Assessing the integrity of randomised trials using individual participant data: the IPD Integrity Tool

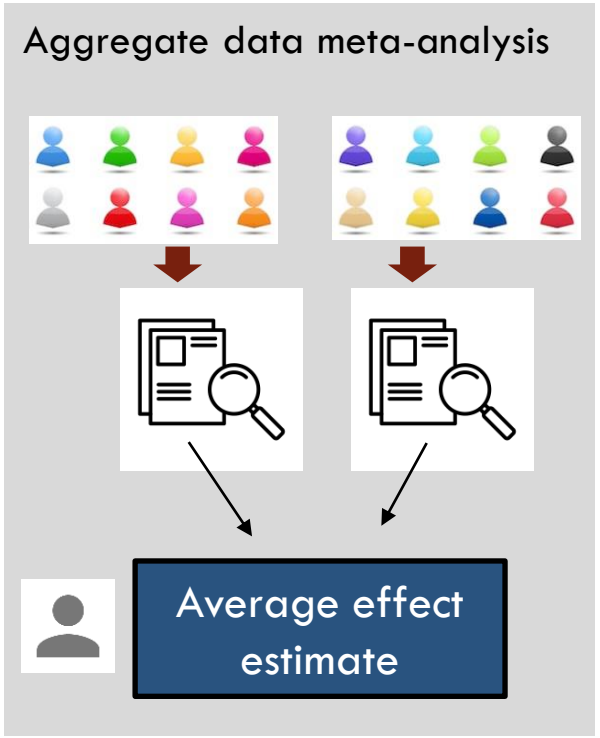
Hunter KE, Aberoumand M, Libesman S, Sotiropoulos JX, Williams JG, Aagerup J, Barba A, Shrestha N, Wang R, Mol BW, Li W, Webster AC, Seidler AL

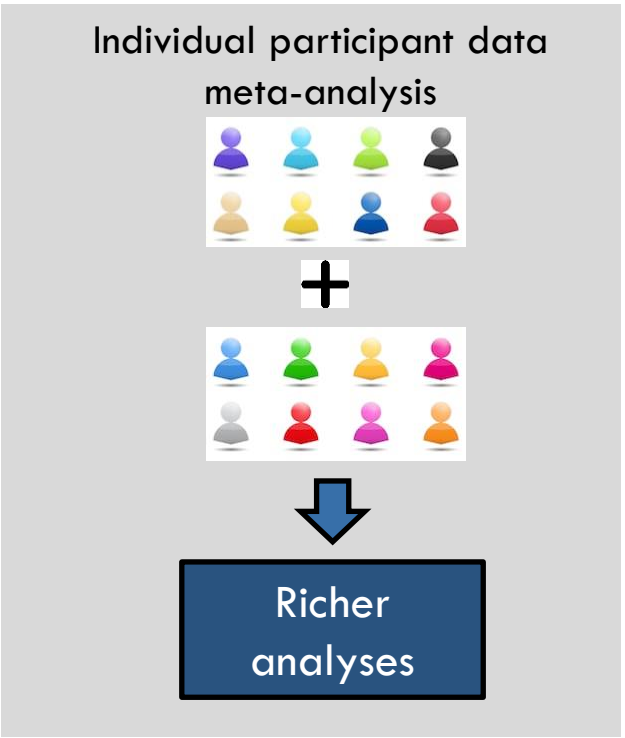
NextGen Evidence Synthesis Team, NHMRC Clinical Trials Centre, The University of Sydney, Australia

IPD
Integrity Tool

My background: individual participant data (IPD) meta-analyses

Individual Participant Data (IPD) meta-analysis involves the central collection of raw data for each participant in the original trials

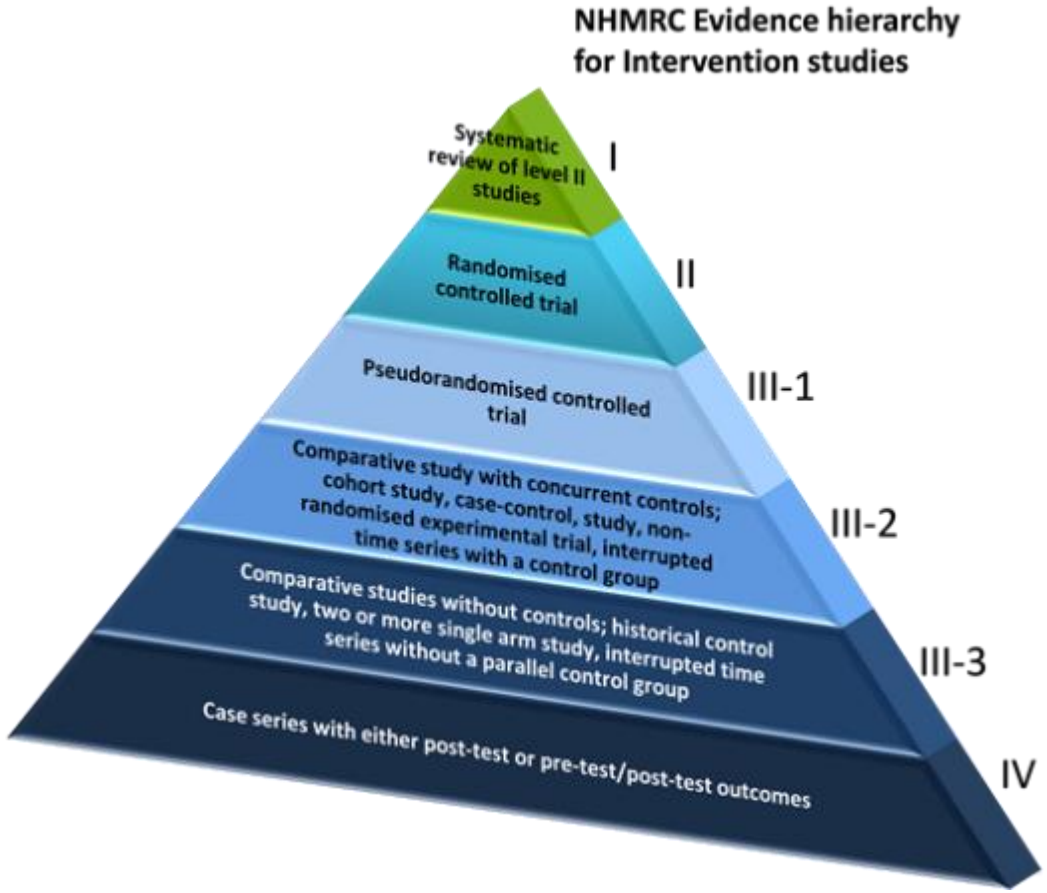




Systematic reviews at the top of the evidence hierarchy

Individual participant data meta-analyses: ‘gold standard’ for evidence synthesis

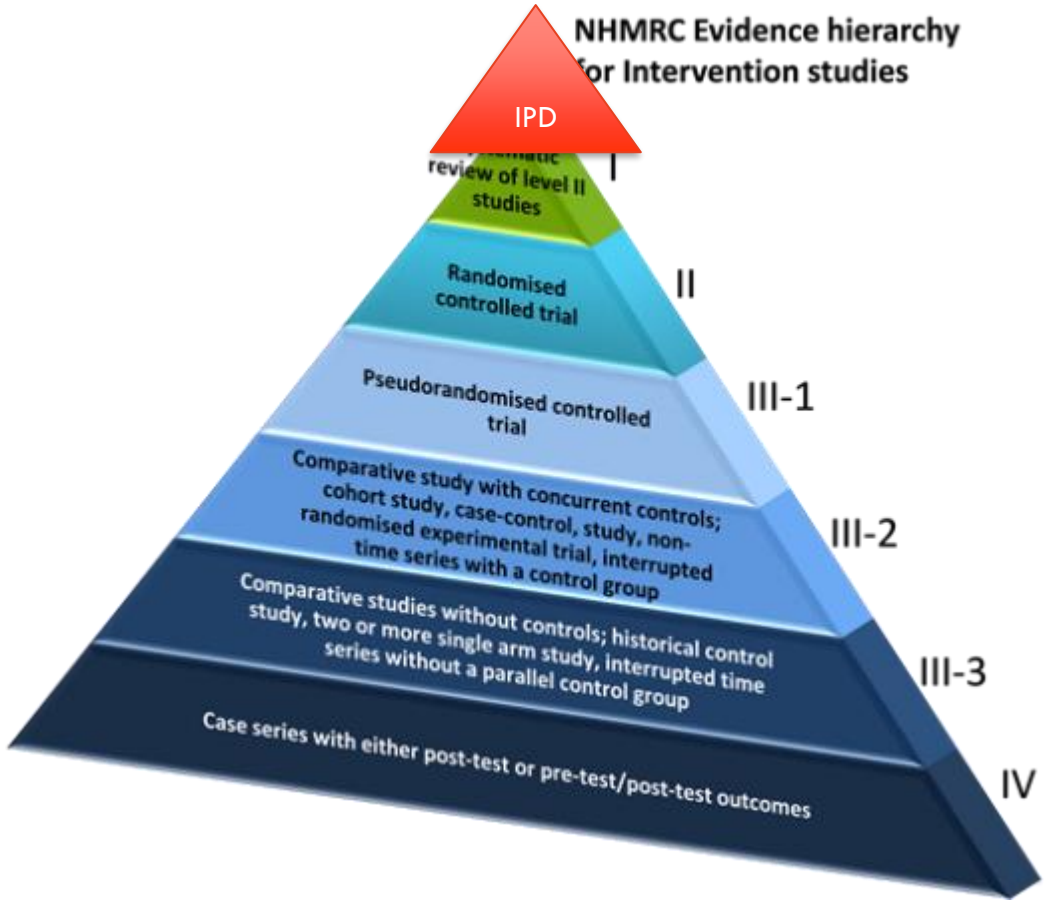
Widely used to inform healthcare policy and practice



Individual participant data meta-analyses at the top of the evidence hierarchy

Individual participant data meta-analyses: ‘gold standard’ for evidence synthesis

Widely used to inform healthcare policy and practice



Integrity crisis

NEWS | 12 December 2023

More than 10,000 research papers were retracted in 2023 – a new record

The number of articles being retracted rose sharply this year. Integrity experts say that this is only the tip of the iceberg.

By Richard Van Noorden

Fake academic papers are on the rise: why they're a danger and how to stop them

Published: March 7, 2024 12.51am AEDT

Fake academic articles can cause significant harm. Nora Carol Photography

Author

Lex Bouter
Professor of Methodology and Integrity, Vrije Universiteit Amsterdam

Disclosure statement

Lex Bouter is the founding chair of the World Conferences on Research Integrity Foundation and co-chair of the 8th WCRI in Athens, 2-5 June 2024.

Partners

View all partners

creative commons

We believe in the free flow of information

Republish our articles for free, online or in print, under Creative Commons licence.

Republish this article

'There's far more scientific fraud than anyone wants to admit

Ivan Oransky and Adam Marcus

Despite recent scandals of research misconduct and error, the academic world still seems determined to look the other way

Marc Tessier-Lavigne, the president of Stanford, announced in July that he would resign, after an independent review cleared him of research misconduct but found flaws in other papers

'The situation has become appalling': fake scientific papers push research credibility to crisis point

Last year, 10,000 sham papers had to be retracted by academic journals, but experts think this is just the tip of the iceberg

Fake research papers could jeopardise drug development, warn academics. Photograph:

NEWS FEATURE | 18 July 2023

Medicine is plagued by untrustworthy clinical trials. How many studies are faked or flawed?

Investigations suggest that, in some fields, at least one-quarter of clinical trials might be problematic or even entirely made up, warn some researchers. They urge stronger scrutiny.

By Richard Van Noorden

How do I assess integrity of trials in my individual participant data meta-analysis?

Emerging tools to assess integrity of studies – none for individual participant data!

> Most tools relate to aggregate data and/or publications


RESEARCH ARTICLE

Research Synthesis Methods WILEY

Identifying and managing problematic trials: A research integrity assessment tool for randomized controlled trials in evidence synthesis

Stephanie Weibel[1] | Maria Popp[1] | Stefanie Reis[1] | Nicole Skoetz[2] | Paul Garner[3] | Emma Sydenham[4]


ELSEVIER

Check for updates

Journal of Clinical Epidemiology 151 (2022) 1–17

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Experts identified warning signs of fraudulent research: a qualitative study to inform a screening tool

Lisa Parker[a], Stephanie Boughton[b], Rosa Lawrence[c], Lisa Bero[c,*]

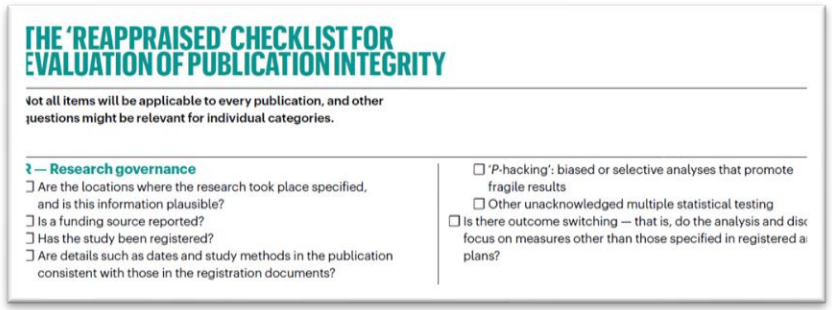
[a]School of Pharmacy, Charles Perkins Centre, The University of Sydney, New South Wales, Australia
[b]Evidence Production and Methods, Cochrane, UK


RESEARCH Open Access

Check for updates

Checklist to assess Trustworthiness in RAndomised Controlled Trials (TRACT checklist): concept proposal and pilot

Ben W. Mol[1,2], Shimona Lai[1], Ayesha Rahim[1], Esmée M. Bordewijk[3], Rui Wang[1], Rik van Eekelen[3,4], Lyle C. Gurrin[5*], Jim G. Thornton[6], Madelon van Wely[3,4,7] and Wentao Li[1]


THE 'REAPPRAISED' CHECKLIST FOR EVALUATION OF PUBLICATION INTEGRITY

Not all items will be applicable to every publication, and other questions might be relevant for individual categories.

? — Research governance

- ☐ Are the locations where the research took place specified, and is this information plausible?
- ☐ Is a funding source reported?
- ☐ Has the study been registered?
- ☐ Are details such as dates and study methods in the publication consistent with those in the registration documents?

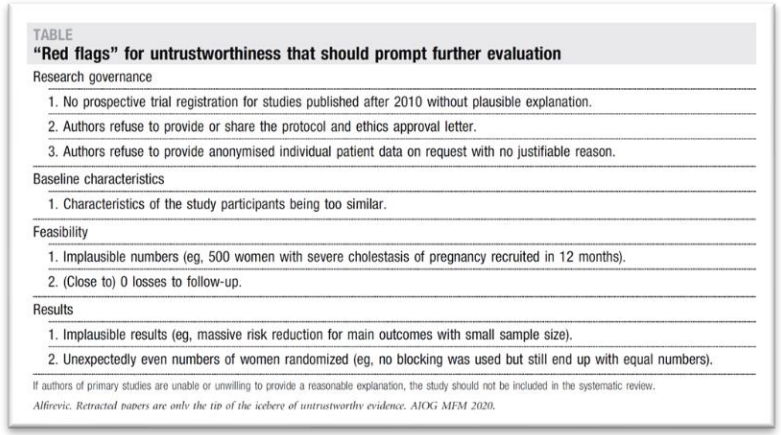
- ☐ 'P-hacking': biased or selective analyses that promote fragile results
- ☐ Other unacknowledged multiple statistical testing
- ☐ Is there outcome switching — that is, do the analysis and disc focus on measures other than those specified in registered a plans?


Cochrane Pregnancy and Childbirth

Version 2.4 (20 July 2021

Identifying and handling potentially untrustworthy trials in Pregnancy and Childbirth Cochrane Reviews

Alfirevic Z, Kellie FJ, Stewart F, Jones L, Hampson L, on behalf of Pregnancy and Childbirth Editorial Board



TABLE

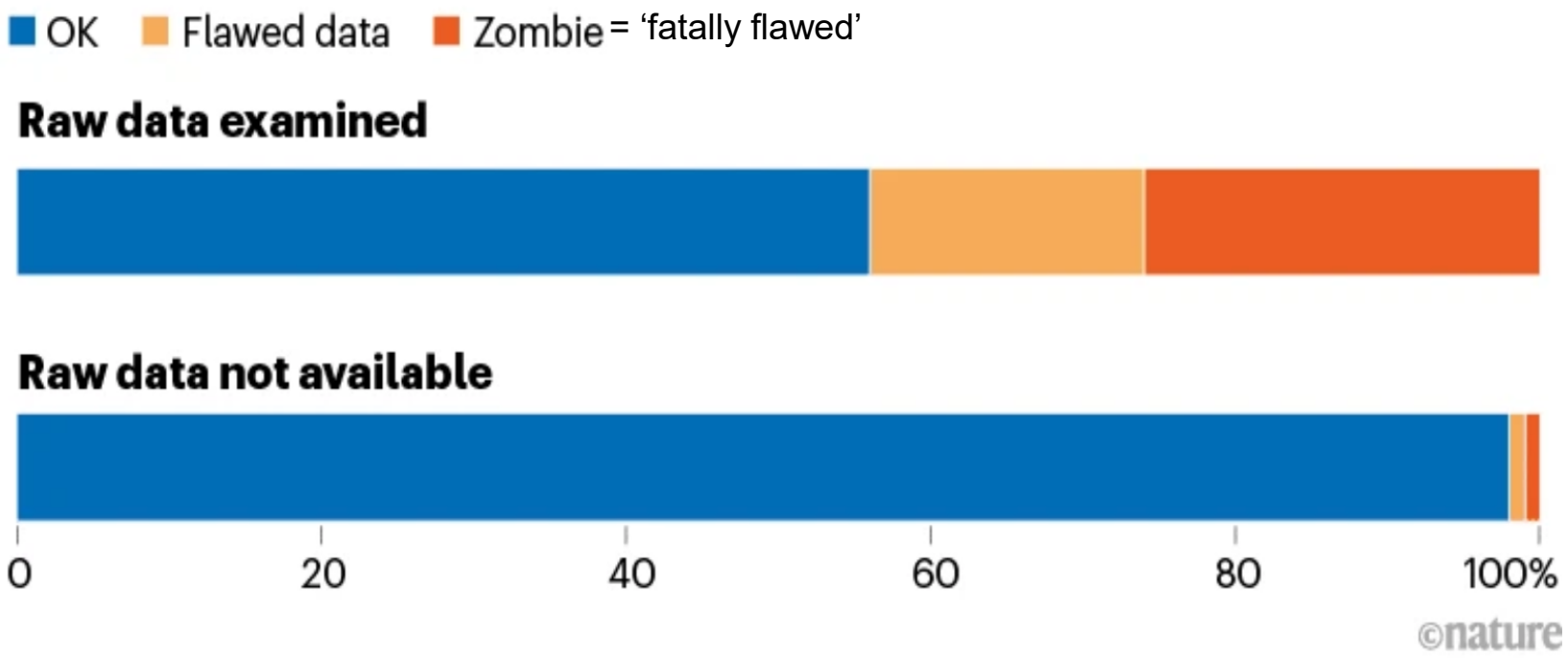
"Red flags" for untrustworthiness that should prompt further evaluation

Research governance
1. No prospective trial registration for studies published after 2010 without plausible explanation.
2. Authors refuse to provide or share the protocol and ethics approval letter.
3. Authors refuse to provide anonymised individual patient data on request with no justifiable reason.
Baseline characteristics
1. Characteristics of the study participants being too similar.
Feasibility
1. Implausible numbers (eg, 500 women with severe cholestasis of pregnancy recruited in 12 months).
2. (Close to) 0 losses to follow-up.
Results
1. Implausible results (eg, massive risk reduction for main outcomes with small sample size).
2. Unexpectedly even numbers of women randomized (eg, no blocking was used but still end up with equal numbers).

If authors of primary studies are unable or unwilling to provide a reasonable explanation, the study should not be included in the systematic review.

Alfirevic. Retracted papers are only the tip of the iceberg of untrustworthy evidence. AJOG MFM 2020.

The power of individual participant data (IPD)



Carlisle, J. B. *Anaesthesia* **76**, 472–479 (2021)

Need for IPD to detect integrity issues

Check for updates

correspondence

The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable

To the Editor — The global demand for purported timelines that are not consistent ... cur. Any study for

'We recommend that meta-analysts who study interventions (…) **should request and personally review IPD in all cases…**

Any study for which authors are **not able or not willing to provide suitably anonymized IPD** should be considered at high risk of bias for incomplete reporting and/or **excluded entirely from meta-syntheses**.'

Lawrence et al, Nature Medicine 2021

Aim

To develop an individual participant data meta-analysis (IPD-MA) integrity tool

iCOMP

Systematic review and network meta-analysis with **i**ndividual participant data on **Co**rd **M**anagement at **P**reterm Birth

THE LANCET

Volume 402, Issue 10418, 9–15 December 2023, Pages 2209-2222

Articles

Deferred cord clamping, cord milking, and immediate cord clamping at preterm birth: a systematic review and individual participant data meta-analysis

Anna Lene Seidler PhD [a], Mason Aberoumand MAppStat [a], Kylie E Hunter MPH [a], Angie Barba MSciMed [a], Sol Libesman PhD [a], Jonathan G Williams PhD [a], Nipun Shrestha PhD [a], Jannik Aagerup MPH [a], James X Sotiropoulos MD [a], Prof Alan A Montgomery PhD [b], Prof Gillian M L Gyte MPhil [c], Prof Lelia Duley MD [b] *, Prof Lisa M Askie PhD [a] *
iCOMP Collaborators†

Show more

THE LANCET

Volume 402, Issue 10418, 9–15 December 2023, Pages 2223-2234

Articles

Short, medium, and long deferral of umbilical cord clamping compared with umbilical cord milking and immediate clamping at preterm birth: a systematic review and network meta-analysis with individual participant data

Anna Lene Seidler PhD [a], Sol Libesman PhD [a], Kylie E Hunter MPH [a], Angie Barba MSciMed [a], Mason Aberoumand MAppStat [a], Jonathan G Williams PhD [a], Nipun Shrestha PhD [a], Jannik Aagerup MPH [a], James X Sotiropoulos MD [a], Prof Alan A Montgomery PhD [b], Gillian M L Gyte MPhil [c], Prof Lelia Duley MD [b] *, Prof Lisa M Askie PhD [a] *
iCOMP Collaborators†

Methods

Literature review & mapping of existing items

Expert advisory group discussion

Checklist v1

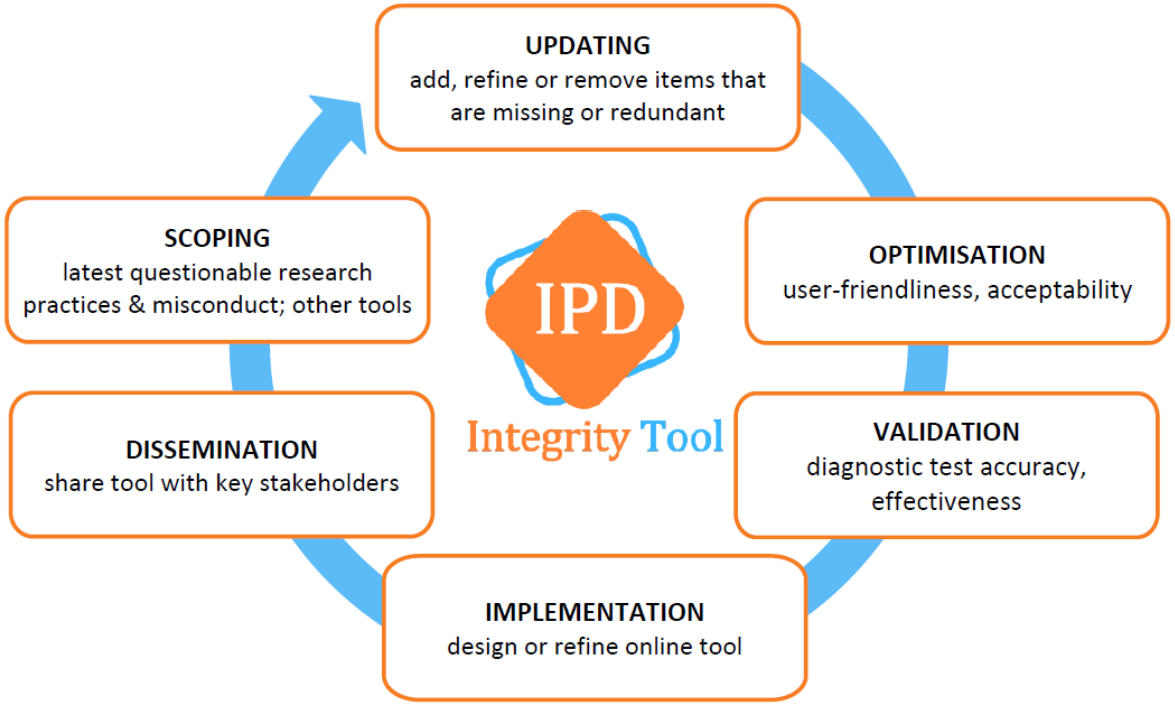
Pilot implementation iCOMP & NETMOTION (73 trials with IPD)

Preliminary validation (13 datasets with and without known integrity issues)

Evaluation

Checklist v2

Testing (TOPCHILD)



The IPD Integrity Tool: for assessing the trustworthiness of randomised trials using IPD

Components of the tool

Rating sheet:
Use to indicate rating for each item (no issues, some/minor issues, many/major issues)

Instructions & decision guide:
explains how to assess each item and provides decision rules to guide rating process

Automation script:
Template script to generate R Markdown report which semi-automates assessment of some items

IPD Integrity Tool

Overview of tool domains

Randomised controlled trials (RCTs)

Aggregate data level integrity domains

1. Retraction notices & expressions of concern
2. Provision of IPD
3. Communication
4. Ethics approval
5. Trial registration / protocol
6. Randomisation
7. Plausibility

Individual participant data (IPD) integrity domains

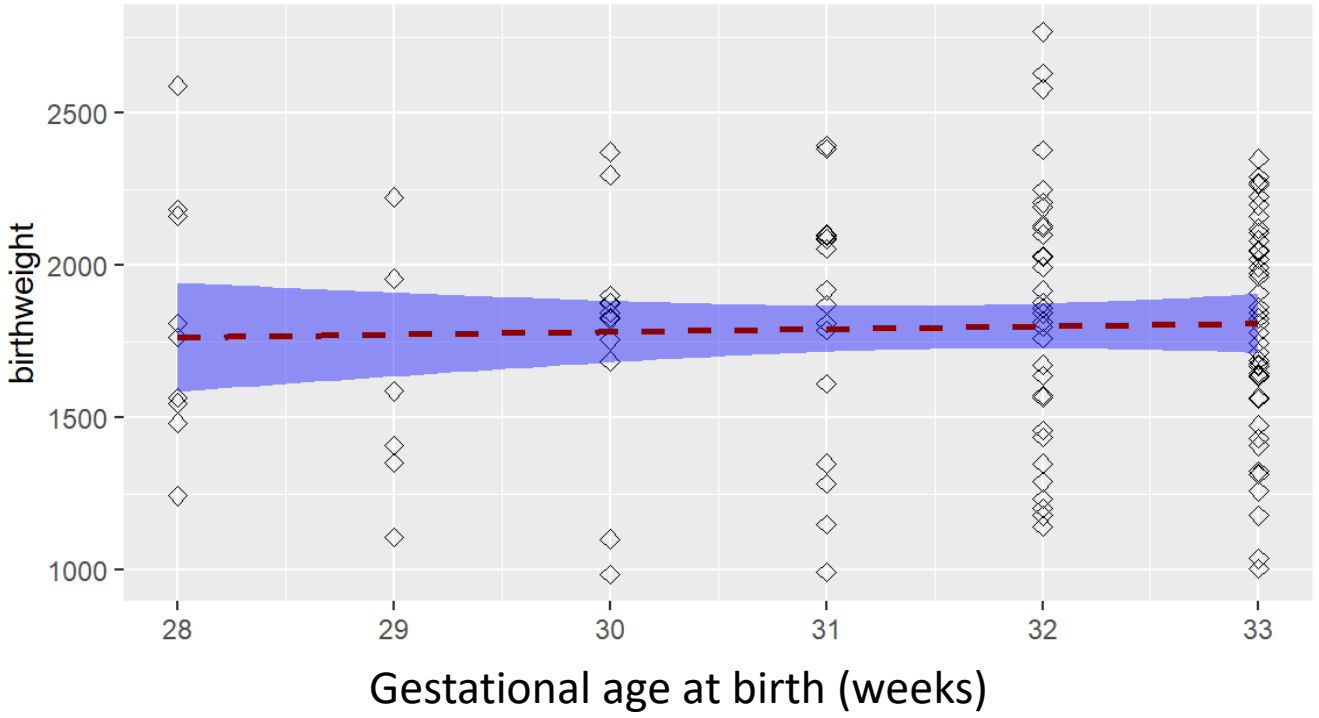
ID	variable1	variable2	...
1	24	1	
2	28	1	
...			

1. Unusual or repeated data patterns
2. Baseline characteristics
3. Correlations
4. Date violations
5. Patterns of allocation
6. Internal inconsistencies
7. External inconsistencies
8. Plausibility of data

Domain 3: Correlations

Are expected correlations present?

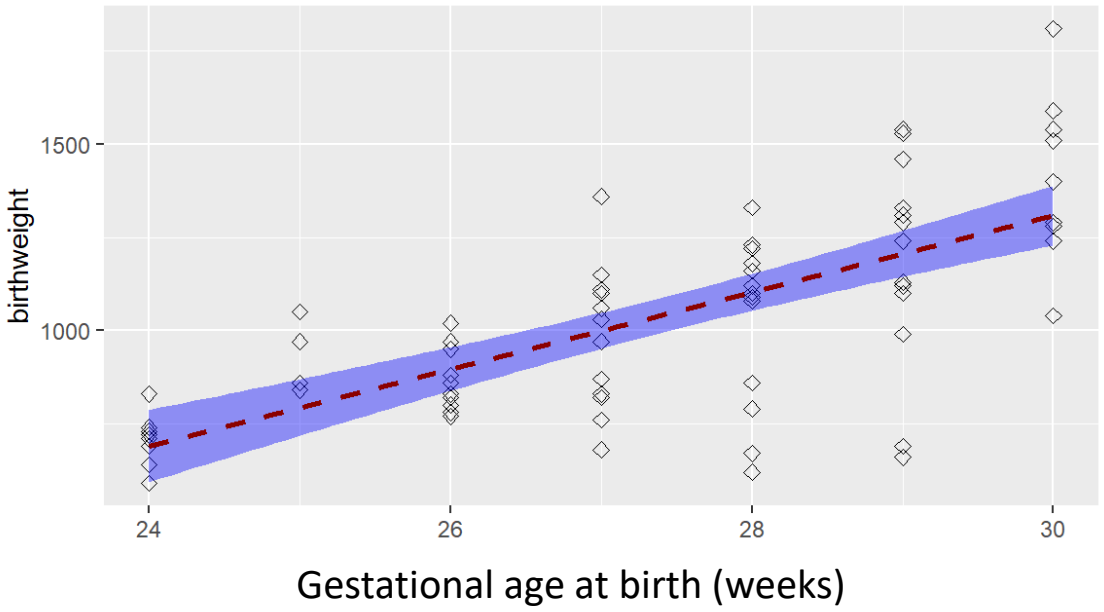
Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
•Correlation between variables is as expected	•Correlations appear too weak or too strong, or are in the wrong direction	•No association between variables known to be highly correlated	




```
Pearson correlation estimate: 0.04
```

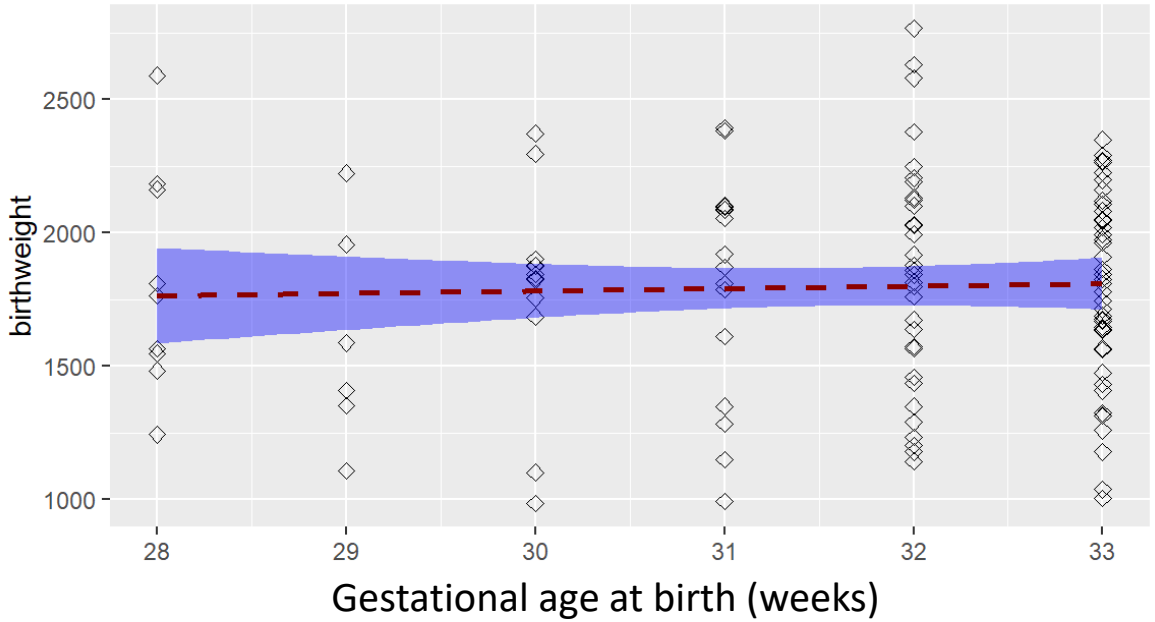
Are expected correlations present?

Trial A. Expected correlation present




```
Pearson correlation estimate: 0.7
```

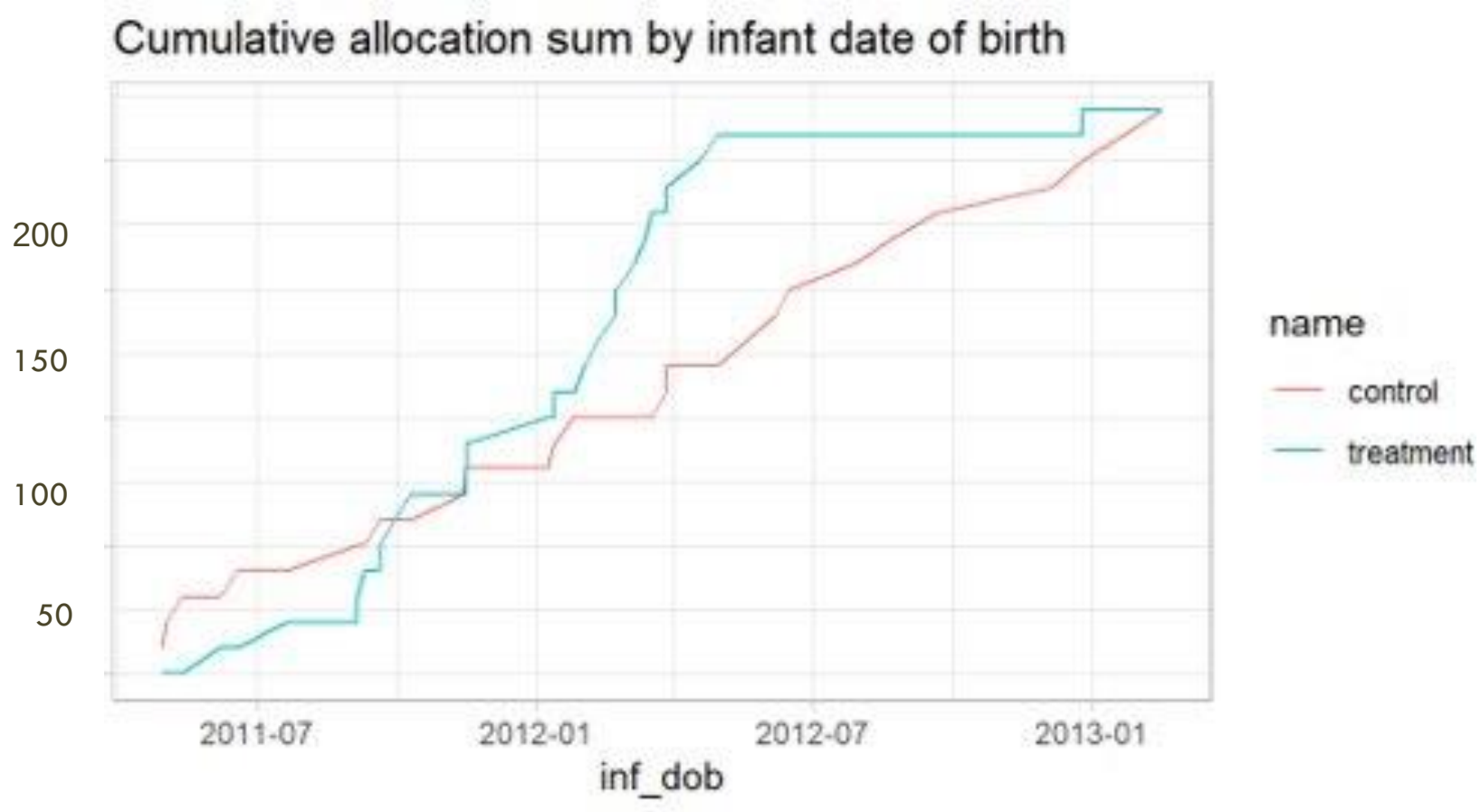
Trial B. Expected correlation NOT present



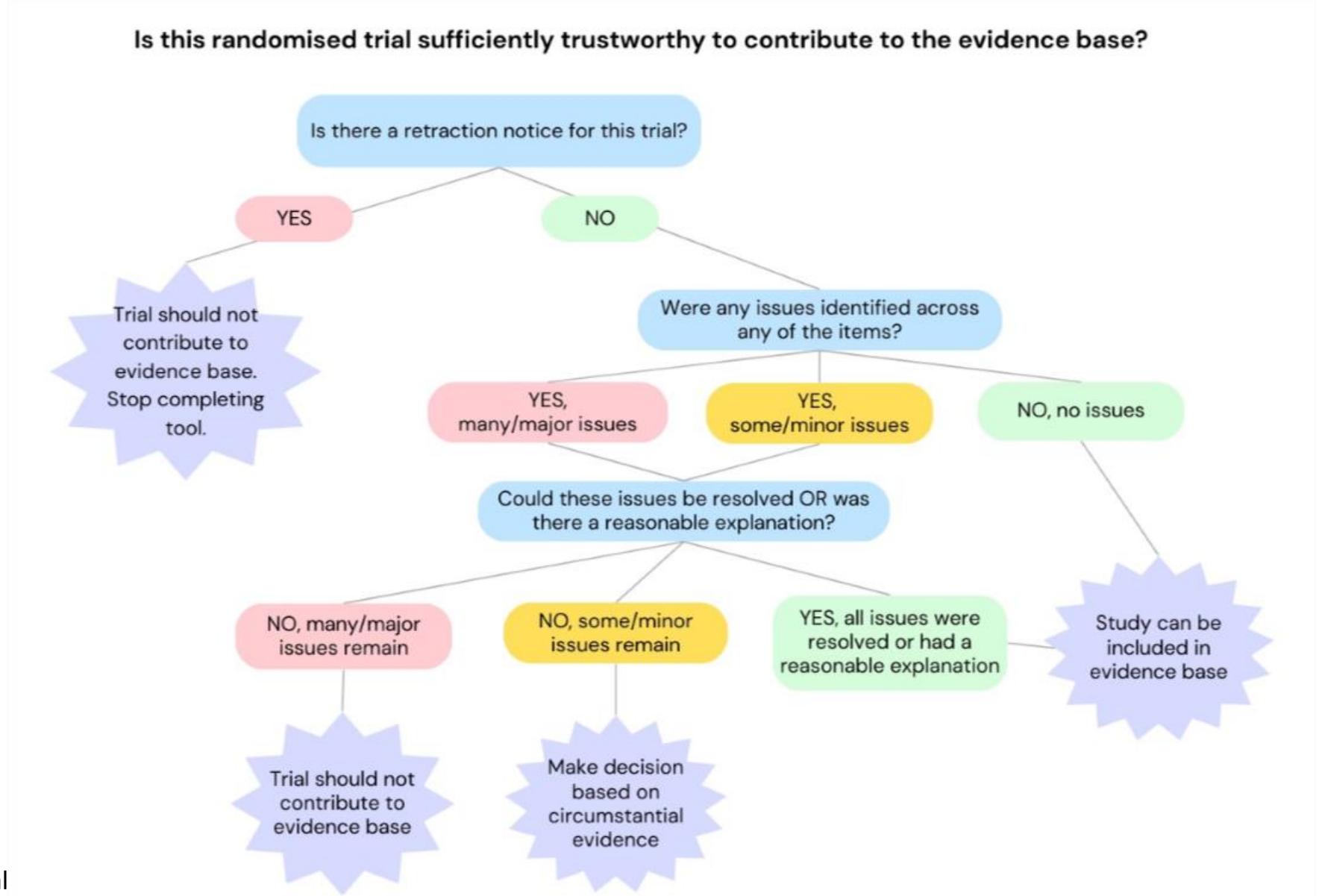

```
Pearson correlation estimate: 0.04
```

Domain 5: Patterns of allocation

Is randomisation appropriate?



Overall assessment: decision-making process



Case study

iCOMP

- 58/64 trials contributing IPD had at least one potential integrity issue identified – mostly minor inconsistencies or errors that were resolved via consultation.
- 3/64 IPD trials **excluded** due to integrity concerns

Conclusion

– The IPD Integrity Tool enables users to assess the integrity of RCTs via examination of IPD

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

Follow this preprint

Development of the Individual Participant Data (IPD) Integrity Tool for assessing the integrity of randomised trials using individual participant data

KE Hunter, M Aberoumand, S Libesman, JX Sotiropoulos, J Williams, W Li, J Aagerup, BW Mol, R Wang, A Barba, N Shrestha, AC Webster, AL Seidler

doi: https://doi.org/10.1101/2023.12.11.23299797

2 manuscripts close to publication

1) Development of tool
2) Instructions on how to use tool

Get in touch if you would like to access our tool!

When to use the IPD Integrity Tool

Scenario	Who uses the tool	What tool is used for
1.Individual participant data meta-analysis (where IPD are available for all or some trials)	IPD-MA project team	Guides decision on whether to include a trial in meta-analysis
2.Questionable trial identified during conduct of aggregate data meta-analysis and IPD are requested to assess trustworthiness	AD-MA project team	Guides decision on whether to include a trial in meta-analysis
3.Questionable study submitted for publication and IPD are requested by editors to assess trustworthiness	Journal editors	Guides decision on whether to consider a manuscript for publication
4.Trustworthiness concerns raised about a published study, and IPD are requested by editors to investigate	Journal editors	Guides decision on whether to retract a publication or issue an expression of concern
5.Routine IPD checks for editors to screen submitted trials	Journal editors	Guides decision on whether to consider a manuscript for publication

The

Open questions

- How to deal with untrustworthy studies in a collaboration?
- Threshold for data exclusion? How strict should we be?
- The role of Artificial Intelligence in data fabrication?

Thank you!

lene.seidler@sydney.edu.au
@LeneSeidler

Kylie Hunter
Kylie.hunter@sydney.edu.au

THE UNIVERSITY OF SYDNEY

NHMRC Clinical Trials Centre

The IPD Integrity Tool

Components of the tool

Instructions & decision guide: explains how to assess each item and provides decision rules to guide rating process

Rating sheet: Use to indicate rating for each item (no issues, some/minor issues, many/major issues)

Automation script: Template script to generate R Markdown report which semi-automates assessment of some items

IPD Integrity Tool

Individual participant data level integrity checks

Note: *(R) denotes assessments that may be semi-automated using the R markdown template*

Integrity domain and items	How to assess	Response options			Exceptions: may downgrade severity of issue(s)
		No issues	Some/minor issue(s)	Many/major issue(s)	
1. Unusual or repeated data patterns					
1.1 Repeating patterns *within* baseline variables	Sort and visually assess the data for repeating patterns within baseline variables. Assess in dataset order, randomisation order, and also separately for study groups	- No repeating data patterns identified	- Some repeating data patterns identified, but may be consistent with chance	- Repeating data patterns identified that are extremely unlikely to have occurred by chance, e.g. trialist copy and pasted every 10 rows	- Poor granularity of measures and rounding may lead to repetition of values, e.g. age rounded to years with narrow eligibility range
1.2 Repeating data patterns *across* baseline variables	Sort each baseline variable from smallest to largest, and look for patterns across variables ***R markdown***	- No repeating data patterns identified	- Some repeating data patterns identified, but may be consistent with chance or plausible correlation between variables	- Repeating data patterns identified that are extremely unlikely to have occurred by chance, e.g. all newborns with a length of 30cm have identical birthweight	- Poor granularity of measures and rounding may lead to repetition of values, e.g. when gestational age at birth is rounded to weeks, and birthweight is rounded to the nearest 500g
1.3 Repeating data patterns across baseline variables *and rare variables*	As above, but focus on repetition across any rare variables present in dataset ***R markdown***	- No repeating data patterns identified	- Some repeating data patterns identified, but may be consistent with chance	- Repeating data patterns identified that are extremely unlikely to have occurred by chance, e.g. all children who suffer an adverse event have the same sex, birthweight and age at enrolment	
1.4 Bias in the terminal (rightmost) digits	Plot and examine bar charts of the terminal digit for select continuous variables (avoid variables that tend to be rounded or that lack precision)	- Terminal digits follow a uniform or expected distribution	- Biased or non-uniform distribution of terminal digits	- Extremely biased or unexpected distribution of terminal digits - Conspicuous absence of a single digit across a large	- Poor granularity of measures, e.g. broad categorisation of continuous measures or use of less precise measurement instruments

Integrity domain and items	Rating (no issues; some/minor issues; many/major issues)	Justification for rating
Aggregate data/publication-level checks		
1. Retraction notices and expressions of concern		
1.1 Retraction notice – study of interest	Select	
1.2 Retraction notice(s) – other study/ies by same authors	Select	
1.3 Expression of concern (EOC) – study of interest	Select	
1.4 Expression of concern – other study/ies by same authors	Select	
2. Provision of individual participant data (IPD)		
2.1 IPD not available or not provided on request	Select	
3. Communication		
3.1 Lack of trialist engagement in communication (see also domain 2. Provision of IPD)	Select	
4. Ethics approval		
4.1 Absent or inadequate ethics approval	Select	
5. Trial registration / protocol		
5.1 Absent or retrospective trial registration +/- publicly available protocol	Select	
6. Randomisation		
6.1 Randomisation - baseline balance/imbalance across groups	Select	
7. Plausibility		
7.1 Implausible recruitment rate	Select	
7.2 Implausible follow-up	Select	
7.3 Implausible results	Select	
7.4 Implausible author group	Select	
OVERALL JUDGEMENT - aggregate data/publication-level checks	Select	
Individual participant data checks		
1. Unusual or repeated data patterns		
1.1 Repeating data patterns *within* baseline variables	Select	
1.2 Repeating data patterns *across* baseline variables	Select	
1.3 Repeating data patterns across baseline variables *and rare variables*	Select	
1.4 Bias in the terminal (rightmost) digits	Select	
2. Baseline characteristics		
2.1 Excessively homogeneous distribution of binary baseline variables, i.e. loss of independence or serial correlation across consecutive observations	Select	
2.2 Excessive imbalances between groups in continuous baseline variables	Select	
2.3 Excessive imbalances in baseline categorical variables between groups	Select	
2.4 Significant difference in variance of continuous baseline variables between groups	Select	
3. Correlations		
3.1 No association between variables known to be highly correlated	Select	
4. Date violations		
4.1 Individual enrolment dates do not fit within study start and end dates	Select	
4.2 Dates (or visits) are not in logical order	Select	

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Appendix 2b_IPD-Integrity-Tool_R-mar...

Source Visual

```
---
title: "IPD Integrity Tool vignette"
author: "NextGen Evidence Synthesis Team, NHMRC Clinical Trials Centre, University of Sydney
output:
  word_document: default
  html_document:
    df_print: paged
  pdf_document: default
editor_options:
  markdown:
    wrap: 72
---
 <br>
 This document provides an example R markdown output or 'vignette' for instructive and illus

### OVERVIEW <br>
Step 1: Load packages  <br>

Step 2: Load dataset  <br>

Step 3: Format the data  <br>

Step 4: Select and label relevant variables  <br>

Step 5: Run integrity tests  <br>



### Step 1: Load packages
Run the code in the 'LOAD PACKAGES' chunk. No changes required.

```{r LOAD PACKAGES,include=FALSE}

### install the packages below if they are not already installed
list.of.packages <-
  c(
    "tidyverse",
    "lubridate",
    "readr",
    "viridis",
```

The NHMRC Clinical Trials Centre, The University of Sydney
```

Domain 1: Unusual or repeated data patterns

WHAT: Scrutinise data for repeating patterns within and across baseline variables and rare variables, terminal digit bias

WHY: generating truly random numbers is very difficult for humans

HOW TO ASSESS: Are there repeating data patterns that are extremely unlikely to have occurred by chance?

iCOMP commentary

"the highest standards for a meta-analysis"

"sophisticated and validated statistical methods" to identify possible falsified data, that "has not been common in meta-analysis and should set a new standard"

Data patterns

Repeating patterns *within* baseline variables

infant_id	birthweight (grams)
1	1940
2	2500
3	2100
4	1850
5	2450
6	1940
7	2500
8	2100
9	1850
10	2450
11	1940
12	2500
13	2100
14	1850
15	2450
16	1940
17	2500
18	2100
19	1850
20	2450

Sheldrick K, "Seven signs of fraud in individual participant data". *NSW Health Statewide Biobank Seminar Series*, Oct 2021.

Carlisle JB, "False individual patient data and zombie randomised controlled trials submitted to Anaesthesia". *Anaesthesia* 2021, 76:472-9.

Data patterns

Repeating patterns *within* baseline variables

infant_id	birthweight (grams)
1	1940
2	2500
3	2100
4	1850
5	2450
6	1940
7	2500
8	2100
9	1850
10	2450
11	1940
12	2500
13	2100
14	1850
15	2450
16	1940
17	2500
18	2100
19	1850
20	2450

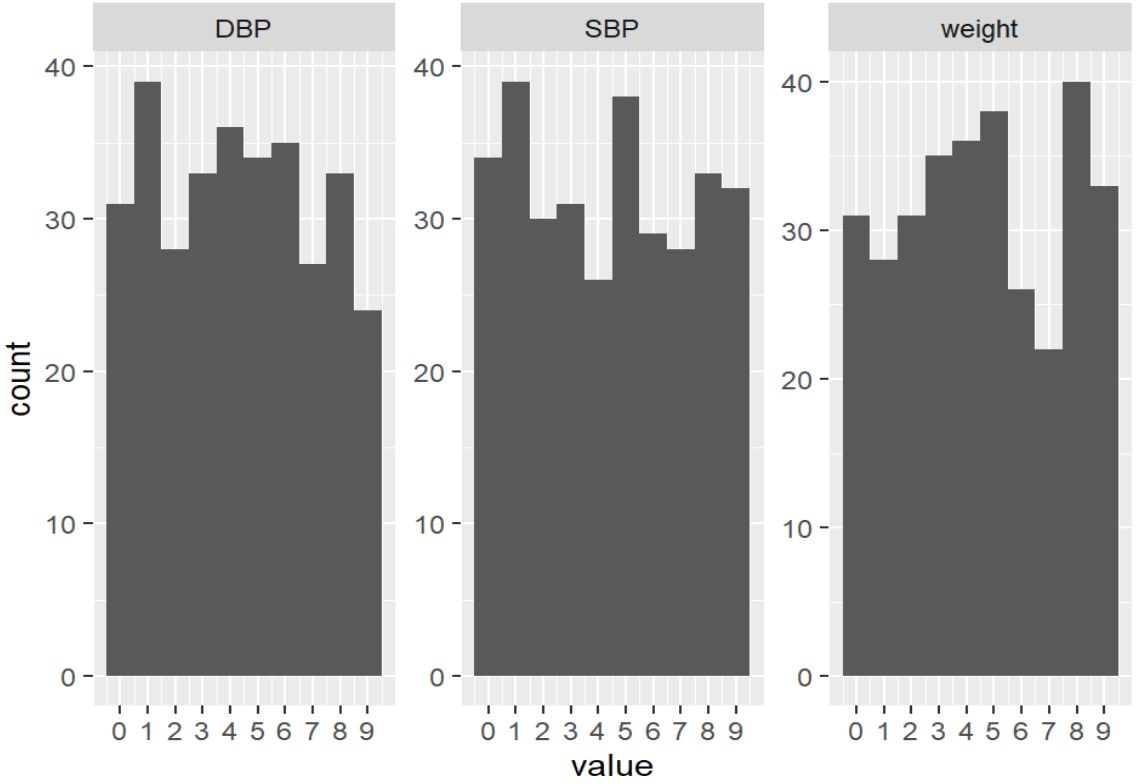
Sheldrick K, "Seven signs of fraud in individual participant data". *NSW Health Statewide Biobank Seminar Series*, Oct 2021.

Carlisle JB, "False individual patient data and zombie randomised controlled trials submitted to Anaesthesia". *Anaesthesia* 2021, 76:472-9.

1.4 Unusual or repeated data patterns: terminal digit bias

Do the plots appear to follow the expected distribution?

Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
- Terminal digits follow a uniform or expected distribution	- Biased or non-uniform distribution of terminal digits	- Extremely biased or unexpected distribution of terminal digits - Conspicuous absence of a single digit across a large number of observations	- Poor granularity of measures, e.g. broad categorisation of continuous measures or use of less precise measurement instruments



DBP = diastolic blood pressure, SBP = systolic blood pressure

Domain 2: Baseline characteristics

WHAT: look for excessively different or excessively similar baseline characteristics between groups that are implausible or beyond what is expected by chance

WHY: Generally, in RCTs, baseline characteristics such as age and sex should be balanced between groups, albeit perfect balance is unrealistic.

particularly important for prognostic factors which may influence outcomes

HOW TO ASSESS: statistical tests

2.1 Excessively homogeneous distribution of binary baseline variables, i.e. loss of independence or serial correlation across consecutive observations

If group allocation is genuinely random, we would not expect a participant's baseline values to be dependent on the previous participant. It is difficult to fabricate a dataset to match expected variation in values. The *Wald-Wolfowitz runs test* examines whether baseline data occurs in a random manner based on row order (if organised chronologically).

var	runs	n1	n2	n	statistic	p.value	method	alternative
Diabetes	85	50	270	320	-0.080	0.936	Runs Test	nonrandomness
Smoking	147	168	152	320	-1.527	0.127	Runs Test	nonrandomness

Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
- No significant p values, i.e. all ≥0.05	- One significant p value (i.e. <0.05)	- Multiple significant p values (i.e. <0.05)	- Variable(s) with significant p values have a low rate of occurrence, i.e. are rare

Domain 3: Correlations

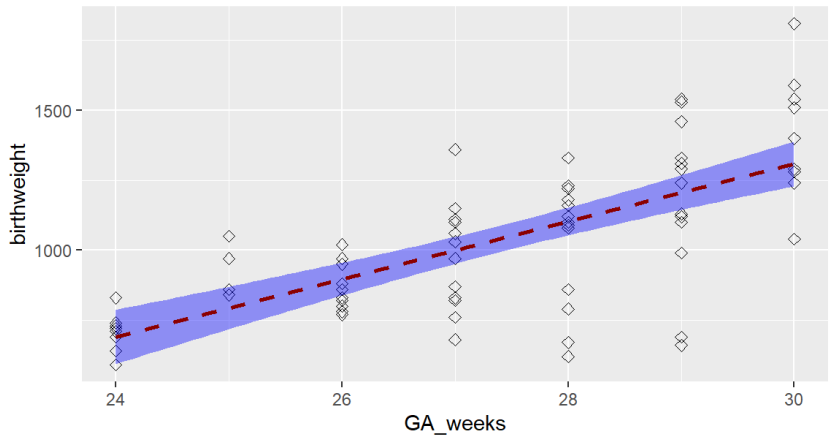
WHAT: examines whether expected relationships between variables are present, e.g. we would expect a child's height to increase with age

WHY: Lack of expected correlations may suggest fabricated data

HOW TO ASSESS: Plot and assess two or three known correlations. Assessment requires contextual knowledge and clinical expertise in the area of study.

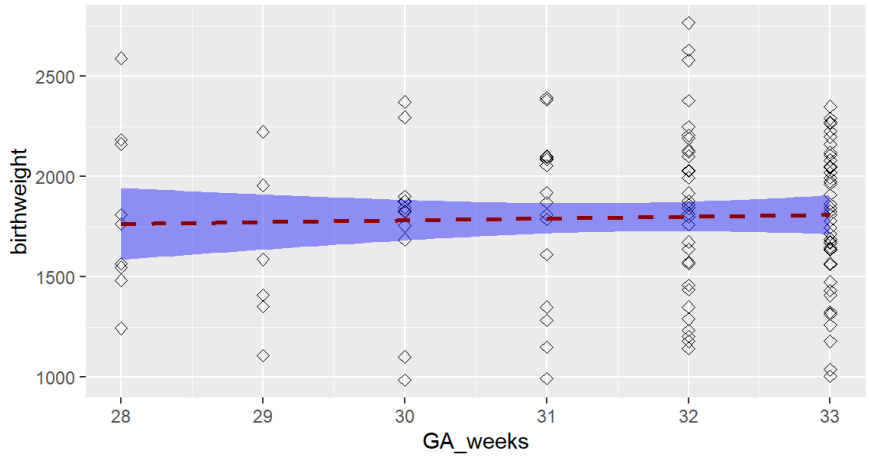
Are expected correlations present?

Trial A. Expected correlation present



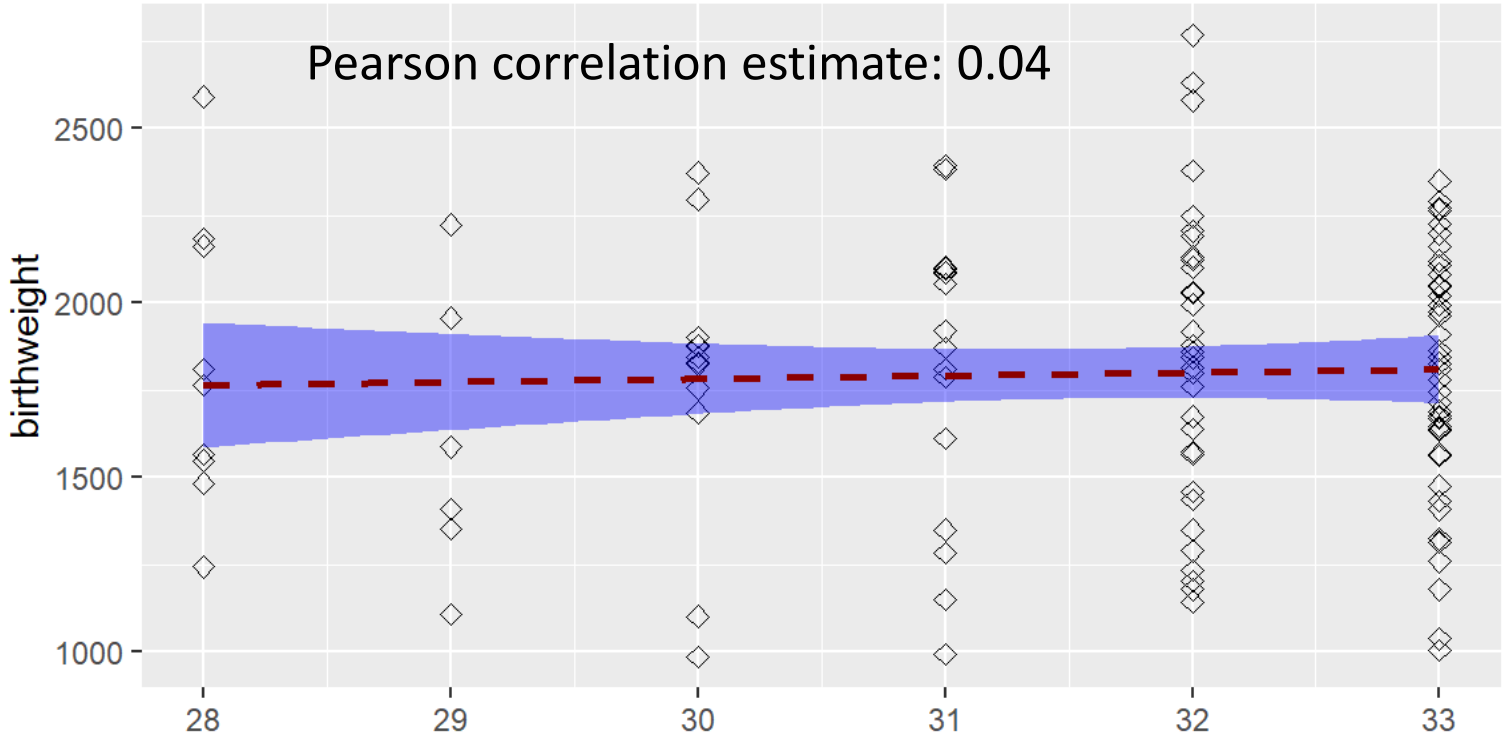

```
Pearson correlation estimate: 0.7
```

Trial B. Expected correlation NOT present




```
Pearson correlation estimate: 0.04
```

3.1 No association between variables known to be highly correlated



Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
- Correlation between variables is as expected	- Correlations appear too weak or too strong, or are in the wrong direction	- No association between variables known to be highly correlated	

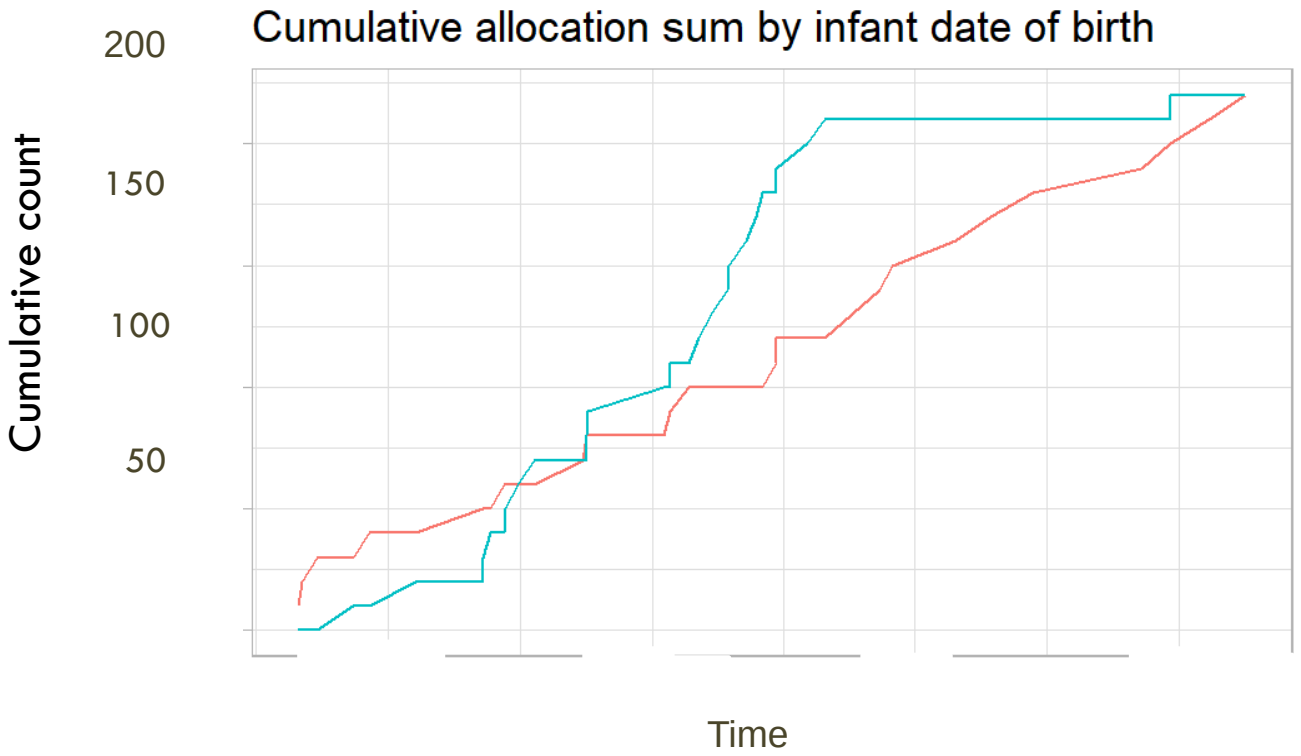
Domain 4: Date violations

WHAT: Date violations describe impossible dates e.g. recruitment outside the recruitment window, a participant's second visit occurred before the first.

WHY: may arise inadvertently or be indicative of integrity violations

HOW TO ASSESS: Check whether dates occur in logical order. Compare the start and end date of each study with individual enrolment dates (may be obtained from publications, trial registration records, or by direct contact with trialists)

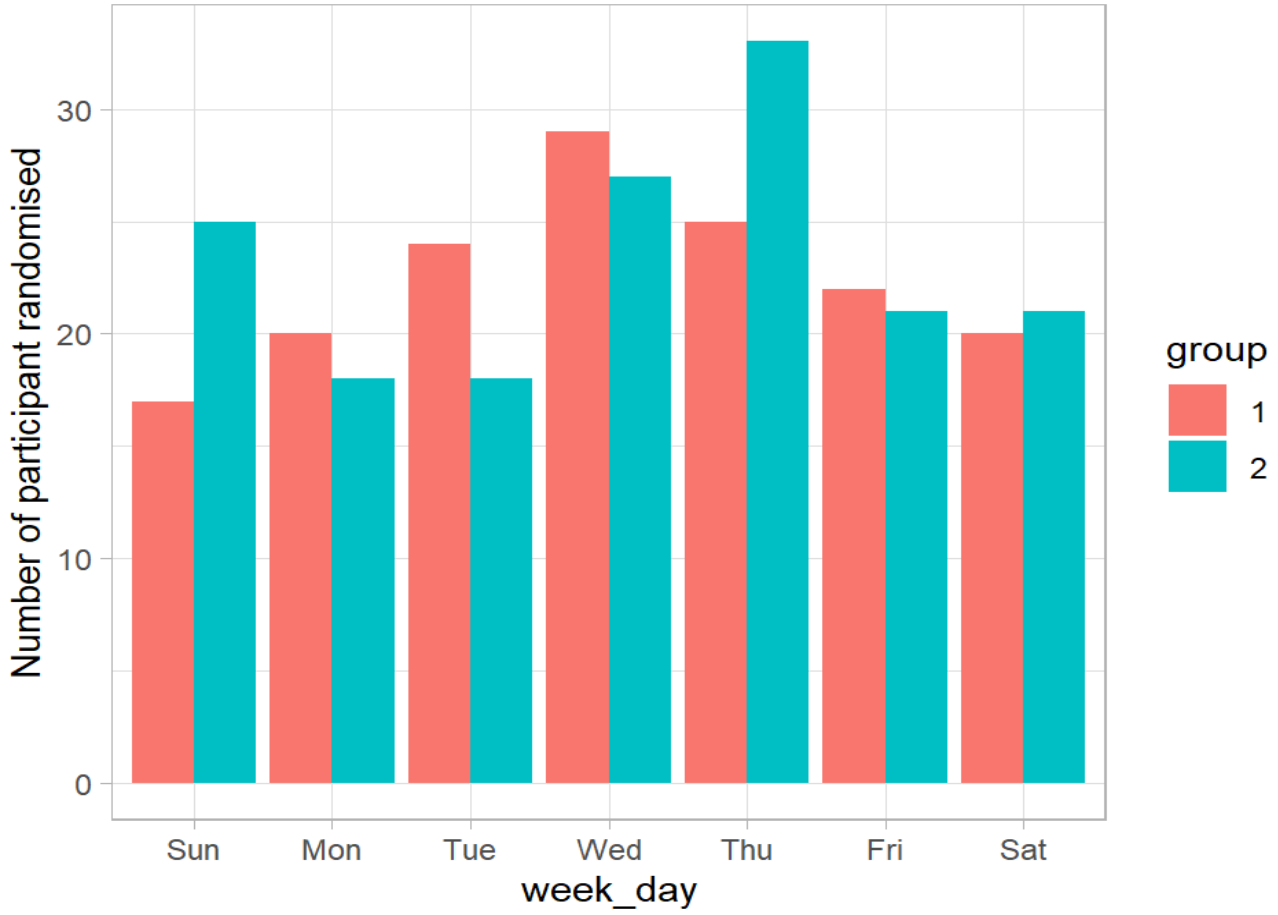
5. Non-random allocation patterns - plot



Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
- Similar numbers in each group and plotted curves do not deviate from each other drastically (1:1 allocation). - If allocation is not 1:1, we would expect curves to track one another but not cross.	-	- Plotted curves deviate drastically from each other	- Smaller trials may have greater separation in curves and less crossing over - Minimisation, blocked or cluster randomisation methods may explain the pattern of sequence generation

5.3 Item 5.3 - Unexpected imbalance in randomisation day of week

Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
- Uniform distribution across groups for each week day, and fewer enrolments on weekends for non-urgent interventions	- Obvious deviations from what is expected, e.g. no participants enrolled on Wednesdays	-	- For urgent interventions, enrolments on weekends may be expected - Trial staff only available on certain days



Domain 6: Internal inconsistencies

WHAT: inconsistent or illogical values across variables within individual participants

WHY: several large or obvious inconsistencies within a dataset may raise doubts about the reliability of the data.

HOW TO ASSESS: Derive logic rules for each variable to be collected, e.g. date of hospital discharge = date of admission + days in hospital; incorporate these rules into statistical checks

Domain 7: External inconsistencies

WHAT: discrepancies between a trial's IPD and published reports

WHY: Several or large unexplained discrepancies raise concerns about the validity and trustworthiness of the data.

HOW TO ASSESS: Plot all variables provided in the IPD dataset and tabulate summary statistics for each, e.g. mean, median, range, etc. Cross-check these against any published trial reports, including appendices and supplements.

Domain 8: Plausibility

WHAT: reasonableness of missing data and event rates

WHY: No or relatively few missing data should trigger concern in most cases (depending on follow-up times and sample size), as should identical missing values across groups; or extreme event rates (particularly for rare adverse events)

HOW TO ASSESS: Check missing values, compare event rates with expected rates based on literature, setting, biological mechanisms, and expert advice..

Domain 8: Plausibility

Which of these are questionable?

Example 1: Intense exercise intervention, 0.5% missing data at 1 year follow up (n=500)

Example 2: In hospital mortality of patients admitted with COVID (n=40, no missings)

Response options			Exceptions: may downgrade severity of issue(s)
No issues	**Some/minor issue(s)**	**Many/major issue(s)**	
•No/few/minor inconsistencies that can often be resolved with trialist	•Implausibly few missing data compared to expected •Identical missing values across groups	•No missing data	•(Close to) 100% follow-up may be achieved for outcomes assessed immediately after intervention delivery

Overall assessment

	How to assess	No concerns	Some concerns	Major concerns
OVERALL ASSESSMENT	Provide an overall rating based on all items	No issues identified, OR any issues adequately resolved or had a reasonable explanation **The study may be considered sufficiently trustworthy to contribute to the evidence base, i.e.. to include in meta-analysis, or to be considered for publication**	≥1 minor issue identified that could not be adequately resolved and had no reasonable explanation **Decision on how to proceed should be based on circumstantial evidence or pending further information**	≥1 major issue identified that cannot be adequately resolved or had a reasonable explanation **The study should NOT be considered trustworthy enough to contribute to the evidence base, i.e. do NOT include in meta-analysis or consider for publication**

IPD – integrity issues

- All studies had multiple integrity issues
- Many issues required individual participant data to detect