

Ottawa Hospital Research Institute

OHRI



IRHO

Institut de recherche de l'Hôpital d'Ottawa

Does journal endorsement of reporting guidelines influence the completeness of reporting of health research?

Larissa Shamseer, MSc Clin Epi

Reporting Guideline Research Coordinator

11 October 2012, Freiburg, DE

AN INSTITUTE OF • UN INSTITUT DE





Project Investigators

- **David Moher**, OHRI, Canada
- **Doug Altman**, The EQUATOR Network, Centre for Statistics in Medicine (CSM), University of Oxford, UK
- **Allison Hirst**, CSM, University of Oxford, UK
- **John Hoey**, Queen's University, Kingston, Canada
- **Anita Palepu**, St. Paul's Hospital, Vancouver, Canada
- **Ken Schulz**, FHI360, North Carolina, USA
- **Iveta Simera**, The EQUATOR Network, Centre for Statistics in Medicine (CSM), University of Oxford, UK
- **Project Funding: CIHR (FRN 111750)**



Research Staff

Staff Leads: Adrienne Stevens & Larissa Shamseer

Hadeel AlYacoob

Raymond Daniel

Mary Gauthier

Mandi Kaur

Misty Pratt

Kavita Singh

Becky Skidmore

Justin Thielman

Sophia Tsouros

Lucy Turner

Erica Weinstein

Fatemeh Yazdi

Review Objectives

To evaluate whether journal endorsement of reporting guidelines influences the completeness of the reported literature by comparing:

1. Completeness of reporting of studies published in journals endorsing reporting guidelines **before and after endorsement**.
2. Completeness of reporting of studies published in journals that **have and have not endorsed** reporting guidelines.

Methods

- **Search:** (1) reporting guidelines (2) evaluations of RGs
- **Screening:** determine relevant evaluations, journal endorsement status for comparisons
- **Validity assessment:** 6 pre-specified criteria
- **Outcome extraction and analyses**
 - obtain data for all included studies
 - Completeness of reporting – RRs/SMDs (99% CIs), where possible
(by item, total score, flow diagram, descriptive)
 - Methodological quality - narrative
 - Unwanted effects - narrative

Methods (2)

- **Statistical heterogeneity:** I^2 statistic
- **Subgroup analyses:**
 - Extent of endorsement
 - Variations in checklist items
 - Official and unofficial extensions
- **Sensitivity analyses:**
 - Six-month endorsement period
 - Study outliers

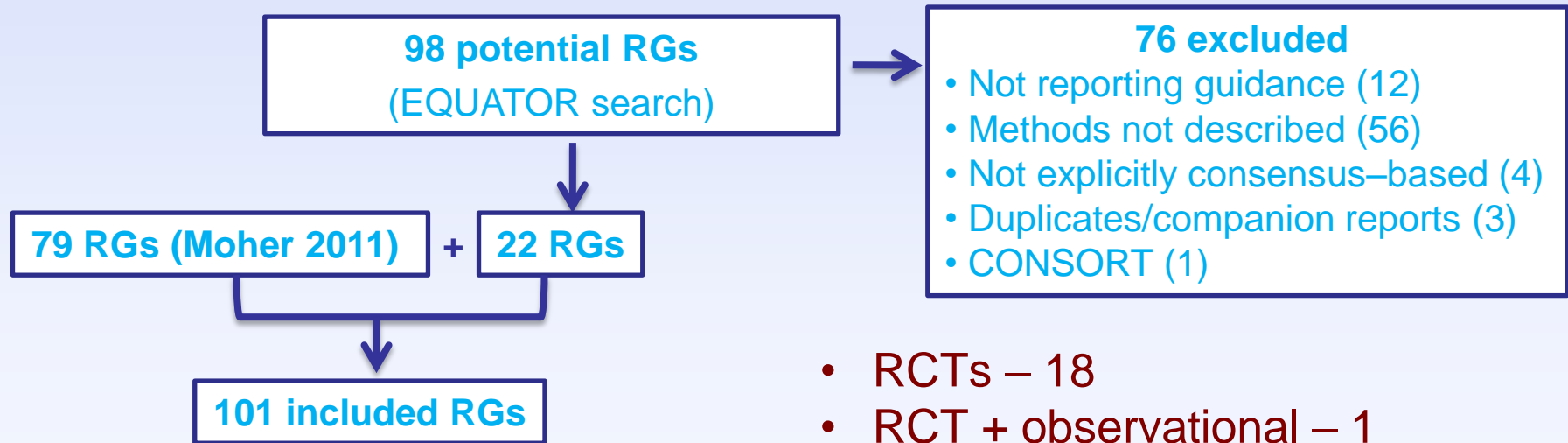
Eligibility criteria - inclusions

- **Primary intent: completeness of reporting**
 - Exclude methodologic quality / quality assessment
 - Exclude SRs addressing clinical Q
- **Evaluates one of 101 RGs of interest**
 - Exclude those that do not explicitly state assessing RG
- **At least one study in journal that endorses RG**
 - Exclude evaluations of conference abstracts
 - Contacted authors
 - Follow-up with journals

Eligibility criteria - exclusions

- Inappropriate use of RG
 - Eg, STROBE for abstracts
 - Eg, STARD for reviews
- Assesses studies before RG published
 - Eg, used PRISMA on reviews published before 2009
- Interrupted series designs, controlled before-after analyses

PRISMA flow - Included RGs



- General – 5
- Animal – 2
- Basic science – 1
- Lab/pre-clinical – 10
- Case reports – 1
- Observational – 9
- Non-randomized – 1
- Prospective studies – 5
- General, clinical trials – 14
- Various content areas – 13

- RCTs – 18
- RCT + observational – 1
- RCTs, quasi, SRs, m/a – 1
- SRs + m/a – 3
- Diagnostic studies – 4
- Economics – 5
- Evaluation studies – 2
- Qualitative – 2
- QI studies – 1
- Standardized patient reports – 1
- Methodology – 1
- Validation studies – 1

**Database records
(n = 17225)**

**Other source records
(n = 40)**

PRISMA flow – Included Evaluations

**After duplicates removed
(n = 15262)**

**Duplicate & multiple reports
quarantined (n=22)**

Records screened (n = 15240)

Records excluded (n = 14096)

Full-text - round 1 (n = 1144)

**Full-text excluded, with
reasons (n = 826)**

Full-text - round 2 (n = 318)

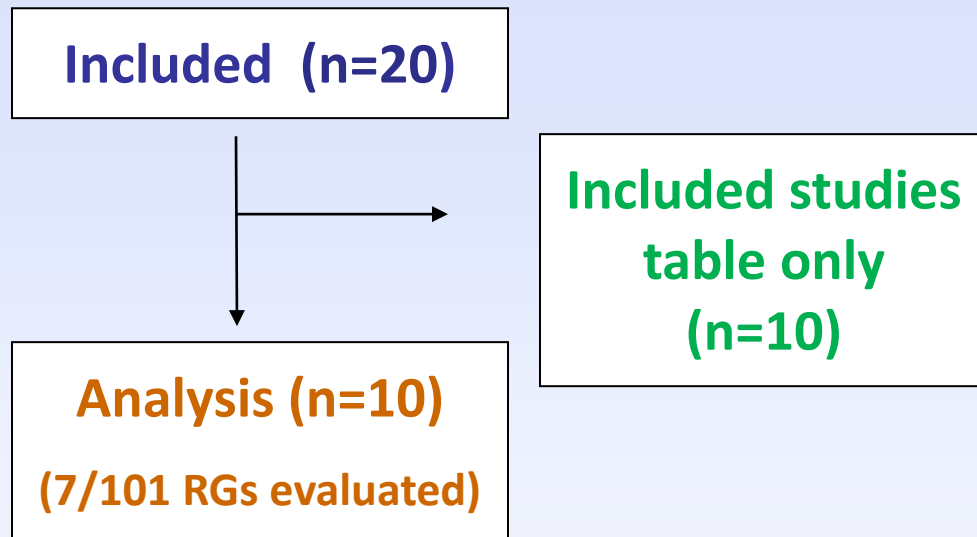
**Full-text excluded, with
reasons (n = 159)**

**Contact authors & journals /
Modified checklists (n = 159)**

**Full-text excluded, with
reasons (n = 139)**

Included (n=20)

Final included evaluations



1. CONSORT for harms, 2004 (n=3): **Haidich 2011, Turner 2011, Lee 2008**
2. CONSORT for herbal medicines, 2006 (n=1): **Ernst 2011**
3. BMJ economic checklist, 1996 (n=2): **Herman 2005, Jefferson 1998**
4. QUOROM, 1999 (n=3): **Hind 2007, Biondi Zoccai 2006, Poolman 2007**
5. STARD, 2003 (n=8): **Freeman 2009, Mahoney 2007, Selman 2011, Smidt 2006, Coppus 2006, Johnson 2007, Krzych 2009, Paranjothy 2007**
6. STRICTA, 2002 (n=1): **Hammerschlag 2011**
7. STROBE, 2007 (n=2): **Parsons 2011, Delaney 2010**

Characteristics of studies

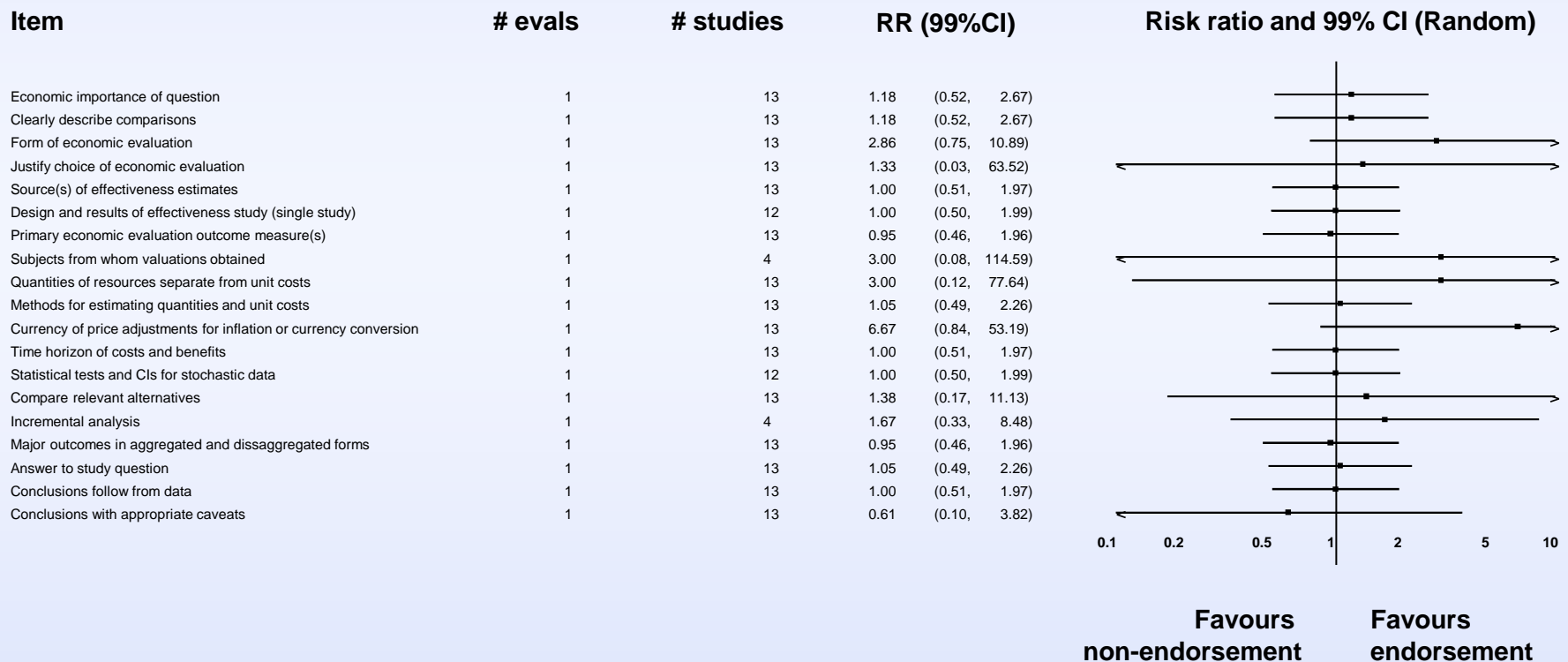
- 7 countries:
 - UK, US, Canada, Netherlands, Greece, Italy, Poland
- Study funding:
 - Not reported, Government, Charity, Industry, No funding, Other
- Range of interventions
 - Diagnostic studies, Complementary medicine, Drug therapies, Therapeutic interventions (general), Surgery, Biochem and Lab Methods, Unspecified
- Range of medical specialties
 - Unspecified, Medicine (General/Internal), Obstet/Gyne, Orthopedics, Cardiac/Cardiovascular, Endocrinology & Metabolism, Hematology, Neurology, Reproductive, Urology/Nephrology

Included evaluations

Guideline	Endorsers v. Non-endorsers			Before v. After Endorsement		
	# evals	# journals	# studies	# evals	# journals	# studies
BMJ econ	2	2 vs. 11	3 vs. 16	1	1	1 vs. 8
CONSORT - Harms	3	8 vs. 108	31 vs. 267	1	1	1 vs. 2
CONSORT - Herbals	1	1 vs. 3	1 vs. 4	-	-	-
QUOROM	2	2 vs. 11	2 vs. 12	1	1	13 vs. 15
STARD	8	26 vs. 71	133 vs. 168	2	8	98 vs. 79
STRICTA (2002)	1	3 vs. 64	17 vs. 129	1	2	11 vs. 4
STROBE	2	3 vs. 9	10 vs. 42	1	2	9 vs. 11

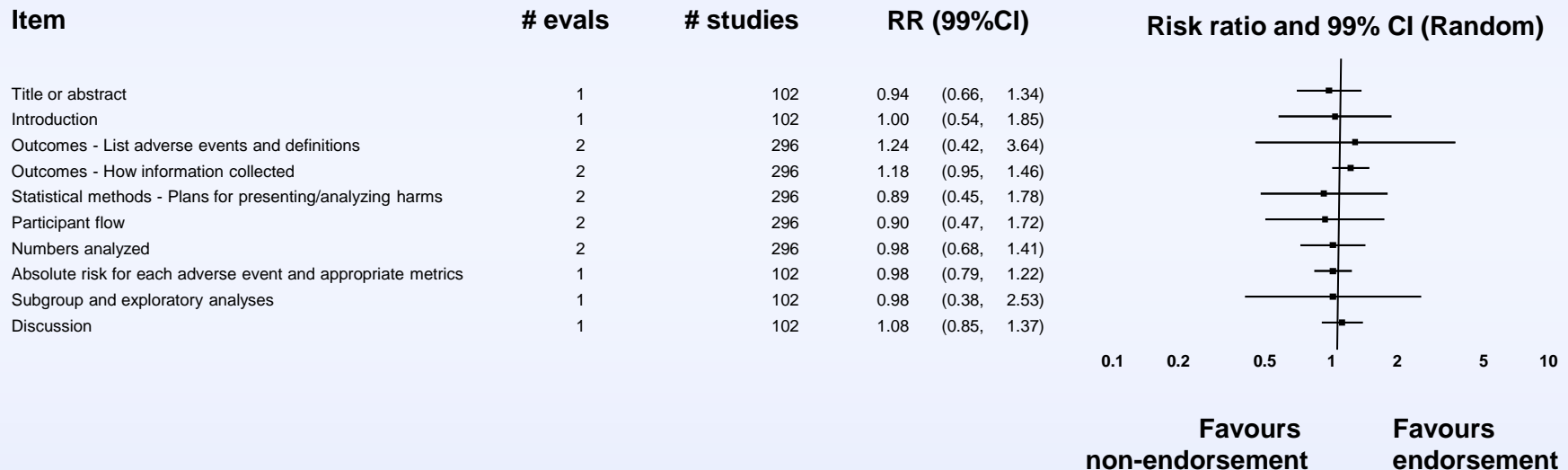
Endorser vs. Non-endorser

BMJ Economics guidelines



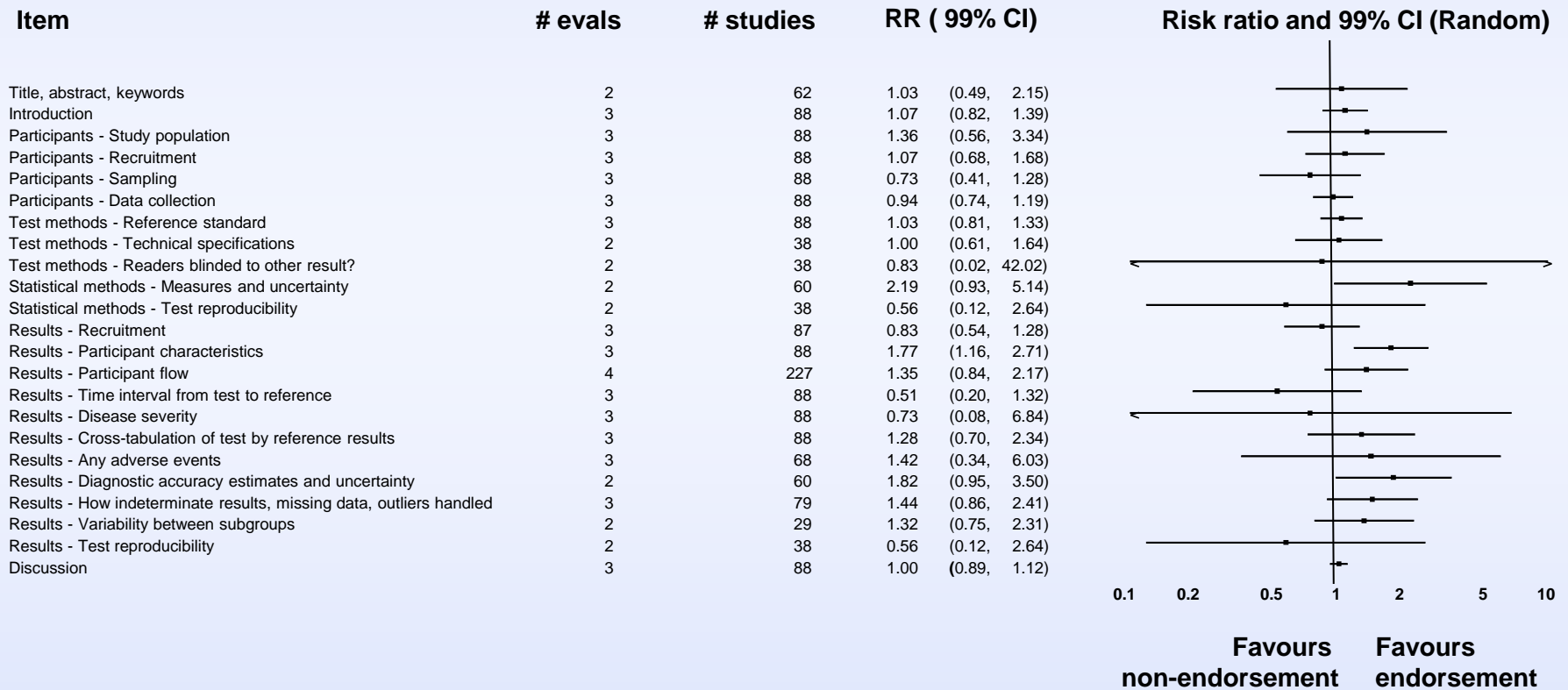
Endorser vs. Non-endorser

CONSORT for Harms



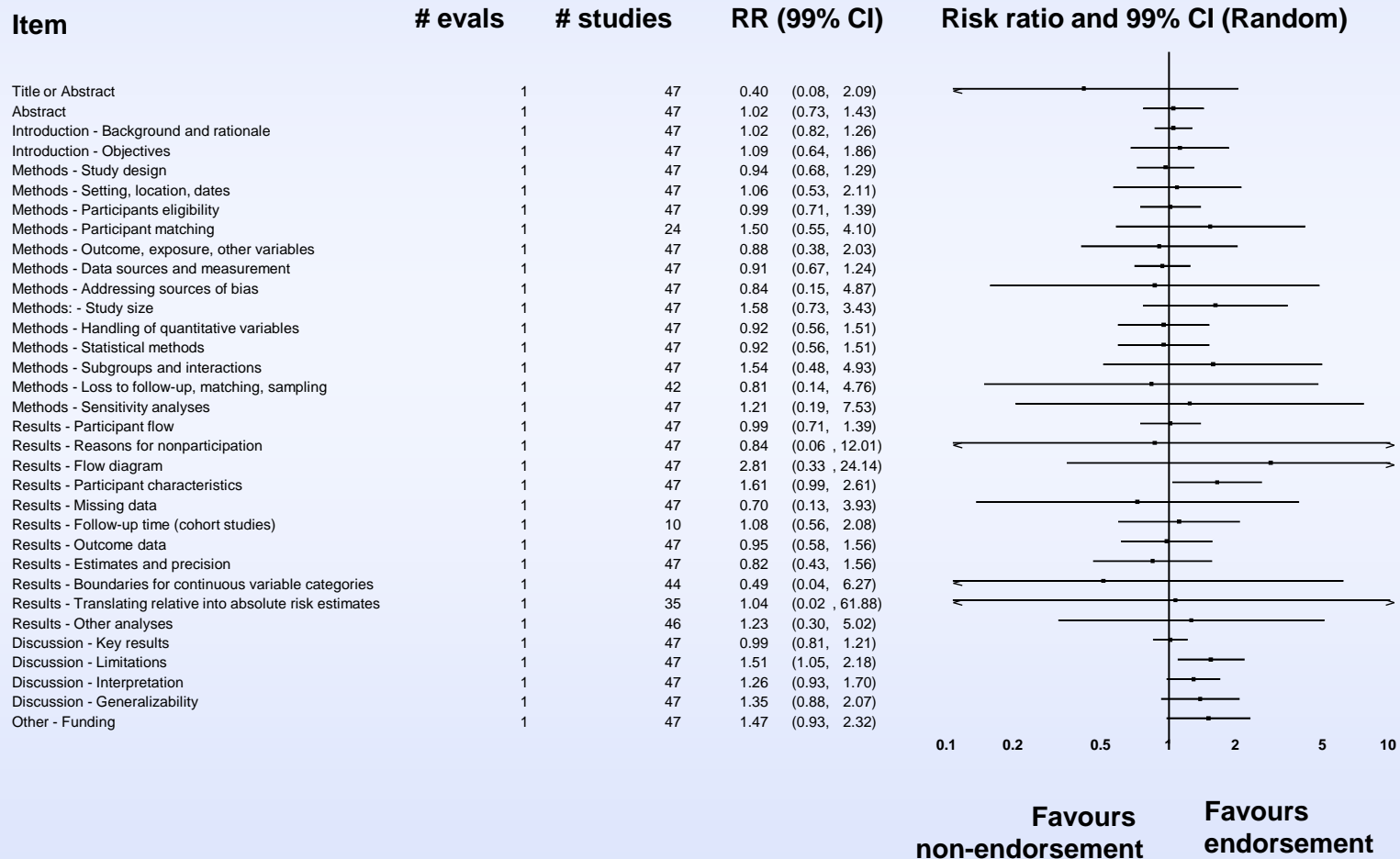
Endorser vs. Non-endorser

STARD



Endorser vs. Non-endorser

STROBE



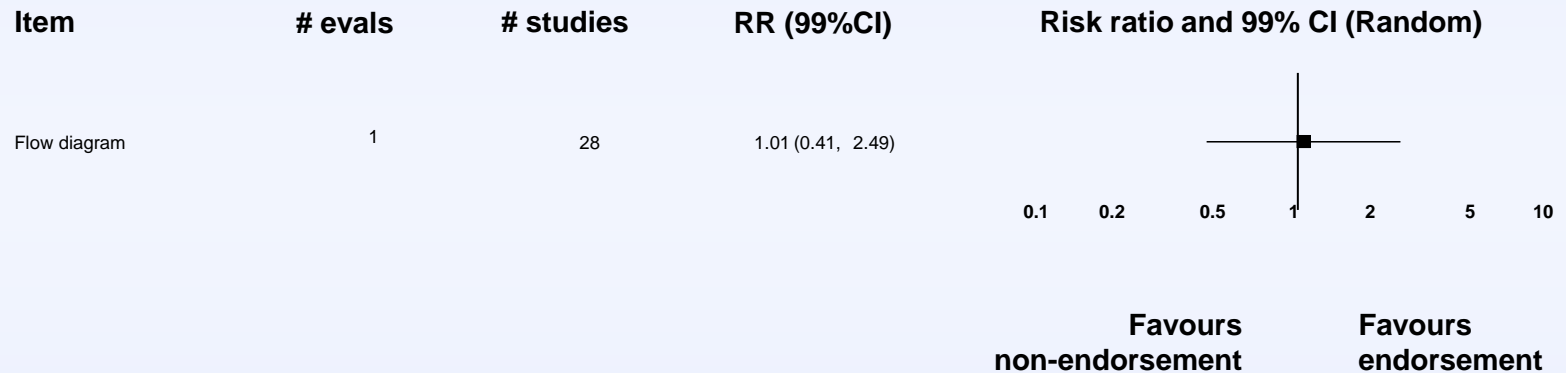
Endorser vs. Non-endorser

Total sum score of items

Guideline	Evaluation	Studies	SMD (99% CI)
BMJ economics guideline, 1996	1	2 vs. 11	3.47 (0.58 to 6.35)
CONSORT for harms, 2004	2	30 vs. 266	-0.06 (-0.59 to 0.47)
STARD, 2003	3	88	0.55 (-0.08 to 1.18)
STRICTA, 2002	1	17 vs. 129	1.35 (-0.11 to 2.81)
STROBE, 2007	1	9 vs. 38	0.33 (-0.63 to 1.29)

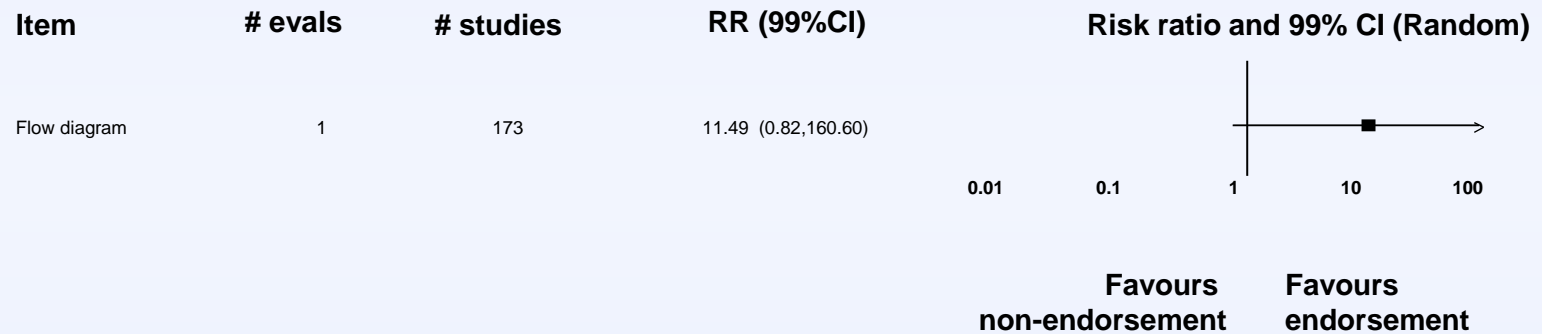
After vs. Before Endorsement

QUOROM



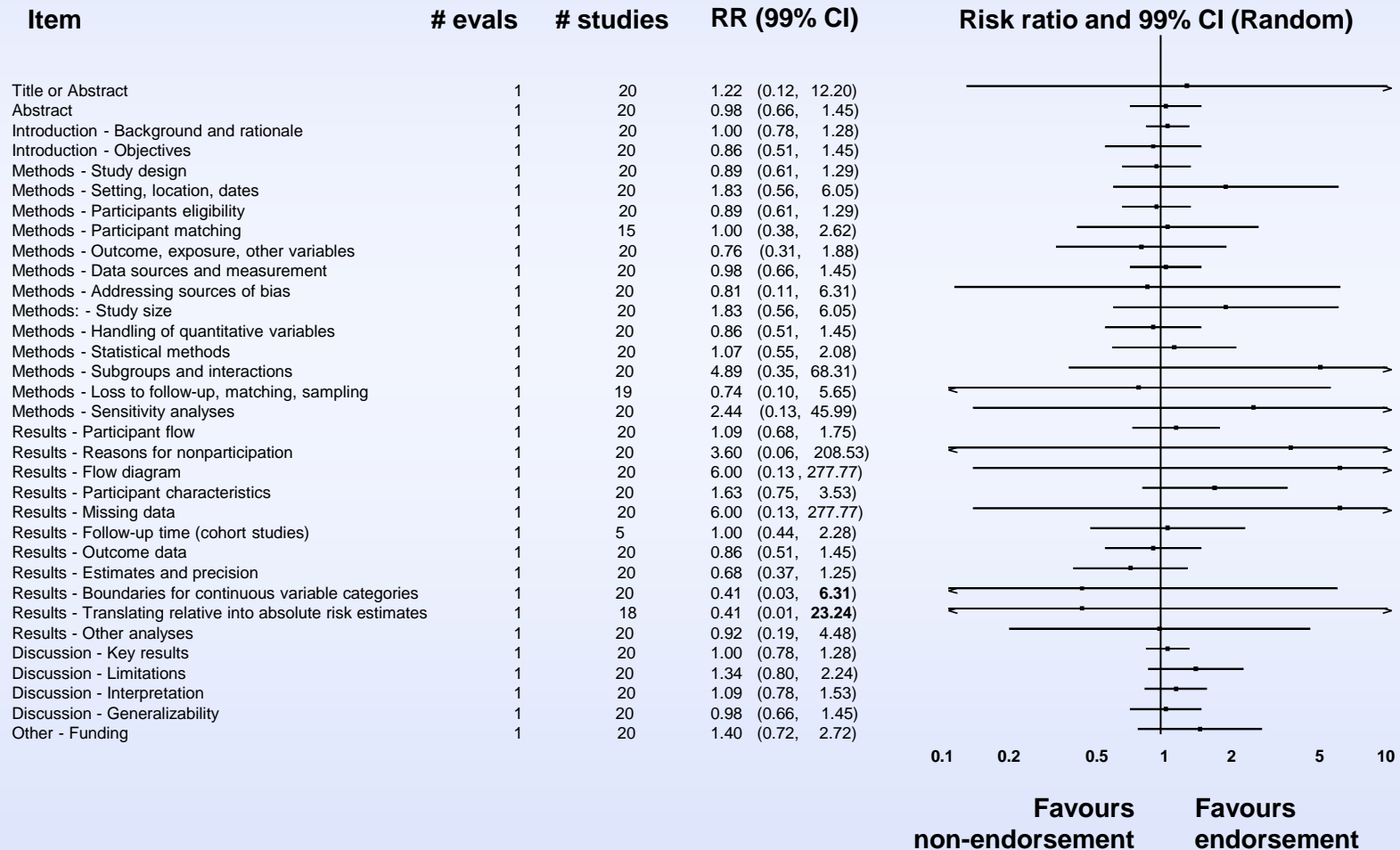
After vs. Before Endorsement

STARD



After vs. Before Endorsement

STROBE



After vs. before endorsement

Total sum score of items

Guideline	Evaluation	Studies	SMD (99% CI)
STRICTA, 2002	1	11 vs. 4	0.63 (-0.91 to 2.17)
STROBE, 2007	1	9 vs. 11	0.26 (-0.90 to 1.42)

Secondary outcomes

- Methodological quality of evaluations' included studies
 - 6 of 20 studies reported information
 - 2 studies used Oxman-Guyatt
 - 1 studies used Cochrane Risk of Bias tool
 - 1 study assessed generalizability and reporting
 - 1 study used QUADAS but could not extract as per our comparisons
 - 1 study assessed internal and external validity
- Unwanted effects from use of guideline
 - No studies assessed

Results

- Only 7/101 RGs evaluated
- Few existing evaluations, each assessing completeness of reporting of a small number of studies; subgroup & sensitivity analyses not possible
 - Few studies assessed methodological quality
 - No studies assessed unwanted effects

vs.

- CONSORT SR (50 evaluations, >16000 RCTs)
 - 25 of 27 items favoured endorsement vs. non-endorsement with 5 items statistically significant
 - 5 items stat sig for after-before endorsement

Strengths and Limitations

- **Strengths:**
 - First cumulative evaluation of RG effectiveness
 - Used best comparisons from CONSORT SR
 - Reworked authors' data to enable comparisons
- **Limitations:**
 - Were all potential evaluations located?
 - For some, 'partial' reporting may have been included in 'complete' data by authors
 - Not able to obtain all data – authors' availability to answer queries, some items considered 'not applicable'

Discussion

- Why so few evaluations?
 - We couldn't put a drug to market without evaluating it!
 - Potentially more evaluations that we wished to include, but could not
- What is the intervention?
 - Journal endorsement (not just guideline publication) is where the rubber-meets-the-road
 - Extent of endorsement – needs to be clear and have an implementation strategy

Published protocol

Shamseer L, Stevens A, Skidmore B, Turner L, Altman DG, Hirst A, Hoey J, Palepu A, Simera I, Schulz KF and Moher D. Does journal endorsement of reporting guidelines influence the completeness of reporting of health research? A systematic review protocol. *Systematic Reviews* 2012, 1:24 doi:10.1186/2046-4053-1-24

<http://www.systematicreviewsjournal.com/content/1/1/24/abstract>



Changes from protocol

- Did not include CONSORT
 - Refer to CONSORT SR
- Did not include variations in checklist items
- Changes to validity assessment
 - 2 items clarified wording
 - Confounding

Collate reports /
Contact authors,
journals/
Assess modified
checklists (n=159)

Full-text excluded (n=139):

- Evaluates CONSORT (n=45)
- Multiple report of excluded study (n=27)
- No response from author to determine eligibility (n=19)
- No journals endorsed the RG (n=14)
- Authors used modified checklist (n=9)
- List of studies or journals not provided (n=6)
- Inappropriate use of RG (n=3)
- Comparison not possible given study dates and endorsement information (n=3)
- Assessed studies before RG published (n=2)
- Full report or information not in English or French (n=2)
- Could not locate author contact information (n=2)
- All journals endorsed the RG, comparison not relevant (n=2)
- Did not address completeness of reporting (n=1)
- Full report or information not available from the author (n=1)
- Reporting guideline publication (n=1)
- Could not locate journal website or instructions to authors (n=1)
- Not comparison of interest (n=1)

Included (n=20)

Validity – Endorsing vs Non-endorsing

Author, Year	Studies	Two or more assessors for reporting	No evidence of selective reporting	Comprehensive search strategy	Balance of studies per journal
BMJ economic guidelines, 1996					
Herman, 2005	2 vs. 11	Unclear	High	Low	High
Jefferson, 1998	1 vs. 5	Unclear	Unclear	High	High
CONSORT extension for harms, 2004					
Haidich, 2011	25 vs. 77	High	High	High	Low
Lee, 2008	1 vs. 1	High	High	High	High
Turner, 2011	5 vs. 189	Low	High	Low	Low
CONSORT extension for herbal interventions, 2006					
Ernst, 2011	1 vs. 4	Unclear	High	Low	High
QUOROM, 1999					
Biondi-Zoccai, '06	1 vs. 6	High	High	High	High
Poolman, 2007	1 vs. 6	High	Unclear	Low	High
STRICTA, 2002					
Hammerschlag, '11	17 vs. 129	Low	High	Low	Unclear
STROBE, 2007					
Delaney, 2010	1 vs. 4	High	Unclear	Low	High
Parsons, 2011	9 vs. 38	Low	Unclear	Low	Low

Validity – Endorsing vs Non-endorsing

Author, Year	Studies	Two or more assessors for reporting	No evidence of selective reporting	Comprehensive search strategy	Balance of studies per journal
STARD, 2003					
Coppus, 2006	8 vs. 19	Low	High	Low	High
Freeman, 2009	3 vs. 9	Unclear	High	High	High
Johnson, 2007	1 vs. 10	High	High	Low	High
Krzych, 2009	4 vs. 21	Unclear	High	Low	High
Mahoney, 2007	6 vs. 20	High	High	Low	High
Paranjothy, 2007	1 vs. 8	High	High	Low	High
Selman, 2011	15 vs. 35	High	Low	Low	Low
Smidt, 2006	95 vs. 46	High	High	Low	Low

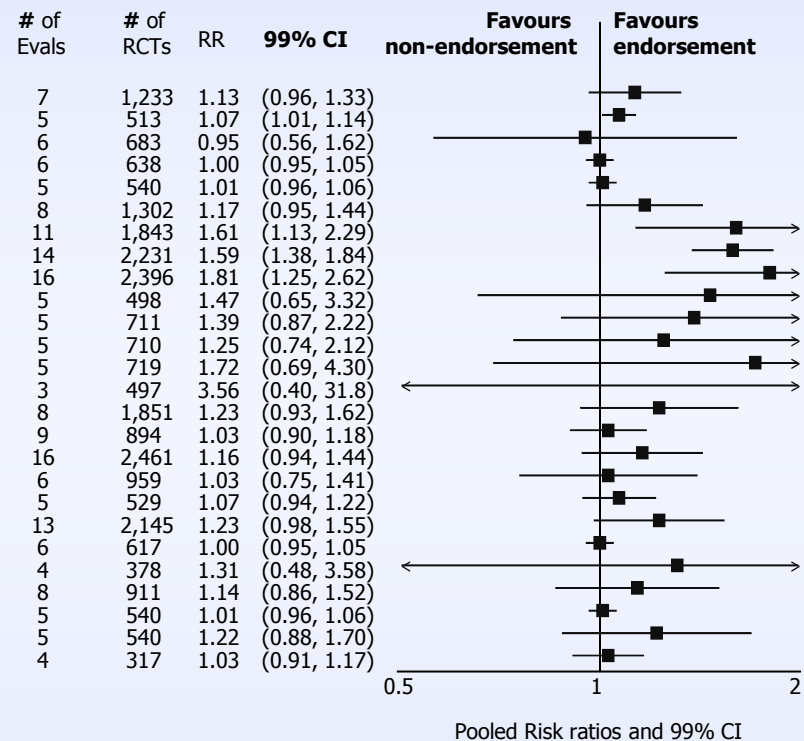
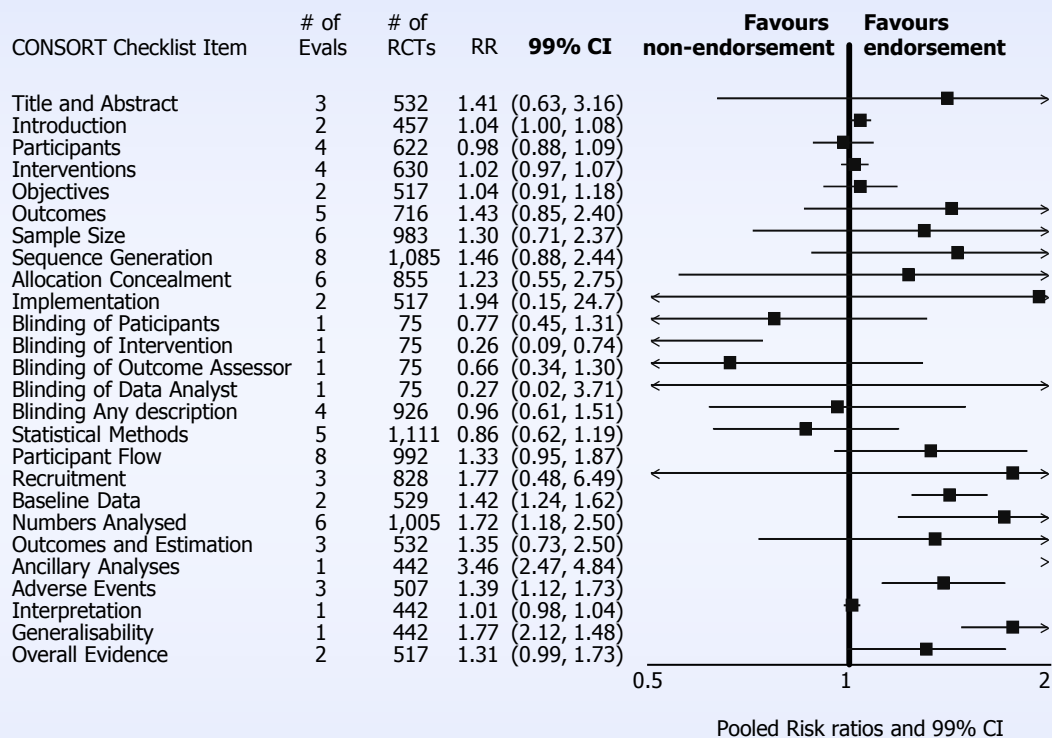
Validity – After vs. before endorsement

Author, Year	Studies	Two or more assessors for reporting	No evidence of selective reporting	Comprehensive search strategy	Balance of studies per journal	Sampling took place after RG publication
BMJ economic guidelines, 1996						
Jefferson, 1998	1 vs. 8	Unclear	Unclear	High	High	High
CONSORT extension for harms, 2004						
Lee, 2008	1 vs. 2	High	High	High	Low	Low
QUOROM, 1999						
Hind, 2007	13 vs. 15	Low	High	Unclear	High	High
STARD, 2003						
Selman, 2011	3 vs. 1	High	Low	Low	Low	High
Smidt, 2006	95 vs. 78	High	High	Low	Low	Low
STRICTA, 2002						
Hammerschlag, 2011	11 vs. 4	Low	High	Low	High	Low
STROBE, 2007						
Parsons, 2011	9 vs. 11	Low	Unclear	Low	Low	Low

CONSORT SR – by item

After vs. Before

Endorser vs. Non



- Stat sig: Introduction, Baseline data, Numbers analyzed, Ancillary analyses, Adverse events, Generalizability, Sum score

- Stat sig: Alloc conceal, Introduction, Sample size, Sequence generation, Sum score