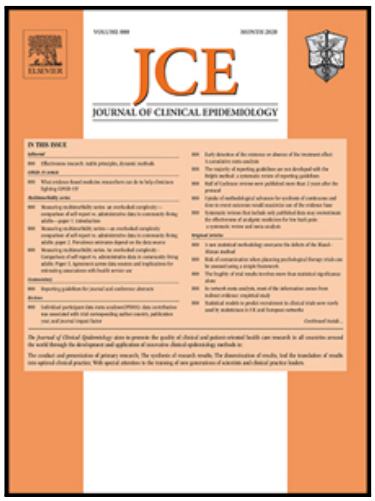

Prespecification of subgroup analyses and examination of treatment-subgroup interactions in cancer individual participant data meta-analyses are suboptimal

Ya Gao , Ming Liu , Shuzhen Shi , Mingming Niu , Jiang Li , Junhua Zhang , Fujian Song , Jinhui Tian

PII: S0895-4356(21)00197-9
DOI: https://doi.org/10.1016/j.jclinepi.2021.06.019
Reference: JCE 10544

To appear in: *Journal of Clinical Epidemiology*

Accepted date: 22 June 2021

Please cite this article as: Ya Gao , Ming Liu , Shuzhen Shi , Mingming Niu , Jiang Li , Junhua Zhang , Fujian Song , Jinhui Tian , Prespecification of subgroup analyses and examination of treatment-subgroup interactions in cancer individual participant data meta-analyses are suboptimal, *Journal of Clinical Epidemiology* (2021), doi: https://doi.org/10.1016/j.jclinepi.2021.06.019

Prespecification of subgroup analyses and examination of treatment-subgroup interactions in cancer individual participant data meta-analyses are suboptimal

Ya Gao[1†], Ming Liu[1†], Shuzhen Shi[1], Mingming Niu[2], Jiang Li[3], Junhua Zhang[4], Fujian Song[5]*, Jinhui Tian[1,6]*

[1]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

[2]Evidence-Based Nursing Center, School of Nursing, Lanzhou University, Lanzhou, China

[3]National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking UnionMedical College, Beijing, China

[4]Evidence-Based Medicine Center, Tianjin University of Traditional Chinese Medicine, Tianjin, China

[5]Public Health and Health Services Research, Norwich Medical School, University of East Anglia, Norwich, UK

[6]Key Laboratory of Evidence-Based Medicine and Knowledge Translation of Gansu Province, Lanzhou, China

[†]These authors contributed equally to this work

*** Correspondence to:** Jinhui Tian. Evidence-Based Medicine Centre, School of Basic Medical Sciences, Lanzhou University, No.199, Donggang West Road, Lanzhou City, 730000, Gansu Province, China. (Email: tianjh@lzu.edu.cn)

Fujian Song. Public Health and Health Services Research, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK. (Email: fujian.song@uea.ac.uk)

Abstract

Objectives: This study aimed to explore the prespecification and conduct of subgroup analyses in cancer individual participant data meta-analyses (IPDMAs).

Study Design and Setting: We searched PubMed, Embase, Cochrane Library, and Web of Science to identify IPDMAs of randomized controlled trials evaluating intervention effects for cancer. We evaluated how often cancer IPDMAs prespecify subgroup analyses and statistical approaches for examining treatment-subgroup interactions and handling continuous subgroup variables.

Results: We included 89 IPDMAs, of which 41 (46.1%) reported a statistically significant treatment-subgroup interaction ($P<0.05$) in at least one subgroup analysis. 47 (52.8%) IPDMAs prespecified methods for conducting subgroup analyses and the remaining 42 (47.2%) did not prespecify subgroup analyses. Of the 47 IPDMAs prespecified subgroup analyses, 19 performed the planned subgroup analyses, 21 added subgroup analyses, 7 reduced subgroup analyses. Eighty IPDMAs examined treatment-subgroup interactions, but 72 IPDMAs did not provide enough information to determine whether an appropriate approach that avoided aggregation bias was used. 85 IPDMAs that used continuous variables in subgroup analyses categorized continuous variables and only 1 IPDMA examined non-linear relationships.

Conclusions: Many cancer IPDMAs did not prespecify subgroup analyses, nor did they fully perform planned subgroup analyses. Lack of details for the test of

treatment-subgroup interactions and examination of non-linear interactions was suboptimal.

Keywords: Individual participant data meta-analysis; Neoplasm; Subgroup analysis; Treatment-subgroup interaction; Prespecification; Methodology

Running title: Characteristics of subgroup analysis in cancer individual participant data meta-analyses

What is new?

Key findings

- Prespecification of subgroup analyses in cancer individual participant data meta-analyses (IPDMAs) were inadequate.
- Lack of details for the test of treatment-subgroup interactions and the use of deft approach was suboptimal.
- Cancer IPDMAs more frequently categorized continuous variables but rarely examined the non-linear interactions.

What this adds to what was known?

- This study explored the prespecification of subgroup analyses in cancer IPDMAs, whether the planned subgroup analyses were conducted, methods used for the test of treatment-subgroup interactions, and statistical approaches for examining continuous subgroup variables.

What is the implication and what should change now?

- The presentation and examination of treatment-covariate interactions and the proper handling of continuous variables need substantial improvement.
- Existing methodological and reporting guidance should be strictly followed to select an appropriate method for the conduct and reporting of subgroup analysis in IPDMAs.
- Journal editors and peer reviewers should ensure that the results of subgroup analyses are fully reported and help finds inappropriate subgroup analyses.

1. Introduction

Individual participant data meta-analysis (IPDMA) is considered the "gold standard" of meta-analysis since it has many advantages over traditional aggregated data meta-analysis [1-6]. IPDMAs allow the possibility of standardizing subgroup definitions and outcomes across studies, identifying subgroups that are more likely to

benefit from treatment, increasing the flexibility to search for subgroups based on a combination of patient and disease characteristics, allowing thorough investigations of the main effect of each covariate, producing more accurate and credible subgroup findings, avoiding incorrect results caused by aggregation bias (also known as ecological bias), and permitting a more comprehensive assessment of whether differences in effect estimates between subgroups are spurious [7-11]. Due to these advantages, numerous IPDMAs have been conducted to examine treatment-covariate interactions at the participant level for a clinical topic to guide the practice of precision medicine [9, 12]. Subgroup analysis is a statistical analysis to explore whether intervention effects vary according to patient groups, way of giving an intervention, or approach to measuring an outcome (e.g. males versus females, old versus young) [13-15]. Subgroup analysis is frequently used to examine differences in treatment effects between patients with different characteristics, which facilitates the development of treatment decisions for individual patients based on different expected treatment effects [16-19]. The results of subgroup analyses may have a significant impact on clinical and public health decision-making [20]. However, inappropriate subgroup analyses can mislead clinical practice.

Previous studies have evaluated subgroup analyses in individual clinical trials and revealed that trials rarely prespecify subgroup analyses, discrepancies exist between subgroup analyses planned in protocol and presented in trial reports, and justification and reporting of subgroup analyses are lacking [21-23]. Existing studies have also investigated whether there are differences in analytical methods regarding

subgroup analyses between IPDMAs and conventional meta-analyses [24], how frequency IPDMAs had significant treatment-subgroup interaction, whether authors consider the most obvious subgroup differences to provide support for stratified treatment decisions [19], and summarized methods used to assess treatment-covariate interactions and developed guidance on method selection in different circumstances [25]. However, how often IPDMAs prespecify subgroup analyses, conduct planned subgroup analyses, and use daft (across-trial interaction alone), deluded (within-trial and across-trial interactions combined), or deft (within-trial interaction alone) approach to assess the treatment-subgroup interactions remain unclear.

Cancer is the second leading cause of death worldwide, has substantial variability in populations and treatments [26, 27]. To provide individualized treatment, several IPDMAs have been implemented in the cancer field to determine subgroup treatment effects based on different patient characteristics. Thus, it is important to ensure that subgroup analyses in cancer IPDMAs provide reliable evidence for patients and healthcare professionals. This study aimed to explore the prespecification of subgroup analyses in cancer IPDMAs, whether the planned subgroup analyses were conducted, conducting of treatment-subgroup interactions, and statistical approaches for examining continuous subgroup variables.

2. Methods

2.1. Eligibility criteria

We included IPDMAs of randomized controlled trials (RCTs) that evaluated the treatment effects of health care interventions for cancer patients. IPDMAs should

conduct at least one subgroup analysis and report results of subgroup analyses. Interventions can be drugs, radiotherapies, surgeries, health technologies, or a combination thereof. We limited the publication language to English. No restriction was put on the publication year.

Studies including the following were excluded: (1) Systematic reviews (SRs) or meta-analyses did not incorporate individual participant data; (2) IPDMAs that included both RCTs and nonrandomized trials or only included nonrandomized trials; (3) IPDMAs did not focus on health care interventions such as IPDMAs of etiology or diagnostic test accuracy; (4) IPDMAs only clarified the methods for conducting subgroup analyses but did not report results of subgroup analyses; (5) IPD network meta-analyses, methodological studies, review protocols, abstracts, conference proceedings, letters, and editorials.

2.2. Literature search

We performed a comprehensive literature search in PubMed, Embase.com, Cochrane Library, and Web of Science up to October 12, 2019. Keywords we used included “IPD”, “individual patient”, “individual participant”, “individual data”, “patient level”, “individual person”, “meta-analysis”, “metaanalysis”, and “systematic review”. The detailed search strategy of PubMed is presented in Appendix Word 1. We also reviewed the reference lists of included IPDMAs and relevant reviews to identify additional potentially relevant studies.

2.3. Study selection process

We used EndNote X8 (Thomson Reuters (Scientific) LLC Philadelphia, PA, US) to manage the retrieved records and remove duplicates. Two reviewers (Y.G. and J.H.T.) independently screened titles and abstracts of identified records to determine if they met the inclusion criteria. Full reports of all potentially relevant papers were further evaluated by the same two reviewers to reach final eligibility. If an IPDMA has been updated, we would only incorporate the latest version. Conflicts were resolved by consensus or referral to a third reviewer (F.J.S).

2.4. Data abstraction and data items

A predefined data abstraction form was developed using Microsoft Excel 2016 (Microsoft Corp, Redmond, WA, www.microsoft.com) and was revised after piloting on a random of five included IPDMAs. The abstracted data included first author, publication year, country of the corresponding author, journal name, topic focused, whether IPDMAs had a priori protocol, funding sources (industry, non-industry, no-funding, or not reported), number of included RCTs, number of included patients, number of RCTs used in data analysis, interventions, primary outcomes, basic characteristics of patients reported in IPDMAs, whether basic characteristics included factors considered in subgroup analyses, whether prespecified methods of subgroup analyses, subgroup analysis factors described in the protocols or Methods section of IPDMAs, subgroup factors reported in the Results section of IPDMAs, whether examined the interaction between treatment effect and subgroup, whether used within-trial or across-trial information (daft, deluded, deft) for assessing

treatment-subgroup interactions, whether subgroup factors with statistically significant (P<0.05) treatment-subgroup interactions, whether used continuous variables in the subgroup analysis, how continuous variables were examined in their interaction with treatment, and whether examined non-linear relationships for continuous variables. We only focused on subgroups of relative effects. One author (Y.G., M.L, S.Z.S, or M.M.N) conducted data abstraction and another author (J.L or J.H.Z) checked the extracted data. Disagreements were settled by consensus or by a discussion with a third reviewer (F.J.S or J.H.T).

2.5. Statistical analysis

We conducted a descriptive analysis using frequency and percentage to summarize the characteristics of the included IPDMAs and statistical approaches for examining subgroups in IPDMAs. To determine whether there are differences in factors used for subgroup analyses between IPDMAs with different cancers and treatments, we categorized subgroup factors for each IPMDA according to cancers and interventions investigated. We compared the number of subgroup factors to assess the differences in subgroup analyses between the protocols or Methods section and the Results section of IPDMAs and presented the results with frequency and percentage.

3. Results

3.1. Screening results

The systematic literature search yielded a total of 20652 records. After de-duplication, 6123 records were screened by titles and abstracts, and 5951 were

considered irrelevant. Further retrieving and assessing full-texts of the remaining 172 records, we excluded 83 IPDMAs, of which 58 did not report results of subgroup analyses. Finally, 89 IPDMAs met the eligibility criteria and were included for analyses. The flowchart of the study screening process is presented in Figure 1. The full lists of included IPDMAs can be found in Appendix Word 2.

3.2. General characteristics

The included IPDMAs were published between 1995 and 2018, while the most productive year was 2017, with eight IPDMAs (Figure 2). The UK (36, 40.4%) published the largest number of IPDMAs, followed by France (13, 14.6%) and Italy (9, 10.1%), Figure 3. The included IPDMAs investigated many types of cancer, with lung cancer, leukemia, and breast cancer being frequently studied (Figure 4). Table 1 presents the general characteristics of IPDMAs. Of the 89 IPDMAs, 37 (41.6%) had a protocol, 85.4% IPDMAs included less than 21 RCTs, and the number of RCTs used for data analysis per IPDMA no more than 40. Nine (10.1%) IPDMAs used the one-stage approach for data synthesis, 3 (3.4%) used the two-stage approach, and 77 (86.5%) did not report the one-stage or two-stage approach. Commonly studied interventions included drugs (75, 84.3%), radiotherapy (28, 31.5%), and surgery (21, 23.6%). More than half (49, 55.1%) IPDMAs had one primary outcome, 24 (27.0%) had two primary outcomes, 5 (5.6%) had three primary outcomes, and 11 (12.4%) had four or five primary outcomes. The most commonly used primary outcome was overall survival (78, 87.6%), followed by disease-free survival (20, 22.5%), and progression-free survival (17, 19.1%), Appendix Table 1.

3.3. Testing for treatment-covariate interactions and handling of continuous variables

Of the 89 IPDMAs, 80 (89.9%) examined the interaction between treatment effect and subgroup, of which 41 (46.1%) IPDMAs reported a statistically significant (P<0.05) treatment-subgroup interaction in at least one subgroup analysis, 39 (43.8%) IPDMAs did not find a statistically significant (P<0.05) treatment-subgroup interaction in subgroup analyses, and the remaining 9 (10.1%) IPDMAs reported that no different effect in subgroup analyses but did not mention treatment-subgroup interactions. Of the 80 IPDMAs that examined treatment-subgroup interactions, 4 IPDMAs used the deft approach to assess the interactions, 4 IPDMAs used the deluded approach, but 72 IPDMAs failed to provide enough information to determine whether a daft, deluded, or deft analysis was carried out. Only 1 (1.1%) IPDMA adjusted prespecified prognostic factors for subgroup analyses. We identified 103 subgroup factors in the 89 IPDMAs, of which 28 factors were found with statistically significant (P<0.05) treatment-subgroup interactions in some IPDMAs and 75 factors did not have statistically significant (P<0.05) treatment-subgroup interactions or did not have different effects in subgroup analyses in all included IPDMAs. The top five most frequently used factors in subgroup analyses were age (80, 89.9%), sex (64, 71.9%), stage (45, 50.6%), performance status (35, 39.3%), and histology (25, 28.1%). Statistically significant (P<0.05) treatment-subgroup interactions were often observed for the factors of age (17, 19.1%), stage (12, 13.5%), performance status (9, 10.1%), sex (6, 6.7%), and nodal status (4, 4.5%). Details regarding subgroup factors can be found in Appendix Table 2. We found that 85 (95.5%) IPDMAs used

continuous variables in the subgroup analysis and the number of continuous variables per IPDMA ranged from 1 to 5. There were 57 (67.1%), 20 (23.5%), 4 (4.7%), 1 (1.2%), 2 (3.5%), and 1 (1.2%) IPDMA used 1, 2, 3, 4, 5, and 7 continuous variables in subgroup analyses, respectively. Commonly investigated continuous variables included age (80, 89.9%), white blood cell count (7, 7.9%), tumour size (7, 7.9%), beta2 microglobulin (3, 3.4%), and body mass index (3, 3.4%). All the 85 IPDMAs categorized continuous variables, but only one IPDMA reported how cut-points were selected for continuous variables. Considering age, 24 IPDMAs dichotomized age, 23 IPDMAs categorized age into three groups, 24 IPDMAs categorized age into four groups, 8 IPDMAs categorized age into five or six groups, and 1 IPDMA did not report how to categorized age. Of the 85 IPDMAs, only 1 (1.2%) examined non-linear relationships for continuous variables (Table 2).

3.4. Subgroup factors of different cancers and interventions

Appendix Table 3 presents the subgroup factors of different cancers and interventions. Cancers with more than 15 subgroup factors included breast cancer, leukemia, and lung cancer. Hematologic malignancies, pancreatic cancer, and prostate cancer only had one or two subgroup factors. We did not found statistically significant treatment-subgroup interactions in colon cancer, esophageal cancer, gastric cancer, malignant gliomas, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, and prostate cancer. There was some variation in the subgroup factors investigated between different cancers. For a certain type of cancer, the subgroup factors also differed by interventions.

3.5. The relationship between the basic characteristics and subgroup factors of IPDMAs

Of the 89 IPDMAs, 80 (89.9%) reported the basic characteristics of patients included in IPDMAs, and 114 characteristics were identified in the 80 IPDMAs. The top five commonly reported basic characteristics were age (75, 84.3%), sex (60, 67.4%), stage (48, 53.9%), performance status (40, 44.9%), and tumor site (35, 39.3%). Of the 80 IPDMAs that provided basic characteristics, 63 (78.8%) IPDMAs described all factors used in subgroup analyses in their basic characteristics. There were 1, 2, 3, and 4 subgroup factors failed to report in the basic characteristics section in 8 (10.0%), 4 (5.0%), 4 (5.0%), and 1 (1.2%) IPDMA, respectively (Table 3).

3.6. Subgroup analyses planned or reported

Of the 89 IPDMAs that provided subgroup analysis results, 47 (52.8%) clarified the planned subgroup factors in the Methods section or the protocols and the remaining 42 (47.2%) did not prespecify subgroup analyses. The number of factors used in planned subgroup analyses ranged from 1 to 7. There were 11 (12.4%), 3 (3.4%), 3 (3.4%), 8 (9.0%), 11 (12.4%), 5 (5.6%), and 6 (6.7%) IPDMAs plan to perform subgroup analyses considering 1, 2, 3, 4, 5, 6, and 7 factors, respectively. 89 IPDMAs reported 1 to 16 subgroup factors in the Results section, with most IPDMAs (71, 79.8%) explored 3 to 7 subgroup factors (Table 4). Among the 47 IPDMAs that prespecified subgroup analyses in their protocols or Methods section, only 19 (40.4%) IPDMAs fully performed the planned subgroup analyses in their Results section. Compared to the factors reported in the protocols or Methods section, 21 (44.7%)

IPDMAs conducted more subgroup analyses in the Results section, of which 12 (25.5%) IPDMAs added 1 or 2 factors, 6 (12.8%) IPDMAs added 3 or 4 factors, and 3 (6.4%) IPDMAs added 5 or 6 factors. Seven (14.9%) IPDMAs reduced subgroup factors in the Results section, and the number of reduced factors ranged from 1 to 4 (Table 5).

4. Discussion

4.1. Main findings

We conducted a comprehensive literature search and included 89 cancer IPDMAs that reported results of subgroup analyses published between 1995 and 2018. We investigated the prespecification and reporting of subgroup analyses, conducting of treatment-subgroup interactions, and statistical approaches for examining continuous subgroup variables. Overall, 47.2% of the IPDMAs did not prespecify factors used for subgroup analyses. For IPDMAs that prespecify methods of subgroup analyses, about 60.0% IPDMAs did not fully follow their planned subgroup analyses. There were considerable inconsistencies between the subgroup factors planned in protocols or Methods section and those presented in Results section of cancer IPDMAs. Eighty IPDMAs examined treatment-subgroup interactions, but only 4 IPDMAs used the deft approach to test the interactions and most IPDMAs did not provide enough information to determine whether an appropriate approach that avoided aggregation bias was used. Approximately 95.5% IPDMAs used continuous variables in the subgroup analysis, all of these IPDMAs categorized continuous

variables, and only 1 IPDMA examined non-linear relationships for continuous variables.

4.2. Interpretation

Among the 89 IPDMAs, we identified 103 subgroup factors, but only 28 factors were found with statistically significant treatment-subgroup interactions. IPDMAs with different relative treatment effects between subgroups can provide some opportunities for stratifying medicine in clinical topics [19]. The focus when examining subgroup effect in IPDMAs is on the statistical significance of the coefficient of the interaction term, typically in the context of relative effect measures (e.g., odds ratio, relative risk, or hazard ratio). Translation to decision-making for individual patients is then not straightforward because these decisions are primarily driven by absolute differences in absolute risks for the outcome between competing therapeutic options. This means that statistically significant results do not always lead to different decisions (e.g. more favorable in women, but still beneficial for men) or statistically non-significant results could still imply differences in benefit. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement is a useful tool that can guide us when translating findings of subgroup analyses into clinical practice [15, 28]. Furthermore, when we apply the results of subgroup analyses on clinical practice, we should consider the impact of possible Type I and II errors [29]. If methods used for test treatment-subgroup interactions are not appropriate, important interactions may be missed [9]. Previous studies have developed guidelines for interpreting subgroup analyses that can help us assess the

credibility of claims of different responses to treatment in clearly definable subgroups of patients [13, 30, 31].

Research protocols for SRs help to increase the transparency of review objectives and methods and can reduce the risk of researcher bias and outcome reporting bias [32-34]. A priori protocol can improve the methodological quality of SRs, whereas the lack of protocols may lead to the post-modification of methodology [35, 36]. However, only 41.6% of IPDMAs had a priori protocol. Furthermore, 47.2% of IPDMAs did not clarify the factors used for subgroup analyses in the protocols or Methods section. These findings indicated that almost half of the subgroup analyses reported in IPDMAs were post hoc, without specification in a priori protocol or the Methods section. However, the credibility of any significant subgroup difference arising from the post hoc rather than a priori hypothesis is questionable [30]. Considering IPDMAs that prespecified subgroup factors, approximately 60.0% did not fully perform the planned subgroup analyses, nor did they describe the justifications or rationales for these inconsistencies. Overall, cancer IPDMAs had serious deficiencies in the prespecification of subgroup analyses.

To avoid false-positive subgroup effects, the number of subgroup analyses in a study cannot be too large. If a study tests a large number of hypotheses, the strength of reasoning to confirm any hypothesis may be decreased, even if the researchers have prespecified their hypotheses [30]. However, our study revealed that 34.8% of the 89 IPDMAs performed more than five subgroup analyses, and even one IPDMA performed sixteen subgroup analyses. Simultaneously, we should also not

overemphasize false-positive subgroup effects to reduce the number of subgroup analyses, leading to a waste of research data and resources [18, 21, 37].

4.3. Comparison of our findings with other studies

To our knowledge, a previous epidemiological study evaluated the conduct and results of subgroup analyses in IPDMAs [19]. This study showed 36.6% of the 279 IPDMAs that reported at least one subgroup analysis had a statistically significant treatment-subgroup interaction for the primary outcome [34]. Another study found that many IPDMAs performed subgroup analyses, but the overall treatment effect was more emphasized than the subgroup effect, and IPDMAs often did not report adequate information on the methods of subgroup analyses [24]. Our study found that 53.9% of the 89 cancer IPDMAs did not report a statistically significant treatment-subgroup interaction, only 26.9% of the 103 factors were found with a statistically significant treatment-subgroup interaction, and most IPDMAs did not describe methods for subgroup analyses, which are similar to the findings of previous studies. Furthermore, most cancer IPDMAs did not perform the planned subgroup analyses, which is also a flaw in the Cochrane intervention reviews regarding age-treatment subgroup analyses [38]. These defects should be given more attention by future reviewers of IPDMAs.

4.4. Strengths and limitations

As far as we know, this is the first study to investigate whether the methods of subgroup analyses of cancer IPDMAs were prespecified and the consistencies between the planned subgroup analyses and final reported results of subgroup analyses. However, our study also has some limitations. First, although we conducted

a comprehensive literature search and included a large number of IPDMAs, the included IPDMAs are all concentrated on RCTs in the cancer field and published in English. Therefore, the findings of this study may not apply to IPDMAs in other fields, IPDMAs published in other languages, or IPDMAs of other types such as case-control studies and cohort studies [39]. Second, our data depended on the information reported in the included IPDMAs, so we could not rule out the possibility that the methods of some subgroup analyses were prespecified but not reported due to non-significant subgroup effects or influences of reviewers' and editors' opinions [40]. Third, prespecification in the context of reviews is more troublesome than it is for individual RCTs where we can verify that the protocol has been established before the inclusion of the first patient. In the context of IPDMAs, the aggregate data and the individual patient data could have been examined before the protocol was written, making some subgroup analyses are no longer truly pre-specified. Fourth, we could not obtain protocols of some IPDMAs due to limited information, which may affect the proportion concerning the prespecification of subgroup analyses. Fifth, we did not assess the methodological and reporting quality of included IPDMAs as the quality of these IPDMAs would not affect the results of our study. However, it is interesting for future research to assess whether the methodological and reporting quality of IPDMAs will affect the reporting of subgroup analyses.

4.5. Implications for research and practice

IPDMAs provide an ideal framework with greatly improved statistical power and is therefore very suitable for subgroup analyses [12, 41]. Our study showed that

nearly half of cancer IPDMAs did not prespecify subgroup analyses and only a few IPDMAs performed all planned subgroup analyses. Preferred Reporting Items for Systematic Review and Meta-Analyses of Individual Participant Data (PRISMA-IPD) recommends that a systematic review report should record all subgroup analyses performed and whether these analyses are prespecified [42]. A Measurement Tool to Assess Systematic Reviews-2 (AMSTAR-2) suggests that review methods should be established before the conduct of the review and any significant deviations from the protocol should be justified [43]. Thus, future IPDMAs should be performed and reported according to these statements. When performing subgroup analysis, authors should first prespecified the subgroup analyses in a priori protocol, and carry out the planned subgroup analyses according to the protocol. Besides, authors should explain and justify any changes from the protocol, such as providing rationales for the subgroup analyses that were not conducted, clarifying the reasons for additional subgroup analyses and subgroup analyses of post hoc [38].

Treatment-subgroup interactions examination was observed in 89.9% of cancer IPDMAs, but only 4 IPDMAs used the deft approach to test the interactions and most IPDMAs did not provide enough information to determine whether a daft, deluded, or deft approach was used. Testing for interaction using across-trial information alone or the combination of across-trial and within-trial information may lead to aggregation bias [44, 45]. Existing guidance recommended that interaction estimates should be based solely on within-trial information, which can avoid aggregation bias and facilitate the clear presentation and readily interpretation of results [9, 25, 46]. When

examining the interaction between continuous variables and treatment effects, categorization of continuous variables using cut-points should be avoided, as this will reduce power and may affect the magnitude and statistical significance of results [47-49]. The recommended approach is to examine nonlinear relationships for continuous variables [9]. However, our research indicated all cancer IPDMAs that investigated continuous variables categorized corresponding continuous variables, and only one IPDMA examined non-linear relationships for continuous variables. Therefore, existing guidance should be strictly followed for future IPDMAs. Journal editors and peer reviewers should ensure that results of subgroup analyses are fully reported, and help finds inappropriate subgroup analyses, as well as checking the protocols and IPDMA reports to identify discrepancies in subgroup analyses between protocols and final reports to ensure subgroup analyses are appropriately conducted and reported before accepting IPDMAs reports for publications.

5. Conclusions

Many cancer IPDMAs did not prespecify subgroup analyses, nor did they fully perform planned subgroup analyses. Lack of details for the test of treatment-subgroup interactions and the use of deft approach is suboptimal. Cancer IPDMAs more frequently categorized continuous variables but rarely examined the non-linear interactions. Future IPDMAs should address these flaws. Existing guidance should be strictly followed to select an appropriate method for the conduct and reporting of subgroup analysis.

Abbreviations

IPDMAs: individual participant data meta-analyses; SR: Systematic review; RCT: randomized controlled trial; AMSTAR-2: Assessment of Multiple Systematic Reviews-2; PRISMA-IPD: Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data.

Declarations

Not applicable.

Acknowledgments

The authors thank all investigators and supporters involved in this study.

Funding

This work was supported by the Gansu Province Science and Technology Plan Funded Project (20CX4ZA027) and the Beijing Nova Program (No. Z201100006820070) from Beijing Municipal Science & Technology Commission.

Role of the Funding Source

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Availability of data and materials

All data generated or analyzed during this study are included in this published study and attachments.

Authors' contributions

YG, FJS, and JHT planned and designed the study. YG and JHT developed search strategies. YG, ML, SZS, MMN, JL, and JHZ screened potential studies and extracted data from the included studies. YG and JHT managed the data and performed the statistical analysis. FJS and JHT conducted the arbitration under disagreement and ensured that there were no errors. JHZ, FJS, and JHT provided methodological support and helped to interpret findings. YG wrote the first draft. YG, FJS, and JHT revised the draft. All authors approved the final version of the manuscript.

What is new?

Key findings

- Prespecification of subgroup analyses in cancer individual participant data meta-analyses (IPDMAs) were inadequate.
- Lack of details for the test of treatment-subgroup interactions and the use of deft approach was suboptimal.
- Cancer IPDMAs more frequently categorized continuous variables but rarely examined the non-linear interactions.

What this adds to what was known?

- This study explored the prespecification of subgroup analyses in cancer IPDMAs, whether the planned subgroup analyses were conducted, methods used for the test of treatment-subgroup interactions, and statistical approaches for examining continuous subgroup variables.

What is the implication and what should change now?

- The presentation and examination of treatment-covariate interactions and the proper handling of continuous variables need substantial improvement.
- Existing methodological and reporting guidance should be strictly followed to select an appropriate method for the conduct and reporting of subgroup analysis in IPDMAs.
- Journal editors and peer reviewers should ensure that the results of subgroup analyses are fully reported and help finds inappropriate subgroup analyses.

Author Statement

Prespecification of subgroup analyses and examination of treatment-subgroup interactions in cancer individual participant data meta-analyses were suboptimal

I have made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND

I have drafted the work or revised it critically for important intellectual content; AND
I have approved the final version to be published; AND

I agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

The detailed author contributions are as follows:

YG, FJS, and JHT planned and designed the study. YG and JHT developed search strategies. YG, ML, SZS, MMN, JL, and JHZ screened potential studies and extracted data from the included studies. YG and JHT managed the data and performed the statistical analysis. FJS and JHT conducted the arbitration under disagreement and ensured that there were no errors. JHZ, FJS, and JHT provided methodological support and helped to interpret findings. YG wrote the first draft. YG, FJS, and JHT revised the draft. All authors approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Veroniki AA, Straus SE, Soobiah C, Elliott MJ, Tricco AC. A scoping review of indirect comparison methods and applications using individual patient data. BMC medical research methodology. 2016;16:47.

[2] Simmonds M, Stewart G, Stewart L. A decade of individual participant data meta-analyses: A review of current practice. Contemporary clinical trials. 2015;45:76-83.

[3] Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Evaluation & the health professions. 2002;25:76-97.

[4] Gao Y, Shi S, Li M, Luo X, Liu M, Yang K, et al. Statistical analyses and quality of individual participant data network meta-analyses were suboptimal: a cross-sectional study. BMC medicine. 2020;18:120.

[5] Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of clinical epidemiology. 2002;55:86-94.

[6] Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ (Clinical research ed). 2017;357:j1390.

[7] Belias M, Rovers MM, Reitsma JB, Debray TPA, IntHout J. Statistical approaches to identify subgroups in meta-analysis of individual participant data: a simulation study. BMC medical research methodology. 2019;19:183.

[8] Koopman L, van der Heijden GJ, Hoes AW, Grobbee DE, Rovers MM. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. International journal of technology assessment in health care. 2008;24:358-61.

[9] Riley RD, Debray TPA, Fisher D, Hattle M, Marlin N, Hoogland J, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. Statistics in medicine. 2020;39:2115-37.

[10] Tudur Smith C, Marcucci M, Nolan SJ, Iorio A, Sudell M, Riley R, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. The Cochrane database of systematic reviews. 2016;9:Mr000007.

[11] D'Amico R, Pifferi S, Leonetti C, Torri V, Tinazzi A, Liberati A. Effectiveness of antibiotic prophylaxis in critically ill adult patients: systematic review of randomised controlled trials. BMJ (Clinical research ed). 1998;316:1275-85.

[12] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ (Clinical research ed). 2010;340:c221.

[13] Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ (Clinical research ed). 2010;340:c117.

[14] Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blumle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ (Clinical research ed). 2014;349:g4539.

[15] Kent DM, van Klaveren D, Paulus JK, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement: Explanation and Elaboration. Annals of internal medicine. 2020;172:W1-W25.

[16] VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. Annals of internal medicine. 2011;154:680-3.

[17] Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet (London, England). 2005;365:176-86.

[18] Song F, Bachmann MO. Cumulative subgroup analysis to reduce waste in clinical research for individualised medicine. BMC medicine. 2016;14:197.

[19] Schuit E, Li AH, Ioannidis JPA. How often can meta-analyses of individual-level data individualize treatment? A meta-epidemiologic study. International journal of epidemiology. 2019;48:596-608.

[20] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. BMJ (Clinical research ed). 2011;342:d1569.

[21] Fan J, Song F, Bachmann MO. Justification and reporting of subgroup analyses were lacking or inadequate in randomized controlled trials. Journal of clinical epidemiology. 2019;108:17-25.

[22] Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in medicine. 2002;21:2917-30.

[23] Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blümle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ (Clinical research ed). 2014;349:g4539.

[24] Koopman L, van der Heijden GJ, Glasziou PP, Grobbee DE, Rovers MM. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. Journal of clinical epidemiology. 2007;60:1002-9.

[25] Fisher DJ, Copas AJ, Tierney JF, Parmar MK. A critical review of methods for the assessment of patient-level interactions in individual participant data

meta-analysis of randomized trials, and guidance for practitioners. Journal of clinical epidemiology. 2011;64:949-67.

[26] GBD 2017 Mortality Collaborators. Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet (London, England). 2018;392:1684-735.

[27] Gargon E, Williamson PR, Blazeby JM, Kirkham JJ. Improvement was needed in the standards of development for cancer core outcome sets. Journal of clinical epidemiology. 2019;112:36-44.

[28] Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Annals of internal medicine. 2020;172:35-45.

[29] Schuit E, Roes KC, Mol BW, Kwee A, Moons KG, Groenwold RH. Meta-analyses triggered by previous (false-)significant findings: problems and solutions. Systematic reviews. 2015;4:57.

[30] Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. Jama. 2014;311:405-11.

[31] Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses.

CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2020;192:E901-e6.

[32] Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Systematic reviews. 2016;5:79.

[33] Koensgen N, Rombey T, Allers K, Mathes T, Hoffmann F, Pieper D. Comparison of non-Cochrane systematic reviews and their published protocols: differences occurred frequently but were seldom explained. Journal of clinical epidemiology. 2019;110:34-41.

[34] Page MJ, McKenzie JE, Forbes A. Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. Journal of clinical epidemiology. 2013;66:524-37.

[35] Gao Y, Cai Y, Yang K, Liu M, Shi S, Chen J, et al. Methodological and reporting quality in non-Cochrane systematic review updates could be improved: a comparative study. Journal of clinical epidemiology. 2020;119:36-46.

[36] Allers K, Hoffmann F, Mathes T, Pieper D. Systematic reviews with published protocols compared to those without: more effort, older search. Journal of clinical epidemiology. 2018;95:102-10.

[37] Varadhan R, Segal JB, Boyd CM, Wu AW, Weiss CO. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. Journal of clinical epidemiology. 2013;66:818-25.

[38] Liu P, Ioannidis JPA, Ross JS, Dhruva SS, Luxkaranayagam AT, Vasiliou V, et al. Age-treatment subgroup analyses in Cochrane intervention reviews: a meta-epidemiological study. BMC medicine. 2019;17:188.

[39] Gao Y, Yang K, Cai Y, Shi S, Liu M, Zhang J, et al. Updating systematic reviews can improve the precision of outcomes: a comparative study. Journal of clinical epidemiology. 2020;125:108-19.

[40] Hannink G, Gooszen HG, van Laarhoven CJ, Rovers MM. A systematic review of individual patient data meta-analyses on surgical interventions. Systematic reviews. 2013;2:52.

[41] Mistry D, Stallard N, Underwood M. A recursive partitioning approach for subgroup identification in individual patient data meta-analysis. Statistics in medicine. 2018;37:1550-61.

[42] Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. Jama. 2015;313:1657-65.

[43] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or

non-randomised studies of healthcare interventions, or both. BMJ (Clinical research ed). 2017;358:j4008.

[44] Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study G. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics in medicine. 2002;21:371-87.

[45] Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statistics in medicine. 2008;27:1870-93.

[46] Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ (Clinical research ed). 2017;356:j573.

[47] Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. International journal of epidemiology. 1999;28:964-74.

[48] Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in medicine. 2006;25:127-41.

[49] Altman DG, Royston P. Statistics notes - The cost of dichotomising continuous variables. BMJ (Clinical research ed). 2006;332:1080.

Figure legend

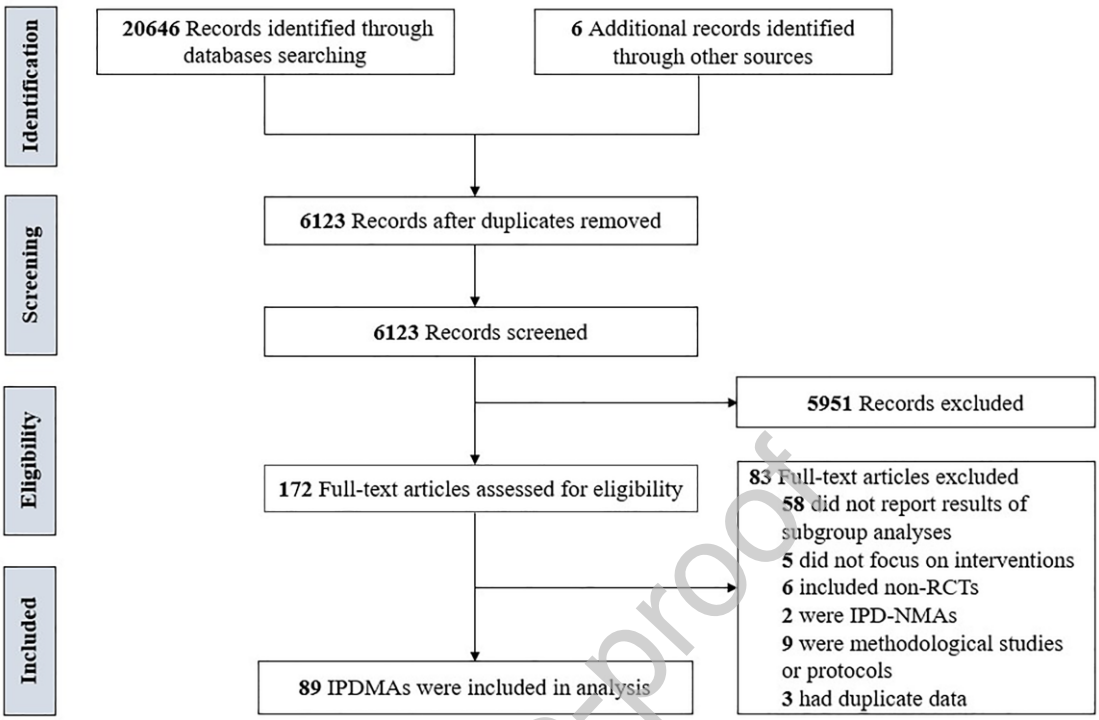

Figure 1. The flowchart of the screening process. RCTs, randomized controlled trials; IPDMAs, individual participant data meta-analyses; IPD-NMAs, individual participant data network meta-analyses.

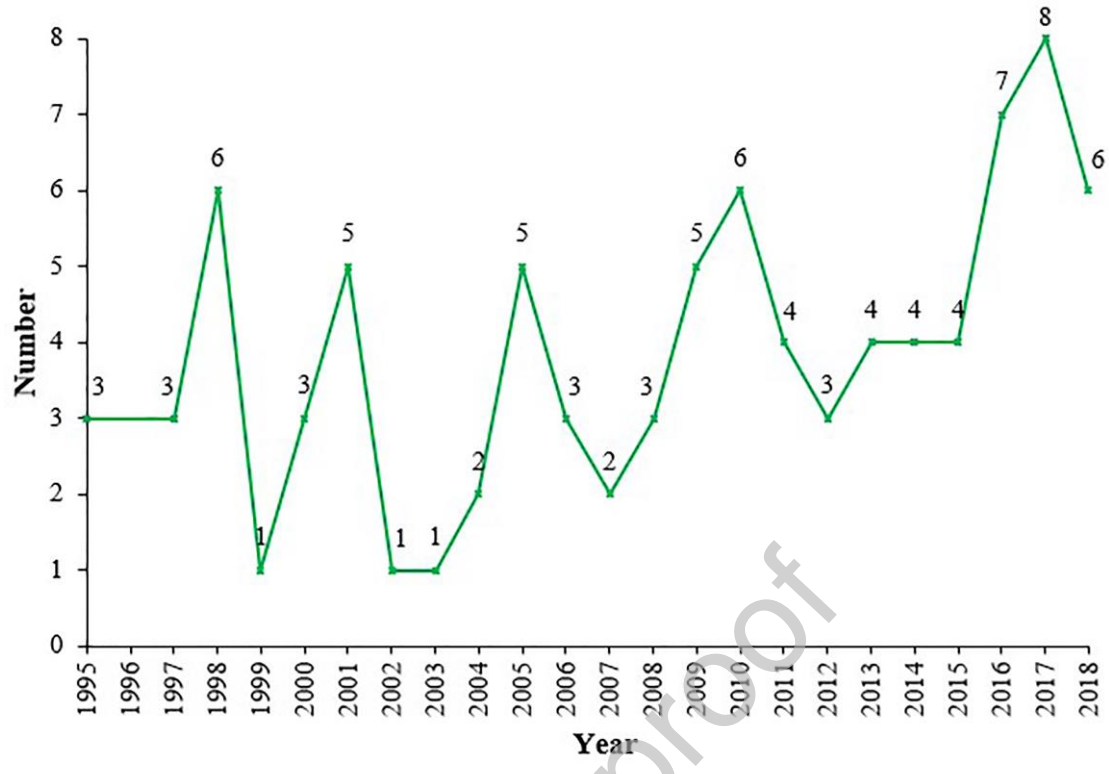

Figure 2. Publication years of included IPDMAs

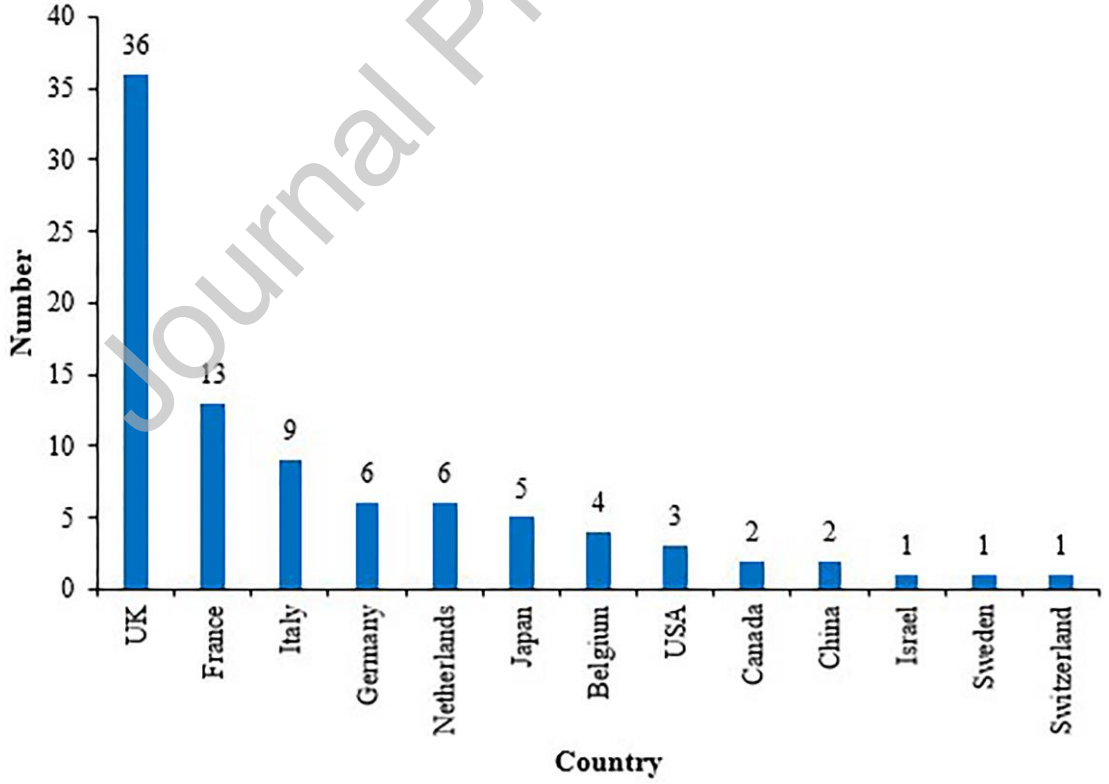

Figure 3. Country of included IPDMAs

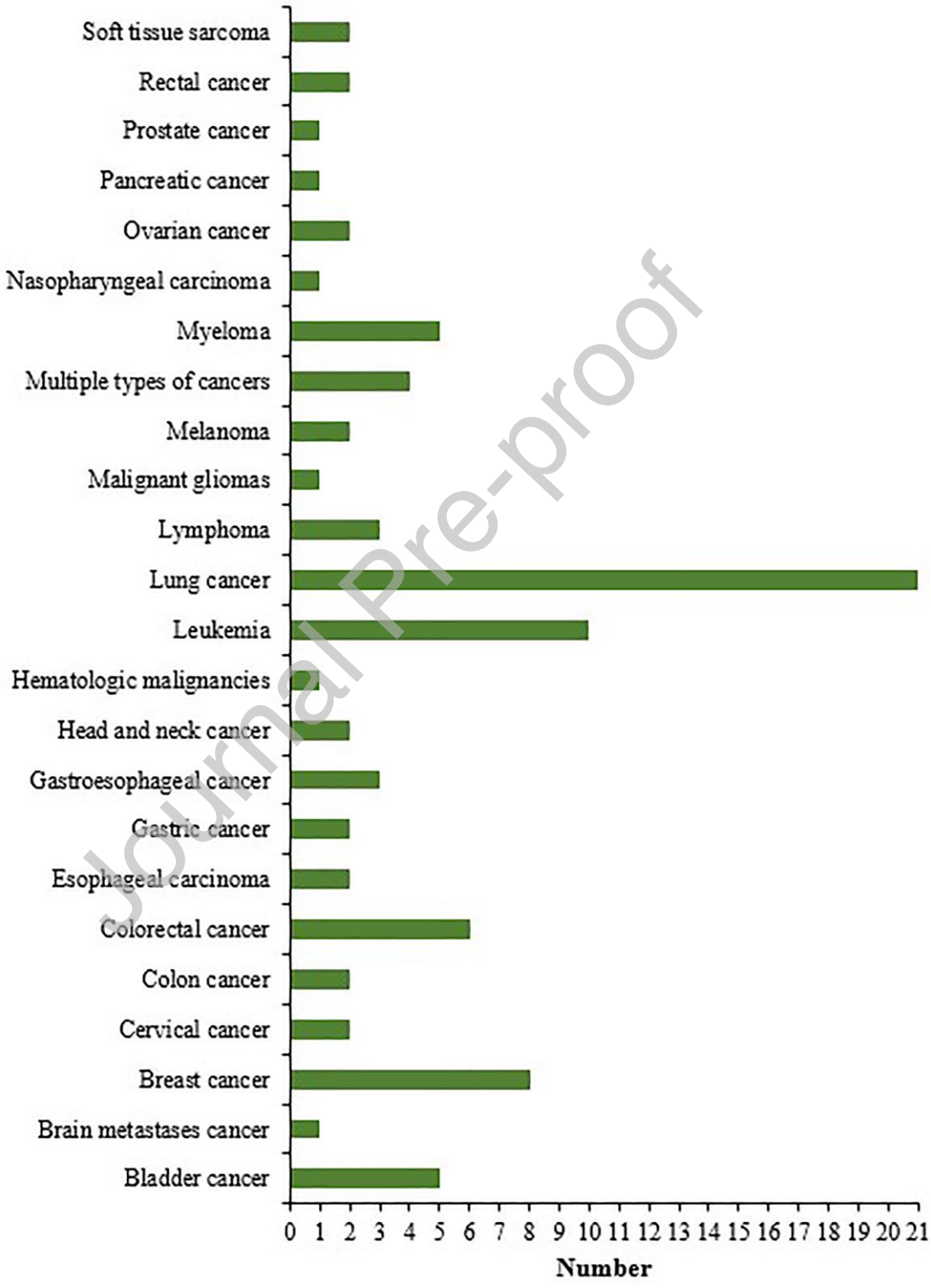

Figure 4. Cancer categories of included IPDMAs

Table 1. Characteristics of included IPDMAs

Category	Frequency	Proportion (%)
Protocol	37	41.6
Cochrane	12	13.5
Supplementary	4	4.5
Article	4	4.5
Contact author	8	9.0
No specific access method	9	10.1
Funding sources		
Industry	8	9.0
Industry + non-industry	4	4.5
Not report	15	16.9
Non-funding	14	15.7
Non-industry	48	53.9
Number of included patients		
500-1000	20	22.5
1001-2000	17	19.1
2001-3000	16	18.0
3001-5000	18	20.2
5001-10000	14	15.7
10001-40000	4	4.5

Number of included RCTs		
≤10	54	60.7
11-20	22	24.7
21-50	10	11.2
>50	3	3.4
Number of RCTs used in data analysis		
≤10	66	74.2
11-20	14	15.7
21-40	9	10.1
One-stage or two-stage approach		
One-stage	9	10.1
Two-stage	3	3.4
Not reported	77	86.5
Intervention		
Drug	49	55.1
Radiotherapy	5	5.6
Surgery	2	2.2
Drug + surgery + radiotherapy	8	9.0
Drug + radiotherapy	11	12.4
Drug + surgery	7	7.9
Radiotherapy + surgery	4	4.5
Other	3	3.4

Table 2. Testing for treatment-subgroup interactions and handling of continuous variables

Category	Frequency	Proportion (%)
Test for treatment-subgroup interactions (n=89)		
Yes	80	89.9
No	9	10.1
Method used for analysis of treatment-subgroup interactions (n=80)		
Deft (within-trial information)	4	5.0
Deluded (within-trial and across-trial information)	4	5.0
Unclear	72	90.0
Used continuous variables in the subgroup analysis (n=89)		
Yes	85	95.5
No	4	4.5
Number of continuous variables used for subgroup analysis (n=85)		
1	57	67.1
2	20	23.5
3	4	4.7
4	1	1.2
5	2	2.4
7	1	1.2
Categorized continuous variables (n=85)		
Yes	85	100.0
No	0	0.0
Reported how cut-points were selected for continuous variables (n=85)		
Yes	1	1.2

No	84	98.8
Examined non-linear relationships for continuous variables (n=85)		
Yes	1	1.2
No	84	98.8

Table 3. The relationship between basic characteristics and subgroup factors

Category	Frequency	Proportion (%)
Reported basic characteristics		
Yes	80	89.9
No	9	10.1
Whether basic characteristics included factors considered in subgroup analyses (n=80)		
Include all the factors	63	78.8
Did not include 1 factor	8	10.0
Did not include 2 factors	4	5.0
Did not include 3 factors	4	5.0
Did not include 4 factors	1	1.2

Table 4. Subgroup factors planned in protocols or Methods section and reported in Results section [N(%)]

Category	Protocols or methods section	Results section
Reported factors used for subgroup analyses?		
Yes	47(52.8)	89(100.0)
No	42(47.2)	0(0.0)
Number of factors used for subgroup analysis		
1	11(12.4)	5(5.6)
2	3(3.4)	4(4.5)
3	3(3.4)	13(14.6)
4	8(9.0)	14(15.7)
5	11(12.4)	22(24.7)
6	5(5.6)	10(11.2)
7	6(6.7)	12(13.5)
8	0(0.0)	2(2.2)
9	0(0.0)	4(4.5)
10	0(0.0)	1(1.1)
12	0(0.0)	1(1.1)
16	0(0.0)	1(1.1)

Table 5. Changes in the number of subgroup factors between the Results section and protocols or Methods section (n=47)

Category	Frequency	Proportion (%)
No change	19	40.4
Add 1 factor	8	17.0
Add 2 factors	4	8.5
Add 3 factors	1	2.1
Add 4 factors	5	10.6
Add 5 factors	2	4.3
Add 6 factors	1	2.1
Reduce 1 factor	5	10.6
Reduce 2 factors	1	2.1
Reduce 4 factors	1	2.1