A, Gelber RD, Piccart-Gebhart MJ, et al. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. *Lancet*. 2013;382:1021–8. [https://doi.org/10.1016/S0140-6736\(13\)61094-6](https://doi.org/10.1016/S0140-6736(13)61094-6). Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't.
11. Vora A, Goulden N, Mitchell C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol*. 2014;15:809–18. [https://doi.org/10.1016/S1470-2045\(14\)70243-8](https://doi.org/10.1016/S1470-2045(14)70243-8). Randomized Controlled Trial Research Support, Non-U.S. Gov't.
12. Penichoux J, Michiels S, Bouche O, et al. Taking into account successive treatment lines in the analysis of a colorectal cancer randomised trial. *Eur J Cancer*. 2013;49:1882–8. <https://doi.org/10.1016/j.ejca.2013.02.006>. Research Support, Non-U.S. Gov't.
13. Cuzick J, Sestak I, Baum M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol*. 2010;11:1135–41. [https://doi.org/10.1016/S1470-2045\(10\)70257-6](https://doi.org/10.1016/S1470-2045(10)70257-6). Clinical Trial Randomized Controlled Trial Research Support, Non-U.S. Gov't.
14. Zietman AL, Bae K, Slater JD, et al. Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/American College of Radiology 95–09. *J Clin Oncol*. 2010;28:1106–11. <https://doi.org/10.1200/JCO.2009.25.8475>. Clinical Trial, Phase I Clinical Trial, Phase II Comparative Study Randomized Controlled Trial Research Support, N.I.H., Extramural.
15. Clamp AR, James EC, McNeish IA, et al. Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCG phase 3 randomised controlled trial. *Lancet*. 2019;394:2084–95. [https://doi.org/10.1016/S0140-6736\(19\)32259-7](https://doi.org/10.1016/S0140-6736(19)32259-7). Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't.
16. Minnaar CA, Kotzen JA, Ayeni OA, et al. The effect of modulated electro-hyperthermia on local disease control in HIV-positive and -negative cervical cancer women in South Africa: Early results from a phase III randomised controlled trial. *PLoS ONE*. 2019;14:e0217894. <https://doi.org/10.1371/journal.pone.0217894>. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't.
17. Park K, Tan EH, O'Byrne K, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncol*. 2016;17:577–89. [https://doi.org/10.1016/S1470-2045\(16\)30033-X](https://doi.org/10.1016/S1470-2045(16)30033-X). Clinical Trial, Phase II Randomized Controlled Trial.
18. James ND, Sydes MR, Mason MD, et al. Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. *Lancet Oncol*. 2012;13:549–58. [https://doi.org/10.1016/S1470-2045\(12\)70088-8](https://doi.org/10.1016/S1470-2045(12)70088-8). Clinical Trial, Phase II Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't.
19. Miles D, Im YH, Fung A, et al. Effect of docetaxel duration on clinical outcomes: exploratory analysis of CLEOPATRA, a phase III randomized controlled trial. *Ann Oncol*. 2017;28:2761–7. <https://doi.org/10.1093/annonc/mdx406>. Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't.
20. Middleton G, Silcocks P, Cox T, et al. Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): an open-label, randomised, phase 3 trial. *Lancet Oncol*. 2014;15:829–40. [https://doi.org/10.1016/S1470-2045\(14\)70236-0](https://doi.org/10.1016/S1470-2045(14)70236-0). Clinical Trial, Phase III Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't.
21. Gianni L, Dafni U, Gelber RD, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol*. 2011;12:236–44. [https://doi.org/10.1016/S1470-2045\(11\)70033-X](https://doi.org/10.1016/S1470-2045(11)70033-X). Clinical Trial, Phase III Randomized Controlled Trial Research Support, Non-U.S. Gov't.
22. Dimopoulos MA, Goldschmidt H, Niesvizky R, et al. Carfilzomib or bortezomib in relapsed or refractory multiple myeloma (ENDEAVOR): an interim overall survival analysis of an open-label, randomised, phase 3 trial. *Lancet Oncol*. 2017;18:1327–37. [https://doi.org/10.1016/S1470-2045\(17\)30578-8](https://doi.org/10.1016/S1470-2045(17)30578-8). Clinical Trial, Phase III Comparative Study Multicenter Study Randomized Controlled Trial Research Support, Non-U.S. Gov't.
23. Hodi FS, Chesney J, Pavlick AC, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol*. 2016;17:1558–68. [https://doi.org/10.1016/S1470-2045\(16\)30366-7](https://doi.org/10.1016/S1470-2045(16)30366-7). Clinical Trial, Phase II Multicenter Study Randomized Controlled Trial.
24. Srivastava T, Prabhu VS, Li H, et al. Cost-effectiveness of pembrolizumab as second-line therapy for the treatment of locally advanced or metastatic urothelial carcinoma in Sweden. *Eur Urol Oncol*. 2020;3:663–70. <https://doi.org/10.1016/j.euo.2018.09.012>. Research Support, Non-U.S. Gov't.
25. Bhadhuri A, Insinga R, Guggisberg P, et al. Cost effectiveness of pembrolizumab vs chemotherapy as first-line treatment for metastatic NSCLC that expresses high levels of PD-L1 in Switzerland. *Swiss Med Wkly*. 2019;149:w20170. <https://doi.org/10.4414/sm.w.2019.20170>. Comparative Study.
26. Miquel-Cases A, Retel VP, Lederer B, et al. Exploratory cost-effectiveness analysis of response-guided neoadjuvant chemotherapy for hormone positive breast cancer patients. *PLoS ONE*. 2016;11:e0154386. <https://doi.org/10.1371/journal.pone.0154386>. Randomized Controlled Trial Research Support, Non-U.S. Gov't.
27. Hoyle M, Peters J, Crathorne L, et al. Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. *Value Health*. 2013;16:288–96. <https://doi.org/10.1016/j.jval.2012.11.001>. Comparative Study Research Support, Non-U.S. Gov't.
28. Nomura S, Hirakawa A, Hamada C. Sample size determination for the current strategy in oncology Phase 3 trials that tests progression-free survival and overall survival in a two-stage design framework. *J Biopharm Stat*. 2018;28:589–611. <https://doi.org/10.1080/10543406.2017.1372775>. Research Support, Non-U.S. Gov't.
29. Osanto S, van der Hulle T. Cabozantinib in the treatment of advanced renal cell carcinoma in adults following prior vascular endothelial growth factor targeted therapy: clinical trial evidence and experience. *Ther Adv Urol*. 2018;10:109–23. <https://doi.org/10.1177/1756287217748867>. Review.
30. Zer A, Prince RM, Amir E, et al. Evolution of randomized trials in advanced/metastatic soft tissue sarcoma: end point selection, surrogacy, and quality of reporting. *J Clin Oncol*. 2016;34:1469–75. <https://doi.org/10.1200/JCO.2015.64.3437>. Review Systematic Review.
31. Hussain M, Tangen C, Higano C, et al. Evaluating intermittent androgen-deprivation therapy phase III clinical trials: the devil is in the details. *J Clin Oncol*. 2016;34:280–5. <https://doi.org/10.1200/JCO.2015.62.8065>. Clinical Trial, Phase III Randomized Controlled Trial.

32. Garcia-Albeniz X, Maurel J, Hernan MA. Why post-progression survival and post-relapse survival are not appropriate measures of efficacy in cancer randomized clinical trials. *Int J Cancer*. 2015;136:2444–7. <https://doi.org/10.1002/ijc.29278>. Research Support, N.I.H., Extramural.
33. Bai X, Liu J, Li L, et al. Adaptive truncated weighting for improving marginal structural model estimation of treatment effects informally censored by subsequent therapy. *Pharm*. 2015;14:448–54. <https://doi.org/10.1002/pst.1719>.
34. Matulonis UA, Oza AM, Ho TW, et al. Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer*. 2015;121:1737–46. <https://doi.org/10.1002/cncr.29082>. Research Support, Non-U.S. Gov't Review.
35. Bria E, Massari F, Maines F, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. *Crit Rev Oncol Hematol*. 2015;93:50–9. <https://doi.org/10.1016/j.critrevonc.2014.08.001>. Research Support, Non-U.S. Gov't Review.
36. Fojo T, Amiri-Kordestani L, Bates SE. Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. *J Natl Cancer Inst*. 2011;103:1738–40. <https://doi.org/10.1093/jnci/djr386>. Research Support, N.I.H., Intramural Review.
37. Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol*. 2011;29:2439–42. <https://doi.org/10.1200/JCO.2011.34.6056>. Review.
38. Yeh C, Bates SE. Two decades of research toward the treatment of locally advanced and metastatic pancreatic cancer: Remarkable effort and limited gain. *Semin Oncol*. 2021;48:34–46. <https://doi.org/10.1053/j.seminoncol.2021.01.001>.
39. Cranmer H, Shields GE, Bullement A. A comparison of partitioned survival analysis and state transition multi-state modelling approaches using a case study in oncology. *J Med Econ*. 2020;23:1176–85. <https://doi.org/10.1080/13696998.2020.1796360>.
40. Chen H, Horita N, Ito K, et al. Hazard ratio of progression-free survival is an excellent predictor of overall survival in phase III randomized controlled trials evaluating the first-line chemotherapy for extensive-disease small-cell lung cancer. *Transl*. 2020;9:1333–42. <https://doi.org/10.21037/tlcr-20-377>.
41. De Maeseneer DJ, Delafontaine B, Rottey S. Checkpoint inhibition: new treatment options in urologic cancer. *Acta Clin Belg*. 2017;72:24–8. <https://doi.org/10.1080/17843286.2016.1260890>. Review.
42. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (2017, Accessed 1 Mar 2023).
43. Batson S, Greenall G, Hudson P. Review of the reporting of survival analyses within randomised controlled trials and the implications for meta-analysis. *PLoS ONE*. 2016;11:e0154870. <https://doi.org/10.1371/journal.pone.0154870>. Review Research Support, Non-U.S. Gov't.
44. Olivier T, Haslam A, Prasad V. Post-progression treatment in cancer randomized trials: a cross-sectional study of trials leading to FDA approval and published trials between 2018 and 2020. *BMC Cancer*. 2023;23:448. <https://doi.org/10.1186/s12885-023-10917-z>.
45. Texier M, Rotolo F, Ducreux M, et al. Evaluation of treatment effect with paired failure times in a single-arm phase II Trial in oncology. *Comput Math Methods Med*. 2018;2018:1672176. <https://doi.org/10.1155/2018/1672176>.
46. Malka D, Rotolo F, Boige V. First-line treatment in metastatic colorectal cancer: Important or crucial? *Eur J Cancer*. 2017;84:363–6.
47. Petracchi E, Scarpi E, Passardi A, et al. Effectiveness of bevacizumab in first- and second-line treatment for metastatic colorectal cancer: ITACa randomized trial. *Ther Adv Med Oncol*. 2020;12:1758835920937427. 20200723. <https://doi.org/10.1177/1758835920937427>.
48. NICE. Cost comparison. Addendum to the Guide to the methods of technology appraisal <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/methods-guide-addendum-cost-comparison.pdf> (Accessed 29 Jul 2020).
49. ISCRTN. ISCRTN Registry, <https://www.isrctn.com/search?q=&filters=trial>Status%3Aongoing%2CconditionCategory%3ACancer> (Accessed 6 Sept 2022).
50. NICE. Guidance, NICE advice and quality standards: 2021 - 2022, <https://www.nice.org.uk/guidance/published?from=2021-01-01&to=2022-01-01&ndt=Guidance&ngt=Technology%20appraisal%20guidance&ps=9999> (Accessed 6 Sept 2022).
51. NICE. Guidance, NICE advice and quality standards: 2010 - 2011, <https://www.nice.org.uk/guidance/published?ps=2500&from=2010-01-01&to=2011-01-01&ndt=Guidance&ngt=Technology+appraisal+guidance> (Accessed 6th Sept 2022).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

