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
Towards an interdisciplinary approach to peer review
Wonderscreen®
(Lo screening meraviglioso)

By
Thomas Jefferson Jr MD
Founder, President and CEO
PharmaTom Inc

Towards an interdisciplinary approach to peer review
Wonderscreen®

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

Towards an interdisciplinary approach to peer review
Wonderscreen®

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)
Wonderscreen®

Extensively trialled!!!!

• 2 RCTs on readability of instructions
• 1 RCT on attitudes to Wonderscreen® by male residents of Goa aged 65 and above
Wonderscreen®

Extensively trialled!!!!

• 2 RCTs on dissemination of instructions by electronic vs paper means
• 1 comparative study on validity of Wonderscreen®

Towards an interdisciplinary approach to peer review
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
<table>
<thead>
<tr>
<th>Outcome / definition</th>
<th>Ideal indicator</th>
<th>Surrogate indicators</th>
</tr>
</thead>
</table>
| **Important**        | - Changes in health status  
                      - Changes in healthcare delivery | - Citation rates  
                      - Media coverage  
                      - Correspondence |
| Study findings have a major impact on health or healthcare |                |                     |
| **Useful**           | - Contributes significantly within a systematic review of the topic  
                      - Narrows CIs around estimates of effect | - Contributes to non-systematic reviews or guidelines  
                      - Citation rates  
                      - Correspondence |
| Study contributes significantly to the scientific debate or knowledge on a subject |                |                     |
| **Relevant**         | - Topic is relevant and consistent with the aims and readership of the journal confirmed by survey | - Citation rates  
                      - Correspondence  
                      - Internet hit rates |
| Topic is relevant to the journal’s aims and readers |                |                     |
| **Methodologically sound** | - Study findings are replicated several times across different settings | - Closeness of fit between methods and 'evidence-based' methodological checklist  
                      - Correspondence |
<p>| Methods used are able to answer the study question |                |                     |</p>
<table>
<thead>
<tr>
<th>Outcome / definition</th>
<th>Ideal indicator</th>
<th>Surrogate indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ethically sound</strong></td>
<td>- No divergence between reality and the report.</td>
<td>- Study received ethical clearance</td>
</tr>
<tr>
<td>Unnecessary harm to humans or animals has been avoided</td>
<td>- Rights of humans and animals safeguarded</td>
<td>- No complaints from participants</td>
</tr>
<tr>
<td>Study has been carried out and reported honestly</td>
<td>- Privacy and informed consent maintained throughout</td>
<td>- No duplicate publication</td>
</tr>
<tr>
<td></td>
<td>- Raw data match presented data</td>
<td></td>
</tr>
<tr>
<td><strong>Complete</strong></td>
<td>- There is no selective presentation of data</td>
<td>- The text is complete</td>
</tr>
<tr>
<td>All relevant information is presented</td>
<td>- All relevant references are cited</td>
<td>- The publication is complete (ie not salami-sliced)</td>
</tr>
<tr>
<td><strong>Accurate</strong></td>
<td>- Measurements truly reflect magnitude of findings</td>
<td>- The figures add up</td>
</tr>
<tr>
<td>Presented information is a true reflection of what went on</td>
<td>- Raw data match presented data</td>
<td>- Corrections</td>
</tr>
<tr>
<td></td>
<td>- References are accurate</td>
<td></td>
</tr>
</tbody>
</table>
Reboxetine vs placebo and/or SSRIs for depression

<table>
<thead>
<tr>
<th>Trial</th>
<th>Reboxetine (n/N)</th>
<th>Placebo (n/N)</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients with adverse events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>014</td>
<td>84/126</td>
<td>78/128</td>
<td></td>
</tr>
<tr>
<td>091</td>
<td>24/28</td>
<td>13/28</td>
<td></td>
</tr>
<tr>
<td>015</td>
<td>71/112</td>
<td>58/112</td>
<td></td>
</tr>
<tr>
<td>046</td>
<td>239/264</td>
<td>208/254</td>
<td></td>
</tr>
<tr>
<td>047</td>
<td>225/258</td>
<td>201/252</td>
<td></td>
</tr>
<tr>
<td>050</td>
<td>138/150</td>
<td>117/150</td>
<td></td>
</tr>
<tr>
<td>045</td>
<td>68/89</td>
<td>52/87</td>
<td></td>
</tr>
<tr>
<td>049</td>
<td>98/106</td>
<td>77/104</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>947/1133</td>
<td>804/1115</td>
<td></td>
</tr>
</tbody>
</table>

Total heterogeneity: $I^2=44.0\%, P=0.085$; total effect: $P=0.001$

|       | Withdrawal owing to adverse events | | |
| 014   | 14/126                          | 15/128        |                     |
| 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 heterogeneity: $I^2=38.4\%, P=0.124$; total effect: $P=0.001$

Towards an interdisciplinary approach to peer review

Edying et al BMJ 2010
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
Towards an interdisciplinary approach to peer review
EMA's release of regulatory data: trust but verify

Towards an interdisciplinary approach to peer review
Paper needed to print oseltamivir study WP16263 (courtesy of Peter Doshi)
Towards an interdisciplinary approach to peer review
**Selection**

RIATAR (Audit record: shows what’s in and what’s out and why)
Report analyses per protocol
Identify analyses which are NOT per protocol
All available as web appendices

**Restored publication**
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

Towards an interdisciplinary approach to peer review
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 commercially-funded studies
• Tool for academic promotion system

Towards an interdisciplinary approach to peer review
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
Editorial peer review for improving the quality of reports of biomedical studies: a Cochrane review
Tom Jefferson on behalf of the PIRATES
(Liz Wager, Frank Davidoff, Phil Alderson)
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

**Excluded**: 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:

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

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

<table>
<thead>
<tr>
<th>Masking</th>
<th>9 studies</th>
<th>Time taken; Constructiveness; Courtesy; Acceptance rates; Authors’ views; Use of supporting evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactions with reviewers</td>
<td>Callaham; Strayhorn; Neuhauser</td>
<td>Acceptance rates; Congruence with editors’ views; Timeliness</td>
</tr>
<tr>
<td>Checklists</td>
<td>Gardner; Jefferson</td>
<td>Study design; stats presentation; Quality of econ submissions (no effect)</td>
</tr>
<tr>
<td>Internet (open) review</td>
<td>Bingham</td>
<td>Timeliness; Etiquette; Use of supporting references</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Bias</td>
<td>Ernst</td>
<td>Bias against unconventional treatments</td>
</tr>
<tr>
<td>Before/after (accepted papers)</td>
<td>Goodman; Perie</td>
<td>Readability; Readers’ views; Experts’ views</td>
</tr>
<tr>
<td>Studies in P-R cf non P-R jnls</td>
<td>Elvik</td>
<td>Retrospective, non-randomised cohort</td>
</tr>
</tbody>
</table>
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 process-centred
• 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 commercially-funded 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 importance/ methods
- Protocol review methods
- Ethical review ethical soundness
- Informal review relevance / context
- **Journal review** quality of reporting accuracy, complet., copy-editing