#### Opening the black box of peer review



**Tom Jefferson** 

jefferson.tom@gmail.com

#### Content

- Origins
- Evidence of effects
- Gaps in the evidence base
- What is quality?
- Reporting bias & the publications industry
- Way forward and RIAT
- Alternatives to peer review





(Lo screening meraviglioso)

By

Thomas Jefferson Jr MD Founder, President and CEO PharmaTom Inc



- Universal screening device
- Distinguishes the good from the bad
- Franchising network
- Global market (60,000 + sites plus 1,000 more per year)



Extensively trialled!!!!

- 9 RCTs (n=2,540) testing whether users could guess which packet the instruction were in
- 2 before & after studies on checklist for instructions (n=568)



Extensively trialled!!!!

- 2 RCTs on readibility of instructions
- 1 RCT on attitudes to Wonderscreen<sup>®</sup> by male residents of Goa aged 65 and above



Extensively trialled!!!!

- 2 RCTs on dissemination of instructions by electronic vs paper means
- 1 comparative study on validity of Wonderscreen<sup>®</sup>



### An Invitation

#### Wonderscreen®

My company would be honoured if you would join us and become one of the franchising sites



## Peer review should identify studies which are:

- Important
- Useful
- Relevant
- Methodologically sound
- Ethical
- Complete
- Accurate

Outcome / definition	Ideal indicator	Surrogate indicators
Important	- Changes in health status	- Citation rates
Study findings have a	- Changes in healthcare	- Media coverage
major impact on health or	delivery	- Correspondence
healthcare		
Useful	- Contributes significantly	- Contributes to non-
Study contributes	within a systematic review	systematic reviews or
significantly to the	of the topic	guidelines
scientific debate or	- Narrows CIs around	- Citation rates
knowledge on a subject	estimates of effect	- Correspondence
Relevant	- Topic is relevant and	- Citation rates
Topic is relevant to the	consistent with the aims	- Correspondence
journal's aims and	and readership of the	- Internet hit rates
readers	journal confirmed by survey	
Methodologically sound	- Study findings are	- Closeness of fit between
Methods used are able to	replicated several times	methods and 'evidence-
answer the study question	across different settings	based' methodological
		checklist
		- Correspondence

Outcome / definition	Ideal indicator	Surrogate indicators
Ethically sound Unnecessary harm to humans or animals has been avoided Study has been carried out and reported honestly	<ul> <li>No divergence between reality and the report.</li> <li>Rights of humans and animals safeguarded</li> <li>Privacy and informed consent maintained throughout</li> <li>Raw data match presented data</li> </ul>	<ul> <li>Study received ethical clearance</li> <li>No complaints from participants</li> <li>No duplicate publication</li> </ul>
<b>Complete</b> All relevant information is presented	<ul> <li>There is no selective presentation of data</li> <li>All relevant references are cited</li> </ul>	<ul> <li>The text is complete</li> <li>The publication is complete (ie not salamisliced)</li> </ul>
Accurate Presented information is a true reflection of what went on	<ul> <li>Measurements truly reflect magnitude of findings</li> <li>Raw data match presented data</li> <li>References are accurate</li> </ul>	<ul> <li>The figures add up</li> <li>Corrections</li> </ul>

## Reboxetine vs placebo and/or SSRIs for depression Unpublished Published

Trial	Reboxetine (n/N)	Placebo (n/N)	Odds ratio (95% CI)
Patients with	adverse events		
014	84/126	78/128	
091	24/28	13/28	
015	71/112	58/112	
046	239/264	208/254	<b></b>
047	225/258	201/252	<b></b>
050	138/150	117/150	
045	68/89	52/87	
049	98/106	77/104	
Total	947/1133	804/1115	-
Total heteroge	neity: I <sup>2</sup> =44.0%, P=0.085; tot	al effect: P(0.001	
Withdrawalow	wing to adverse events		
014	14/126	15/128	c
091	1/28	1/28	≺ →
015	11/112	7/112	
046	26/264	9/254	· · · · · · · · · · · · · · · · · · ·
047	20/258	10/252	
050	27/150	12/150	
045	15/89	7/87	
049	23/106	3/104	
Total	137/1133	64/1115	-
Total heteroge	neity: I <sup>2</sup> =38.4%, P=0.124; tot	al effect: P<0.001	
		0.	10 0.33 0.50 1 2 3 10

Edying et al BMJ 2010

Towards an interdisciplinary approach to peer review

**Control worse** 

**Reboxetine worse** 

Abandoned trials - Distortion

Drugs for which negative outcomes (adverse events or lack of efficacy) were discovered using company data

- Reboxetine (Edronax; Pharmacia-Pfizer)
- Oseltamivir (Tamiflu; Roche)
- Gabapentin (Neurontin; Parke-Davis-Pfizer)
- Rofecoxib (Vioxx; Merck)
- Rosiglitazone (Avandia; GSK)
- Oseltamivir (Tamiflu, Roche)

Source: Doshi, Del Mar & Jefferson PLOSMed 2012

Information that was missed without access to internal company files on Tamiflu

- Total number of trials done on topic
- Adverse events not reported in articles
- Adverse events classified as "complications"
- Trials published 10 years after completion
- Trial details vital to interpretation
- Authorship of reports

Source: Doshi et al PloS Med 2012



#### EMA's release of regulatory data: trust but verify

http://www.ema.europa.eu/docs/en\_GB/document\_li brary/Other/2014/10/WC500174796.pdf





#### Paper needed to print oseltamivir study WP16263 (courtesy of Peter Doshi)

D	
50	

8545	
8000	
7000	
6000	
5000	
4000	
3000	
2000	
1000	1 12

#### RIAT



#### RIAT

**RIAT Audit Record (RIATAR)** 

A tool for documenting the transformation from regulatory documents to journal publication, based on the CONSORT 2010 checklist of information to include when reporting a randomised trial\*





## Salient aspects of the current editorial peer review system

- quality assurance through experts' opinions
- managing competition for publication space
- the scholarly task of improving scientific knowledge

Are we being honest about the aims of journal peer review?

- Protect journal's reputation (*it ain't me guv*)
- Make journal more interesting
- Reduce work of in-house editors
- Provide acceptability for commerciallyfunded studies
- Tool for academic promotion system

## What are the alternatives?

- No change
- Free for all (electronic, paper)
- Pre-publication/post-publication
- Closed (autarchic) p.r.
- Data extraction
- CSR linked commentaries abandoning competition for space



## Background

- Peer review is seen as a key process in guaranteeing quality of published material
- "Every scientist has a story to tell about the inequities of the peer review system" – Drummond Rennie
- Do the benefits outweigh the harms?

## Inclusion - types of studies

Reports of original research submitted to biomedical journals:

- randomised/quasi-randomised controlled trials
- interrupted time series
- before and after studies
- other observational studies where there was some attempt to control for confounding
- <u>Excluded</u>: surveys comparing editorial practice or editorial outcomes with characteristics of journals or reviewers

#### Inclusion - types of intervention

- Different ways of
  - Screening submissions
  - Assigning submissions
  - Masking submissions
  - Eliciting internal opinions (i.e. within publisher)
  - Eliciting external opinions
  - Making decisions on whether to publish
  - Feeding back to authors and making revisions
- Combinations of the above
- Anything else we hadn't thought of in the list that might be called peer review

#### Results

- 19 included studies
  - 11 randomised
  - 8 non-randomised

### Discussion

- Small amount of research compared to the use and power of ed. peer review
- Concentration of research on processes, both for questions and the outcomes measured
- Limitation to biomedical publications

#### Conclusions

- Very limited evidence that peer review improves quality of publications
- No evidence that blinding/masking has a major effect, and it is difficult to achieve
- Checklists may improve consistency

#### Difficulties encountered

- Definition of objectives of peer review
- Definition of processes
- Definition of outcomes acceptable degree of surrogacy

#### (1) P.r. should identify submissions that are:

Outcome / definition	Ideal indicator	Surrogate indicators
Important	- Changes in health status	- Citation rates
Study findings have a major	- Changes in healthcare delivery	- Media coverage
impact on health or healthcare		- Correspondence
Useful	- Contributes significantly within a	- Contributes to non-systematic
Study contributes significantly	systematic review of the topic	reviews or guidelines
to the scientific debate or	- Narrows CIs around estimates of	- Citation rates
knowledge on a subject	effect	- Correspondence
Relevant	- Topic is relevant and consistent with	- Citation rates
Topic is relevant to the	the aims and readership of the journal	- Correspondence
journal's aims and readers	confirmed by survey	- Internet hit rates
Methodologically sound	- Study findings are replicated several	- Closeness of fit between
Methods used are able to	times across different settings	methods and 'evidence-based'
answer the study question		methodological checklist
		- Correspondence

#### (2) P.r. should identify submissions that are:

Outcome / definition	Ideal indicator Surroy		gate indicators	
Ethically sound	- No divergence between reality and		- Study received ethical	
Unnecessary harm to	the report.		clearance	
humans or animals has	- Rights of humans and animals		- No complaints from	
been avoided	safeguarded		participants	
Study has been carried out	- Privacy and informed consent		- No duplicate	
and reported honestly	maintained throughout		publication	
	- Raw data match presented data			
	- Number preference check is neg	gative		
Complete	- There is no selective presentation	on of	- The text is complete	
All relevant information is	data		- The publication is	
presented	- All relevant references are cited		complete (ie not	
			salami-sliced)	
Accurate	- Measurements truly reflect mag	nitude	- The figures add up	
Presented information is a	of findings		- Corrections	
true reflection of what went	- Raw data match presented data			
on	- References are accurate			

### Outcomes and quality measures

#### Importance of findings

- Ideal indicator: *change in health status*
- 1st rank surrogate: citation rate
- 2nd rank surrogate: *correspondence*
- 3rd rank surrogate: *reviewer agreement*
- Process centred: use of checklist

## The reality

Masking	9 studies	Time taken; Constructiveness; Courtesy; Acceptance rates; Authors' views; Use of supporting evidence
Interactions with reviewers	Callaham; Strayhorn; Neuhauser	Acceptance rates; Congruence with editors' views; Timeliness
Checklists	Gardner; Jefferson	Study design; stats presentation; Quality of econ submissions (no effect)

Internet (open) review	Bingham	Timeliness; Etiquette; Use of supporting references
Bias	Ernst	Bias against unconventional treatments
Before/after (accepted papers)	Goodman; Pierie	Readability; Readers' views; Experts' views
Studies in P-R <i>cf</i> non P-R jnls	Elvik	Retrospective, non-randomised cohort

#### Review showed that:

- Most studies have been process-centred and used surrogate outcome measures
- One study with broader aims had serious methodological weaknesses
- Two studies showing effects of peer review considered only accepted papers
- Most aspects of journal peer review remain untested and unproven

#### Conclusions

- Unless we define the aims of peer review we cannot measure its quality
- Studies have largely been processcentred
- Current practice is largely empirical
- Journal peer review is only one part of the scientific process
- It may not be the best model for all types of biomed publishing

Are we being honest about the aims of journal peer review?

- Protect journal's reputation (*it ain't me guv*)
- Make journal more interesting
- Reduce work of in-house editors
- Provide acceptability for commerciallyfunded studies
- Tool for academic promotion system

### What are the alternatives?

- No change
- Free for all (electronic, paper)
- Pre-publication/post-publication
- Closed (autarchic) p.r.
- Data extraction

## Overall effect of peer review

- Elvik 1998
  - Comparing studies in peer reviewed journals with similar studies in other journals
  - No clear differences in study validity
- Goodman 1994
  - Before and after study at Annals of Internal Medicine on 111 manuscripts
  - Improved quality of reporting, but reliability of scoring low
- Pierie 1996
  - Assessment by journal readers of quality of submitted and accepted versions of 50 articles
  - Improved overall quality

# Effect of blinding/masking in peer review

- 9 studies
  - No convincing evidence that blinding/masking improves the quality of the publication
  - Evidence that reviewers produce more courteous reports when their name is to be revealed
  - Blinding is probably difficult to achieve

### Usefulness of checklists

- Gardner 1990
  - Statistical refereeing with the use of a checklist improved statistical quality
- Jefferson 1998
  - Publication of BMJ guidelines for economic submissions
  - No evidence of improved quality of economic submissions

#### Presumed aims of peer review

- Select 'good' research
- Improve:
  - usefulness
  - comprehensibility
  - accuracy
  - relevance

for healthcare workers

- Reject research / reports that are:
  - misleading
  - unsound
  - weak/ trivial
  - ?fraudulent
  - ?redundant

## Journal peer review is part of the scientific process

- Funding review in
- Protocol review
- Ethical review
- Informal review
- Journal review

importance/ methods methods

ethical soundness

relevance / context

quality of reporting accuracy, complet., copy-editing