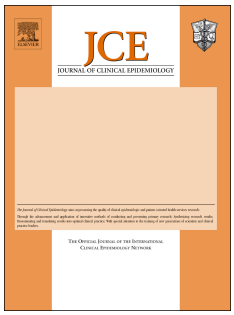

Non-linear effects and effect modification at the participant-level in IPD meta-analysis part 1: Analysis methods are often substandard

Nadine Marlin, Peter J. Godolphin, Richard Hooper, Richard D. Riley, Ewelina Rogozinska

PII: S0895-4356(23)00101-4

DOI: https://doi.org/10.1016/j.jclinepi.2023.04.013

Reference: JCE 11070

To appear in: *Journal of Clinical Epidemiology*

Received Date: 16 January 2023

Revised Date: 20 March 2023

Accepted Date: 26 April 2023

Please cite this article as: Marlin N, Godolphin PJ, Hooper R, Riley RD, Rogozinska E, Non-linear effects and effect modification at the participant-level in IPD meta-analysis part 1: Analysis methods are often substandard, *Journal of Clinical Epidemiology* (2023), doi: https://doi.org/10.1016/j.jclinepi.2023.04.013.

Non-linear effects and effect modification at the participant-level in IPD meta-analysis part 1: Analysis methods are often substandard

Authors: Nadine Marlin[a], Peter J Godolphin[b], Richard Hooper[a], Richard D Riley[c], Ewelina Rogozinska[b]

a) Methodology Research Unit, Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London
58 Turner Street, London E1 2AB, UK

b) MRC Clinical Trials Unit at University College London, Institute of Clinical Trials and Methodology, London, UK
90 High Holborn, London WC1V 6LJ, UK

c) Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham
Birmingham, UK. B15 2TT

Corresponding author: Nadine Marlin, Methodology Research Unit, Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London, 58 Turner Street, London E1 2AB, UK

Email: n.marlin@qmul.ac.uk, Phone: +44 (0) 20 7882 7327

Abstract

Objective: To review analysis methods used for linear effect modification (LEM), non-linear associations (NL) and non-linear effect modification (NLEM) at the participant-level in individual participant data meta-analyses (IPDMA).

Study Design and Setting: We searched Medline, Embase, Web of Science, Scopus, PsycINFO and the Cochrane Library to identify IPDMA of randomized controlled trials (PROSPERO CRD42019126768). We investigated if and how IPDMA examined LEM, NL and NLEM, including whether aggregation bias was addressed and if power was considered.

Results: We screened 6466 records, randomly sampled 207 and identified 100 IPDMA of LEM, NL or NLEM. Power for LEM was calculated a priori in 3 IPDMA. Of 100 IPDMA, 94 analysed LEM, 4 NLEM and 8 NL. One-stage models were favoured for all three (56%, 100%, 50% respectively). Two-stage models were used in 15%, 0% and 25% of IPDMA with unclear descriptions in 30%, 0% and 25%, respectively. Only 12% of one-stage LEM and NLEM IPDMA provided sufficient detail to confirm they had addressed aggregation bias.

Conclusion: Investigation of effect modification at the participant-level is common in IPDMA projects, but methods are often open to bias or lack detailed descriptions. Non-linearity of continuous covariates and power of IPDMA are rarely assessed.

Word count: 198

Key words:

- Individual participant data meta-analysis
- Effect modification
- interaction
- Non-linear
- Sample size

Running title: Non-linear effects and effect modification in IPD meta-analysis

1. Introduction

Evidence synthesis is increasingly used to identify characteristics of patients that respond better to an intervention than others, often termed personalised medicine. Compared to a single trial, the larger sample size afforded by meta-analysis usually provides greater power to reliably determine if a treatment is beneficial and whether the effect of treatment is modified by patient-level covariates.

Meta-analysis can be performed on aggregated summary measures or on individual participant data (IPD). The benefits of analysing on the individual level are well known [2-4]. Issues such as outcome reporting bias, accounting for prognostic factors, and aggregation bias can be alleviated with IPD meta-analysis (IPDMA), and complex relationships including linear effect modification (LEM), non-linear covariate-outcome associations (NL) and non-linear effect modification (NLEM) can be assessed more reliably. Terminology varies in the literature [5] (Box 1). Allowing for such complexity during analysis facilitates the identification of patient subgroup effects, if they exist, and can improve the precision of effect sizes. Leijten et al included non-linear effect modification in their IPDMA and showed that contrary to expectation, children with more severe conduct problems benefitted more from the IY parenting program [6]

In a single trial, methods for analysing effect modification and non-linearity are well established. Effect modification is commonly assessed by splitting the data or including interaction terms into the analysis. Common methods for addressing non-linearity include categorization [7, 8], fractional or ordinary polynomials [9, 10] and splines, usually restricted cubic [10]. Extending these to meta-analyses is not always straightforward as variation between studies must be incorporated. In recent years, the methodology available to perform such complex analyses has seen considerable development. Therefore, it is essential to understand what methods are currently used and how they are implemented in IPDMA.

In this article, we present findings of the first of two reviews on IPDMAs. In part 1, we review IPDMAs of randomised controlled trials (RCTs) published between 2015-2020, with the aims (1) to describe, if and how LEM, NL and/or NLEM are assessed at the individual level; (2) what methods are being employed when examining LEM, NL and/or NLEM at the individual level; and (3) whether these methods meet current methodological standards. The second review (part 2) will focus on summarising available methodology guidance so researchers can identify the most relevant method for their IPDMA [11].

Here we present some brief explanations of commonly used terms in the literature.

LEM, NL and NLEM:

Interaction: The combined effect of two factors is different than their individual effects. During analysis a multiplicative term is included into the model in addition to the individual factors.

Effect modification: It is a type of interaction between a binary intervention indicator and a covariate called the effect modifier. The effect of an intervention differs depending on the level of the modifier characteristic. During analysis an interaction term between the intervention indicator and covariate is included in the model. If the covariate is categorical, the term is also used when the effect is estimated within subsets of data.

Subgroup effect: The effect of the intervention within a defined subset of patients usually defined by categorical characteristics. The term subgroup effect is used for analyses including interaction terms or analyses within subsets of data.

Non-linearity: Estimates are not consistent across varying levels of patient characteristics, either in an effect modification or covariate-outcome relationship.

IPDMA approaches and distributional assumptions:

Two-stage IPDMA: The effect of interest is analysed in each trial separately and the estimates combined using meta-analysis techniques.

One-stage IPDMA: Data from all trials are analysed together while accounting for clustering by trial

Common / Fixed effects: The true effect is assumed to be the same across trials. Differences seen in individual trial estimates are only due to sampling error.

Random effects: The true effects in each trial are assumed to follow a normal distribution allowing for between study variation.

Effects stratified by trial: The effect in each trial is independent from those in other trials.

Box 1: Terminology for individual participant data meta-analysis of complex relationships

2. Methods

This literature review was guided by a prospectively registered protocol (CRD42019126768) and recommendations on the conduct of methodological reviews [12]. Reporting has been according to PRISMA-ScR where possible [13].

2.1. Literature review

We conducted a comprehensive literature search without language restrictions in Medline (via PubMed), Embase, Web of Science, Scopus, PsycINFO and the Cochrane Library. We searched for IPDMAs of RCTs and methodology publications on IPDMA, published between 01 January 2015 and 04 November 2020 (Appendix A). This time limit was chosen to build on previous reviews of IPDMA [1, 14]. We describe the results of the review of published methodology in part 2 [11]. A case study comparing the methods in an example IPD will be published separately.

2.2. Eligibility criteria

Articles reporting IPDMA were eligible if they included RCTs only and the main objective was estimation of an intervention effect. We restricted to IPDMA of RCTs due to their generally higher reporting standards, and to 'simplify' additional analysis and modelling issues that would otherwise arise using observational, non-randomised data. We excluded articles if clustering by trial was ignored during analysis, the paper only described network meta-analysis, or where the full text was not accessible.

2.3. Screening

After removal of duplicates, one researcher (NM) screened titles and abstracts and grouped the references into (1) potentially eligible IPDMAs, (2) potentially relevant IPDMA methods papers and (3) non-relevant papers.

Among the potentially eligible IPDMAs the same researcher (NM) then randomly sampled articles for a full-text review until 100 eligible IPDMAs that considered LEM, NL and/or NLEM had been identified. At least one other researcher (PG or ER) independently confirmed that each of the 100 articles analysed LEM, NL and/or NLEM at participant-level. The random sampling procedure was conducted as outlined in the study protocol (PROSPERO CRD42019126768) A random sample of 100 was deemed sufficiently large to identify most methods used by researchers while keeping double data extraction feasible within a limited time frame [12].

2.4. Data extraction

Data were extracted from all eligible IPDMAs identified during the sampling process, regardless of whether they assessed LEM, NL or NLEM, thereby enabling investigation of the frequency that such complex relationships are assessed.

Data were extracted using a prospectively developed data collection form that was reviewed after the first ten extractions (Appendix B). Initially we planned to double extract 10% of the included IPDMAs, however due to vague and unclear descriptions in the included articles, all articles were independently extracted by at least two members of the review team (NM, ER, PG). Remaining discrepancies were discussed between the group and, if necessary, further assessed by a senior researcher (RR).

We extracted general information on the IPDMAs, the analysis method employed by the authors and whether sample size and multiplicity had been considered (Box 2). If necessary, we checked associated documents such as protocols, analysis plans or previous publications on the same dataset.

General: Date of extraction, First author, Research group/collaboration, Year of publication, Medical Field, Number of datasets analysed, Number of participants included/analysed per dataset, Dataset identification, PROSPERO ID

Main effect analysis: General IPDMA approach (one-/two-stage, common/random/unclear intervention effect); General modelling approach (logistic, survival, etc.)

Outcome: Format of primary outcome, Is any outcome a composite

Power considerations: Sample size calculation (Power calculation, Post hoc power assessment, General talk of increased power when using IPD compared to single trial only); Multiple testing adjustments

Complex associations at the individual level: Non-linear effects; Effect modification; Subgroup analysis without testing

Details on non-linear effects analysis: Analysis approach; Justification for analysis approach; Any non-linear effect modification; If so, analysis approach used for non-linear effect modification

Details on effect modification analysis: Format of effect modifier (at least one analysed on continuous scale, all continuous effect modifiers categorised, categorical effect modifiers only); Number of effect modifiers; Number of outcomes for effect modification; Analysis approach for effect modification; Classification of analysis approach using terminology by Fisher et al [1] (Deft, One-stage unclear if within & across information separated out, Deluded, Daft, Unclear); Inclusion of additional covariates besides treatment-covariate interaction

Box 2: Data extraction (Not all data is reported in this article. For further findings see Godolphin et al. preprint [15]

For each analysis, we extracted information on whether the trials had been analysed separately and the effect measures combined ("two-stage") or individual level data from all trials had been analysed together while accounting for clustering by trial ("one-stage") [16]. We also extracted whether the true intervention effect was assumed to be the same across trials ("common or fixed effects") or whether the true trial-specific intervention effects were assumed to follow a normal distribution ("random effects") [16]. If interaction terms were included in the models, we also attempted to extract whether the model term assessing effect modification was assumed to be common or random.

We made the following data extraction rules with regards to one- or two-stage and common- or random-effects modelling:

1. A one-stage common effect approach had been employed if the methods mentioned "model stratified for trial" [16, 17]
2. We inferred the method used from software packages described or software commands used. For example, Review Manager or Stata commands 'metan' or 'ipdmetan' would have been used for two-stage approaches [18, 19]
3. Descriptions such "a random effect for study" or "random effects model" were models with random coefficient for treatment. If "random intercept" was mentioned, it was a common effect model.

We were not able to not extract whether effects of other parameters in the model (such as the intercept, covariates, or residual variances) were modelled as common, random, or stratified. This information was generally not reported.

For the effect modification analysis, we also extracted information on whether aggregation bias had been properly addressed by separating within- and across-trial variation (Table 1). In one-stage models this is done by centering the covariate around its study-specific mean or stratifying the parameters outside the interaction by trial [3, 20]. In two-stage models of interaction terms aggregation bias is addressed automatically [16, 20].

Table 1: Classification of effect modification approaches

IPDMA approach	**Effect modification approach**	**Within and across trial variation separated**
Two-stage	Analysis of interaction terms (subgroup analyses that do not involve modelling an interaction term do not count) [21]	Yes
	Analysis of subgroup effects	No
One-stage	Model including interaction terms between treatment and covariate and centering around study-specific mean or stratification is reported	Yes
	Model including interaction terms between treatment and covariate but no mentioning of centering or stratification	Unclear
	Model including interaction terms between treatment and covariate, where the covariate is centered around its overall mean or stating that it has not been centered	No
Aggregate level	Meta-regression of aggregated patient-level covariate measures	No

2.5. Data synthesis

Data from the included studies were summarised as frequencies or presented in histograms. No quantitative analysis was performed.

3. Results

3.1. Classification of articles

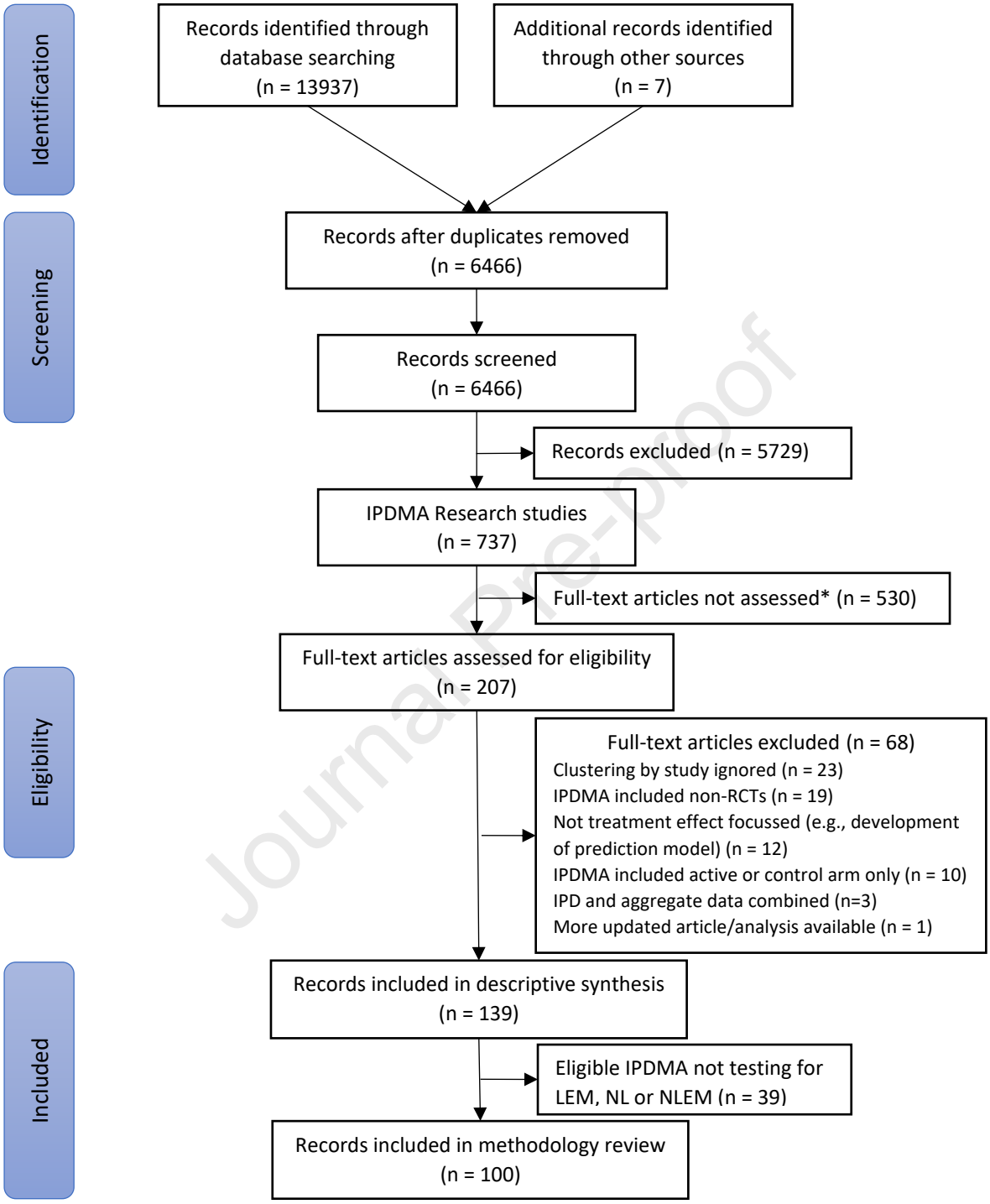

*Figure 1: Flow diagram (*This is in the line with the procedure described a priori whereas only a subset of identified IPD research studies was considered in full text. IPDMA… Individual participant data meta-analysis, RCT… Randomised controlled trial, LEM... linear effect modification, NL… non-linear associations, NLEM… non-linear effect modification)*

Database searches identified 6466 unique records including 737 potentially eligible IPDMAs. Of these, 207 were randomly sampled for full-text review to identify 139 eligible IPDMAs, 100 (70%) of which considered LEM, NL or NLEM and 39 (30%) which did not (Figure 1).

3.2. Characteristics of 100 IPDMAs analysing LEM, NL or NLEM

3.2.1. Description

Most of the IPDMAs identified came from cancer, cardiovascular or neurology research (Table 1, Appendix C Table 4)).

The majority of IPDMAs identified datasets systematically through literature searches (53/100, 53%) while 14 (14%) did not report how they identified the datasets (Table 1, Appendix C Table 4). Reporting standards were moderate with almost a quarter of article titles not indicating a meta-analysis of IPD and less than a quarter of publications reporting PROSPERO registration (Table 1, Appendix C Table 4).

IPDMAs included between 2 to 34 RCTs (median 5) with individual trials contributing between 7 and 44567 participants (smallest trial median 91 to largest trial median 709) to their respective IPDMA (Figure 2 and 3). The median total number of participants included in an IPDMA was 2186.

A priori power calculations, where power is assessed before IPD collection, were uncommon and only identified in more recent publications from 2017 onwards (Table 1). All five were performed knowing the exact studies and their total sample which ranged from 70 to 4965. Post hoc power assessments, where the achieved power is calculated based on the effect size estimated during data analysis, were slightly more common (8/100).

Only 3 studies *a priori* calculated the power required to analyse effect modification at the individual level. Similar trends were seen with adjustments for multiplicity where 94% did not adjust or did not mention the issue.

Table 2: General characteristics of the included IPD research studies (Data are frequency (%) unless stated otherwise)

Variable	Category	Number of IPD meta-analyses or statistic (N=100)
Year published	2015	14 (14%)
	2016	8 (8%)
	2017	17 (17%)
	2018	19 (19%)
	2019	22 (22%)
	2020	20 (20%)
Medical field	Cancer	11 (11%)
	Cardiovascular	29 (29%)
	Mental health	6 (6%)
	Neurology	14 (14%)
	Women's Health	7 (7%)
	Other	33 (33%)
Dataset identification	Systematic	53 (53%)
	Non-systematic	33 (33%)
	Not reported	14 (14%)
Title identifies article as meta-analysis of individual participant data	Yes	76 (76%)
	No	24 (24%)
PROSPERO registration	Yes	23 (23%)
	None reported	77 (77%)
Sample size consideration – Main effect	A priori power calculation	5 (5%)

	Post hoc power assessment	8 (8%)
	None or general discussion of power in IPDMA	87 (87%)
Sample size consideration – Effect modification	Power calculation	3 (3%)
Multiple testing considered	Yes	6 (6%)
	No	21 (21%)
	Unclear or not applicable	1 (1%)
	Not reported	72 (72%)

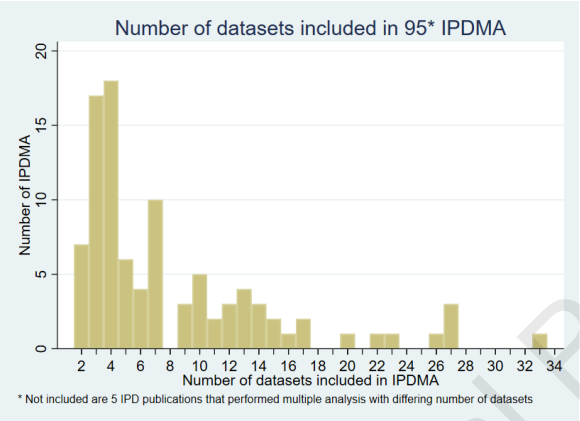

Figure 1: Number of datasets in each individual participant dataset meta-analysis

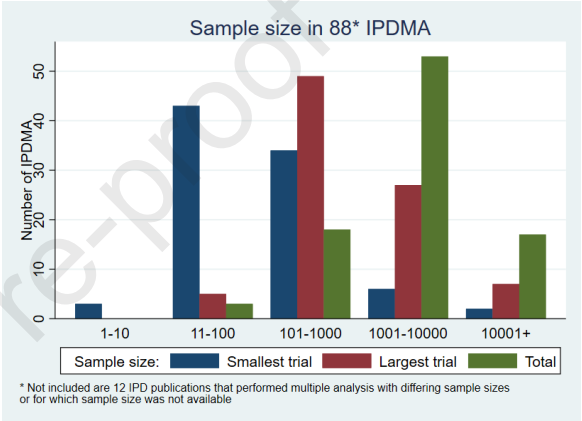

Figure 2: Number of participants in each individual participant dataset meta-analysis

3.2.2. Comparison to 39 IPD meta-analyses that did not model LEM, NL or NLEM

The median total number of participants included in analysis was about twice as large for IPDMAs investigating LEM, NL and/or NLEM compared to those that do not (2186 vs 1190, Appendix C Table 4).

Sample size considerations including multiple testing were more common in IPDMAs analysing LEM, NL and/or NLEM compared to those that do not. No other major differences in the key characteristics between those studies were found (Appendix C Table 4).

3.3. Modelling LEM, NL or NLEM at the individual level

Two IPDMAs evaluated non-linear covariate-outcome associations only, 88 assessed effect modification only and ten IPDMAs did both. Of those ten, four considered non-linear effect modifications while six had separate models for assessing effect modification and non-linear associations (Table 3). Categorisation of continuous effect modifiers was common (Table 2). Of 84 studies that modelled continuous effect modifiers (continuous or categorised continuous in Table 2), 59 (70%) only presented analysis of the categorised effect modifier.

Table 3: Forms of effect modification modelled in each IPDMA

Effect modifiers	Non-linear effects (n=12)	No non-linear effects (n=88)	Total
At least one effect modifier analysed as continuous	8 (4+4)*	17	**25**
All continuous effect modifiers were categorised	1	58	**59**
Only categorical effect modifiers analysed	1	13	**14**
None	2	-	**2**

* 4 studies assessed non-linear effect modification, 4 studies modelled effect modification and non-linear terms separately

The majority of IPDMAs tested for effect modification by including interaction terms into their one- or two-stage IPDMA model (62/94, Table 3). Of those that had used one-stage approaches only 11% (6/54) reported centering the covariate by its study-specific mean, which is a valid approach for separating within- and across-trial variation[3, 20]. 40 one-stage IPDMAs of effect modification did not mention either approach while 8 reported not to have centered their effect modifiers. Ten reported testing subgroup effects. We only identified one analysis of effect modification on the aggregate study level, which was suitably termed a "daft" approach by Fisher et al [1]. 24 didn't report their approach or the description was insufficient.

Non-linear effect modification was modelled in one study using the multivariable fractional polynomial interaction procedure to identify the best fitting non-linear form [22]. Three studies pre-specified for the interactions to include quadratic terms of the potential effect modifying covariate. All four reported using one-stage models but only one reported dealing with aggregation bias by centering the effect modifier around its study-specific mean. One stated that the covariate had not been centered while two did not mention the issue at all.

Non-linear covariate-outcome effects were modelled in eight studies. Two used three-knot splines, one positioned the knots at tertiles of the covariate and one did not report knot locations. Two studies included unspecified polynomial terms of the covariate in their statistical model, one included a quadratic term, and one log-transformed their outcome. The remaining two studies did not present sufficient detail, simply stating they had used non-linear mixed models or employed a variety of non-linear methods.

Table 4: Meta-analysis approach for Linear effect modification (LEM), non-linear covariate-outcome associations (NL) or non-linear effect modification (NLEM) in 100 IPD research studies (studies can appear multiple times)

LEM, NL or NLEM	IPDMA approach	Analysis of	Within and across trial variation separated	Nr of IPDs
Linear effect modification (n=94)*	Two-stage (n=15)	Interaction term(s)[£]	Yes	8
		Subgroup effects	No	6
		log rank tests	No	1
	One-stage (n=54)	Interaction term(s) [£]	Yes[#]	6
		Interaction term(s) [£]	Unclear[@]	40
		Interaction term(s) [£]	No[$]	8
	Unclear	Subgroup effects	No	3
	Aggregate level	Subgroup effects	No	1
	Insufficient detail			24

Non-linear effect modification (n=4)	One-stage (n=4)	Interaction term(s), quadratic covariate	Yes[#]	1
		Interaction term(s), quadratic covariate	Unclear[@]	1
		Multivariable fractional polynomials interaction (degree 1)	Unclear[@]	1
		Interaction term(s), quadratic covariate	No[$]	1
Non-linear effects (n=8)	Two-stage	Restricted cubic splines		2
	One-stage (n=4)	Linear and polynomials terms with backward selection		1
		Fractional polynomials		1
		Quadratic term (pre-specified)		1
		Log-transformed outcome (based on model performance)		1
	Insufficient detail			2

*… 3 studies reported multiple approaches falling into different categories
£… Interaction term(s) between a binary intervention indicator and covariate (i.e. effect modifier)
#... Article reported centering of the effect modifier around the study-specific mean
@... Article did not report on dealing with aggregation bias by centering or stratification
$... Article reported that aggregation bias was not dealt with or dealt with incorrectly

4. Discussion

4.1. Main findings

In this article, we have reviewed the methodology employed in IPD research studies to address LEM, NL or NLEM. Investigating such associations with sufficient power can support the reliable identification of patient subgroups that benefit the most from an intervention. Our random sample of IPDMAs showed that analysis of linear effect modification at the individual-level is common in IPDMA of RCTs, with only a few investigating non-linearity.

1. Planning the analysis of effect modification - Sample size

Many IPDMAs analyse effect modification, whether linear or non-linear, but very few power for it. Post hoc assessments of power are more common. Zhang and colleagues showed that such assessments do not capture the true power to detect a desired effect. They can be misleading and are therefore discouraged by many authors [23-25]. The only three studies calculating power for the effect modification analysis *a priori* came from the same research team and due to lack of detail we could not reproduce the calculation.

Admittedly, control over the sample size in an IPDMA may be limited. However, researchers can assess before onset of the project whether their "promised" data will allow detection of a clinically relevant effect. This may impact decisions on whether it is worth embarking on the time-consuming tasks of obtaining, cleaning, harmonizing, and analysing the IPD [26]. It may also allow the researchers to focus their efforts on obtaining IPD for a subset of trials if there is clear justification (such as when the IPDMA needs to be completed quickly due to a very urgent clinical need for evidence).

Power calculations, done prior to IPD meta-analysis, are currently not part of PRISMA-IPD guidelines but can support promising IPDMA of effect modification or deter futile ones. It is also important to note that analysing large numbers of effect modifiers could increase the risk of spurious findings. Godolphin et al [15] show that on average six modifying covariates (range 1 to 28) were analysed for

on average two outcomes (range 1 to 16). As far as we are aware, there is currently no guidance how to power an IPDMA if multiple effect modifiers are of interest. We suggest assessing the power for each of a small set of effect modifiers deemed of most interest (primary analyses).

2. IPDMA methodology

Of 207 articles considered in full text we excluded over 10% (23/207) as they simply pooled the individual datasets during analysis without accounting for clustering by trial. This flawed practice can lead to misleading effect estimates and conclusions and should not be used [27].

One-stage models were clearly favoured for analysis of effect modification and/or non-linearity. Analysing all data in one step may appear more elegant but requires more care than the traditional two-step approach especially when analysing effect modification. Within- and across-trial variation need to be separated in a one-stage approach, for example by centering the effect modifier around their study specific mean and then including the interaction terms into the model [3]. Only seven of 58 one-stage analyses of effect modification reported such centering, while nine stated that they had not centered the effect modifier and 42 did not mention either approach so it is reasonable to assume it wasn't done during analysis. Therefore, almost three quarters of IPDMA analysing effect modification using a one-stage approach provided insufficient details to assess their appropriateness. It is important to note that many methodology articles were published during the study period (2015 – 2020) and therefore awareness of the issues might have been limited especially in the earlier IPDMA studies.

Non-linearity is rarely considered. Researchers still prefer to categorise continuous effect modifiers rather than analysing on the continuous scale and allowing for non-linearity. Admittedly, in some cases this may have been due to how the data were recorded in the original trials or how they were shared. Categorisation may be a useful investigatory tool [28], but dangers and pitfalls of categorisation as the primary analysis have been extensively shown for both single studies [7, 8] and for IPDMAs [29].

3. Reporting of IPDMA

PRISMA IPD (published in 2015) standardised the reporting of IPDMAs of randomised trials and provided an easy-to-follow guideline and checklist that should ensure high quality reporting of any such study [30]. However, even the reporting of basic information, such as the study type in the title of the IPDMA was missing for a substantial number of articles.

The overall IPD approach is usually extractable, (i.e. the use of common-effect or random-effects models or one- or two-stage approach), but modelling details such as the centering of covariates and handling of nuisance parameters (e.g. study-specific intercepts and adjustment factor effects), or for studies that utilised multiple approaches, what results were produced by each analysis was often not reported or unclear. We strongly support previous suggestions to publish software code or writing out the formal model specification (e.g. regression equation) to improve understanding and reproducibility especially of one-stage models [31, 32]. Some may resist due to the extra effort required to review manuscripts [33].

4.2. Limitations

This literature review has some limitations. Firstly, our results may be limited as we did not consider all IPD studies published 2015 – 2020 and instead used random sampling until we identified 100 eligible articles. However, we did perform a comprehensive search across a variety of databases with inclusive search terms, thus identifying a representative sample to randomly select articles from.

Secondly, in our random sample we identified 32 IPDMA that were conducted by researchers from the same team, used the same dataset or utilised shared protocols (14 groups). A number of these IPDMAs shared the same analysis strategy, likely reducing the variation in methods identified in this review.

Thirdly, we identified few analyses of non-linear covariate-outcome relationships. This is likely due to our inclusion criteria and restricting IPD research studies to those focussing on treatment effect estimation. Assessment of non-linear covariate outcome relationships are more common in prognostic factor studies or development of prediction models.

Finally, due to unclear reporting, we had to make some data extraction rules regarding analysis strategies. These might not necessarily reflect the interpretation of other researchers. However, these rules were agreed within a review team with extensive experience of IPDMA.

5. Conclusion

Analysis of complex relationships, in particular effect modification, is common in IPDMAs but rarely done appropriately. Most continuous covariates are categorised and analysed in a one-stage model that amalgamates within- and across-trial information, potentially introducing aggregation bias at the patient-level. To have a meaningful impact on stratified and precision medicine research, effect modification analyses need to be conducted correctly, sufficiently powered and reported transparently.

6. References

1. Fisher, D.J., et al., *Meta-analytical methods to identify who benefits most from treatments: Daft, deluded, or deft approach?* BMJ (Online), 2017. **356**.
2. Chen, B. and A. Benedetti, *Quantifying heterogeneity in individual participant data meta-analysis with binary outcomes.* Systematic Reviews, 2017. **6**(1).
3. Riley, R.D., 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**(15): p. 2115-2137.
4. Tierney JF, S.L., Clarke M, *Chapter 26: Individual participant data.*, in *Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022)*, T.J. Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, Editor. 2022, Cochrane.
5. Schandelmaier, S., et al., *Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses.* CMAJ, 2020. **192**(32): p. E901-E906.
6. Leijten, P., et al., *Individual Participant Data Meta-analysis: Impact of Conduct Problem Severity, Comorbid Attention-Deficit/Hyperactivity Disorder and Emotional Problems, and Maternal Depression on Parenting Program Effects.* Journal of the American Academy of Child and Adolescent Psychiatry, 2020. **59**(8): p. 933-943.
7. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* BMJ, 2006. **332**(7549): p. 1080.
8. Royston, P., D.G. Altman, and W. Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Stat Med, 2006. **25**(1): p. 127-41.
9. Royston, P. and W. Sauerbrei, *Multivariable Model-Building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. 2008, West Sussex, England: Wiley.
10. Harrell, F.E.J., *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2015, Heidelberg, Germany: Springer.
11. Marlin, N., et al., *Examining Non-Linear Effects and Effect Modification in IPD Meta-Analysis Part 2: Guidance is Available*. preprint 2023. Available at https://dx.doi.org/10.2139/ssrn.4333136
12. Mbuagbaw, L., et al., *A tutorial on methodological studies: the what, when, how and why.* BMC Med Res Methodol, 2020. **20**(1): p. 226.
13. Tricco, A.C., et al., *PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.* Ann Intern Med, 2018. **169**(7): p. 467-473.
14. Debray, T.P., et al., *Get real in individual participant data (IPD) meta-analysis: a review of the methodology.* Research synthesis methods, 2015. **6**(4): p. 293-309.
15. Godolphin, P.J., et al., *Use of multiple covariates in assessing treatment-effect modifiers: A methodological review of individual participant data meta-analyses*. preprint 2023. Available at https://doi.org/10.48550/arXiv.2302.12801
16. Burke, D.L., J. Ensor, and R.D. Riley, *Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ.* Statistics in Medicine, 2017. **36**(5): p. 855-875.
17. Morris, T.P., et al., *Meta-analysis of Gaussian individual patient data: Two-stage or not two-stage?* Statistics in Medicine, 2018. **37**(9): p. 1419-1438.
18. Fisher, D.J., *Two-stage individual participant data meta-analysis and generalized forest plots.* Stata Journal, 2015. **15**(2): p. 369-396.
19. Harris, R.J., et al., *metan: fixed- and random-effects meta-analysis.* Stata Journal, 2008. **8**(1): p. 3-28.
20. Hua, H., et al., *One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information.* Statistics in Medicine, 2017. **36**(5): p. 772-789.

21. Godolphin, P.J., et al., *Estimating interactions and subgroup-specific treatment effects in meta-analysis without aggregation bias: A within-trial framework.* Res Synth Methods, 2022.
22. Kasenda, B., et al., *Multivariable fractional polynomial interaction to investigate continuous effect modifiers in a meta-analysis on higher versus lower PEEP for patients with ARDS.* BMJ open, 2016. **6**(9): p. e011148.
23. Zhang, Y., et al., *Post hoc power analysis: is it an informative and meaningful analysis?* Gen Psychiatr, 2019. **32**(4): p. e100069.
24. Dziadkowiec, O., *Use of Statistical Power Analysis in Prospective and Retrospective Research.* J Obstet Gynecol Neonatal Nurs, 2021. **50**(2): p. 119-121.
25. Heinsberg, L.W. and D.E. Weeks, *Post hoc power is not informative.* Genet Epidemiol, 2022. **46**(7): p. 390-394.
26. Stewart, L.A. and J.F. Tierney, *To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data.* Eval Health Prof, 2002. **25**(1): p. 76-97.
27. Abo-Zaid, G., et al., *Individual participant data meta-analyses should not ignore clustering.* J Clin Epidemiol, 2013. **66**(8): p. 865-873 e4.
28. White, I.R., et al., *Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods.* Statistics in Medicine, 2019. **38**(3): p. 326-338.
29. Ensor, J., et al., *Simulation-based power calculations for planning a two-stage individual participant data meta-analysis.* BMC medical research methodology, 2018. **18**(1): p. 41.
30. Stewart, L.A., et al., *Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement.* JAMA, 2015. **313**(16): p. 1657-65.
31. Simmonds, M., G. Stewart, and L. Stewart, *A decade of individual participant data meta-analyses: A review of current practice.* Contemporary Clinical Trials, 2015. **45**: p. 76-83.
32. Assel, M. and A.J. Vickers, *Statistical Code for Clinical Research Papers in a High-Impact Specialist Medical Journal.* Ann Intern Med, 2018. **168**(11): p. 832-833.
33. Rougier, N.P. and K. Hinsen, *Code reviewing puts extra demands on referees.* Nature, 2018. **556**(7701): p. 309.

Appendix A Search strategies for identification of IPD meta-analysis and methodology publications

MEDLINE (pubmed version) – search performed 04th November 2020

1 "individual participant"
2 "individual participants"
3 "individual patient"
4 "individual patients"
5 "individual participant-level"
6 "individual patient-level"
7 "individual person"
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 Meta-Analysis as Topic [MH]
10 Meta-Analysis [MH]
11 Meta-Analysis [ALL]
12 Metaanalysis [ALL]
13 "Pooled analysis" [ALL]
14 9 or 10 or 11 or 12 or 13
15 8 and 14
16 "Individual patient data analysis"
17 15 or 16
18 limit 17 to PUBLICATION year = "2015 – 2020"

EMBASE, Web of Science, PsycINFO – searches performed 21st October 2020

1 "individual participant"
2 "individual participants"
3 "individual patient"
4 "individual patients"
5 "individual participant-level"
6 "individual patient-level"
7 "individual person"
8 1 or 2 or 3 or 4 or 5 or 6 or 7
9 Meta-Analysis [ALL]
10 Metaanalysis [ALL]
11 "Pooled analysis" [ALL]
12 9 or 10 or 11
13 8 and 12
14 "Individual patient data analysis"
15 13 or 14
16 limit 15 to PUBLICATION year = "2015 – 2020"

Scopus – search performed 21st October 2020

TITLE-ABS-KEY ((((((("individual participant") OR "individual patient") OR "individual person")) AND (((meta-analysis) OR metaanalysis) OR "pooled analysis"))) OR "individual patient data analysis")) AND PUBYEAR > 2014

1 "individual participant"
2 "individual patient"
3 "individual person"

4 1 or 2 or 3
5 Meta-Analysis
6 Metaanalysis
7 “Pooled analysis”
8 5 or 6 or 7
9 4 and 8
10 “Individual patient data analysis”
11 9 or 10
12 limit 11 to PUBLICATION year = “2015 – 2020”

The Cochrane Library – search performed 21st October 2020

1 individual NEXT participan* [ALL]
2 individual NEXT patien* [ALL]
3 individual NEXT person* [ALL]
4 1 or 2 or 3
5 Meta-Analysis as Topic [MeSH]
6 Meta-Analysis [MeSH]
7 Meta-Analysis [ALL]
8 Metaanalysis [ALL]
9 “Pooled analysis” [ALL]
10 5 or 6 or 7 or 8 or 9
11 4 and 10
12 “Individual patient data analysis” [ALL]
13 11 or 12
14 limit 13 to year = “2015 – 2019”

only extract protocols, reviews and trials

For trials the search is performed on year added to database (CENTRAL)

Appendix B - Data extraction form for IPDMA studies

Section A: General information

Administrative	
Date extracted	
Name of person extracting	
First author	
Research group / collaboration name	
IPD MA characteristics	
Publication Year	
Medical Field	
Number of datasets analysed	
Number of participants in those datasets	Min: Max:
Number of participants analysed from those datasets	Min: Max: Total:
Dataset identification	☐Systematic review COCHRANE ☐Literature search (*includes if an existing IPD dataset was used but updated with a literature search) ☐ Data sharing platform ☐ Collaboration – Prospective IPD with shared protocol ☐ Collaboration – existing IPD dataset (*includes studies using data from a previously published IPD analysis, e.g. done in the same research unit) ☐ Research program ☐ Trial registry database ☐ Patient registry database ☐ Company database ☐ Not reported
PROSPERO ID (if given)	
Types of study data included	☐RCTs only ☐Any other study types
IPDMA approach	☐1 stage, fixed effects ☐1 stage, random effects ☐2 stage, fixed effects ☐2 stage, random effects ☐Trial effect ignored
General modelling approach	
Data format of primary outcome	☐Binary ☐Categorical ☐Continuous ☐Time to event
Is any outcome reported a composite?	☐Yes ☐No

Sample size considerations	☐Power calculation ☐Post hoc power assessment ☐Limited to general talk of "IPDs" have larger power compared to aggregate Metanalysis or single trials ☐None (*includes if discussion mentions that the study may not have had enough power to detect an effect size)
Is multiple testing accounted for?	☐Yes ☐No ☐Not mentioned
Complex associations	☐Non-linear effects – GO TO SECTION B ☐Effect modification – GO TO SECTION C ☐Subgroup analysis without comparison test ☐Neither *NOTE: effect modification is not considered between treatment effect and trial. This is usually done to test homogeneity and does not involve interest in identification of effect modifiers*.

Section B: Complete only for papers assessing non-linear effects

Analysis approach for investigation of non-linear trends	
Justification for analysis method	

Section C: Complete only for papers assessing effect modification

Type of covariate in the effect modification (tick multiple)	☐Continuous variable (e.g. age, weight) ☐Categorised continuous variable (e.g. age <35 vs age 35+) ☐Categorical variable (e.g. gender, marital status)
Number of effect modifications considered in total	
Number of outcomes looked at for effect modification	
Analysis approach for investigation of effect modification	
Are additional covariates included beyond the treatment-covariate interaction for any effect modification?	

If so, how are these covariates selected?	☐Univariate analysis ☐Stepwise procedure ☐Determined a priori ☐Other, please describe……
If so, how are these covariates included?	☐Additional variables included in the model ☐3-way interaction (2 covariates interacted with treatment) ☐Multiple 2-way interactions included in the model ☐A higher-order factorial covariate constructed from two other categorical covariates ☐Other, please describe……
If so, are reasons reported? What are they?	

Section D: Contact details of IPD meta-analysis team

Corresponding authors name, email, and job/study role	
Principal investigator's name and email. If corresponding author is principal investigator then leave blank	
Lead statistician's name and email. If corresponding author is lead statistician, then leave blank	

Appendix C – Full baseline tables

Table 5: Detailed characteristics of eligible IPD research studies (Data are frequency (%) unless stated otherwise)

Variable	**Category or summary statistic**	**Nr of IPD**		
		No LEM, NL or NLEM* (N=39)	**LEM, NL or NLEM* (N=100)**	**All IPD (N=139)**
Year published	2015	3 (8%)	14 (14%)	17 (12%)
	2016	9 (23%)	8 (8%)	17 (12%)
	2017	4 (10%)	17 (17%)	21 (15%)
	2018	5 (13%)	19 (19%)	24 (17%)
	2019	6 (15%)	22 (22%)	28 (20%)
	2020	12 (31%)	20 (20%)	32 (23%)
Medical field	Addiction/Alcoholism	0 (0%)	1 (1%)	1 (1%)
	CAM (Complementary and alternative medicine)	1 (3%)	2 (2%)	3 (2%)
	Cancer	6 (15%)	11 (11%)	17 (12%)
	Cardiovascular	8 (21%)	29 (29%)	37 (27%)
	Critical Care	1 (3%)	5 (5%)	6 (4%)
	Dentistry	0 (0%)	1 (1%)	1 (1%)
	Diabetes	2 (5%)	0 (0%)	2 (1%)
	Gastroenterology	0 (0%)	1 (1%)	1 (1%)
	Hematology	1 (3%)	0 (0%)	1 (1%)
	Infection	0 (0%)	1 (1%)	1 (1%)
	Infectious Disease	4 (10%)	4 (4%)	8 (6%)
	Mental Health	4 (10%)	6 (6%)	10 (7%)
	Neonatal Health	0 (0%)	3 (3%)	3 (2%)
	Nephrology	1 (3%)	2 (2%)	3 (2%)
	Neurology	3 (8%)	14 (14%)	17 (12%)
	Nutrition	1 (3%)	2 (2%)	3 (2%)
	Public Health	0 (0%)	4 (4%)	4 (3%)
	Respiratory	0 (0%)	2 (2%)	2 (1%)
	Rheumatology	2 (5%)	1 (1%)	3 (2%)
	Sleep Medicine	0 (0%)	1 (1%)	1 (1%)
	Surgery	0 (0%)	2 (2%)	2 (1%)
	Virology	0 (0%)	1 (1%)	1 (1%)
	Women's Health	5 (13%)	7 (7%)	12 (9%)
Dataset identification	**Systematic**:			
	• Literature search	23 (59%)	51 (51%)	74 (53%)
	• Systematic review COCHRANE	0 (0%)	2 (2%)	2 (1%)
	Non-systematic:			
	• Collaboration - Prospective IPD with shared protocol	1 (3%)	6 (6%)	7 (5%)
	• Collaboration - existing IPD dataset	3 (8%)	19 (19%)	22 (16%)
	• Company database	1 (3%)	2 (2%)	3 (2%)
	• Data sharing platform	0 (0%)	3 (3%)	3 (2%)
	• Patient registry database	1 (3%)	0 (0%)	1 (1%)
	• Research program	1 (3%)	0 (0%)	1 (1%)

	• Trial registry database	1 (3%)	3 (3%)	4 (3%)
	Not reported	8 (21%)	14 (14%)	22 (16%)
Title identified article as meta-analysis of individual participant data	Yes	24	76	100
	No	15	24	39
PROSPERO registration	Yes	10 (26%)	23 (23%)	33 (24%)
	None reported	29 (74%)	77 (77%)	106 (76%)

* IPDMA… individual participant data meta-analysis, LEM… linear effect modification, NL… non-linear associations, NLEM… non-linear effect modification

Table 6: Sample size and power considerations in IPD research studies (Data are frequency (%) unless stated otherwise)

Variable	**Category or summary statistic**	**Number of IPD or statistic**		
		No LEM, NL or NLEM* (N=39)	**LEM, NL or NLEM* (N=100)**	**All IPD (N=139)**
Number of datasets	Median (Min – Max)	5 (2 – 41)	5 (2 – 34)	5 (2 – 41)
	2	4 (10%)	7 (7%)	11 (8%)
	3-5	21 (54%)	41 (41%)	62 (45%)
	6-10	8 (21%)	22 (22%)	30 (22%)
	11-20	2 (5%)	18 (18%)	20 (14%)
	More than 20	3 (8%)	7 (7%)	10 (7%)
	Multiple analyses with differing number of datasets	1 (3%)	5 (5%)	6 (4%)
Total number of participants in IPDMA	Median (Min – Max)	1190 (49 – 91779)	2186 (70 – 174000)	1764 (49 – 174000)
Number of participants per dataset	Smallest dataset Median (Min – Max)	80 (3-1399)	91 (7-20011)	84 (3-20011)
	Largest dataset Median (Min – Max)	432 (17 – 16608)	709 (12 – 44567)	645 (12-44567)
Sample size Main effect	*A priori* power calculation	0 (0%)	5 (5%)	5 (4%)
	Post hoc power assessment	3 (8%)	8 (8%)	11 (8%)
	None or general discussion of power in IPDMA	36 (92%)	87 (87%)	123 (88%)
Sample size Effect modification	*A priori* power calculation	N/A	3 (3%)	N/A
Multiple testing considered	Yes	0 (0%)	6 (6%)	6 (4%)
	No	2 (5%)	21 (21%)	23 (17%)
	Unclear or not applicable	1 (3%)	1 (1%)	2 (1%)

	Not reported	36 (92%)	72 (72%)	108 (78%)

* IPDMA… individual participant data meta-analysis, LEM… linear effect modification, NL… non-linear associations, NLEM… non-linear effect modification

Declaration of interests

☐ 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.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Nadine Marlin reports financial support was provided by National Institute for Health and Care Research.

What is new?

Key findings

- Examination of subgroup effects and effect modification is common in individual participant data meta-analysis (IPDMA) but with often inadequate methods. Non-linearity in effect modification of continuous covariates is seldom considered.
- Power requirements for main effects or effect modification are rarely calculated a priori.

What this adds to what is known?

- This review provides an overview of what methods are currently used to address non-linear associations and effect modification in 100 IPDMA of randomised controlled trials between 2015-2020.

What is the implication?

- Analysis of effect modification in IPDMA can be improved by following existing methodological guidance, including separating within-trial and across-trial relationships to remove aggregation bias, by considering non-linear relationships, and by assessing the potential power of the IPDMA project in advance of IPD collection.
- Lack of details in reporting IPDMAs methods could (partly) be addressed by including analysis code or formal model specifications in publications.

Nadine Marlin: Conceptualization; Data curation; Formal analysis; Investigation; Roles/Writing - original draft; Supervision

Peter J Godolphin: Data curation; Investigation; Roles/Writing - review & editing; Supervision

Richard Riley: Conceptualization; Roles/Writing - review & editing

Richard Hooper: Conceptualization; Roles/Writing - review & editing

Ewelina Rogozińska: Data curation; Investigation; Roles/Writing - review & editing; Supervision