

Systematic Reviews and Corrections and Retractions

Jack Wilkinson, Centre for Biostatistics, Division of Population Health, Health Services
Research, and Primary Care



@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

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: Fake data to patient care pipeline

1

Attempt to identify all studies 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

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>

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 a systematic review.

EDITORIAL

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

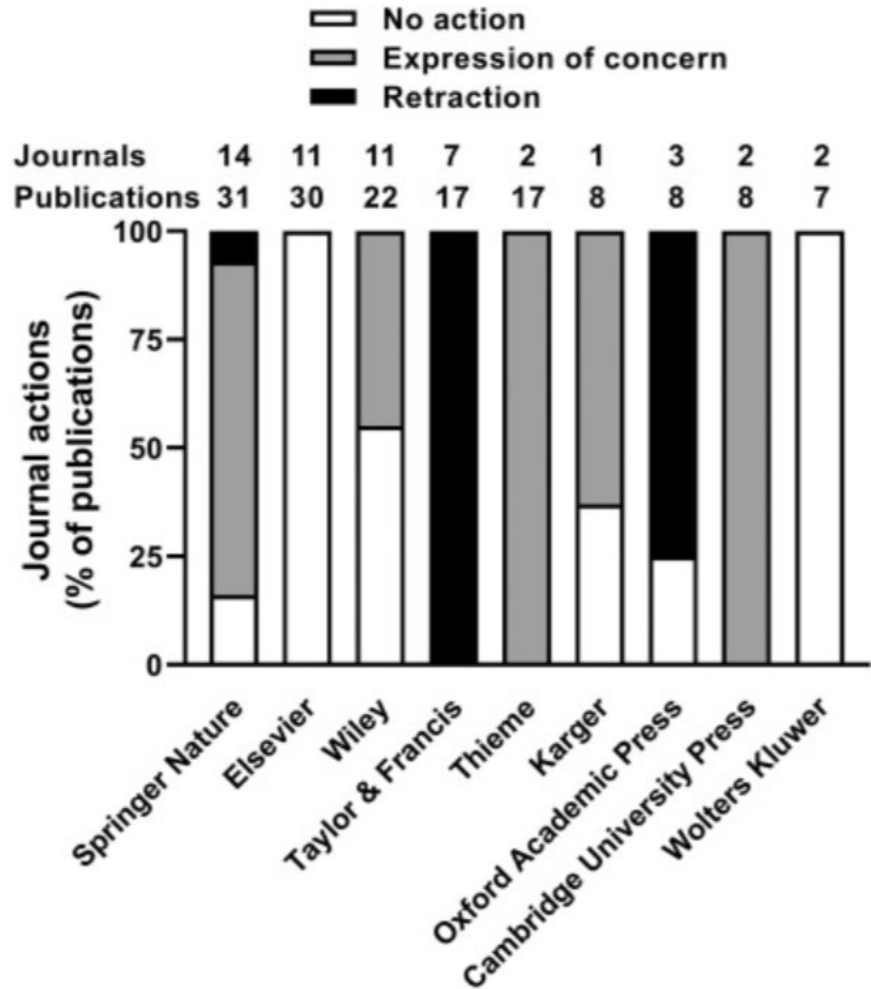
Stephanie L Boughton, Jack Wilkinson, Lisa Bero

Managing potentially problematic studies

<https://bit.ly/3SsJO9F>

- Do not include problematic studies in a systematic review.
- But how do 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

Systematic reviewers may be the first to notice problems!

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

 [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

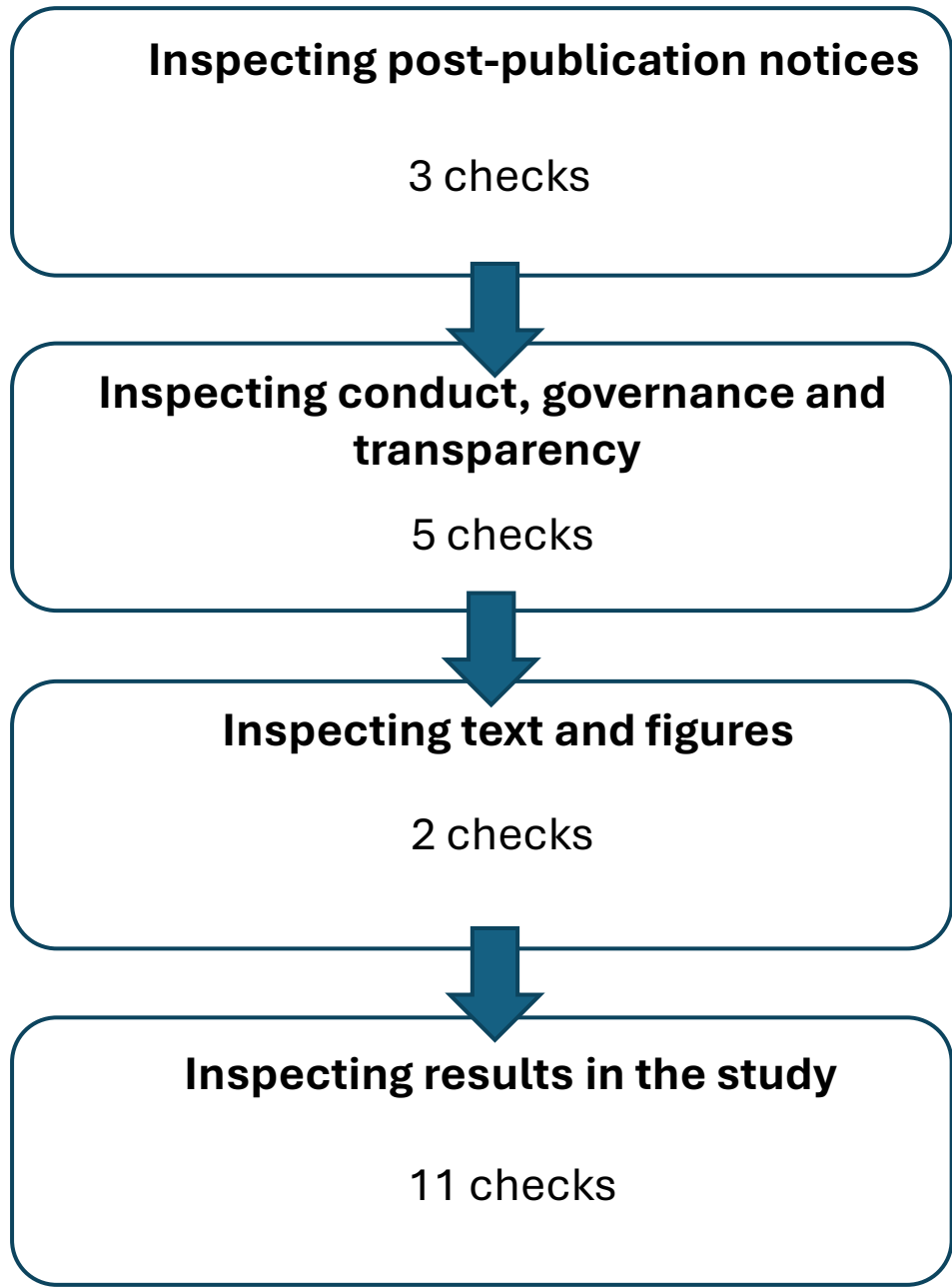
doi: <https://doi.org/10.1101/2025.09.03.25334905>

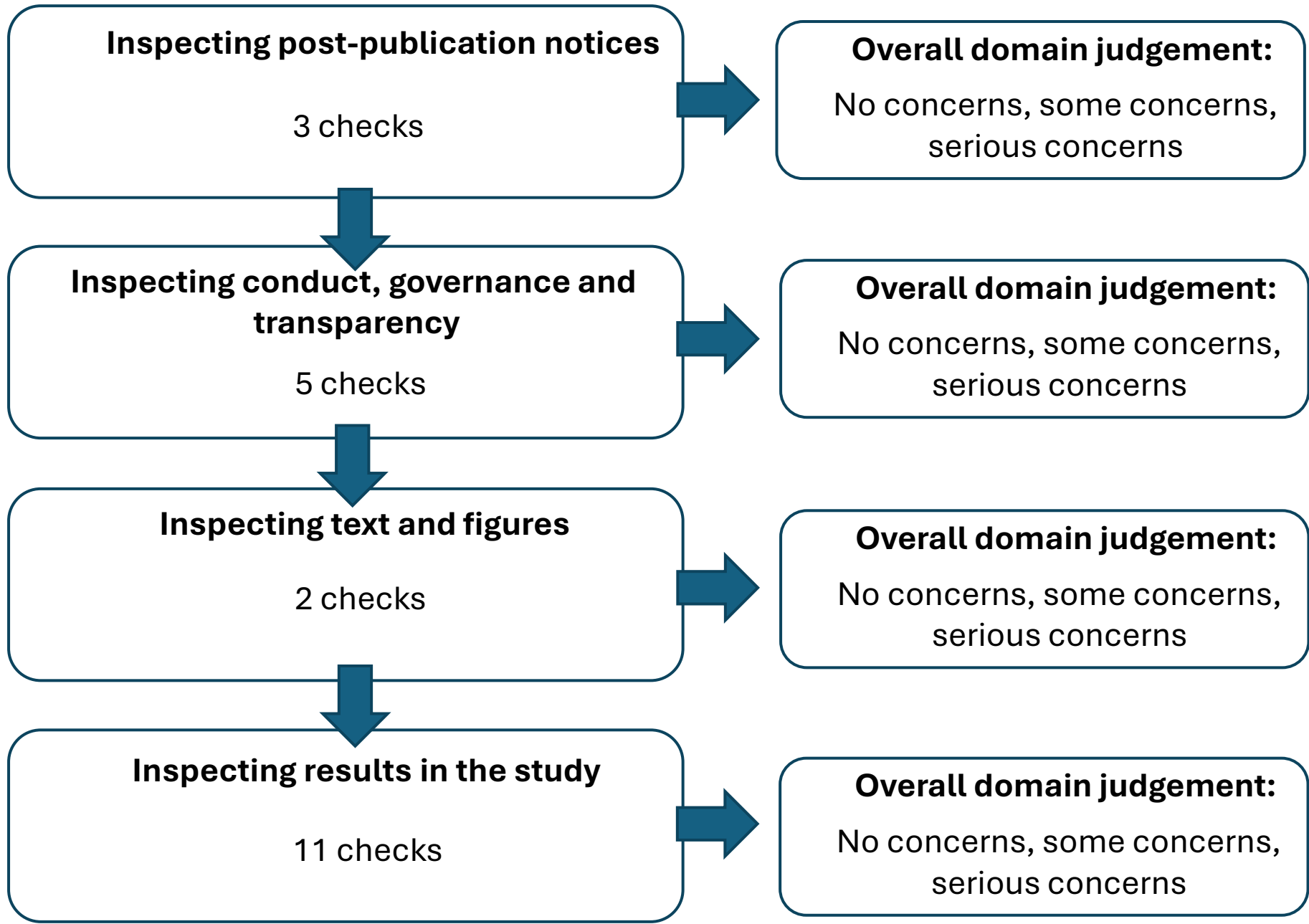
- INSPECT-SR is a tool for assessing trustworthiness of randomised trials in reviews
- Developed using empirical evidence and consensus methods
- An editable template and detailed guidance document are provided: <https://osf.io/b74wj/files>



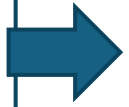
FUNDED BY

NIHR | National Institute for Health and Care Research



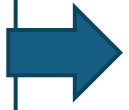


Inspecting post-publication notices
3 checks



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

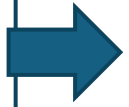


Overall domain judgement:
No concerns, some concerns,
serious concerns



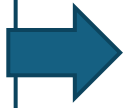
Overall study judgement:
No concerns, some concerns,
serious concerns

Inspecting post-publication notices
3 checks



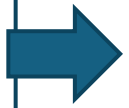
Overall domain judgement:
No concerns, some concerns,
serious concerns

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

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

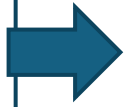


Overall domain judgement:
No concerns, some concerns,
serious concerns



Overall study judgement:
No concerns, some concerns,
serious concerns

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

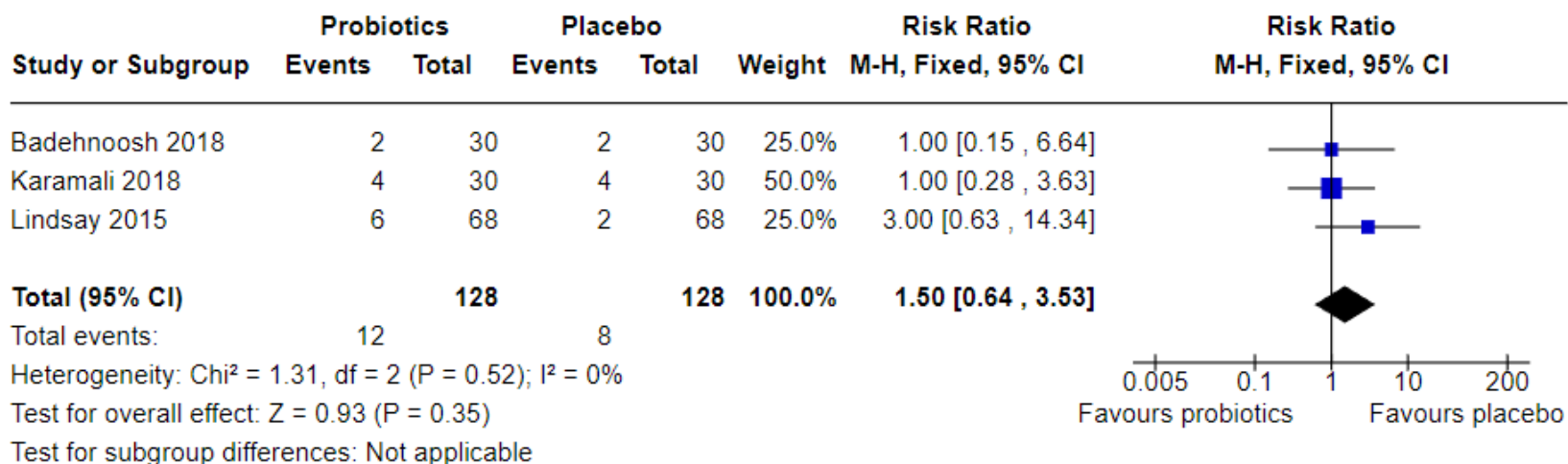
Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

✉ 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>



Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias): All outcomes
Blinding of outcome assessment (detection bias): All outcomes
Incomplete outcome data (attrition bias): All outcomes
Selective reporting (reporting bias)
Other bias

Ahmadi 2016	+	?	+	+	+	?	+
Badehnoosh 2018	+	?	+	+	+	?	+
Hajifaraji 2017	+	?	+	+	+	?	+
Jafarnejad 2016	+	?	+	+	+	?	+
Karamali 2016	+	?	+	+	+	?	+
Karamali 2018	+	?	+	+	+	?	+
Kijmanawat 2019	+	?	+	+	+	?	+
Lindsay 2015	+	+	+	+	+	?	+
Nabhani 2018	+	?	+	+	+	?	+

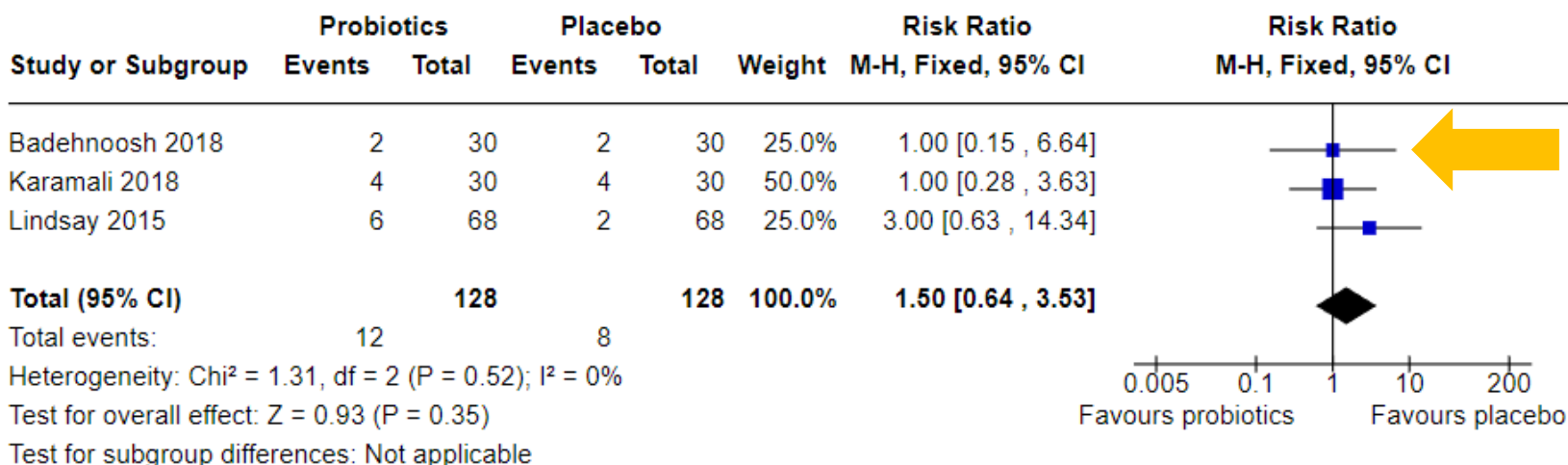
Probiotic treatment for women with gestational diabetes to improve maternal and infant health and well-being

✉ 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>



Random sequence generation (selection bias)
 Allocation concealment (selection bias)
 Blinding of participants and personnel (performance bias): All outcomes
 Blinding of outcome assessment (detection bias): All outcomes
 Incomplete outcome data (attrition bias): All outcomes
 Selective reporting (reporting bias)
 Other bias

Ahmadi 2016	+	?	+	+	+	?	+
Badehnoosh 2018	+	?	+	+	+	?	+
Hajifaraji 2017	+	?	+	+	+	?	+
Jafarnejad 2016	+	?	+	+	+	?	+
Karamali 2016	+	?	+	+	+	?	+
Karamali 2018	+	?	+	+	+	?	+
Kijmanawat 2019	+	?	+	+	+	?	+
Lindsay 2015	+	+	+	+	+	?	+
Nabhani 2018	+	?	+	+	+	?	+

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

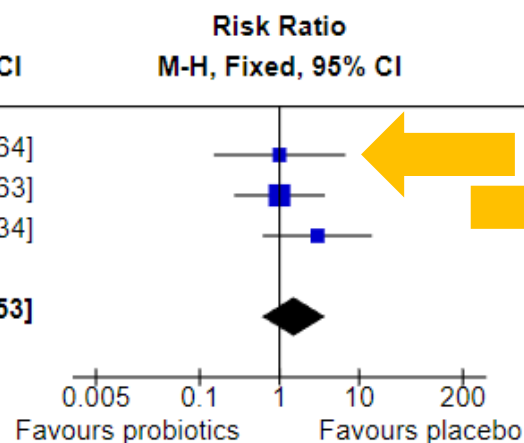
✉ Karaponi AM Okesene-Gafa, Abigail E Moore, Vanessa Jordan, Lesley McCowan, Caroline A Crowther

Authors' declarations of interest

Version published: 24 June 2020 Version history

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

Study or Subgroup	Probiotics		Placebo		Weight	Risk Ratio
	Events	Total	Events	Total		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: Chi ² = 1.31, df = 2 (P = 0.52); I ² = 0%						
Test for overall effect: Z = 0.93 (P = 0.35)						
Test for subgroup differences: Not applicable						



Random sequence generation (selection bias)
 Allocation concealment (selection bias)
 Blinding of participants and personnel (performance bias): All outcomes
 Blinding of outcome assessment (detection bias): All outcomes
 Incomplete outcome data (attrition bias): All outcomes
 Selective reporting (reporting bias)
 Other bias

Ahmadi 2016	+	?	+	+	+	?	+
Badehnoosh 2018	+	?	+	+	+	?	+
Hajifaraji 2017	+	?	+	+	+	?	+
Jafarnejad 2016	+	?	+	+	+	?	+
Karamali 2016	+	?	+	+	+	?	+
Karamali 2018	+	?	+	+	+	?	+
Kijmanawat 2019	+	?	+	+	+	?	+
Lindsay 2015	+	+	+	+	+	?	+
Nabhani 2018	+	?	+	+	+	?	+

Inspecting post-publication notices

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/>

Author(s):	<input type="text" value="Asemi, Zatollah"/>	Country(s):	<input type="text"/>
Title:	<input type="text" value="Type to search"/>		
Reason(s) for Retraction:	<input type="text"/>		
Subject(s):	<input type="text"/>	Article	<input type="text"/>
	<input type="text"/>	Type(s):	<input type="text"/>
Journal:	<input type="text"/>		
Publisher:	<input type="text"/>		
Affiliation(s):	<input type="text"/>		
Notes:	<input type="text"/>		
URL:	<input type="text"/>		

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/>

Author(s):	<input type="text" value="Asemi, Zatollah"/>	Country(s):	<input type="text"/>
Title:	<input type="text" value="Type to search"/>		
Reason(s) for Retraction:	<input type="text"/>		
Subject(s):	<input type="text"/>	Article Type(s):	<input type="text"/>
Journal:	<input type="text"/>		
Publisher:	<input type="text"/>		
Affiliation(s):	<input type="text"/>		
Notes:	<input type="text"/>		
URL:	<input type="text"/>		

Retraction or Other Notices
Title/Subject(s)/Journal --- Publisher/Affiliation(s)/Retraction Watch Post URL(s)
50 Items Displayed Out of 72 Item(s) Found

Inspecting results in the study

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

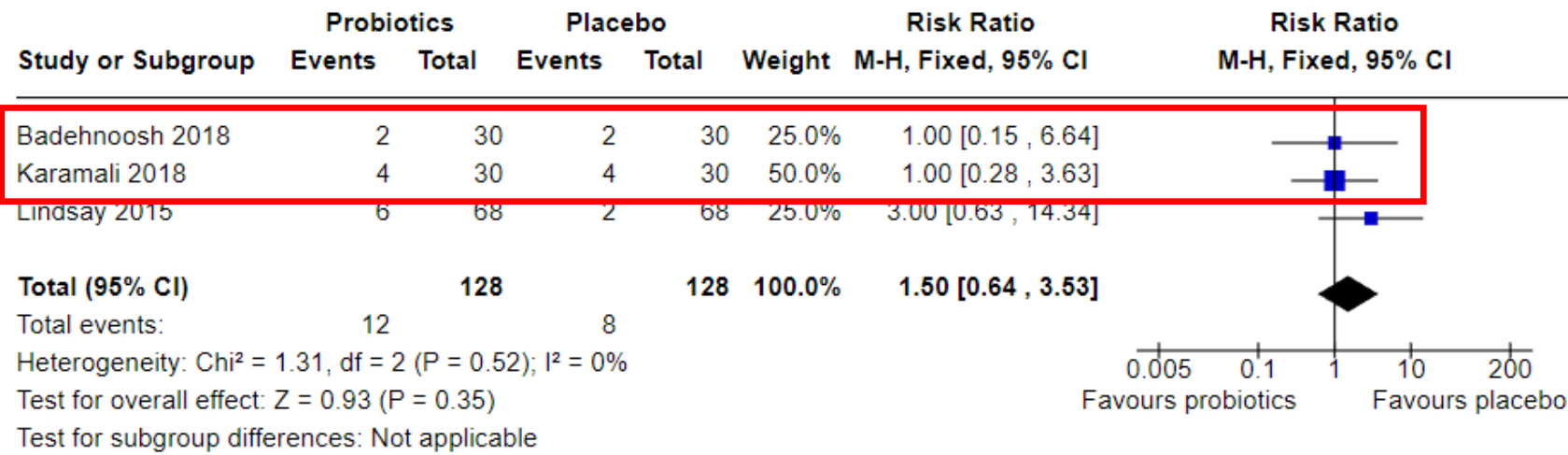
Inspecting results in the study

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, GRIMMER).

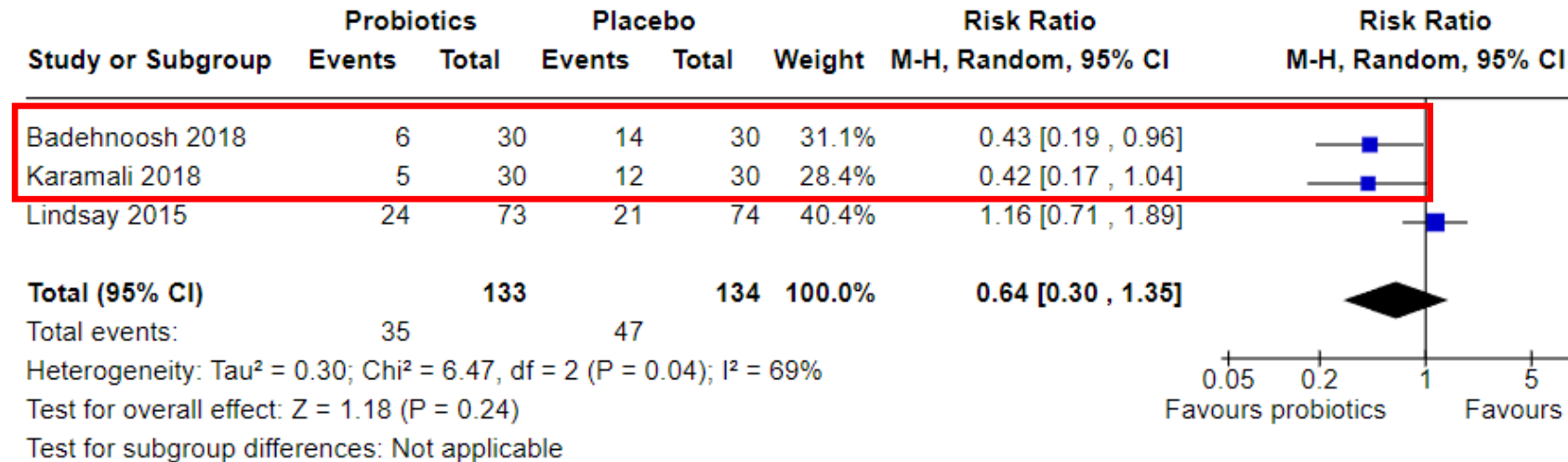
Hypertensive disorders



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

- Karamali 2018 is another trial from same group.

Caesarian deliveries



- Almost identical risk ratios in the two studies (2 here, several other outcomes have identical effect estimates).

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

Key points

- Reviewers look to see whether a study is retracted/ has expression of concern when deciding whether to include in a systematic review
- So it is crucial that these notices are **prompt** and **visible**: expressions of concern when investigations are underway – need for **clarity** and **consistency** in these notices
- Because retraction status is not reliable, reviewers need to assess trustworthiness using other markers
- Assessing trustworthiness takes time – funders need to recognise this and allow additional time/ resource
- Systematic reviews offer a good opportunity to identify problems – can compare across studies
- Systematic reviewers can report concerns to journals/ publishers – (Cochrane guidance, templates)