Too good to be true? Spotting fake trials in the literature

Jack Wilkinson, Centre for Biostatistics, University of Manchester.

@jdwilko.bsky.social

Some of the research discussed in this presentation is funded by the NIHR Research for Patient Benefit programme (NIHR203568, NIHR303741). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

FUNDED BY

NIHR | National Institute for Health and Care Research

Disclaimer

I am not accusing anyone of fraud or other forms of misconduct. A study might be ‘problematic’ or ‘untrustworthy’ due to honest errors.

Declaration of interest

I am Stats Editor for BJOG, Cochrane Gynaecology and Fertility, Fertility and Sterility, Journal of Hypertension. I undertake integrity investigations for various journals, and sometimes I am paid for this.

Systematic reviews of health interventions

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

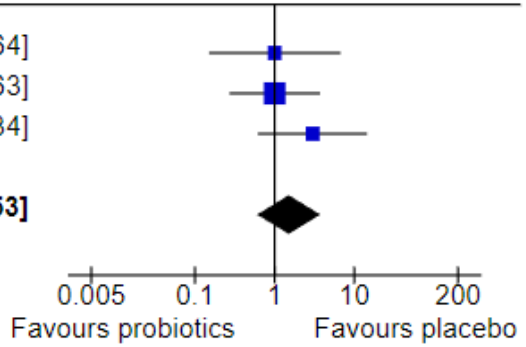
https://doi.org/10.1002/14651858.CD012970.pub2

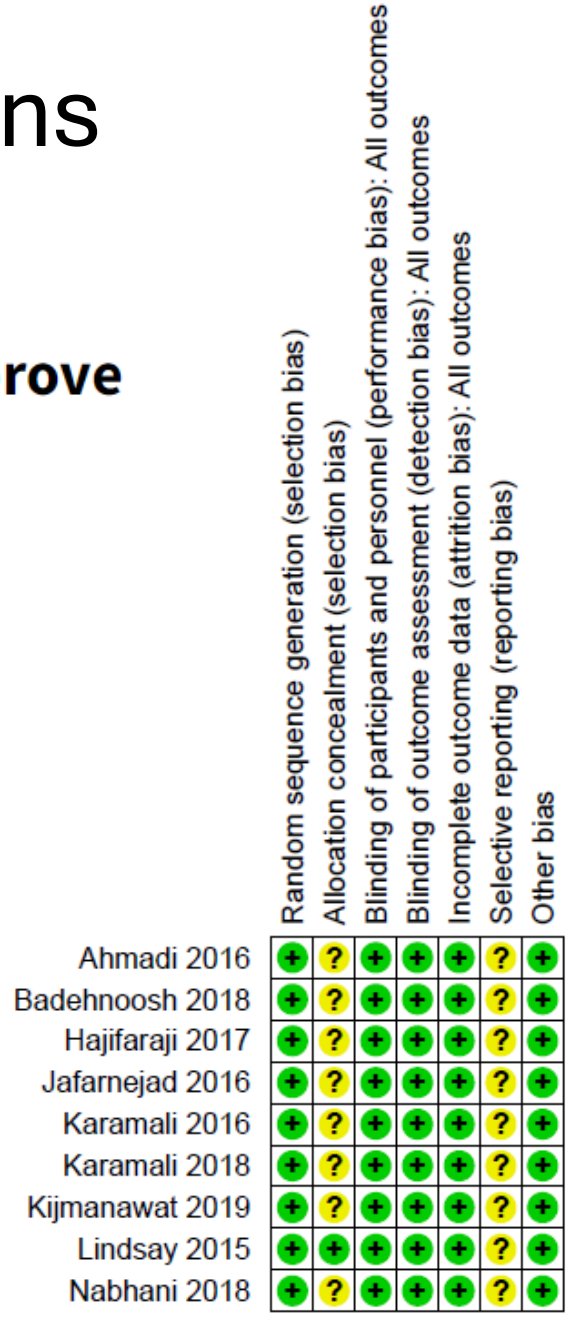
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





Systematic reviews of health interventions

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

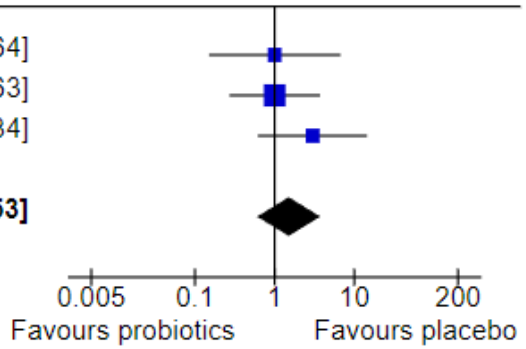
https://doi.org/10.1002/14651858.CD012970.pub2

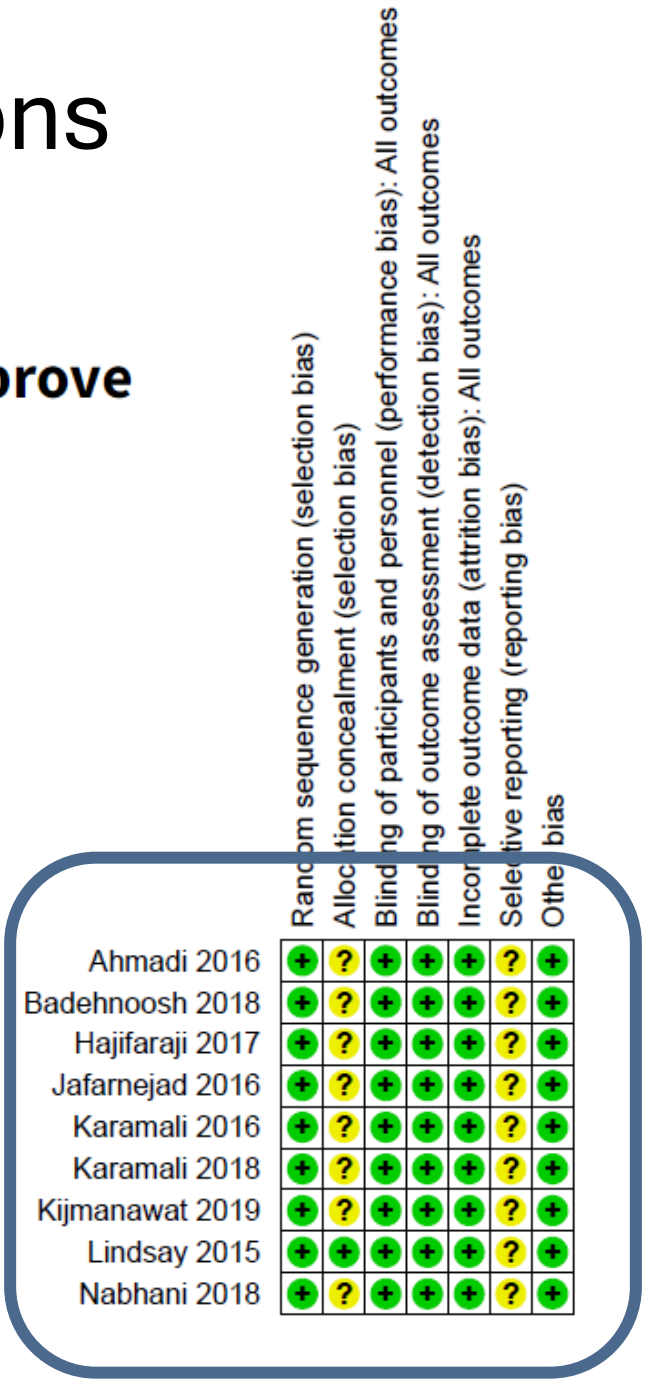
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





Systematic reviews of health interventions

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

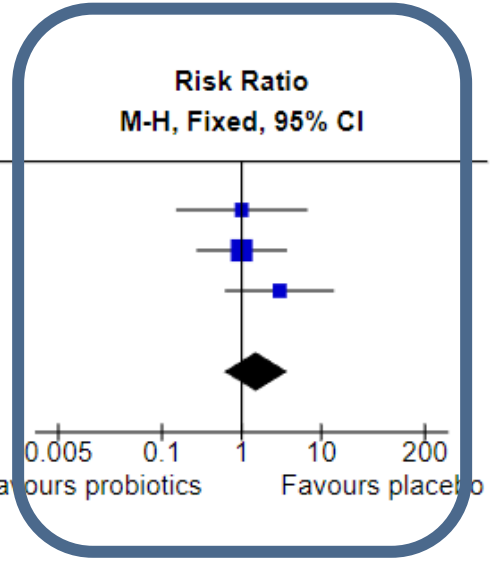
https://doi.org/10.1002/14651858.CD012970.pub2

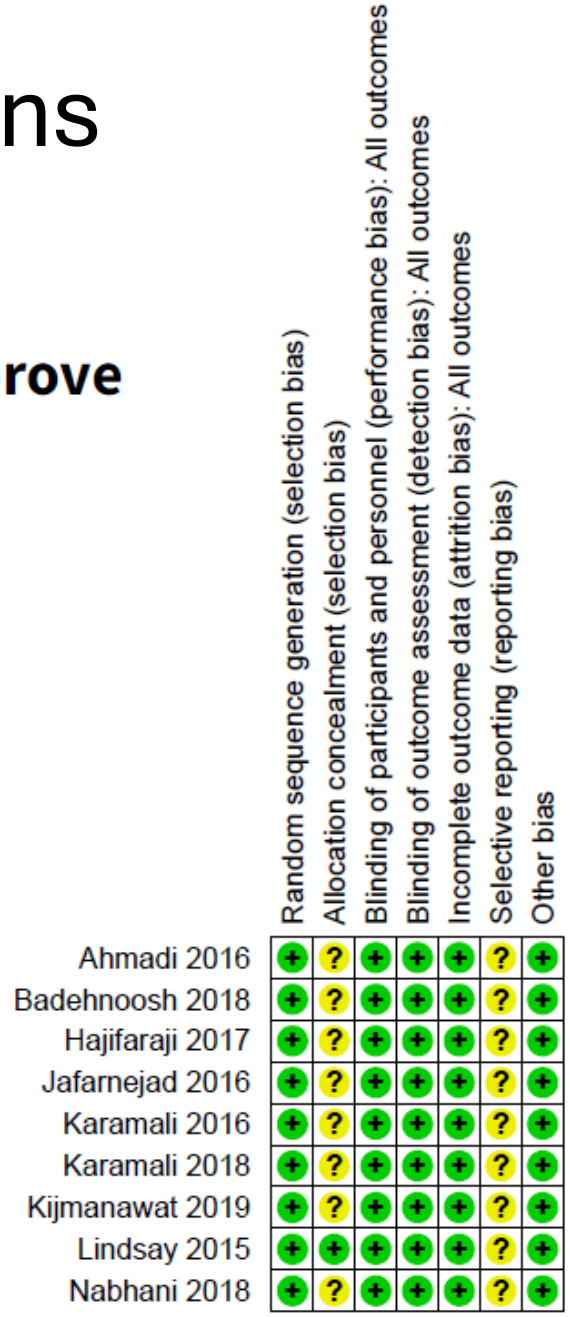
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





Vitamin K and the Prevention of Fractures

Systematic Review and Meta-analysis of Randomized Controlled Trials

Sarah Cockayne, MSc; Joy Adamson, PhD; Susan Lanham-New, PhD; Martin J. Shearer, PhD, MRCPath; Simon Gilbody, DPhil; David J. Torgerson, PhD

Ivermectin for Prevention and Treatment of COVID-19 Infection: A Systematic Review, Meta-analysis, and Trial Sequential Analysis to Inform Clinical Guidelines

Bryant, Andrew MSc[1,*]; Lawrie, Theresa A. MBBCh, PhD[2]; Dowswell, Therese PhD[2]; Fordham, Edmund J. PhD[2]; Mitchell, Scott MBChB, MRCS[3]; Hill, Sarah R. PhD[1]; Tham, Tony C. MD, FRCP[4]

Psychological therapies for the management of chronic pain (excluding headache) in adults (Review)

Williams ACDC, Fisher E, Hearn L, Eccleston C

Vitamin K and the Prevention of Fractures

Systematic Review and Meta-analysis of Randomized Controlled Trials

Sarah Cockayne, MSc; Joy Adamson, PhD; Susan Lanham-New, PhD; Martin J. Shearer, PhD, MRCPath; Simon Gilbody, DPhil; David J. Torgerson, PhD

3 out of 5 trials subsequently identified as **fake**.

Ivermectin for Prevention and Treatment of COVID-19 Infection: A Systematic Review, Meta-analysis, and Trial Sequential Analysis to Inform Clinical Guidelines

Bryant, Andrew MSc[1,*]; Lawrie, Theresa A. MBBCh, PhD[2]; Dowswell, Therese PhD[2]; Fordham, Edmund J. PhD[2]; Mitchell, Scott MBChB, MRCS[3]; Hill, Sarah R. PhD[1]; Tham, Tony C. MD, FRCP[4]

Suggested impressive benefit on mortality due to **fraudulent** trials.

Psychological therapies for the management of chronic pain (excluding headache) in adults (Review)

Williams ACDC, Fisher E, Hearn L, Eccleston C

3 trials from one investigator suggested to be **implausible** (huge effects, no attrition), 2 **retracted**.

Systematic reviews: Fake data to patient care pipeline

1

Attempt to identify all RCTs on the review topic

- Problematic trials will be included

2

Critically appraise study methodology, include in meta-analysis

- Assess risk of bias
- But do not consider authenticity
- Many (not all) fake trials report sound methods

3

Make conclusions, recommendations, on basis of evidence

- SRs seen as gold standard
- Included in guidelines
- Influence patient care

How big is the problem?

Difficult to estimate!

Carlisle 2020: 26% RCTs submitted to Anaesthesia “critically flawed by false data”.

Gopalakrishna 2022: 10% of researchers in life/ medical sciences have committed fabrication or falsification.

Xu 2025: Identified 3902 meta-analyses in 847 systematic reviews synthesising retracted trials.

Cochrane Library
Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

EDITORIAL

When beauty is but skin deep: dealing with problematic studies in systematic reviews

Stephanie L Boughton, Jack Wilkinson, Lisa Bero

Managing potentially problematic studies

https://bit.ly/3SsJO9F

Cochrane Library
Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

EDITORIAL

When beauty is but skin deep: dealing with problematic studies in systematic reviews

Stephanie L Boughton, Jack Wilkinson, Lisa Bero

Managing potentially problematic studies

https://bit.ly/3SsJO9F

- Do not include problematic studies in the systematic review.

Cochrane Library
Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

EDITORIAL

When beauty is but skin deep: dealing with problematic studies in systematic reviews

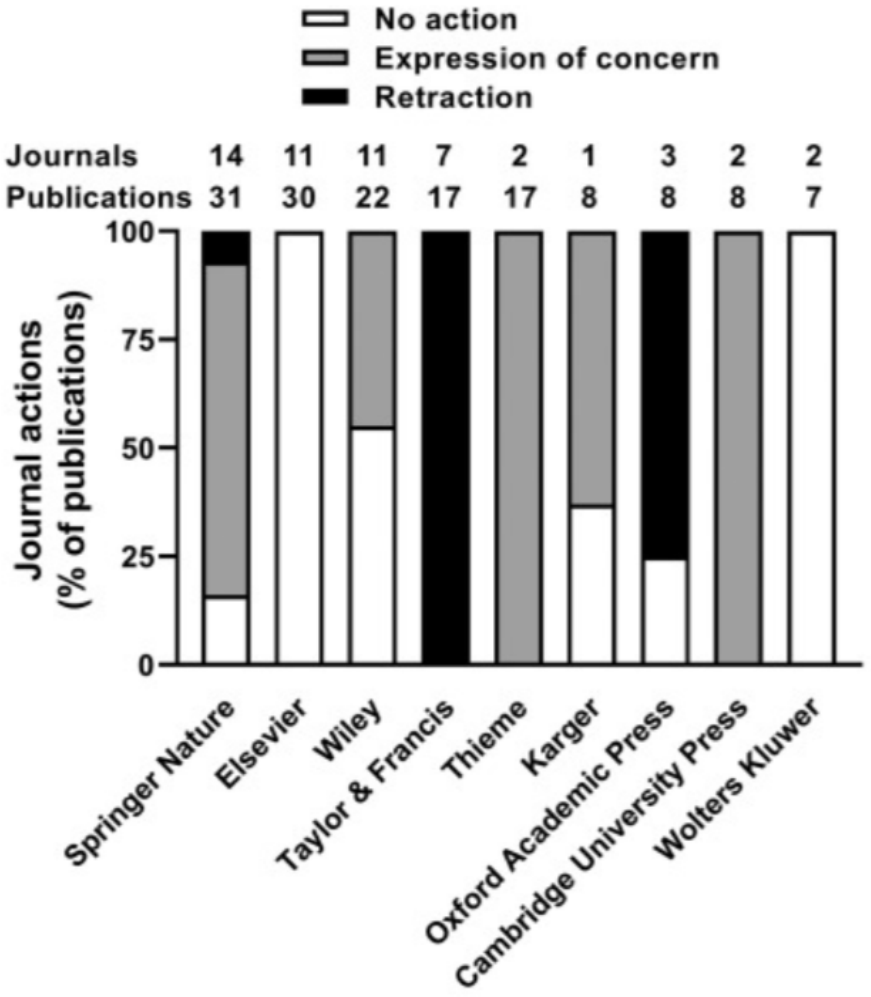
Stephanie L Boughton, Jack Wilkinson, Lisa Bero

Managing potentially problematic studies

https://bit.ly/3SsJO9F

- Do not include problematic studies in the systematic review.
- But how can we identify problematic studies in the first place?

Can’t we just use retraction status?



It typically takes a long time for problematic studies to be retracted

Many are never retracted

Example (left): outcome 5 years after reporting 172 problematic clinical trials from one research group

So we cannot rely on retraction status alone as an indicator of trustworthiness

https://doi.org/10.1016/j.jclinepi.2025.111918

Do we need a tool to identify fraud and misconduct?

Our view: it is neither necessary nor desirable to develop a “diagnostic test for fraud”

Legal consequences, harms

What matters it the trustworthiness of a study (example)

A trustworthiness tool should help the user make a judgement about a study’s trustworthiness, and to articulate a basis for that judgement

Wilkinson and Tovey 2025, JCE: doi.org/10.1016/j.jclinepi.2025.111670

Table I. Patients' characteristics.

	Group A (minimal stimulation) ($n=58$)	Group B (CC) ($n=55$)	p Value
Age (years)	27.7 ± 2.4	26.8 ± 2.3	0.57
Duration of infertility(years)	1.88 ± 0.7	2.14 ± 0.8	0.71
BMI (kg/m^2)	24.9 ± 2.3	25.3 ± 2.6	0.62
Waist-to-hip ratio	0.8 ± 0.04	0.8 ± 0.05	0.38
Menstrual cycle			
Oligomenorrhea	54 (93.1%)	51 (92.7%)	0.43
Amenorrhea	4 (6.9%)	4 (7.3%)	0.61
Hyperandrogenism	26 (44.8%)	23 (41.8%)	0.74
LH (mIU/mL)	12.9 ± 2.3	13.4 ± 2.4	0.52
FSH (mIU/mL)	5.6 ± 1.1	5.7 ± 1.3	0.72
LH/FSH ratio	2.6 ± 1.1	2.6 ± 1.2	0..42
Testosterone (ng/mL)	0.94 ± 0.21	0.92 ± 0.24	0.77
SHBG (nmol/L)	33.62 ± 3.6	35.53 ± 3.1	0.48
Fasting glucose (mg/dL)	91.6 ± 1.5	89.7 ± 1.4	0.76
Fasting insulin (μU/mL)	10.7 ± 2.7	11.2 ± 3.1	0.57
Fasting glucose/ insulin ratio	7.7 ± 3.7	7.2 ± 3.6	0.33
Ovarian volume (mL)	11.3 ± 2.5	11.4 ± 2.4	0.22

BMI, Body mass index; CC, clomiphene citrate; FSH, follicle-stimulating hormone;

Table I. Patients' characteristics.

	Group A (minimal stimulation) ($n=58$)	Group B (CC) ($n=55$)	p Value
Age (years)	27.7 ± 2.4	26.8 ± 2.3	0.57
Duration of infertility(years)	1.88 ± 0.7	2.14 ± 0.8	0.71
BMI (kg/m^2)	24.9 ± 2.3	25.3 ± 2.6	0.62
Waist-to-hip ratio	0.8 ± 0.04	0.8 ± 0.05	0.38
Menstrual cycle			
Oligomenorrhea	54 (93.1%)	51 (92.7%)	0.43
Amenorrhea	4 (6.9%)	4 (7.3%)	0.61
Hyperandrogenism	26 (44.8%)	23 (41.8%)	0.74
LH (mIU/mL)	12.9 ± 2.3	13.4 ± 2.4	0.52
FSH (mIU/mL)	5.6 ± 1.1	5.7 ± 1.3	0.72
LH/FSH ratio	2.6 ± 1.1	2.6 ± 1.2	0..42
Testosterone (ng/mL)	0.94 ± 0.21	0.92 ± 0.24	0.77
SHBG (nmol/L)	33.62 ± 3.6	35.53 ± 3.1	0.48
Fasting glucose (mg/dL)	91.6 ± 1.5	89.7 ± 1.4	0.76
Fasting insulin (μU/mL)	10.7 ± 2.7	11.2 ± 3.1	0.57
Fasting glucose/ insulin ratio	7.7 ± 3.7	7.2 ± 3.6	0.33
Ovarian volume (mL)	11.3 ± 2.5	11.4 ± 2.4	0.22

BMI, Body mass index; CC, clomiphene citrate; FSH, follicle-stimulating hormone;

Do we need a tool to identify fraud and misconduct?

Our view: it is neither necessary nor desirable to develop a “diagnostic test for fraud”

Legal consequences, harms

What matters it the trustworthiness of a study (example)

A trustworthiness tool should help the user make a judgement about a study’s trustworthiness, and to articulate a basis for that judgement

Wilkinson and Tovey 2025, JCE: doi.org/10.1016/j.jclinepi.2025.111670

INSPECT SR

Cochrane

FUNDED BY

NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.

INSPECT SR

Cochrane

FUNDED BY
NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.

Stage 1: Assemble list of checks for problematic studies (previous studies, new survey of 71 experts):
76 checks identified

INSPECT SR

Cochrane

FUNDED BY
NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.

Stage 1: Assemble list of checks for problematic studies (previous studies, new survey of 71 experts): **76 checks identified**

Stage 2: Apply list of checks to 95 RCTs in 50 Cochrane Reviews (feasibility, impact)

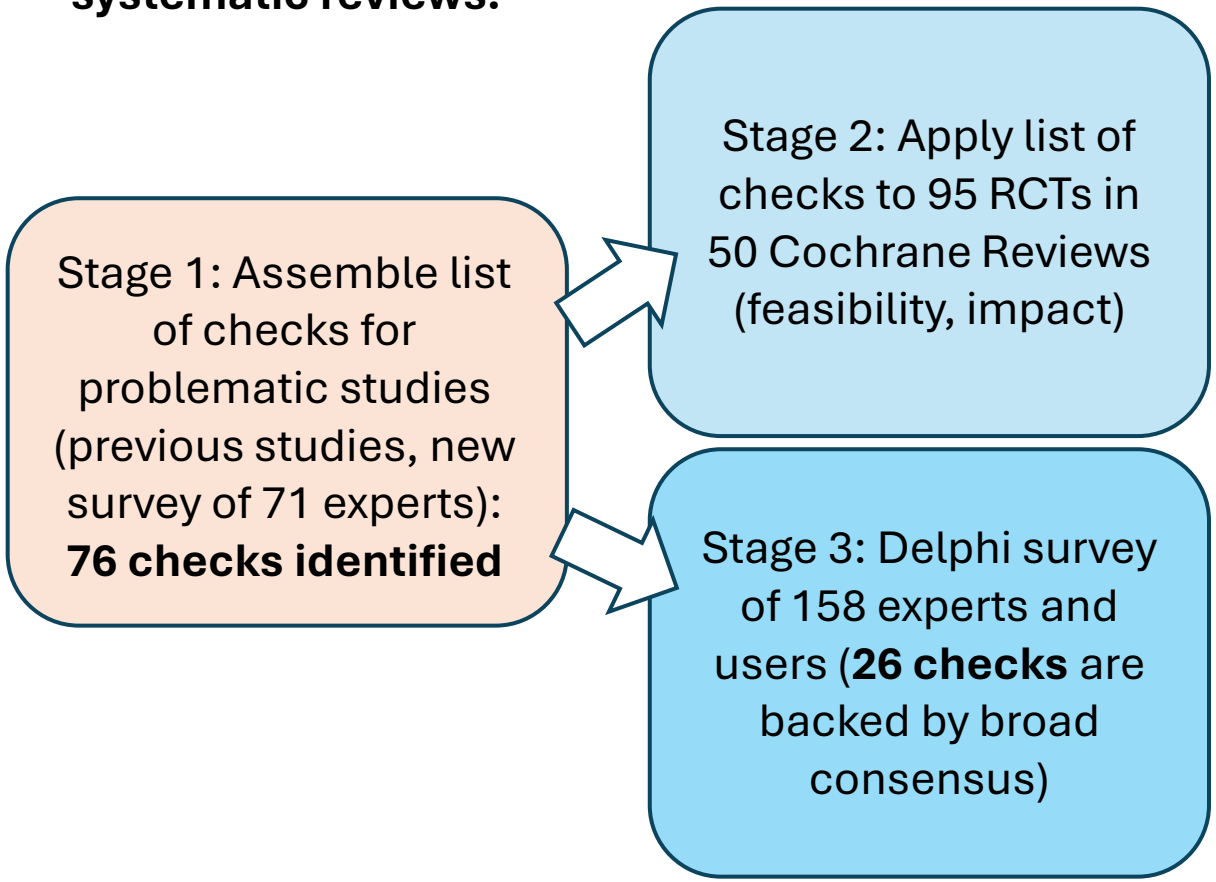
INSPECT SR

Cochrane

FUNDED BY
NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.



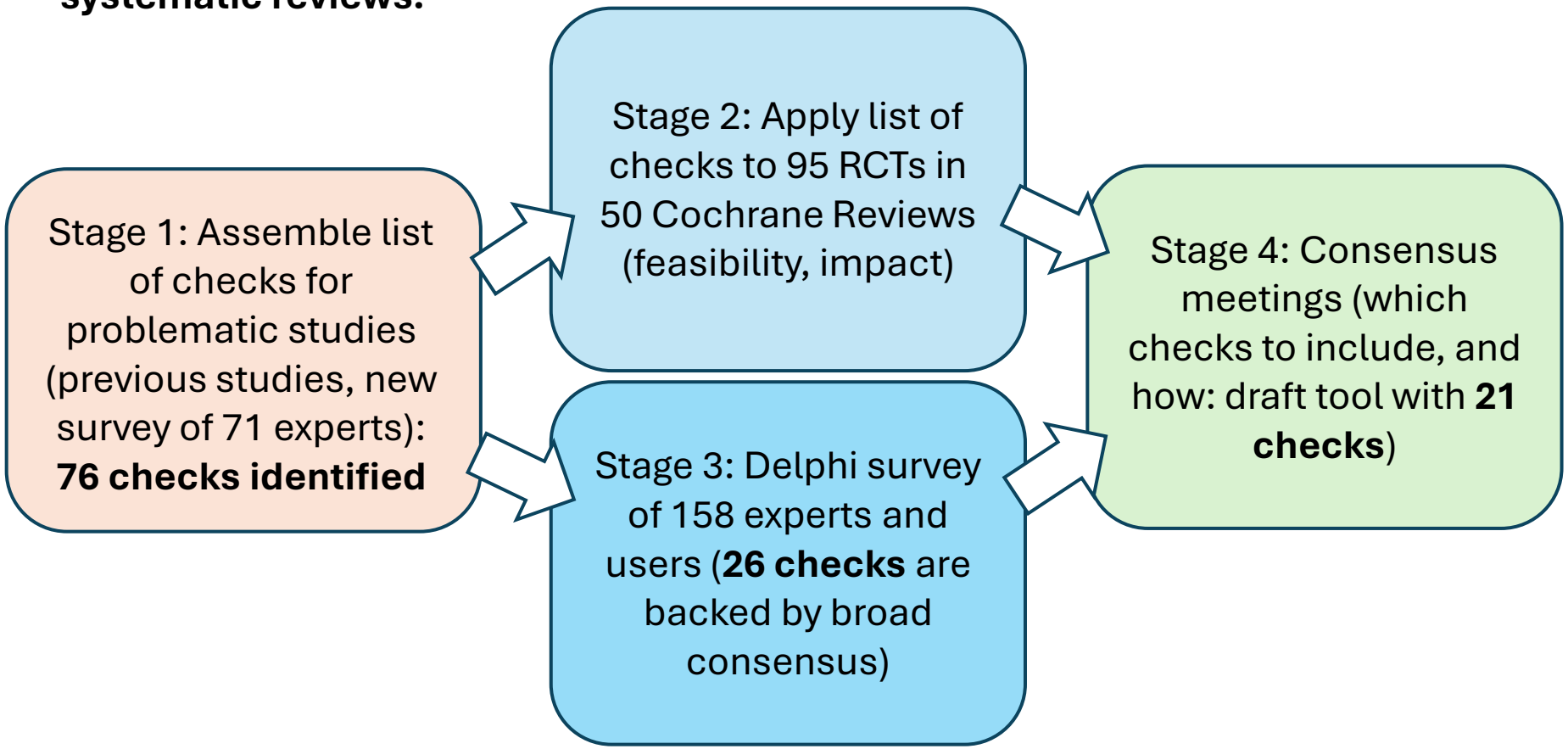
INSPECT SR

Cochrane

FUNDED BY
NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.



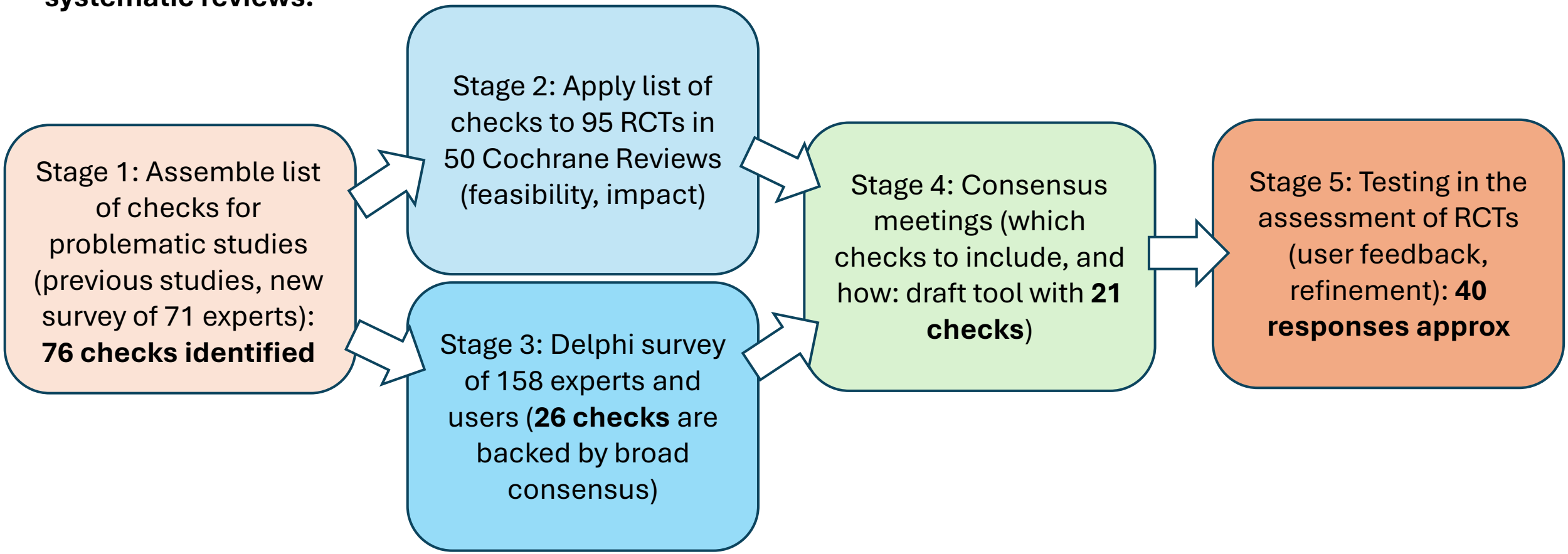
INSPECT SR

Cochrane

FUNDED BY
NIHR | National Institute for Health and Care Research

INveStigating ProblEmatic Clinical Trials in Systematic Reviews

Aim: To develop a tool for identifying problematic randomised controlled trials in the context of health systematic reviews.



Inspecting post-publication notices

3 checks

Inspecting conduct, governance and transparency

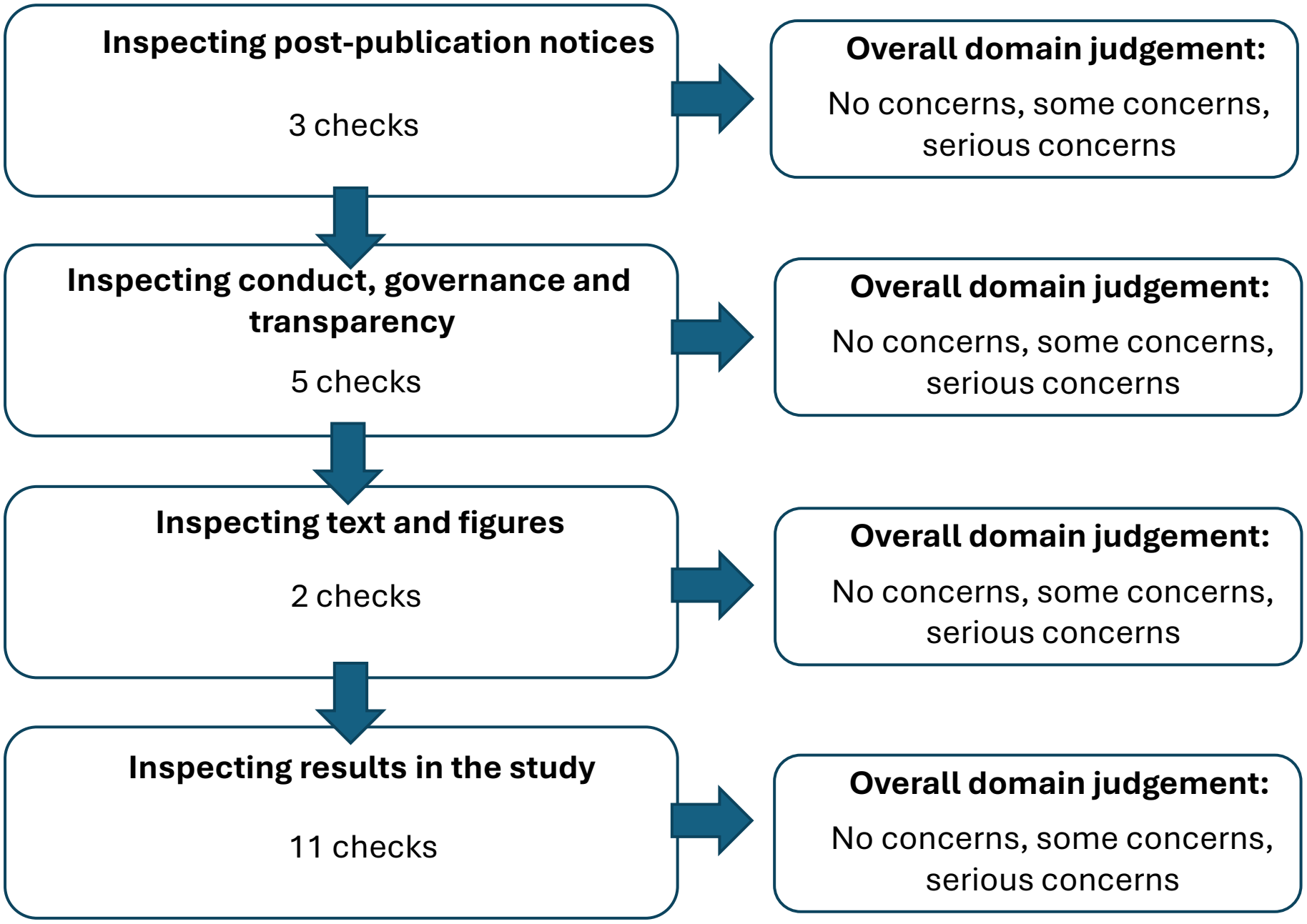
5 checks

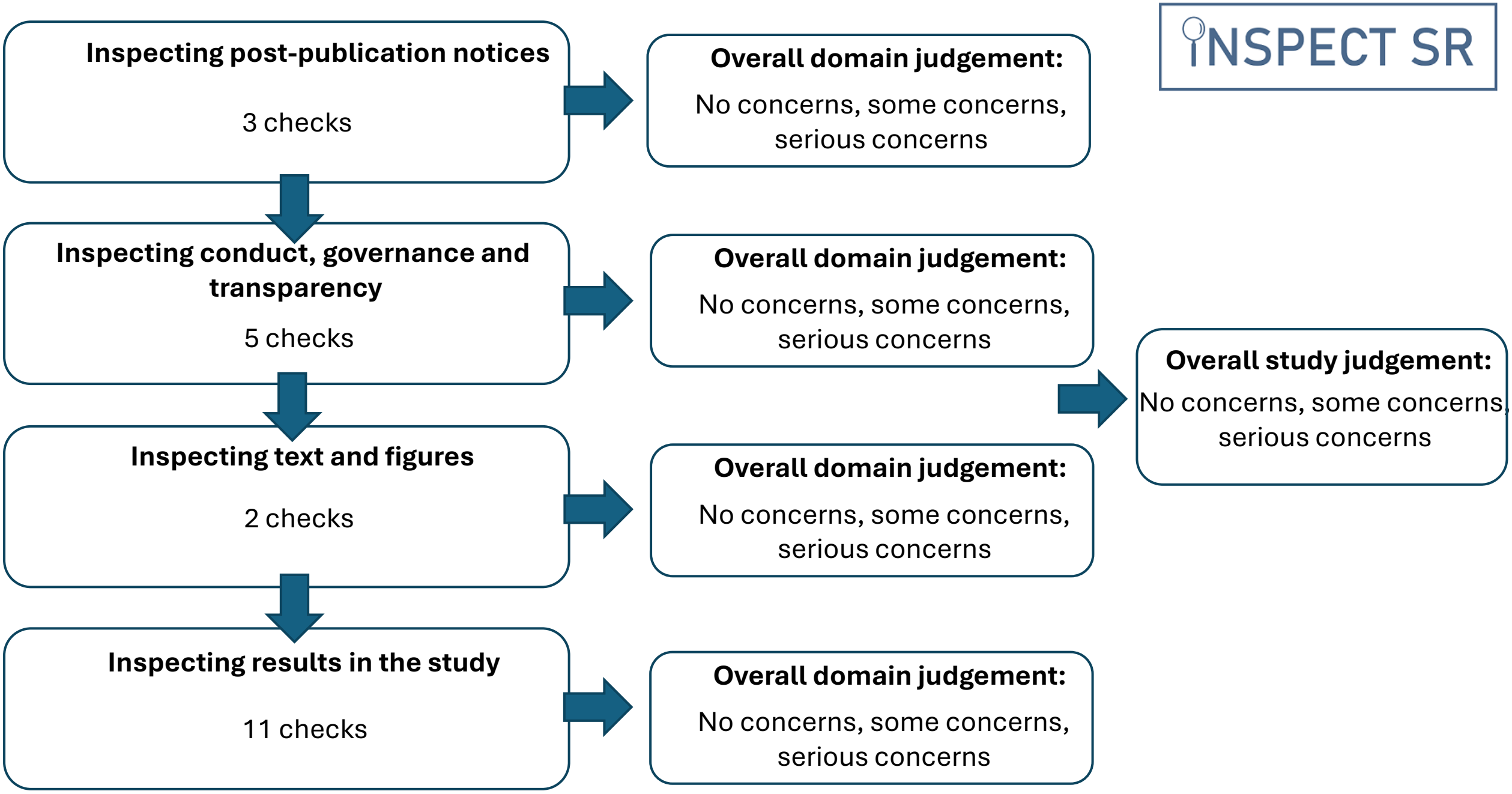
Inspecting text and figures

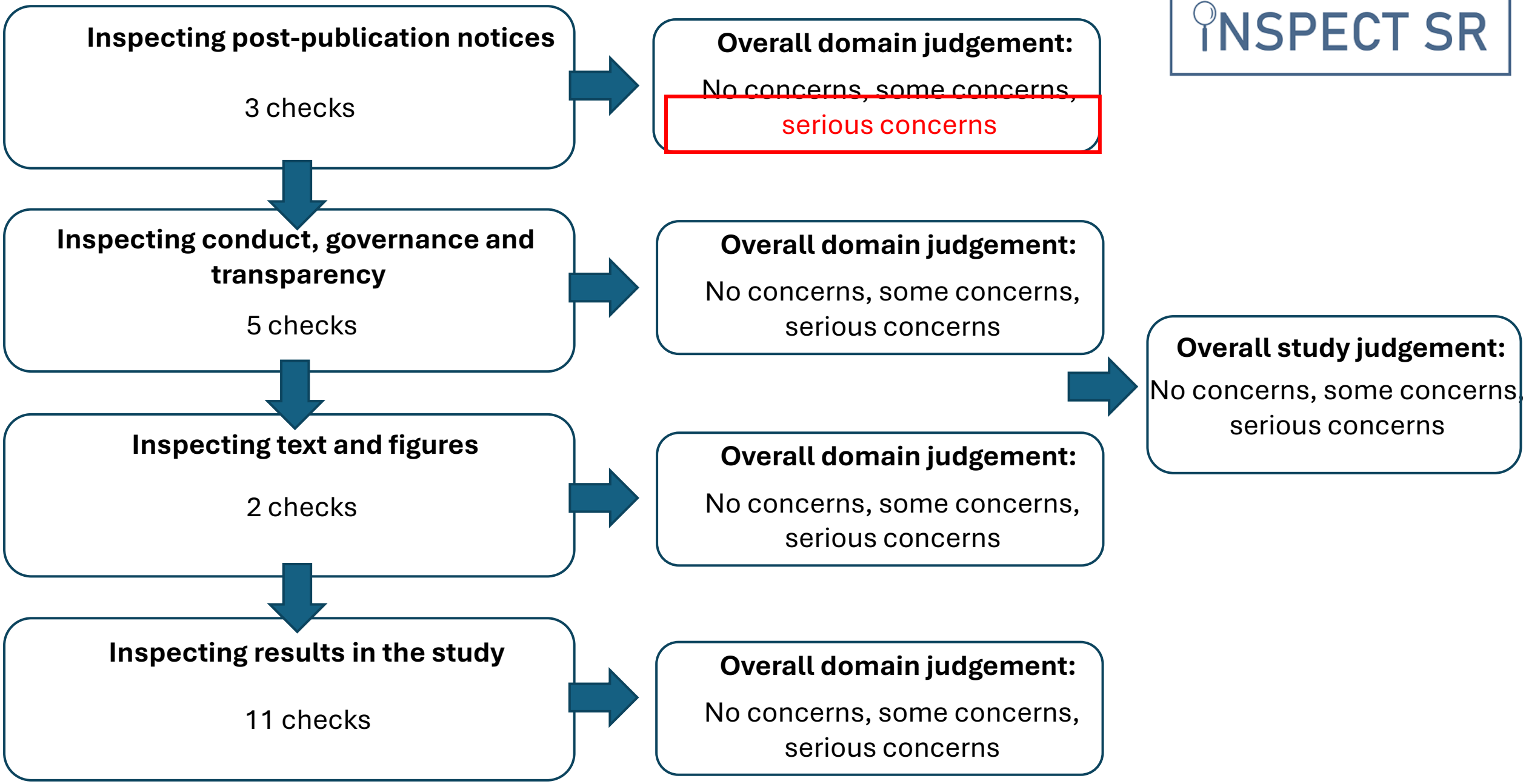
2 checks

Inspecting results in the study

11 checks


Inspecting post-publication notices
3 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting conduct, governance and transparency
5 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting text and figures
2 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting results in the study
11 checks
Overall domain judgement:
No concerns, some concerns, serious concerns


INSPECT SR
Inspecting post-publication notices
3 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting conduct, governance and transparency
5 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting text and figures
2 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Inspecting results in the study
11 checks
Overall domain judgement:
No concerns, some concerns, serious concerns
Overall study judgement:
No concerns, some concerns, serious concerns


INSPECT SR
Inspecting post-publication notices
3 checks
Overall domain judgement:
No concerns, some concerns,
serious concerns
Inspecting conduct, governance and transparency
5 checks
Overall domain judgement:
No concerns, some concerns,
serious concerns
Inspecting text and figures
2 checks
Overall domain judgement:
No concerns, some concerns,
serious concerns
Inspecting results in the study
11 checks
Overall domain judgement:
No concerns, some concerns,
serious concerns
Overall study judgement:
No concerns, some concerns,
serious concerns

INSPECT SR

Inspecting post-publication notices

3 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting conduct, governance and transparency

5 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting text and figures

2 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting results in the study

11 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Overall study judgement:

No concerns, some concerns, serious concerns

INSPECT SR

Inspecting post-publication notices

3 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting conduct, governance and transparency

5 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting text and figures

2 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Inspecting results in the study

11 checks

Overall domain judgement:

No concerns, some concerns, serious concerns

Overall study judgement:

No concerns, some concerns, serious concerns

INSPECT SR

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

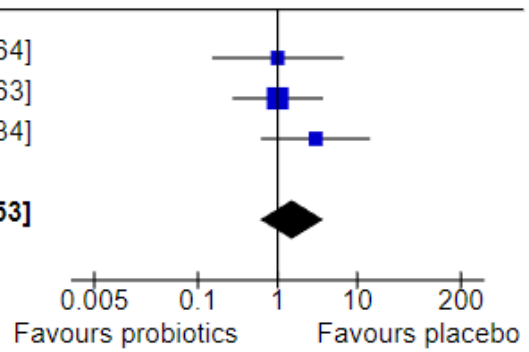
https://doi.org/10.1002/14651858.CD012970.pub2

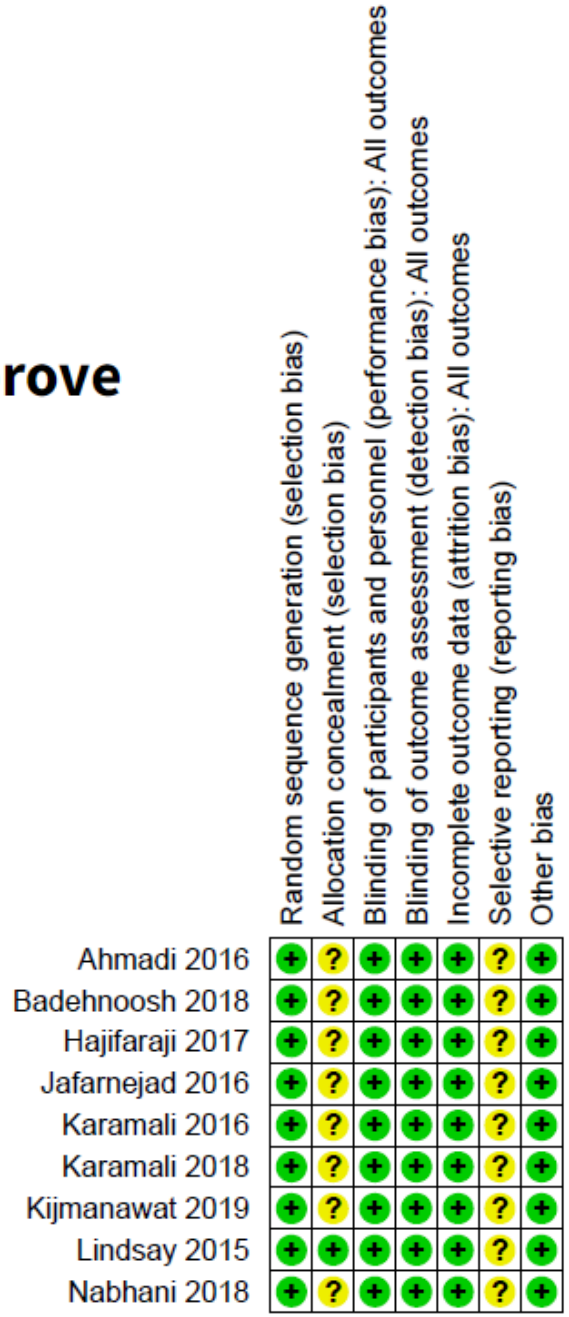
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





INSPECT SR

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

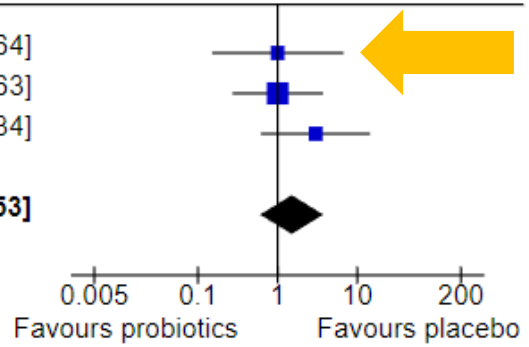
https://doi.org/10.1002/14651858.CD012970.pub2

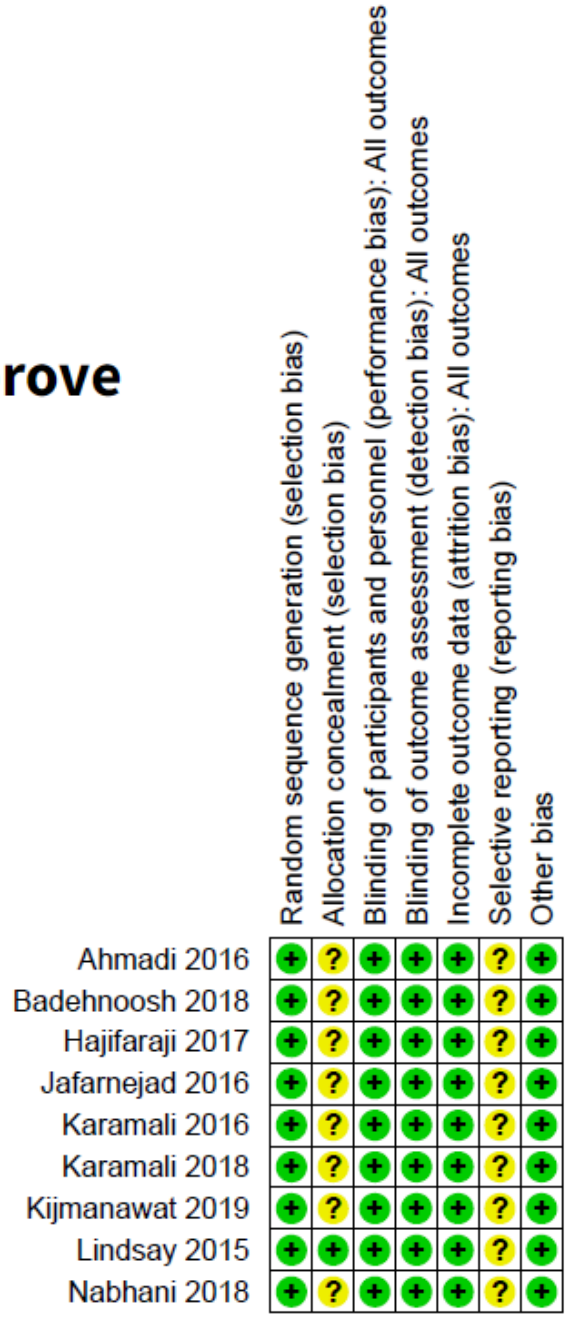
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





INSPECT SR

Cochrane Database of Systematic Reviews | Review - Intervention

Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

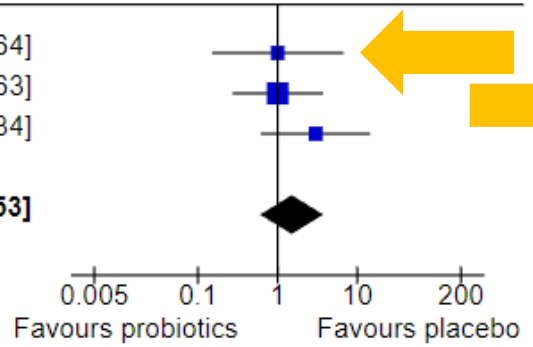
https://doi.org/10.1002/14651858.CD012970.pub2

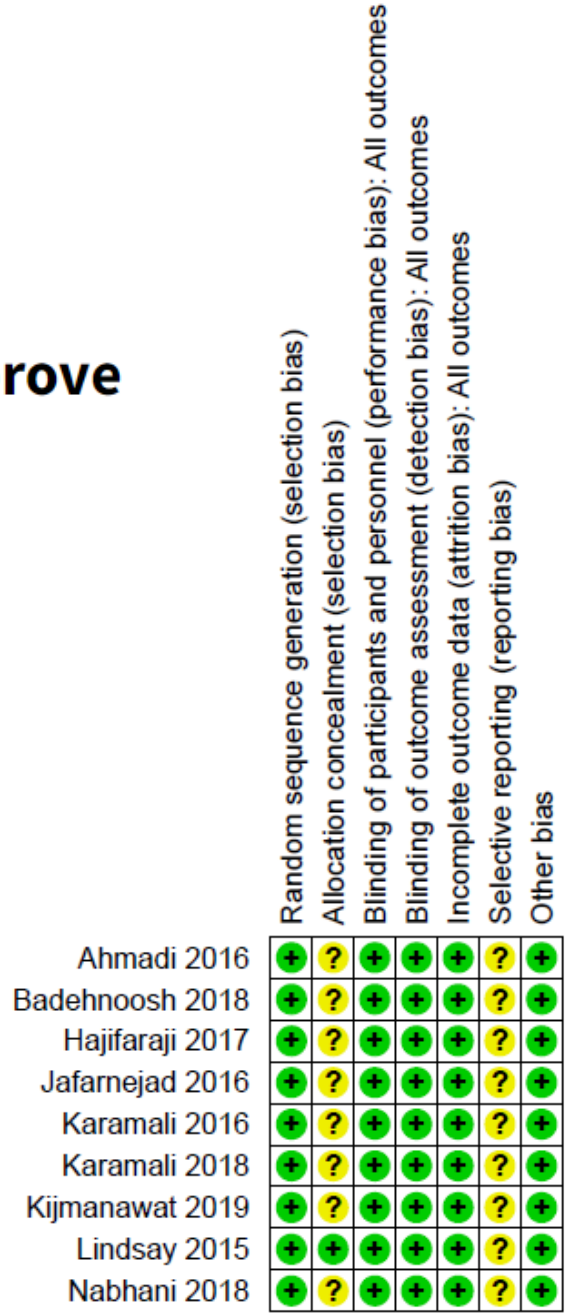
Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

Test for subgroup differences: Not applicable





Inspecting post-publication notices

INSPECT SR

Does the study have an associated retraction?

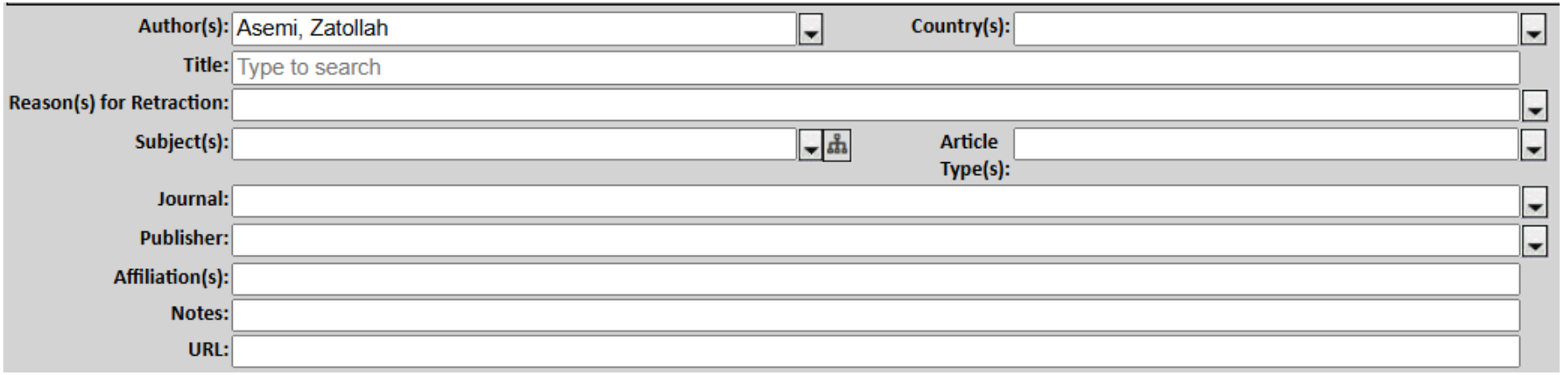
This study has recently been **retracted.**

RETRACTED ARTICLE: The effects of probiotic supplementation on biomarkers of inflammation, oxidative stress and pregnancy outcomes in gestational diabetes

Both the journal editorial team and the external statistical reviewer were unable to confirm the integrity of the trial design and the main outcome of the external review was that the article's conclusions are unreliable. Therefore, as the editorial team no longer have confidence in the reported conclusions the decision has been made to retract the article.

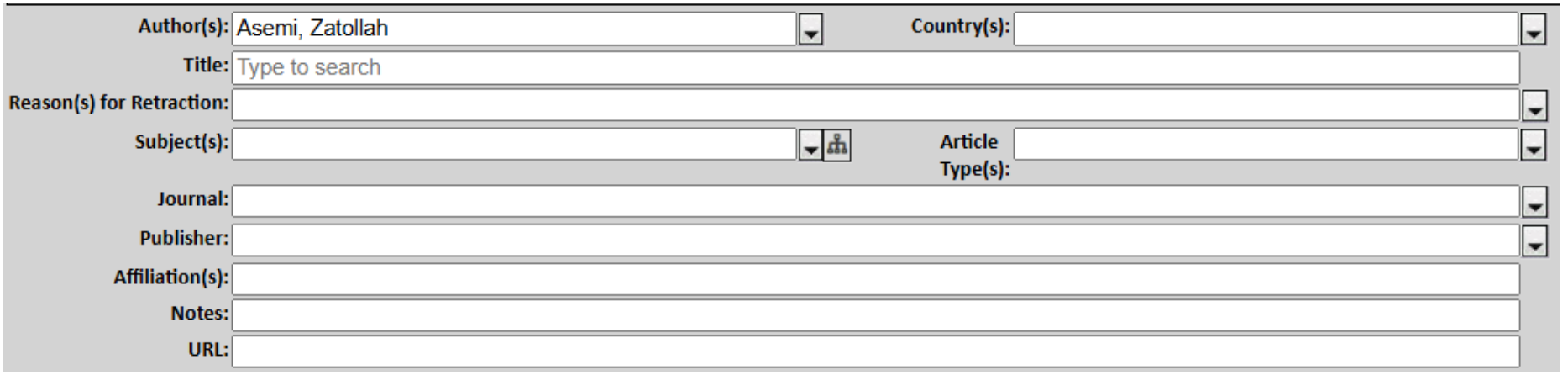
Does consideration of other studies from members of the research team highlight causes for concern?

Search on last author on Retraction Watch database: http://retractiondatabase.org/



Does consideration of other studies from members of the research team highlight causes for concern?

Search on last author on Retraction Watch database: http://retractiondatabase.org/



Inspecting conduct, governance and transparency

INSPECT SR

Are there important inconsistencies between the publication and the registration documents?

- Retrospective registration, with **various inconsistencies** (e.g. recruitment period 6 months vs 1 month)
- Description of the control arm differs between paper and registration (later changed to match):

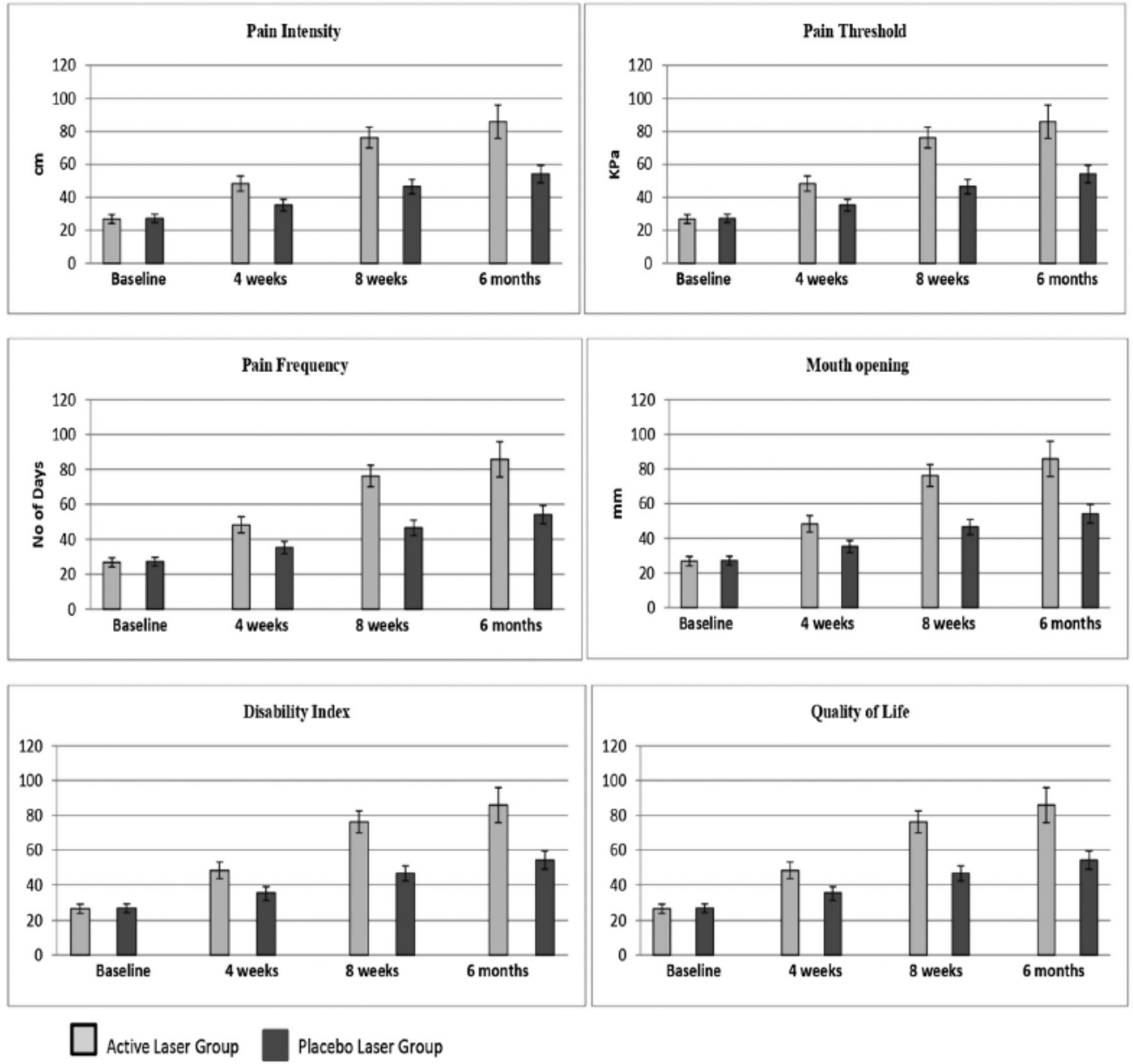
Description - English	Control group: Placebo (Barij Essence, Kashan, Iran), daily, for 6 weeks orally.	Control group: Placebo (Tak Gen Zist, Tehran, Iran), daily, for 6 weeks orally.
Description - Persian	گروه کنترل: پلاسبو (باریج اسانس، کاشان، ایران)، روزانه، به صورت خوراکی برای 6 هفته.	گروه کنترل: پلاسبو (تک ژن زیست، تهران، ایران)، روزانه، به صورت خوراکی برای 6 هفته.

Inspecting text and figures

INSPECT SR

Is there evidence of manipulation or duplication of figures?

INSPECT SR


Pain Intensity
cm
120
100
80
60
40
20
0
Baseline
4 weeks
8 weeks
6 months
Pain Threshold
KPa
Pain Frequency
No of Days
Mouth opening
mm
Disability Index
Quality of Life
Active Laser Group
Placebo Laser Group

Inspecting results in the study

INSPECT SR

Are the means and variances of integer data impossible?

1-min Apgar score	8.93 ± 0.25	8.96 ± 0.18	.561
5-min Apgar score	9.93 ± 0.18	9.96 ± 0.18	.561
Newborns' hyperbilirubinemia (%)	8 (26.7)	2 (6.7)	.080[b]
Newborns' hospitalization (%)	8 (26.7)	2 (6.7)	.080[b]
Newborns' hypoglycemia (%)	3 (10.0)	2 (6.7)	>.999[b]

Inspecting results in the study

INSPECT SR

Are the means and variances of integer data impossible?

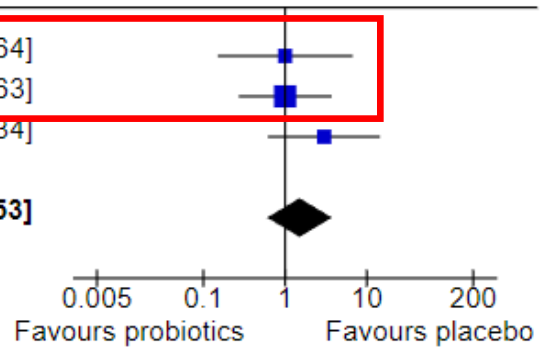
1-min Apgar score	8.93 ± 0.25	8.96 ± 0.18	.561
5-min Apgar score	9.93 ± 0.18	9.96 ± 0.18	.561
Newborns' hyperbilirubinemia (%)	8 (26.7)	2 (6.7)	.080[b]
Newborns' hospitalization (%)	8 (26.7)	2 (6.7)	.080[b]
Newborns' hypoglycemia (%)	3 (10.0)	2 (6.7)	>.999[b]

- Apgar score is a variable which only takes **integer values** (1,2,3,4,5,6,7,8,9,10).
- The highlighted values **cannot occur** for the group sizes in the study.
- 'GRIM' and 'GRIMMER'

Hypertensive disorders

Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Badehnoosh 2018	2	30	2	30	25.0%	1.00 [0.15 , 6.64]
Karamali 2018	4	30	4	30	50.0%	1.00 [0.28 , 3.63]
Lindsay 2015	6	68	2	68	25.0%	3.00 [0.63 , 14.34]
Total (95% CI)		**128**		**128**	**100.0%**	**1.50 [0.64 , 3.53]**
Total events:	12		8			

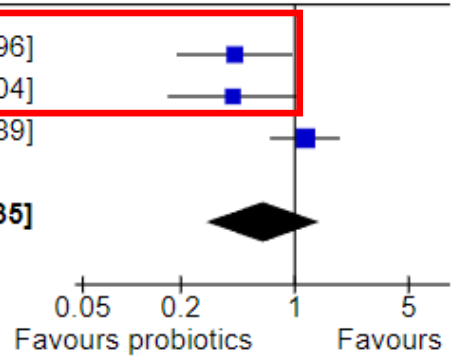
Heterogeneity: Chi² = 1.31, df = 2 (P = 0.52); I² = 0%
Test for overall effect: Z = 0.93 (P = 0.35)
Test for subgroup differences: Not applicable



Caesarian deliveries

Study or Subgroup	Probiotics Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Badehnoosh 2018	6	30	14	30	31.1%	0.43 [0.19 , 0.96]
Karamali 2018	5	30	12	30	28.4%	0.42 [0.17 , 1.04]
Lindsay 2015	24	73	21	74	40.4%	1.16 [0.71 , 1.89]
Total (95% CI)		**133**		**134**	**100.0%**	**0.64 [0.30 , 1.35]**
Total events:	35		47			

Heterogeneity: Tau² = 0.30; Chi² = 6.47, df = 2 (P = 0.04); I² = 69%
Test for overall effect: Z = 1.18 (P = 0.24)
Test for subgroup differences: Not applicable



INSPECT SR

Are any outcome data, including estimated treatment effects, implausible?

- Karamali 2018 is another trial from same group.
- Almost identical risk ratios in the two studies (2 here, several other outcomes have identical effect estimates).

INSPECT SR

Domain	Check	Result
Inspecting post-publication notices	*Does the study have an associated retraction?*	**Yes**, this article has been retracted
	Do other studies by the research team highlight causes for concern?	**Yes**, the senior author has more than 50 notices on Retraction Watch
Inspecting conduct, governance, and transparency	*Are there important inconsistencies between the publication and the registration documents?*	**Yes**, various discrepancies, including recruitment dates and description of control intervention
Inspecting text and figures	*Is there evidence of manipulation or duplication of figures?*	**NA**, no figures in the manuscript
Inspecting results in the study	*Are the means and variances of integer data impossible?*	**Yes**, there are impossible means and variances for Apgar scores
	Are any outcome data, including estimated treatment effects, implausible?	**Yes**, results essentially identical in another RCT from this team, across multiple measures

Applying the draft INSPECT-SR tool would lead to **serious concerns** about the study's trustworthiness.

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

INSPECT SR

Follow this preprint

INSPECT-SR: a tool for assessing trustworthiness of randomised controlled trials

Jack Wilkinson, Calvin Heal, Ella Flemyng, Georgios A. Antoniou, Tony Aburrow, Zarko Alfirevic, Alison Avenell, Virginia Barbour, Vincenzo Berghella, Dorothy V. M. Bishop, Esmée M Bordewijk, Nicholas J. L. Brown, Jana Christopher, Mike Clarke, Darren Dahly, Jane Dennis, Patrick Dicker, Jo Dumville, Helen Frankish, Andrew Grey, Steph Grohmann, Lyle C. Gurrin, Jill A. Hayden, James A.J. Heathers, Kylie E Hunter, Ian Hussey, Lukas Jung, Emily Lam, Toby J. Lasserson, Sarah Lensen, Tianjing Li, Wentao Li, Jianping Liu, Elizabeth Loder, Andreas Lundh, Gideon Meyerowitz-Katz, Ben W. Mol, Florian Naudet, Anna Noel-Storr, Neil E. O'Connell, Lisa Parker, Rita F. Redberg, Barbara K. Redman, Rachel Richardson, Anna Lene Seidler, Kyle Sheldrick, Emma Sydenham, Madelon van Wely, Colby J. Vorland, Rui Wang, Stephanie Weibel, Matthias Wjst, Lisa Bero, Jamie J. Kirkham

doi: https://doi.org/10.1101/2025.09.03.25334905

- INSPECT-SR is now available
- Endorsed by Cochrane: https://www.cochrane.org/authors/methods-cochrane
- An editable template and detailed guidance document are provided: https://osf.io/b74wj and inspect.sr

Upcoming online training workshop:

- **July 23rd** 10am-12pm (London, UK)

Booking link:

https://www.trybooking.com/uk/GNCQ

THIS ARTICLE HAS BEEN RETRACTED | Full Access

Retracted: Randomized controlled trial of elevation of the fetal head with a fetal pillow during cesarean delivery at full cervical dilatation

Correction(s) for this article

Retraction(s) for this article

Subrata L. Seal, Alok Dey, Sannyashi C. Barman, Gourisankar Kamilya, Joydev Mukherji, Joseph L. Onwude

- RCT retracted – National Institute for Health and Care Excellence (NICE) subsequently reversed recommendation of this device.
- Royal College of Obstetrics and Gynaecology retracted Scientific Impact Paper
- See https://doi.org/10.1111/1471-0528.18004

- INSPECT-SR would have flagged problems ahead of retraction
- E.g. Numerical anomalies in the abstract and results

Results

Overall, 120 women were assigned to each group. Major uterine wound extensions occurred in 6 (5.0%) women in the Fetal Pillow group and 39 (32.5%) in the control group (relative risk 0.23, 95% confidence interval 0.11–0.48).

THIS ARTICLE HAS BEEN RETRACTED | Full Access

Retracted: Randomized controlled trial of elevation of the fetal head with a fetal pillow during cesarean delivery at full cervical dilatation

Correction(s) for this article

Retraction(s) for this article

Subrata L. Seal, Alok Dey, Sannyashi C. Barman, Gourisankar Kamilya, Joydev Mukherji, Joseph L. Onwude

- RCT retracted – National Institute for Health and Care Excellence (NICE) subsequently reversed recommendation of this device.
- Royal College of Obstetrics and Gynaecology retracted Scientific Impact Paper
- See https://doi.org/10.1111/1471-0528.18004

- INSPECT-SR would have flagged problems ahead of retraction
- E.g. Numerical anomalies in the abstract and results

Results

Overall, 120 women were assigned to each group. Major uterine wound extensions occurred in 6 (5.0%) women in the Fetal Pillow group and 39 (32.5%) in the control group (relative risk 0.23, 95% confidence interval 0.11–0.48).

Are there errors in statistical results?

Table I. Patients' characteristics.

	Group A (minimal stimulation) ($n=58$)	Group B (CC) ($n=55$)	p Value
Age (years)	27.7 ± 2.4	26.8 ± 2.3	0.57
Duration of infertility(years)	1.88 ± 0.7	2.14 ± 0.8	0.71
BMI (kg/m^2)	24.9 ± 2.3	25.3 ± 2.6	0.62
Waist-to-hip ratio	0.8 ± 0.04	0.8 ± 0.05	0.38
Menstrual cycle			
Oligomenorrhea	54 (93.1%)	51 (92.7%)	0.43
Amenorrhea	4 (6.9%)	4 (7.3%)	0.61
Hyperandrogenism	26 (44.8%)	23 (41.8%)	0.74
LH (mIU/mL)	12.9 ± 2.3	13.4 ± 2.4	0.52
FSH (mIU/mL)	5.6 ± 1.1	5.7 ± 1.3	0.72
LH/FSH ratio	2.6 ± 1.1	2.6 ± 1.2	0..42
Testosterone (ng/mL)	0.94 ± 0.21	0.92 ± 0.24	0.77
SHBG (nmol/L)	33.62 ± 3.6	35.53 ± 3.1	0.48
Fasting glucose (mg/dL)	91.6 ± 1.5	89.7 ± 1.4	0.76
Fasting insulin (μU/mL)	10.7 ± 2.7	11.2 ± 3.1	0.57
Fasting glucose/ insulin ratio	7.7 ± 3.7	7.2 ± 3.6	0.33
Ovarian volume (mL)	11.3 ± 2.5	11.4 ± 2.4	0.22

BMI, Body mass index; CC, clomiphene citrate; FSH, follicle-stimulating hormone;

- Look at waist-to-hip ratio, which has been analysed by t-test.
- Do you think there are any problems with these results?

Are there errors in statistical results?

Table I. Patients' characteristics.

	Group A (minimal stimulation) ($n=58$)	Group B (CC) ($n=55$)	p Value
Age (years)	27.7 ± 2.4	26.8 ± 2.3	0.57
Duration of infertility(years)	1.88 ± 0.7	2.14 ± 0.8	0.71
BMI (kg/m^2)	24.9 ± 2.3	25.3 ± 2.6	0.62
Waist-to-hip ratio	0.8 ± 0.04	0.8 ± 0.05	0.38
Menstrual cycle			
Oligomenorrhea	54 (93.1%)	51 (92.7%)	0.43
Amenorrhea	4 (6.9%)	4 (7.3%)	0.61
Hyperandrogenism	26 (44.8%)	23 (41.8%)	0.74
LH (mIU/mL)	12.9 ± 2.3	13.4 ± 2.4	0.52
FSH (mIU/mL)	5.6 ± 1.1	5.7 ± 1.3	0.72
LH/FSH ratio	2.6 ± 1.1	2.6 ± 1.2	0..42
Testosterone (ng/mL)	0.94 ± 0.21	0.92 ± 0.24	0.77
SHBG (nmol/L)	33.62 ± 3.6	35.53 ± 3.1	0.48
Fasting glucose (mg/dL)	91.6 ± 1.5	89.7 ± 1.4	0.76
Fasting insulin (μU/mL)	10.7 ± 2.7	11.2 ± 3.1	0.57
Fasting glucose/ insulin ratio	7.7 ± 3.7	7.2 ± 3.6	0.33
Ovarian volume (mL)	11.3 ± 2.5	11.4 ± 2.4	0.22

BMI, Body mass index; CC, clomiphene citrate; FSH, follicle-stimulating hormone;

- Look at waist-to-hip ratio, which has been analysed by t-test.
- Do you think there are any problems with these results?

- We *might* expect a p-value of 1 because the means are equal. However!
- Must remember that the summary stats in the paper are **rounded**.
- These data are consistent with very small p-values (e.g. p<0.00001) so there is no contradiction here.
- Could have means of 0.75 and 0.85, smaller SDs.
- Have to check whether p-values are **consistent with** rounded summary data (see INSPECT-SR guidance check 4.9)

Are there errors in statistical results?

Table I. Patients' characteristics.

	Group A (minimal stimulation) ($n=58$)	Group B (CC) ($n=55$)	p Value
Age (years)	27.7 ± 2.4	26.8 ± 2.3	0.57
Duration of infertility(years)	1.88 ± 0.7	2.14 ± 0.8	0.71
BMI (kg/m^2)	24.9 ± 2.3	25.3 ± 2.6	0.62
Waist-to-hip ratio	0.8 ± 0.04	0.8 ± 0.05	0.38
Menstrual cycle			
Oligomenorrhea	54 (93.1%)	51 (92.7%)	0.43
Amenorrhea	4 (6.9%)	4 (7.3%)	0.61
Hyperandrogenism	26 (44.8%)	23 (41.8%)	0.74
LH (mIU/mL)	12.9 ± 2.3	13.4 ± 2.4	0.52
FSH (mIU/mL)	5.6 ± 1.1	5.7 ± 1.3	0.72
LH/FSH ratio	2.6 ± 1.1	2.6 ± 1.2	0..42
Testosterone (ng/mL)	0.94 ± 0.21	0.92 ± 0.24	0.77
SHBG (nmol/L)	33.62 ± 3.6	35.53 ± 3.1	0.48
Fasting glucose (mg/dL)	91.6 ± 1.5	89.7 ± 1.4	0.76
Fasting insulin (μU/mL)	10.7 ± 2.7	11.2 ± 3.1	0.57
Fasting glucose/ insulin ratio	7.7 ± 3.7	7.2 ± 3.6	0.33
Ovarian volume (mL)	11.3 ± 2.5	11.4 ± 2.4	0.22

BMI, Body mass index; CC, clomiphene citrate; FSH, follicle-stimulating hormone;

It turns out that loads of the results in the paper **are incompatible** with the p-values even taking rounding into account.

All of the results in red are not possible given these summary statistics.

Similar problems in outcome table, including primary outcome:

Clinical pregnancy per patient p = 0.77 vs 0.36

Clinical pregnancy per cycle p = 0.76 vs 0.23

Being able to exactly reproduce p-values from rounded summary data might actually indicate a problem

Inflammatory variables	Group A: control (n = 30)	Group B: esmolol (n = 30)	t test/ χ^{2*}	P value
MAP mmHg	70.2 ± 6.42	72.3 ± 6.7	1.240	0.2201
HR rate/min	115.75 ± 12.86	113.4 ± 12.78	− 0.710	0.4806
CVP CmH2o	11.7 ± 4.1	11.6 ± 2.6	− 0.113	0.9106
ScvO2 %	76.2 ± 5.92	74.3 ± 4.8	− 1.365	0.1774
Lactate	5.3 ± 1.35	5.2 ± 1.6	− 0.262	0.7945
APACHE II	23.5 ± 6.2	24.2 ± 5.4	0.466	0.6427
SOFA score	9.2 ± 3.2	8.8 ± 3.5	− 0.462	0.6458

Data presented as mean ± SD
P value > 0.05 NS; *P value < 0.05 S; **P value < 0.001 HS
There was no statistically significant difference between both groups according to admission data as shown in Table 2

- Given that summary data are rounded, we don't expect to be able to reproduce p-values exactly.
- If we *can* reproduce them all exactly, that might suggest no underlying data were analysed to produce the results
- That could suggest that there is no underlying data. Alternative explanation: they calculated summary statistics, and then fed those into a calculator (an odd thing to do)
- Example: the p-values can be reproduced to 4dp from the rounded summary data

	A	B	C	D	E	F
1	Surname and Nar	Sex	Birthdate	Eye	Date of Surgery	Age at Sur
2	Alison Green	M	05/06/1976	RE	11/08/2008	32
3	Jared Friedman	F	13/04/1997	LE	13/04/2015	17
4	Jessica Cohen	M	21/03/1976	RE	28/12/2008	32
5	Sara Young DDS	F	12/05/1964	LE	03/04/2010	45
6	Whitney Douglas	M	26/01/1974	RE	02/01/2009	34
7	Daniel Watkins	M	13/11/1987	RE	03/11/2006	18
8	Jennifer Johnson	M	05/08/1952	RE	31/07/2010	57
9	Tara Thompson	M	16/06/1999	RE	16/06/2017	18
10	Jennifer Dixon	M	30/11/1977	RE	14/03/2007	29
11	Susan Howe	M	19/04/1955	LE	04/04/2005	49
12	Daniel Mccarthy	F	11/10/1989	LE	30/04/2009	19
13	Bianca Boyd	M	07/11/1951	RE	02/05/2005	53
14	Lisa Kaufman	M	23/11/1996	RE	23/11/2014	17
15	Michael Stephen	M	06/10/1972	RE	21/07/2005	32
16	Robert Hill	F	05/08/1982	RE	18/08/2008	26
17	Caleb Grant	F	28/02/1971	RE	28/10/2008	37
18	Shawn Mccarty	F	28/02/1956	RE	31/05/2006	50
19	Jennifer Callahar	M	13/11/1975	RE	12/04/2006	30
20	Alexis Cortez	F	11/05/1954	LE	21/11/2008	54
21	Daniel Robertsor	M	05/11/1954	RE	21/07/2006	51
22	Susan Howe	F	22/09/1954	RE	11/06/2008	53
23	Vanessa Moore	F	01/10/1952	RE	13/05/2010	57
24	Brittany Poole M	M	15/10/1954	LE	31/07/2007	52
25	Angela Cunningh	F	06/06/1991	RE	06/06/2009	18
26	Joseph Robinson	M	07/10/1981	LE	02/10/2006	24
27	Jordan Reid	F	02/11/1983	LE	14/06/2005	21
28	Jordan Dunlap	M	01/04/1980	LE	05/02/2006	25
29	Dr. Susan Townse	M	19/04/1984	RE	14/09/2006	22
30	Rebecca Malone	M	03/03/1997	RE	03/03/2015	17

Can you spot any problems in this RCT dataset?

https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2811505

	A	B	C	D	E	F
1	Surname and Nar	Sex	Birthdate	Eye	Date of Surgery	Age at Sur
2	Alison Green	M	05/06/1976	RE	11/08/2008	32
3	Jared Friedman	F	13/04/1997	LE	13/04/2015	17
4	Jessica Cohen	M	21/03/1976	RE	28/12/2008	32
5	Sara Young DDS	F	12/05/1964	LE	03/04/2010	45
6	Whitney Douglas	M	26/01/1974	RE	02/01/2009	34
7	Daniel Watkins	M	13/11/1987	RE	03/11/2006	18
8	Jennifer Johnson	M	05/08/1952	RE	31/07/2010	57
9	Tara Thompson	M	16/06/1999	RE	16/06/2017	18
10	Jennifer Dixon	M	30/11/1977	RE	14/03/2007	29
11	Susan Howe	M	19/04/1955	LE	04/04/2005	49
12	Daniel Mccarthy	F	11/10/1989	LE	30/04/2009	19
13	Bianca Boyd	M	07/11/1951	RE	02/05/2005	53
14	Lisa Kaufman	M	23/11/1996	RE	23/11/2014	17
15	Michael Stephen	M	06/10/1972	RE	21/07/2005	32
16	Robert Hill	F	05/08/1982	RE	18/08/2008	26
17	Caleb Grant	F	28/02/1971	RE	28/10/2008	37
18	Shawn Mccarty	F	28/02/1956	RE	31/05/2006	50
19	Jennifer Callahar	M	13/11/1975	RE	12/04/2006	30
20	Alexis Cortez	F	11/05/1954	LE	21/11/2008	54
21	Daniel Robertsor	M	05/11/1954	RE	21/07/2006	51
22	Susan Howe	F	22/09/1954	RE	11/06/2008	53
23	Vanessa Moore	F	01/10/1952	RE	13/05/2010	57
24	Brittany Poole M	M	15/10/1954	LE	31/07/2007	52
25	Angela Cunningh	F	06/06/1991	RE	06/06/2009	18
26	Joseph Robinson	M	07/10/1981	LE	02/10/2006	24
27	Jordan Reid	F	02/11/1983	LE	14/06/2005	21
28	Jordan Dunlap	M	01/04/1980	LE	05/02/2006	25
29	Dr. Susan Townse	M	19/04/1984	RE	14/09/2006	22
30	Rebecca Malone	M	03/03/1997	RE	03/03/2015	17

https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2811505

Can you spot any problems in this RCT dataset?

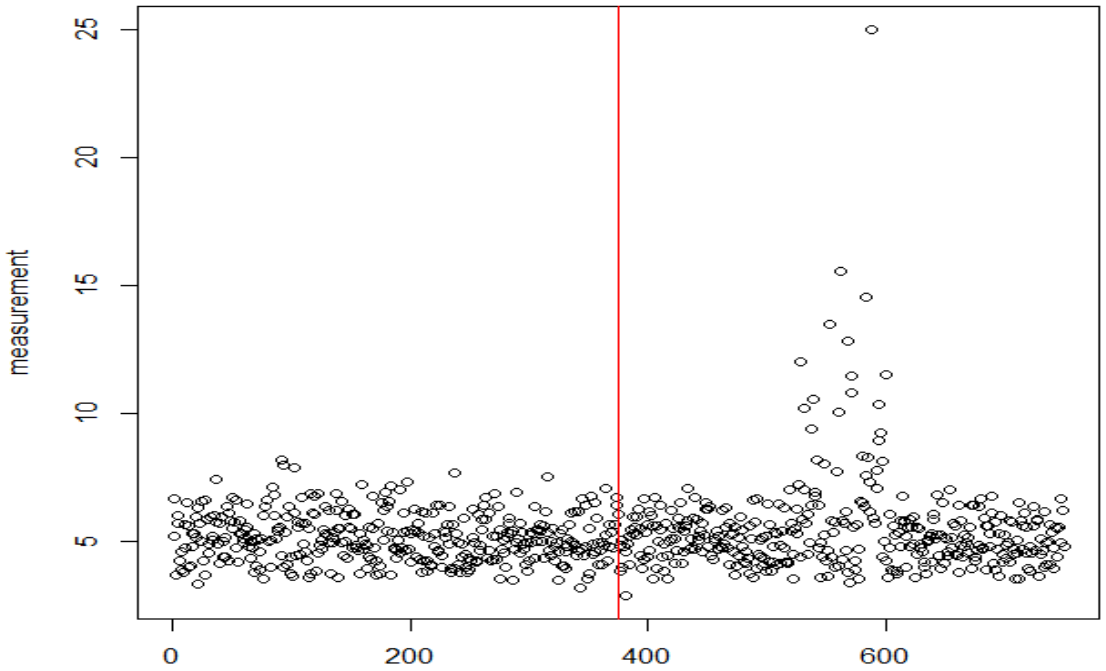
- No correspondence between name and sex at all (e.g. Alison, Jessica, Whitney, Jennifer, Bianca all Male).
- This dataset was actually created using artificial intelligence (Chat GPT)
- Multiple eyes per person (fine), but we have people with two right eyes, people with different birthdays for each eye, and one person with five eyes
- (see M Wjst https://tinyurl.com/3dmzfz8r)

Example – dubious data

2-arm RCT

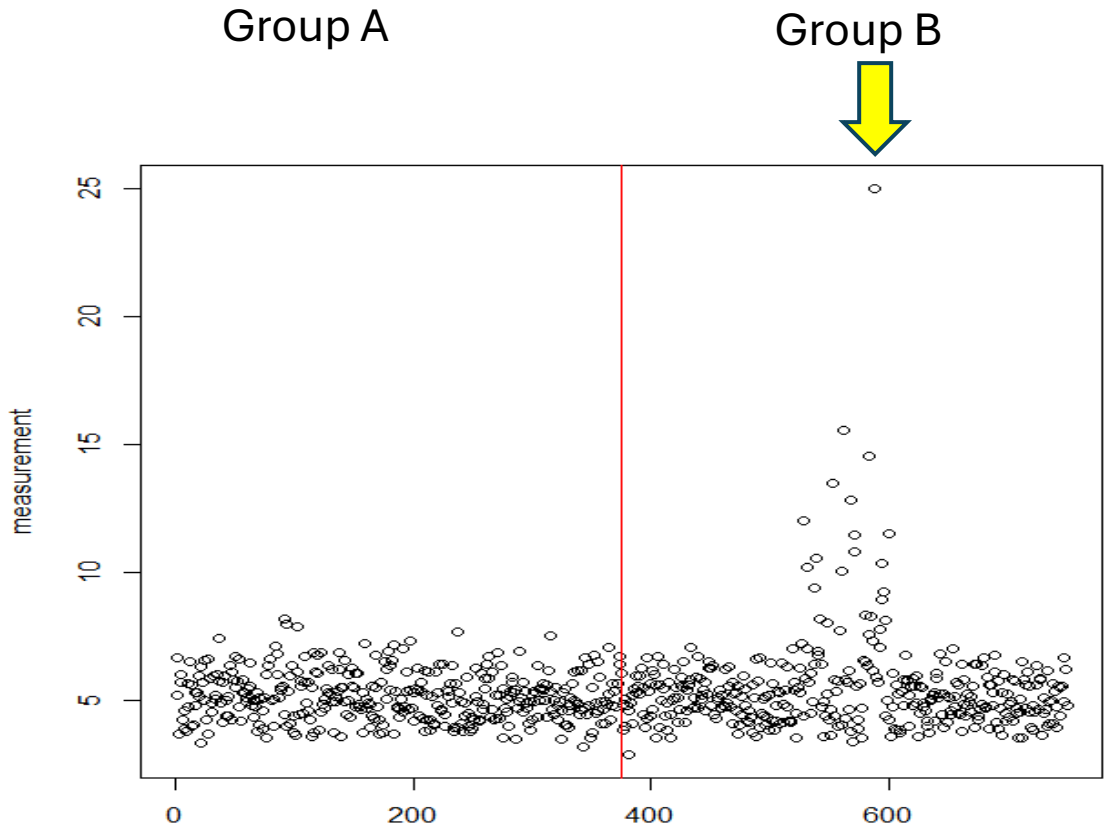
Group A

Group B



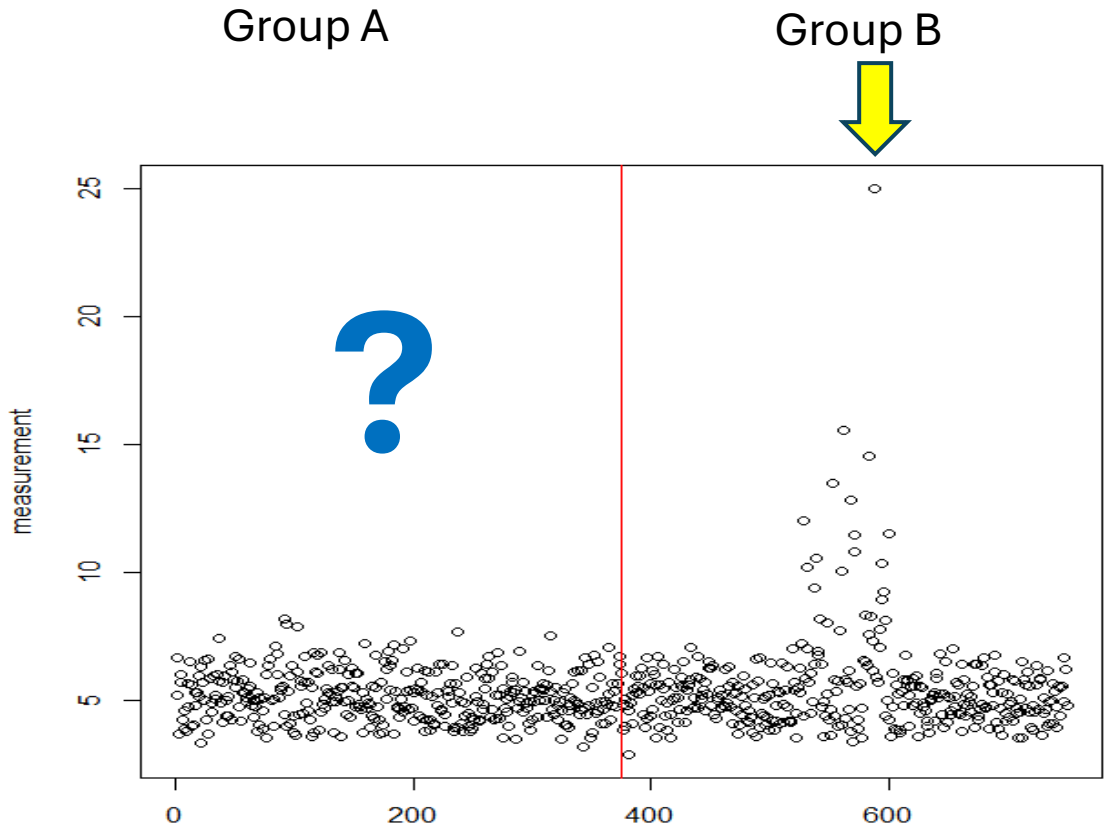
Example – dubious data

2-arm RCT

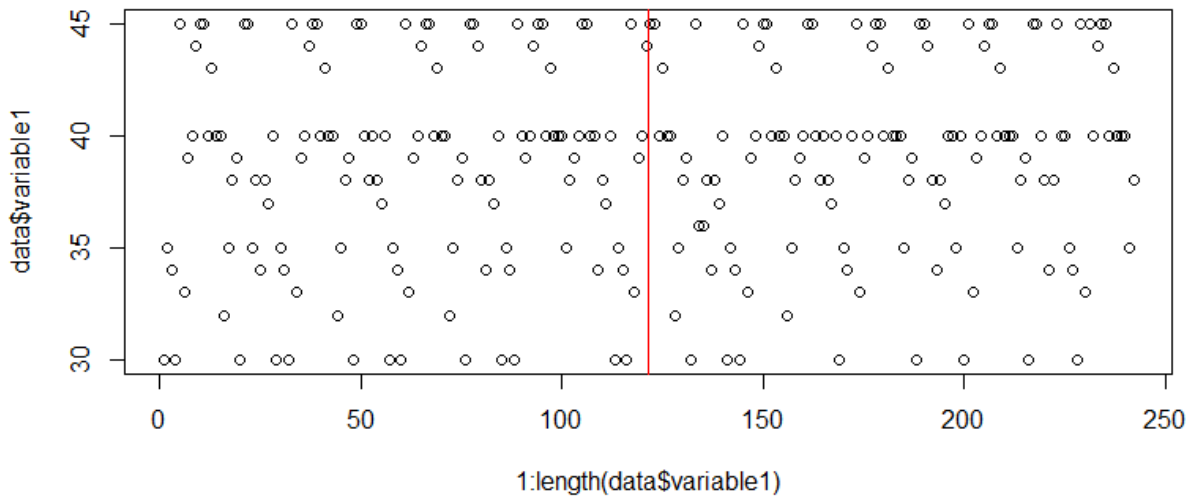


Example – dubious data

2-arm RCT

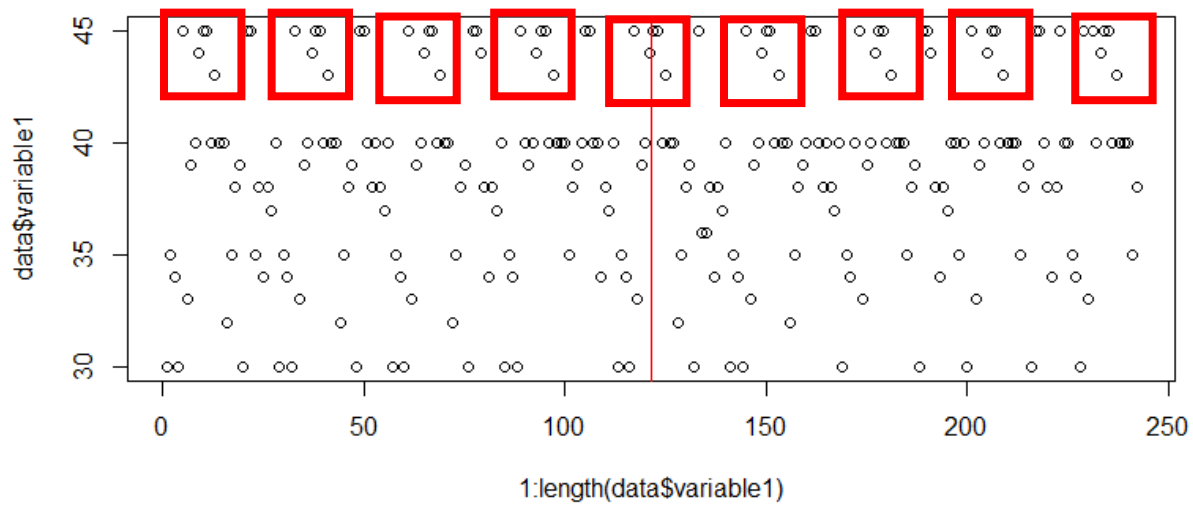


Other problems can be revealed by these plots



Take a moment – can you spot any problems?

Other problems can be revealed by these plots



Correlation across rows in the dataset

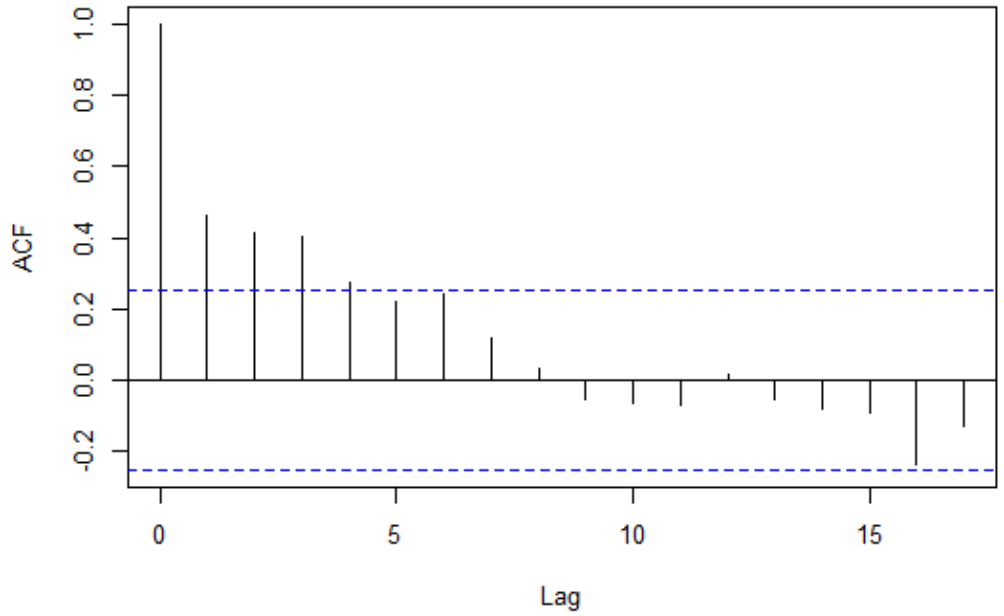
INSPECT IPD

- We don’t expect to see substantial correlation between the baseline values of successive participants.
- E.g. each participant’s duration of infertility shouldn’t be related to the duration of infertility of the person recruited after them, or to the next person’s, or the next person’s etc.
- We do expect correlation across rows if someone has typed (fabricated) values into the column – people are poor random number generators.
- Certainly don’t expect this to differ between randomised groups.

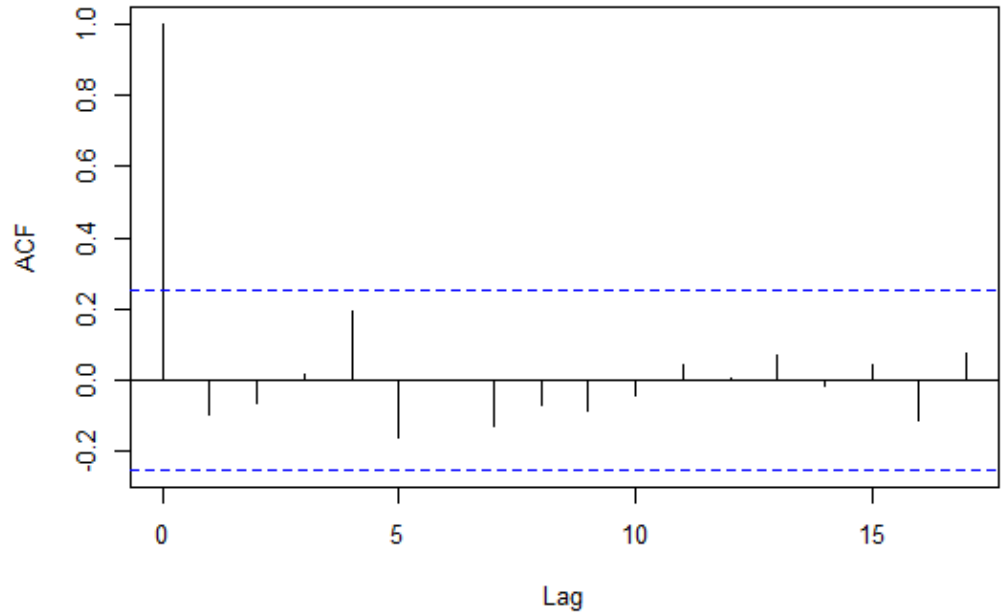
Study ID	InfertilityDuration
1	12
2	8
3	8
4	3
5	1
6	3
7	9
8	3
9	7
10	6
11	5
12	1
13	4
14	6
15	2

Autocorrelation plot

Treatment group



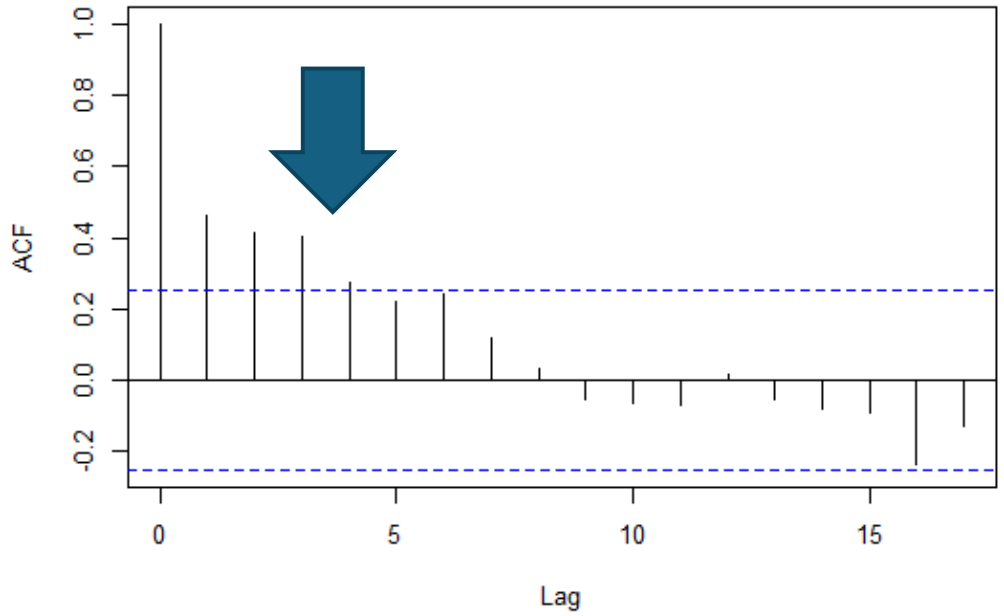
Control group



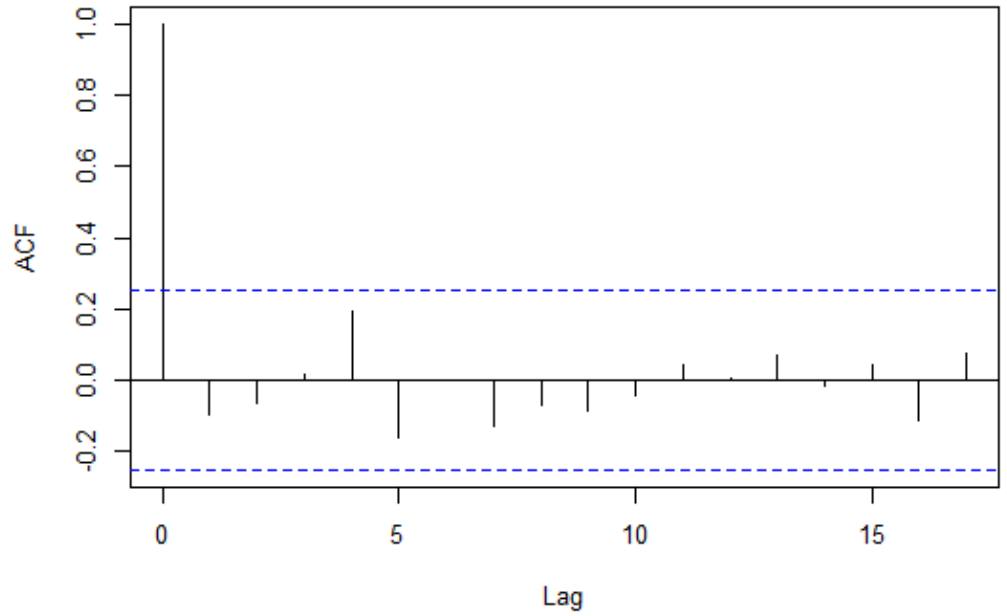
- Plot of correlation between duration of infertility values 1 row apart (Lag =1), 2 rows apart (Lag = 2), 3 rows apart etc.

Autocorrelation plot

Treatment group



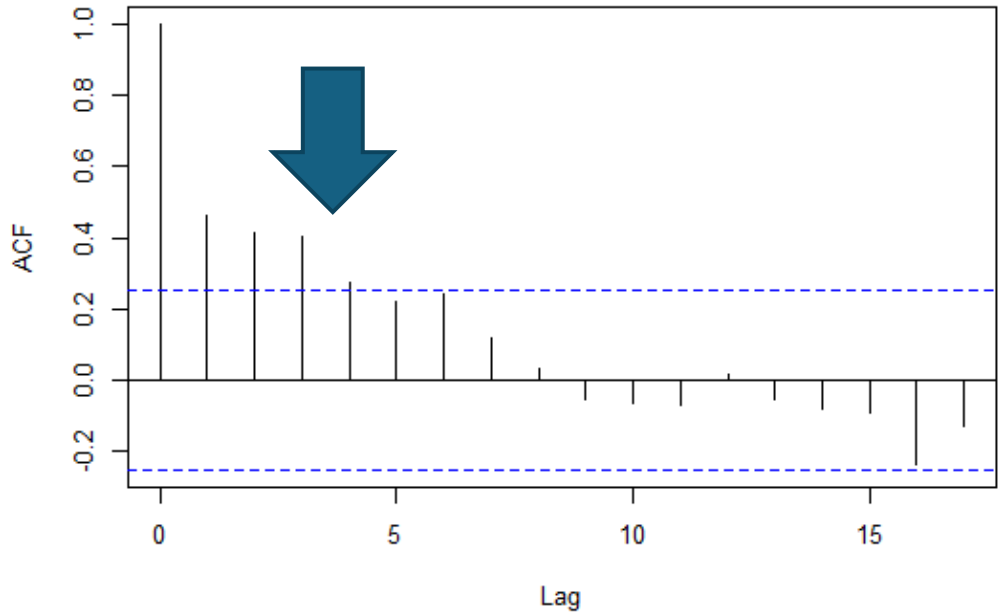
Control group



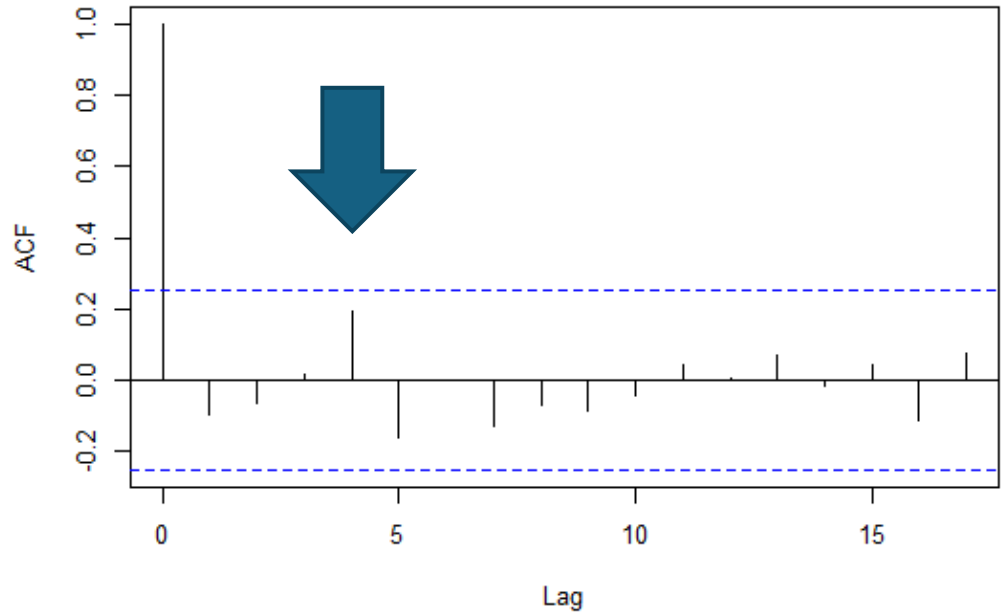
- Plot of correlation between duration of infertility values 1 row apart (Lag =1), 2 rows apart (Lag = 2), 3 rows apart etc.
- In treatment group, there is a correlation between successive rows, which decays as we get further apart.

Autocorrelation plot

Treatment group



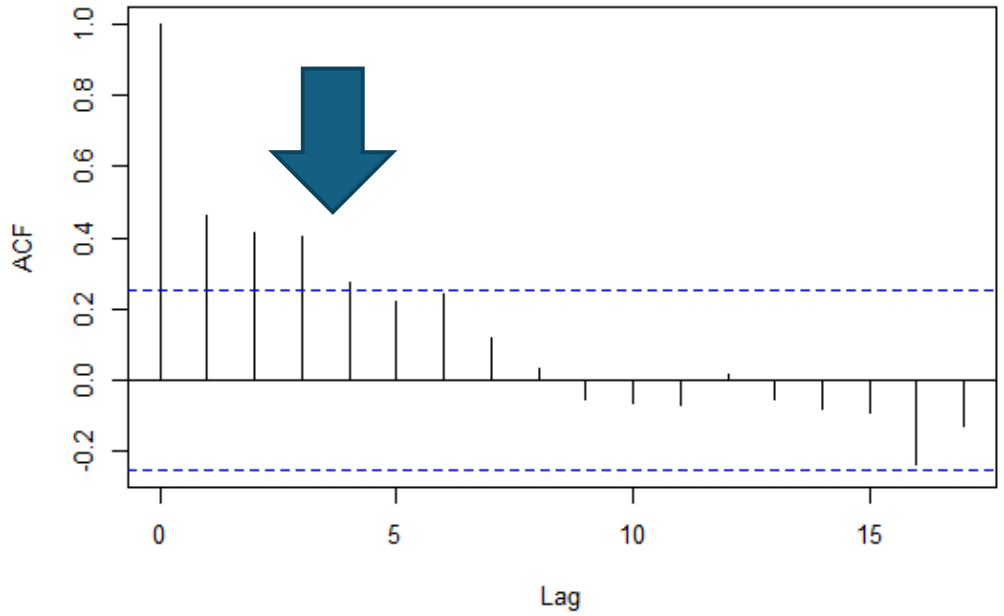
Control group



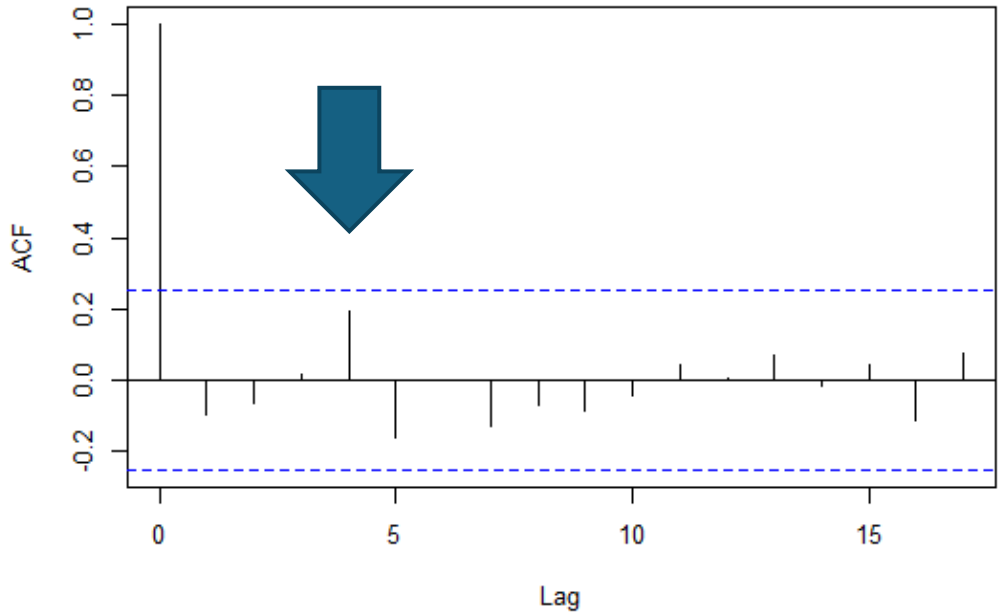
- Plot of correlation between duration of infertility values 1 row apart (Lag =1), 2 rows apart (Lag = 2), 3 rows apart etc.
- In treatment group, there is a correlation between successive rows, which decays as we get further apart.
- Control group looks like genuine data – no serial correlation between rows.

Autocorrelation plot

Treatment group



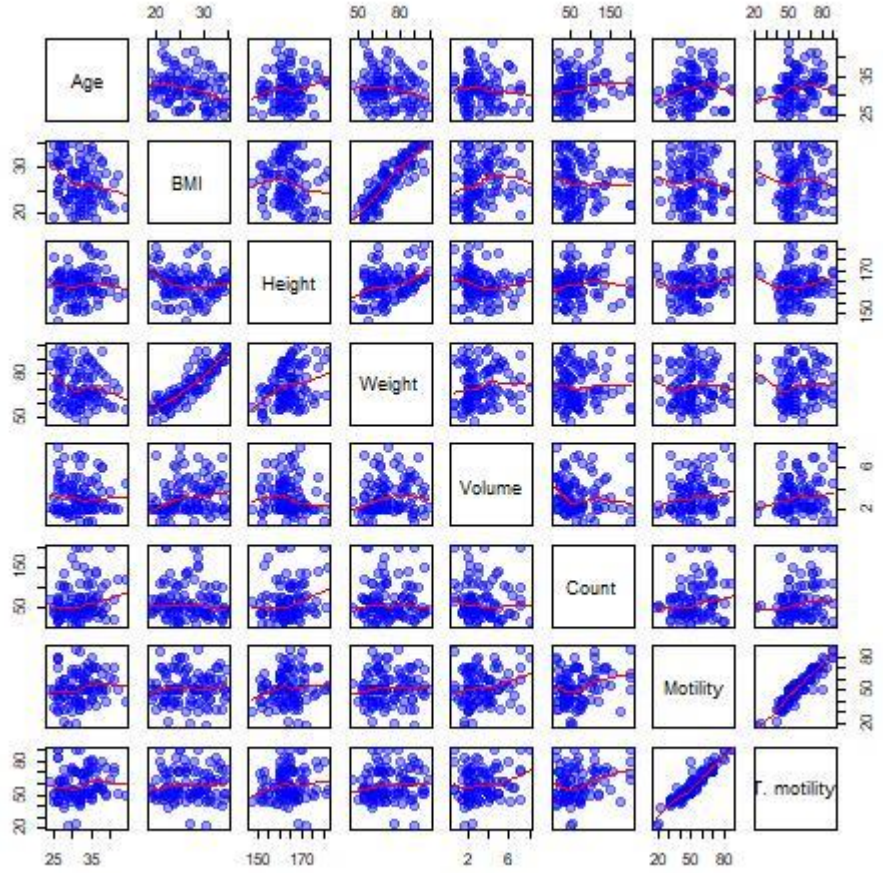
Control group

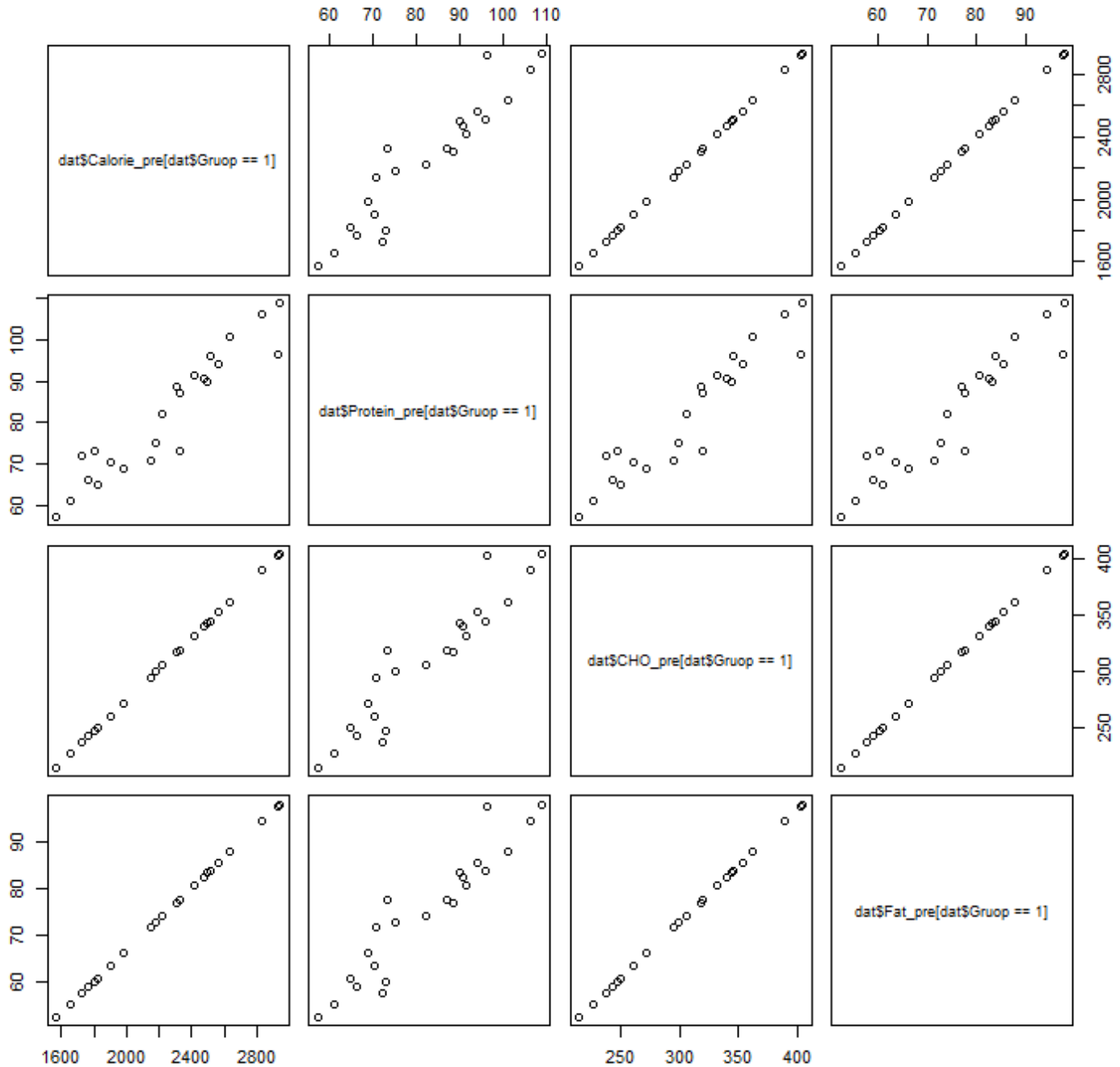


- Plot of correlation between duration of infertility values 1 row apart (Lag =1), 2 rows apart (Lag = 2), 3 rows apart etc.
- In treatment group, there is a correlation between successive rows, which decays as we get further apart.
- Control group looks like genuine data – no serial correlation between rows.
- The treatment group correlation is suspicious, and the difference between groups is more so.

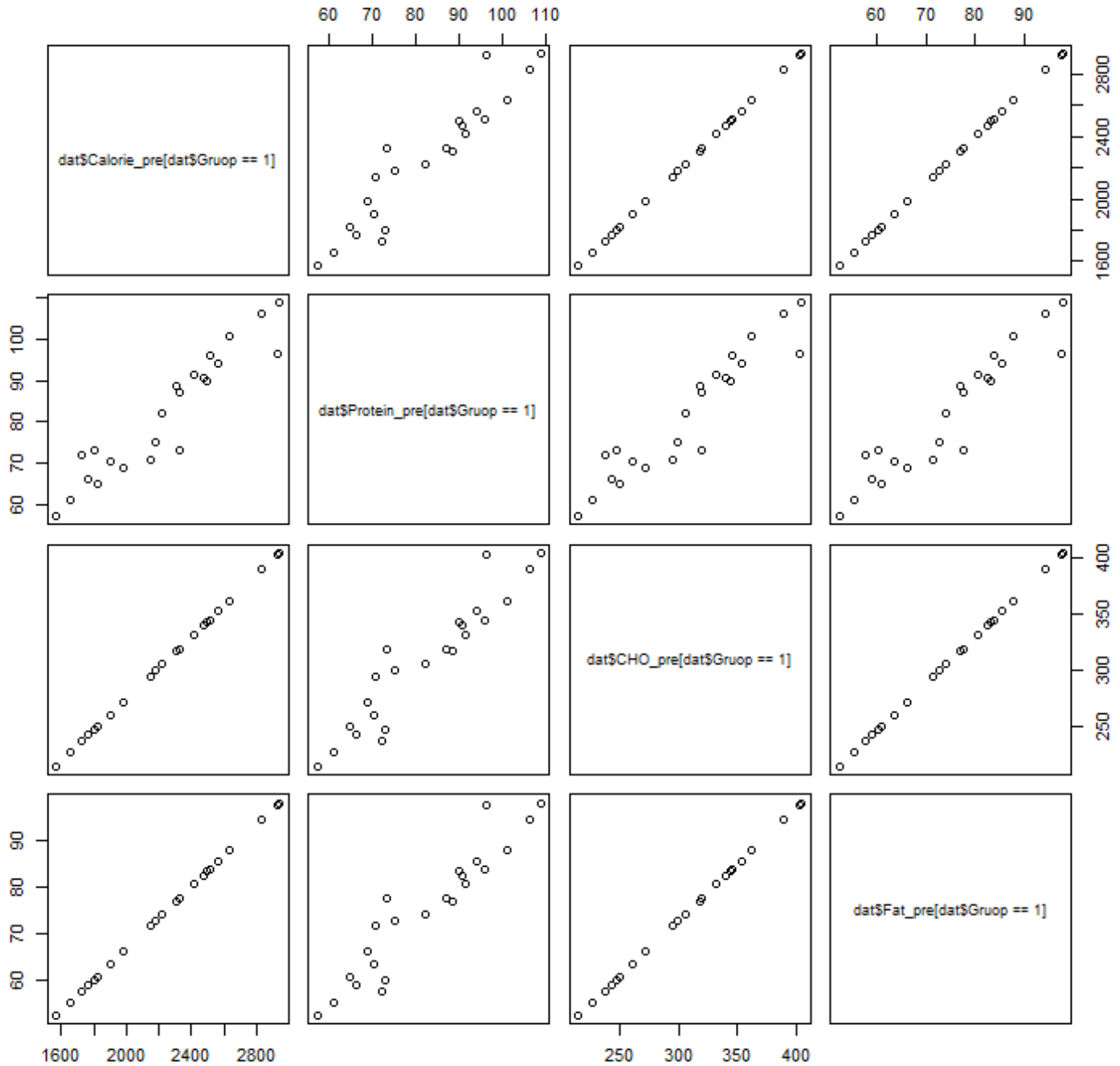
Relationships between variables

- Are expected relationships between variables present?
- Hard to fake (unless you know what you are doing). LLMs don't do a great job.
- Requires contextual knowledge (e.g. should we expect relationship between gestational age and birthweight in a particular trial).
- Don't expect multivariate distribution to differ between randomised groups.





- An example where relationships are too perfect – study of a diet intervention
- Data shown are pre-randomisation dietary variables



- An example where relationships are too perfect – study of a diet intervention
- Data shown are pre-randomisation dietary variables
- E.g. carbohydrate consumption perfectly correlated with fat consumption
- Implies that all participants consume carbohydrate and fat in the same ratio
- Statistical methods such as Principal Components Analysis can be useful in some cases (relationships between variables are too perfect)

The value of raw data

INSPECT IPD

- These could be **easily detected** as being problematic with the IPD
- Suggests journals should request data for review with submissions
- Carlisle – increase in RCTs ‘critically flawed by false data’ when IPD inspected
- Better yet: publication of data with study – detection, deterrence
- How? INSPECT-IPD (Calvin Heal, NIHR Doctoral Research Fellow)

Anaesthesia 2021, 76, 472–479 doi:10.1111/anae.15263

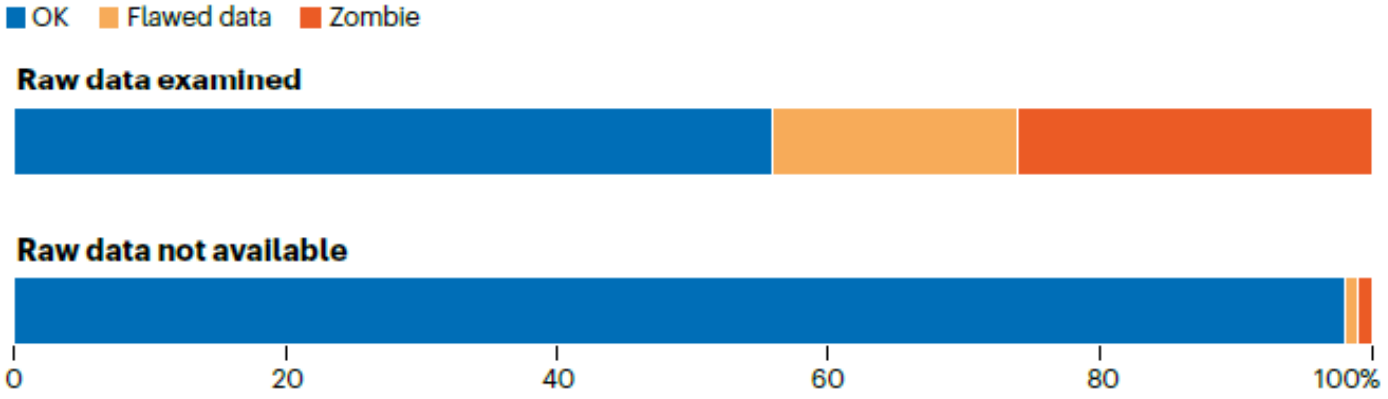
Original Article

False individual patient data and zombie randomised controlled trials submitted to *Anaesthesia*

J. B. Carlisle[1,2]

THE PREVALENCE OF 'ZOMBIE' TRIALS

More than one-quarter of a subset of manuscripts describing randomized clinical trials submitted to the journal *Anaesthesia* between 2017 and 2020 seemed to be faked or fatally flawed when their raw data could be examined, editor John Carlisle reported. He called these 'zombies'. But when their raw data could not be obtained, Carlisle could label only 1% as zombies.

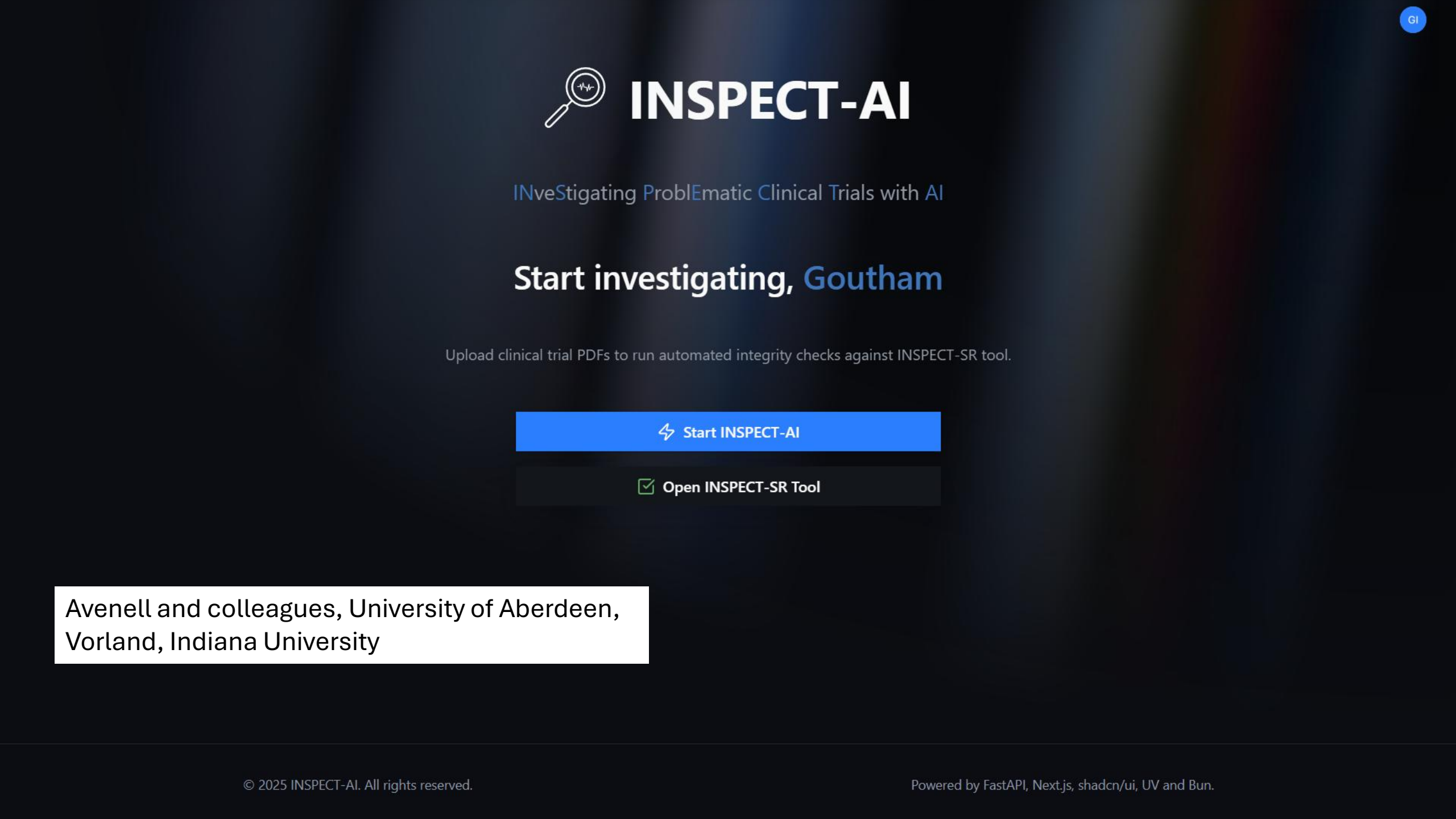


https://www.nature.com/articles/d41586-023-02299-w

The value of raw data

INSPECT IPD

- These could be **easily detected** as being problematic with the IPD
- Suggests journals should request data for review with submissions
- Carlisle – increase in RCTs ‘critically flawed by false data’ when IPD inspected
- Better yet: publication of data with study – detection, deterrence
- How? INSPECT-IPD (Calvin Heal, NIHR Doctoral Research Fellow)



Avenell and colleagues, University of Aberdeen, Vorland, Indiana University

Received: 27 September 2024 | Revised: 6 December 2024 | Accepted: 9 December 2024

DOI: 10.1111/apa.17549

ACTA PÆDIATRICA

WILEY

ORIGINAL ARTICLE

Closed-loop automated oxygen control in late preterm and term, ventilated infants: A randomised controlled trial

Ourania Kaltsogianni[1,2] | Allan Jenkinson[1,2] | Christopher Harris[1,2] | Eleanor Jeffreys[2] | Oishi Sikdar[2] | Anne Greenough[1] | Theodore Dassios[1,2]

[1]Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK

[2]Neonatal Intensive Care Centre, King's College Hospital NHS Foundation Trust, London, UK

Correspondence
Theodore Dassios, Neonatal Intensive Care Centre, King's College Hospital NHS Foundation Trust, Denmark Hill, London SE5 9RS, UK.
Email: theodore.dassios@kcl.ac.uk

Funding information
National Institute for Health Research (NIHR) Biomedical Research Centre; Guy's and St Thomas' NHS Foundation Trust; King's College London

Abstract

Aim: To compare the time spent above the target oxygen saturation range ($SpO_2 > 96\%$) and the duration of supplemental oxygen between ventilated infants receiving closed-loop automated oxygen control (CLAC) or manual oxygen control in late preterm and term ventilated infants.

Methods: Infants were randomised to receive CLAC or manual oxygen control from recruitment and within 24 h of mechanical ventilation until successful extubation.

Results: Forty infants with a median (IQR) gestational age of 37.4 (35.9–38.5) weeks were studied at a corrected postmenstrual age of 37.6 (36.0–38.7) weeks. In infants randomised to CLAC ($n = 18$) the time spent above the target oxygen saturation range was reduced by 20% ($p < 0.001$), and the time spent in the target range (92%–96%) was increased by 32% ($p < 0.001$) and the time spent in hyperoxia was reduced ($p = 0.003$). CLAC reduced the time spent in hypoxemia ($SpO_2 < 85\%$) ($p = 0.017$) and there were fewer manual adjustments to the inspired oxygen concentration (FiO_2) ($p < 0.001$). There was no significant difference in the duration of supplemental oxygen ($p = 0.271$).

Conclusion: CLAC in ventilated infants born at or near term was associated with reduced time spent in hyperoxemia, more time spent in the target oxygen range, and fewer manual adjustments to the FiO_2.

Does the study have an associated expression of concern or other relevant post publication notice?

0	0
EOC/Correction Notices	PubPeer Comments

Expression of Concern & Correction Notices

No Concerns Found

SEARCH METHOD
DOI-based lookup

DOI (USED FOR SEARCH)
10.1111/apa.17549

TITLE
Closed-loop automated oxygen control in late preterm and term, ventilated infants: A randomised controlled trial

PubPeer Community Comments

No PubPeer Comments Found

This publication has no comments on PubPeer

Does the study have an associated expression of concern or other relevant post publication notice?

INSPECT-AI Suggestion: No

Yes | No | Unclear | N/A

Comments

No comments on webpage either

INSPECT-AI · Assessments

Job ID: a125239e-b4de-4757-b906-1e10e28ba4d7

Status: COMPLETED

Report Issue

AA

1 / 7

171%

Check for updates

Received: 27 September 2024 | Revised: 6 December 2024 | Accepted: 9 December 2024

DOI: 10.1111/apa.17549

ORIGINAL ARTICLE

ACTA PÆDIATRICA

WILEY

Closed-loop automated oxygen control in late preterm and term, ventilated infants: A randomised controlled trial

Ourania Kaltsogianni[1,2] | Allan Jenkinson[1,2] | Christopher Harris[1,2] | Eleanor Jeffreys[2] | Oishi Sikdar[2] | Anne Greenough[1] | Theodore Dassios[1,2]

[1]Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK

[2]Neonatal Intensive Care Centre, King's College Hospital NHS Foundation Trust, London, UK

Correspondence
Theodore Dassios, Neonatal Intensive Care Centre, King's College Hospital NHS Foundation Trust, Denmark Hill, London SE5 9RS, UK.
Email: theodore.dassios@kcl.ac.uk

Funding information
National Institute for Health Research (NIHR) Biomedical Research Centre; Guy's and St Thomas' NHS Foundation Trust; King's College London

Abstract

Aim: To compare the time spent above the target oxygen saturation range (SpO_2 > 96%) and the duration of supplemental oxygen between ventilated infants receiving closed-loop automated oxygen control (CLAC) or manual oxygen control in late preterm and term ventilated infants.

Methods: Infants were randomised to receive CLAC or manual oxygen control from recruitment and within 24 h of mechanical ventilation until successful extubation.

Results: Forty infants with a median (IQR) gestational age of 37.4 (35.9–38.5) weeks were studied at a corrected postmenstrual age of 37.6 (36.0–38.7) weeks. In infants randomised to CLAC ($n = 18$) the time spent above the target oxygen saturation range was reduced by 20% ($p < 0.001$), and the time spent in the target range (92%–96%) was increased by 32% ($p < 0.001$) and the time spent in hyperoxia was reduced ($p = 0.003$). CLAC reduced the time spent in hypoxemia (SpO_2 < 85%) ($p = 0.017$) and there were fewer manual adjustments to the inspired oxygen concentration (FiO_2) ($p < 0.001$). There was no significant difference in the duration of supplemental oxygen ($p = 0.271$).

Conclusion: CLAC in ventilated infants born at or near term was associated with reduced time spent in hyperoxemia, more time spent in the target oxygen range, and

Q1.3 Research team concerns Ctrl + /

Do other studies by the research team highlight causes for concern (retraction, expression of concern, relevant post-publication notices)?

7
Authors with No Mentions

0
Authors with Some Mentions

› **Authors without Mentions in Retraction Watch (7)**

Do other studies by the research team highlight causes for concern (retraction, expression of concern, relevant post-publication notices)?

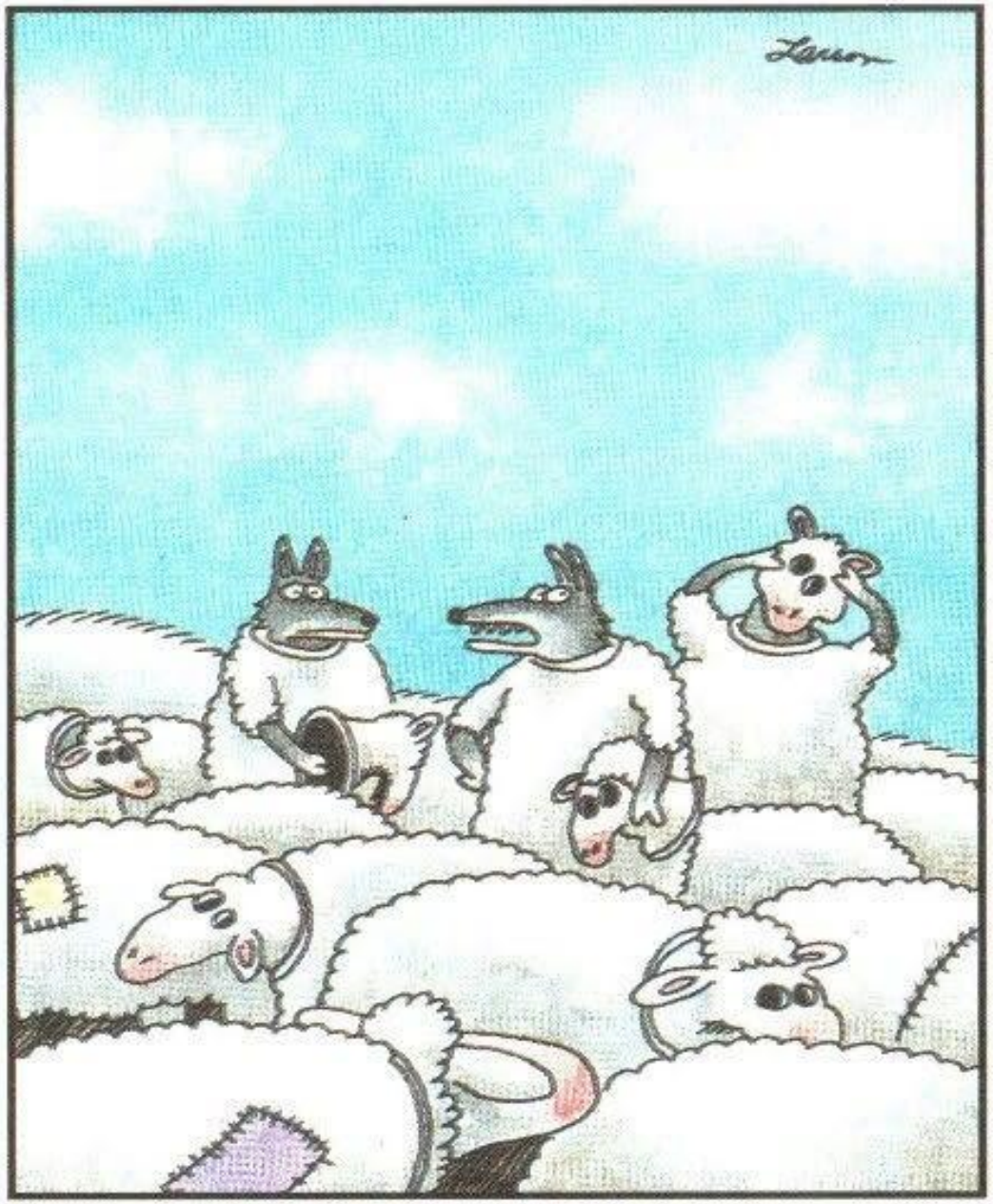
INSPECT-AI Suggestion: No

Yes | No | Unclear | N/A | Clear

Comments

No concerns

Save comment



"Wait a minute! Isn't anyone here a real sheep?"

Source: The Far Side, Gary Larson

What's next?

- INSPECT-SR has been designed for RCTs, post-publication review, assuming no raw data access
- If you have raw data: INSPECT-IPD in development (Calvin Heal)
- Adapting for journal context: INSPECT-JR (Zewen Lu)
- Other study designs (e.g. Nonrandomised studies - INSPECT-NRS)
- Training and continual updating of guidance (new software, new research)

Trustworthiness assessment: everybody's doing it

Created with Google Gemini

- General awareness of research misconduct is probably **higher than ever**
- **This is good** – we want people to routinely consider trustworthiness of evidence
- But this is **new territory** for most people
- And while it isn't too hard to do well, it is **very easy to do it poorly** – so we need to use and promote good methods in our practice and guidance
- **Some** methods endorsed by **some** sleuths are just wrong – (this is not to denigrate the good work of the sleuthing community!) – statistical expertise not well represented
- Let's try to spread awareness of the issue, and also good methods for forensic analysis

INSPECT SR

Thanks, questions?