Case Study 1: “Egg on their faces: the story of human albumin solution”*

Human albumin solution, a blood product, has been used in the treatment of blood loss and burns since the attack on Pearl Harbour over half a century ago.

In the UK alone, an estimated 100,000 patients are treated with human albumin solution each year, at a cost to the NHS of close to 12 million.

In 1996, the global albumin market was worth £900,000.

But is human albumin administration beneficial?

“Egg on their faces: the story of human albumin solution”

To answer this question a systematic review of controlled trials comparing albumin with crystalloid was conducted by the Cochrane Injuries Group.

30 RCTs including 1419 randomised patients identified.

A meta-analysis showed that the risk of death among those treated with albumin was higher than in the comparison groups.

The pooled risk ratio was 1.68 (95% CI 1.26, 2.23)

The data suggested that for every seventeen critically ill patients treated with albumin there is one extra death.

“Egg on their faces: the story of human albumin solution”

“Despite vigorous attempts by the plasma products industry to limit the impact of the systematic review on albumin sales, the use of albumin declined steeply.

Throughout the UK albumin sales fell by 40%.

The decline in albumin use occurred despite vigorous criticism of the review in the letters pages of the BMJ.

The decline in albumin sales is a clear indication that doctors took into account the evidence presented in the systematic review and that many doctors changed their practice in response.”

Patient Survival after Human Albumin Administration
A Meta-Analysis of Randomized, Controlled Trials
Mahlon M. Wilkes, PhD, and Roberta J. Navickis, PhD

Purpose: To test the hypothesis that albumin administration is not associated with excess mortality.

Data Sources: Computer searches of the MEDLINE and EMBASE databases, the Cochrane Library, and Internet documents; hand searching of medical journals; inquiries to investigators and medical directors; and review of reference lists.

Study Selection: Randomized, controlled trials comparing albumin therapy with crystalloid therapy, no albumin, or lower doses of albumin.

Data Extraction: Two investigators independently extracted data. The primary end point was relative risk for death. Criteria used to assess methodologic quality were blinding, method of allocation concealment, presence of mortality as a study end point, and crossover. Small-trial bias was also investigated.

Data Synthesis: Fifty-five trials involving surgery or trauma, burns, hypoalbuminemia, high-risk neonates, ascites, and other indications were included. Albumin administration did not significantly affect mortality in any category of indications. For all trials, the relative risk for death was 1.11 (95% CI, 0.95 to 1.28). Relative risk was lower among trials with blinding (0.73 [CI, 0.48 to 1.12]; n = 7), mortality as an end point (1.00 [CI, 0.84 to 1.18]; n = 17), no crossover (1.04 [CI, 0.89 to 1.22]; n = 35), and 100 or more patients (0.94 [CI, 0.77 to 1.14]; n = 10). In trials with two or more such attributes, relative risk was further reduced.

Conclusions: Overall, no effect of albumin on mortality was detected; any such effect may therefore be small. This finding supports the safety of albumin. The influence of methodologic quality on relative risk for death suggests the need for further well-designed clinical trials.

For author affiliations, current addresses, and contributions, see end of text.
See editorial comment on pp 205-208.
Case study 2: The Vioxx story

On Sept 30, 2004, Merck announced the withdrawal of rofecoxib (Vioxx) because of an increased cardiovascular risk in patients taking the drug for >18 months.

Decision was based on the 3-year results of the unpublished APPROVe study, a RTC of rofecoxib for the prevention of colorectal polyps in patients with a history of colorectal adenomas.

By 2004, rofecoxib had been taken by ~80 million people (sales US$2.5 billion).

Juni et al. did a meta-analysis of 18 RCTs and 11 observational studies.

By the end of 2000 (52 events, 20742 patients) the relative risk from RTCs was 2.30 (95% CI 1.22–4.33, p=0.010), and 1 year later (64 events, 21432 patients) it was 2.24 (1.24–4.02, p=0.007).

Juni et al. concluded that “rofecoxib should have been withdrawn several years earlier; the reasons why manufacturer and drug licensing authorities did not continuously monitor and summarise the accumulating evidence need to be clarified”
Case study 2: The Vioxx story

Relative risk (95% CI) of myocardial infarction

- Ehrich et al. (2001)
- Extension of Ehrich et al. (2001)
- Cannon et al. (2000)
- Day et al. (2000)
- Hawkey et al. (2000)
- Truitt et al. (2001)
- Saag et al. (2000 A)
- Kivitz et al. (2004)
- Extension of Schnitzer et al. (1999)
- Bombardier et al. (2000)
- Geba et al. (2001)
- Truitt et al. (2001 A)
- Lisse et al. (2003)
- Extension of Truitt et al. (2001 A)
- Extension of Geusens et al. (2002)
- Katz et al. (2003)

Combined 2.24 (95% CI 1.24–4.02)
## Case study 2: The Vioxx story

<table>
<thead>
<tr>
<th>Year</th>
<th>Patients</th>
<th>Events</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>523</td>
<td>1</td>
<td>0.916</td>
</tr>
<tr>
<td>1998</td>
<td>615</td>
<td>2</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>1399</td>
<td>5</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>2208</td>
<td>6</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>2983</td>
<td>8</td>
<td>0.649</td>
</tr>
<tr>
<td></td>
<td>3324</td>
<td>9</td>
<td>0.866</td>
</tr>
<tr>
<td>1999</td>
<td>4017</td>
<td>12</td>
<td>0.879</td>
</tr>
<tr>
<td></td>
<td>5059</td>
<td>13</td>
<td>0.881</td>
</tr>
<tr>
<td>2000</td>
<td>5293</td>
<td>16</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>13,269</td>
<td>40</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>14,247</td>
<td>44</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>15,156</td>
<td>46</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>20,742</td>
<td>52</td>
<td>0.010</td>
</tr>
<tr>
<td>2001</td>
<td>20,742</td>
<td>58</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>20,742</td>
<td>63</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>21,432</td>
<td>64</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Relative risk (95% CI) of myocardial infarction

![Graph showing relative risk of myocardial infarction](image-url)
Case study 3. Does male circumcision reduce risk of HIV?

International Journal of STD & AIDS 1999; 10: 8–16

REVIEW ARTICLE

Circumcision and HIV infection: review of the literature and meta-analysis

R S Van Howe MD FAAP
Department of Pediatrics, Marshfield Clinic, Lakeland Center, USA

Summary: Thirty-five articles and a number of abstracts have been published in the medical literature looking at the relationship between male circumcision and HIV infection. Study designs have included geographical analysis, studies of high-risk patients, partner studies and random population surveys. Most of the studies have been conducted in Africa. A meta-analysis was performed on the 29 published articles where data were available. When the raw data are combined, a man with a circumcised penis is at greater risk of acquiring and transmitting HIV than a man with a non-circumcised penis (odds ratio (OR)=1.06, 95% confidence interval (CI)=1.01–1.12). Based on the studies published to date, recommending routine circumcision as a prophylactic measure to prevent HIV infection in Africa, or elsewhere, is scientifically unfounded.
Circumcision increases risk of HIV!

<table>
<thead>
<tr>
<th>Author</th>
<th>Intact+</th>
<th>Intact-</th>
<th>Circ+</th>
<th>Circ-</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>Relative risk</th>
<th>Attributable risk (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Truck drivers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bwayo⁶</td>
<td>92</td>
<td>86</td>
<td>160</td>
<td>612</td>
<td>4.09</td>
<td>2.91-5.76</td>
<td>2.49</td>
<td>59.9</td>
</tr>
<tr>
<td>Bwayo⁷</td>
<td>22</td>
<td>46</td>
<td>37</td>
<td>200</td>
<td>2.59</td>
<td>1.39-4.79</td>
<td>2.07</td>
<td>51.7</td>
</tr>
<tr>
<td><strong>STD clinics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kreiss²³</td>
<td>59</td>
<td>18</td>
<td>254</td>
<td>168</td>
<td>2.17</td>
<td>1.24-3.81</td>
<td>1.27</td>
<td>21.4</td>
</tr>
<tr>
<td>Hira⁸</td>
<td>418</td>
<td>172</td>
<td>10</td>
<td>10</td>
<td>2.43</td>
<td>0.99-5.94</td>
<td>1.42</td>
<td>29.4</td>
</tr>
<tr>
<td>Cameron¹¹</td>
<td>18</td>
<td>61</td>
<td>6</td>
<td>208</td>
<td>10.23</td>
<td>3.89-26.90</td>
<td>8.13</td>
<td>87.7</td>
</tr>
<tr>
<td>Pépin²⁴¹¹</td>
<td>5</td>
<td>42</td>
<td>13</td>
<td>243</td>
<td>2.22</td>
<td>0.75-6.57</td>
<td>2.09</td>
<td>52.3</td>
</tr>
<tr>
<td>Greenblatt¹²</td>
<td>11</td>
<td>28</td>
<td>8</td>
<td>68</td>
<td>3.34</td>
<td>1.21-9.18</td>
<td>2.68</td>
<td>62.7</td>
</tr>
<tr>
<td>Diallo¹⁰</td>
<td>38</td>
<td>46</td>
<td>212</td>
<td>873</td>
<td>3.40</td>
<td>2.16-5.36</td>
<td>2.32</td>
<td>56.8</td>
</tr>
<tr>
<td>Simonsen¹³</td>
<td>17</td>
<td>70</td>
<td>21</td>
<td>232</td>
<td>2.68</td>
<td>1.34-5.37</td>
<td>2.35</td>
<td>57.5</td>
</tr>
<tr>
<td>Tyndall²</td>
<td>85</td>
<td>93</td>
<td>105</td>
<td>527</td>
<td>4.59</td>
<td>3.20-6.58</td>
<td>2.87</td>
<td>65.2</td>
</tr>
<tr>
<td>Nasio³⁶</td>
<td>86</td>
<td>78</td>
<td>137</td>
<td>580</td>
<td>4.67</td>
<td>3.26-6.68</td>
<td>2.74</td>
<td>63.6</td>
</tr>
<tr>
<td>Mehlenda²²</td>
<td>837</td>
<td>3411</td>
<td>38</td>
<td>253</td>
<td>1.63</td>
<td>1.15-2.32</td>
<td>1.51</td>
<td>33.7</td>
</tr>
<tr>
<td>Bolinger²⁵</td>
<td>50</td>
<td>241</td>
<td>1</td>
<td>14</td>
<td>2.90</td>
<td>0.37-22.60</td>
<td>2.58</td>
<td>61.2</td>
</tr>
<tr>
<td>Chiasson²²</td>
<td>36</td>
<td>797</td>
<td>14</td>
<td>542</td>
<td>1.75</td>
<td>0.93-3.27</td>
<td>1.72</td>
<td>41.7</td>
</tr>
<tr>
<td><strong>TB patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sassan³⁰</td>
<td>75</td>
<td>18</td>
<td>415</td>
<td>221</td>
<td>2.22</td>
<td>1.29-3.81</td>
<td>1.24</td>
<td>19.1</td>
</tr>
<tr>
<td><strong>High-risk total</strong></td>
<td>1849</td>
<td>5207</td>
<td>1431</td>
<td>4751</td>
<td>1.18</td>
<td>1.09-1.28</td>
<td>1.13</td>
<td>11.7</td>
</tr>
<tr>
<td><strong>Partner studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hunter³¹</td>
<td>43</td>
<td>330</td>
<td>165</td>
<td>3765</td>
<td>2.97</td>
<td>2.09-4.24</td>
<td>2.75</td>
<td>63.6</td>
</tr>
<tr>
<td>Carael³⁴</td>
<td>90</td>
<td>105</td>
<td>34</td>
<td>45</td>
<td>1.13</td>
<td>0.67-1.92</td>
<td>1.07</td>
<td>6.8</td>
</tr>
<tr>
<td>Chao³⁴</td>
<td>442</td>
<td>4844</td>
<td>75</td>
<td>232</td>
<td>0.28</td>
<td>0.21-0.37</td>
<td>0.34</td>
<td>65.8</td>
</tr>
<tr>
<td>Mose³⁴</td>
<td>24</td>
<td>16</td>
<td>12</td>
<td>17</td>
<td>2.13</td>
<td>0.80-5.62</td>
<td>1.45</td>
<td>31.0</td>
</tr>
<tr>
<td>Allen⁵⁵</td>
<td>275</td>
<td>612</td>
<td>132</td>
<td>324</td>
<td>1.10</td>
<td>0.86-1.41</td>
<td>1.07</td>
<td>6.6</td>
</tr>
<tr>
<td>Sedlin³⁶</td>
<td>32</td>
<td>26</td>
<td>53</td>
<td>40</td>
<td>1.49</td>
<td>0.75-2.98</td>
<td>1.22</td>
<td>18.1</td>
</tr>
<tr>
<td>Konde-Lucœ²⁴⁵</td>
<td>153</td>
<td>1516</td>
<td>6</td>
<td>127</td>
<td>2.13</td>
<td>0.93-4.93</td>
<td>2.03</td>
<td>50.8</td>
</tr>
<tr>
<td><strong>Partner total</strong></td>
<td>1059</td>
<td>7449</td>
<td>457</td>
<td>4550</td>
<td>1.42</td>
<td>1.26-1.59</td>
<td>1.36</td>
<td>26.7</td>
</tr>
<tr>
<td><strong>Random population surveys</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barongo³⁷</td>
<td>55</td>
<td>1356</td>
<td>42</td>
<td>642</td>
<td>0.62</td>
<td>0.41-0.94</td>
<td>0.63</td>
<td>36.5</td>
</tr>
<tr>
<td>Grosskurth³⁸</td>
<td>158</td>
<td>4664</td>
<td>61</td>
<td>1026</td>
<td>0.58</td>
<td>0.43-0.78</td>
<td>0.59</td>
<td>40.9</td>
</tr>
<tr>
<td>Van de Perre⁴⁰</td>
<td>46</td>
<td>224</td>
<td>6</td>
<td>26</td>
<td>0.89</td>
<td>0.35-2.28</td>
<td>0.91</td>
<td>9.1</td>
</tr>
<tr>
<td>Seed⁹</td>
<td>171</td>
<td>422</td>
<td>51</td>
<td>192</td>
<td>1.52</td>
<td>1.07-2.17</td>
<td>1.37</td>
<td>27.2</td>
</tr>
<tr>
<td>Malamba⁴⁴</td>
<td>111</td>
<td>114</td>
<td>21</td>
<td>47</td>
<td>2.17</td>
<td>1.22-3.88</td>
<td>1.60</td>
<td>56.8</td>
</tr>
<tr>
<td>Quigley⁴⁶</td>
<td>101</td>
<td>272</td>
<td>48</td>
<td>121</td>
<td>0.94</td>
<td>0.62-1.40</td>
<td>0.95</td>
<td>4.7</td>
</tr>
<tr>
<td>Urassa 1³⁸</td>
<td>56</td>
<td>1301</td>
<td>42</td>
<td>600</td>
<td>0.61</td>
<td>0.41-0.93</td>
<td>0.63</td>
<td>36.9</td>
</tr>
<tr>
<td>Urassa 2</td>
<td>105</td>
<td>2040</td>
<td>32</td>
<td>426</td>
<td>0.69</td>
<td>0.46-1.03</td>
<td>0.70</td>
<td>29.9</td>
</tr>
<tr>
<td>Urassa 3</td>
<td>38</td>
<td>309</td>
<td>19</td>
<td>158</td>
<td>1.02</td>
<td>0.57-1.83</td>
<td>1.02</td>
<td>2.0</td>
</tr>
<tr>
<td>Urassa 4</td>
<td>112</td>
<td>716</td>
<td>54</td>
<td>692</td>
<td>2.00</td>
<td>1.43-2.82</td>
<td>1.87</td>
<td>46.5</td>
</tr>
<tr>
<td>Urassa 5</td>
<td>101</td>
<td>365</td>
<td>48</td>
<td>136</td>
<td>0.78</td>
<td>0.53-1.66</td>
<td>0.83</td>
<td>16.9</td>
</tr>
<tr>
<td><strong>Random total</strong></td>
<td>1054</td>
<td>11723</td>
<td>424</td>
<td>4066</td>
<td>0.86</td>
<td>0.77-0.97</td>
<td>0.87</td>
<td>12.6</td>
</tr>
</tbody>
</table>

*Intact+=non-circumcised and HIV+. Intact-=non-circumcised and HIV-. Circ+=circumcised and HIV+. Circ-=circumcised and HIV-. OR = 1.06*
Case study 3. Does male circumcision reduce risk of HIV?

Circumcision in men and the prevention of HIV infection: a ‘meta-analysis’ revisited

Nigel O’Farrell MD MRCP¹ and Matthias Egger MD MFPHM²

¹Department of Genitourinary Medicine, Milne Clinic, Bristol Royal Infirmary, Bristol, and
²MRC Health Services Research Collaboration, Department of Social Medicine,
University of Bristol, Bristol, UK

Summary: There is debate on the role of male circumcision in HIV transmission. Most case-control and cohort studies from Africa have shown an association between a lack of circumcision and an increased risk of HIV infection in men. The evidence is conflicting, however, with cross-sectional surveys from Tanzania and Rwanda either showing no relationship or an association in the opposite direction. A recent review and meta-analysis of the literature¹ concluded that the risk of HIV infection was lower in uncircumcised men (combined odds ratio 0.94, 95% confidence interval 0.89 to 0.99). However, the analysis was performed by simply pooling the data from 33 diverse studies, which is an inappropriate method for combining studies.

We re-analysed the data, stratifying by study, and found that an intact foreskin was associated with an increased risk of HIV infection: combined odds ratio 1.43 (1.32 to 1.54) with a fixed effect model and 1.67 (1.25 to 2.24) with a random effect model. There was significant between-study heterogeneity (P<0.0001) which was partly explained by stronger associations in studies in high-risk groups. The results from this re-analysis thus support the contention that male circumcision may offer protection against HIV infection, particularly in high-risk groups where genital ulcers and other STDs ‘drive’ the HIV epidemic. A systematic review is required to clarify this issue. Such a review should be based on an extensive search for relevant studies, published and unpublished, and should include a careful assessment of the design and methodological quality of studies. Much emphasis should be given to the exploration of possible sources of heterogeneity. In view of the continued high prevalence and incidence of HIV in many countries in sub-Saharan Africa, the question of whether circumcision could contribute to prevent infections is of great importance, and a sound systematic review of the available evidence should be performed without delay.
HIV and male circumcision—a systematic review with assessment of the quality of studies

N Siegfried, M Muller, J Deeks, J Volmink, M Egger, N Low, S Walker, and P Williamson

This Cochrane systematic review assesses the evidence for an interventional effect of male circumcision in preventing acquisition of HIV-1 and HIV-2 by men through heterosexual intercourse. The review includes a comprehensive assessment of the quality of all 37 included observational studies. Studies in high-risk populations consisted of four cohort studies, 12 cross-sectional studies, and three case-control studies; general population studies consisted of one cohort study, 16 cross-sectional studies, and one case-control study. There is evidence of methodological heterogeneity between studies, and statistical heterogeneity was highly significant for both general population cross-sectional studies ($\chi^2=132.34$; degrees of freedom [df]=15; $p<0.00001$) and high-risk cross-sectional studies ($\chi^2=29.70$; df=10; $p=0.001$). Study quality was very variable and no studies measured the same set of potential confounding variables. Therefore, conducting a meta-analysis was inappropriate. Detailed quality assessment of observational studies can provide a useful visual aid to interpreting findings. Although most studies show an association between male circumcision and prevention of HIV, these results may be limited by confounding, which is unlikely to be adjusted for.
Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial

Berton Assefa,1,2,3 Akihiro Taniwaki,2,3 Emmah Lapiganto,2,4 Jéssie Solange Tomboumba,2,4 Reni Sitika,2,3 Adrian Pare,5
1 Medical University of Innsbruck, Institute of Human Genetics, Department of Molecular and Cell Biology, Innsbruck, Austria
2 MRC Unit in Nairobi, Kenya
3 Imperial College London, School of Public Health, London, UK
4 University of Nairobi, Kenya
5 University of California, San Diego, California, USA

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: All authors researched data, contributed to drafting the manuscript, and approved the final manuscript.

Randomized, controlled intervention trial conducted in a general population of South Africa to test this hypothesis.

ABSTRACT

Background
Observational studies suggest that male circumcision may provide protection against HIV-1 infection. A randomized, controlled intervention trial was conducted in a general population of South Africa to test this hypothesis.

Methods and Findings
A total of 3,318 uncircumcised men aged 18-24 years were randomized to a control or an intervention group with follow-up visits at months 3, 12, and 24. Male circumcision was offered to the intervention group immediately after randomization and to the control group at the end of the follow-up. The group enrolled data were analyzed in univariate and multivariate analyses. The main outcome measure was HIV incidence. Male circumcision was associated with a significant reduction in HIV incidence (rate ratio, 0.52; 95% CI, 0.35-0.75) compared with the control group. Male circumcision was associated with a significant reduction in HIV incidence (rate ratio, 0.52; 95% CI, 0.35-0.75).

Conclusion
Male circumcision provides a degree of protection against acquiring HIV infection, equivalent to what a vaccine of high efficacy would have achieved. Male circumcision may provide an important way of reducing the spread of HIV infection in sub-Saharan Africa. The ethical and public health benefits of male circumcision in preventing HIV infection need to be considered in the global context.

Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial

Robert C. Bailey, Stephen Aherne, Cassie B. Parker, Greener A. Ayugi, Jon Maclennan, John Koziol, Carolyn A. Williams, Robert C. Campbell, Joseph O. Ondieki, Anthony M. Mwaba

Summary
Background Male circumcision could provide substantial protection against acquisition of HIV-1 infection. Our aim was to determine whether male circumcision had a protective effect against HIV infection, and to assess safety and changes in sexual behaviour related to this intervention.

Methods We did a randomised controlled trial of 2784 men aged 18-24 years in Kisumu, Kenya. Men were randomly assigned to an intervention group (circumcision, n=1393) or a control group (deferred circumcision, 1393), and assessed by HIV testing, medical examinations, and behavioural interviews during follow-ups at 1, 3, 6, 12, 18, and 24 months. HIV incidence was estimated in an intention-to-treat analysis. This trial is registered with ClinicalTrials.gov, with the number NCT00691097.

Findings The trial was stopped early on December 12, 2006, after a third interim analysis reviewed the data by the safety and monitoring board. The median length of follow-up was 21 months (interquartile range, 13-24 months). The main results were inconclusive for 340 (6.4%) participants. 22 in the intervention group and 47 in the control group had tested positive for HIV when the study was stopped. The year HIV incidence was 2.1% (95% CI 1.3-3.9) in the circumcision group and 4.2% (3.0-5.5) in the control group (p=0.003), and in our group men were 0.47 (95% CI 0.1-0.8), which corresponds to a reduction in the risk of acquiring an HIV infection of 53% (22-70). Adjusting for comorbidities in treatment and exclusion of four men found to be seropositive at enrolment, the protective effect of circumcision was 68% (32-77). Adverse events related to the intervention (2 events in 1.5% of those circumcised) resolved quickly. No beneficial gain in prevention was the result of circumcision was observed.

Interpretation Male circumcision significantly reduced the HIV-1 infection rate in young men in Africa. Where appropriate, voluntary, safe, and affordable circumcision services should be integrated with other HIV prevention interventions and provided as expeditiously as possible.
Meta-analysis of 3 RCTs shows strong, consistent effect

Table 1: Study characteristics and outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Setting</th>
<th>Population</th>
<th>n</th>
<th>Intervention</th>
<th>Control</th>
<th>Relative risk (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auvert et al.</td>
<td>Randomized trial</td>
<td>Orange farm, South Africa</td>
<td>Males aged 18–24 years</td>
<td>3128</td>
<td>20/1546</td>
<td>49/1582</td>
<td>0.42 (0.25–0.70)</td>
</tr>
<tr>
<td>Bailey et al.</td>
<td>Randomized trial</td>
<td>Kisumu, Kenya</td>
<td>Males aged 18–24 years</td>
<td>2780</td>
<td>19/1388</td>
<td>46/1392</td>
<td>0.41 (0.24–0.70)</td>
</tr>
<tr>
<td>Gray et al.</td>
<td>Randomized trial</td>
<td>Rakai district, Uganda</td>
<td>Males aged 15–49 years</td>
<td>4936</td>
<td>22/2474</td>
<td>45/2522</td>
<td>0.50 (0.30–0.83)</td>
</tr>
</tbody>
</table>

*3274 randomized, 3128 included in analysis.
**2784 randomized, 2780 included in analysis.

Study name | Statistics for each study | Risk ratio and 95% CI
|------------|---------------------------|-------------------|
| Auvert, RSA| Risk ratio 0.42, Lower limit 0.25, Upper limit 0.70, p-value 0.001 | ![Risk ratio](image)
| Bailey, Kenya | Risk ratio 0.41, Lower limit 0.24, Upper limit 0.70, p-value 0.001 | ![Risk ratio](image)
| Gray, Uganda | Risk ratio 0.50, Lower limit 0.30, Upper limit 0.83, p-value 0.007 | ![Risk ratio](image)
| Combined    | Risk ratio 0.44, Lower limit 0.33, Upper limit 0.60, p-value <0.0001 | ![Risk ratio](image)

Fig. 1: Random effects meta-analysis.
What is evidence-based medicine?

The practice of EBM is the integration of

- individual clinical expertise
  with the
- best available external clinical evidence from systematic research
  and
- patient’s values and expectations

http://www.cebm.net/index.asp
The importance of research synthesis

We need evidence for both clinical practice and for public health decision making

Where does evidence come from?

- An good review is a state-of-the-art synthesis of current evidence on a given research question
- Given the explosion of medical literature, and the fact that time is always scarce, review articles play a big role in decision-making
- According to one estimate, to keep up to date in Internal Medicine, need to read 17 articles a day, 365 days a year!
The importance of research synthesis

Given that most clinicians and public health professionals do not have the time to track down all the original articles, critically read them, and obtain the evidence they need for their questions,

- Systematic reviews and clinical practice guidelines may be their best source of evidence
  
- Several “pre-digested” sources of evidence are currently available
  
- The EBM movement is heavily dependent on these pre-appraised evidence sources
<table>
<thead>
<tr>
<th>Grade of Recommendation</th>
<th>Level of Evidence</th>
<th>Therapy/Prevention, Aetiology/Harm</th>
<th>Prognosis</th>
<th>Diagnosis</th>
<th>Economic analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>A</td>
<td>SR (with homogeneity*) of RCTs</td>
<td>SR (with homogeneity*) of inception cohort studies; or a CPG validated on a test set.</td>
<td>SR (with homogeneity*) of Level 1 diagnostic studies; or a CPG validated on a test set.</td>
<td>SR (with homogeneity*) of Level 1 economic studies</td>
</tr>
<tr>
<td>1b</td>
<td>A</td>
<td>Individual RCT (with narrow Confidence Interval*)</td>
<td>Individual inception cohort study with ≥ 80% follow-up</td>
<td>Independent blind comparison of an appropriate spectrum of consecutive patients, all of whom have undergone both the diagnostic test and the reference standard.</td>
<td>Analysis comparing all (critically-validated) alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.</td>
</tr>
<tr>
<td>1c</td>
<td>A</td>
<td>All or none*</td>
<td>All or none case-series</td>
<td>Absolute SP, SP, SN, Nout*</td>
<td>Clearly as good or better*, but cheaper. Clearly as bad or worse but more expensive. Clearly better or worse at the same cost.</td>
</tr>
<tr>
<td>2a</td>
<td>B</td>
<td>SR (with homogeneity*) of cohort studies</td>
<td>SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs.</td>
<td>SR (with homogeneity*) of Level ≥2 diagnostic studies</td>
<td>SR (with homogeneity*) of Level ≥2 economic studies</td>
</tr>
<tr>
<td>2b</td>
<td>B</td>
<td>Individual cohort study (including low quality RCT; e.g., &lt;80% follow-up)</td>
<td>Retrospective cohort study or follow-up of untreated control patients in an RCT; or CPG not validated in a test set.</td>
<td>Independent blind comparison but either in non-consecutive patients, or confined to a narrow spectrum of study individuals (or both), all of whom have undergone both the diagnostic test and the reference standard; or a diagnostic CPG not validated in a test set.</td>
<td>Analysis comparing a limited number of alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.</td>
</tr>
<tr>
<td>2c</td>
<td>B</td>
<td>“Outcomes” Research</td>
<td>“Outcomes” Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>A</td>
<td>SR (with homogeneity*) of case-control studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>A</td>
<td>Individual Case-Control Study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>Case-series (and poor quality cohort and case-control studies*)</td>
<td>Case-series (and poor quality prognostic cohort studies*)</td>
<td>Reference standard was not applied to all study patients</td>
<td>Analysis with no sensitivity analysis</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles”</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles”</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles”</td>
<td>Expert opinion without explicit critical appraisal, or based on economic theory</td>
</tr>
</tbody>
</table>
Guidelines and recommendations: GRADE

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide.

http://www.gradeworkinggroup.org/
Doing New Research? Don’t Forget the Old
Nobody should do a trial without reviewing what is known

Mike Clarke

On May 2, 1898, George Gould used his address to the founding meeting of the Association of Medical Librarians in Philadelphia to present a vision of the future of health information. ‘I look forward,’ he said, ‘to such an organisation of the literary records of medicine that a puzzled worker in any part of the civilised world shall in an hour be able to gain a knowledge pertaining to a subject of the experience of every other man in the world’ [1]. Has his vision been realised? With one or more search engines? Almost certainly, as the speed of the search increased through these four

Box 1. Practical Suggestions for Researchers

• Conduct a systematic review of your research question before embarking on a new study, or identify a relevant review done by someone else.
• Design your study to take account of the relevant successes and failures of the prior studies, and of the evidence within them.
• Discuss the findings of your study in the context of an updated systematic review of relevant research.
• Publish the systematic review within, alongside, or shortly after the report of your study.
• Provide information from your study to others doing systematic reviews of similar topics.


Copyright: © 2004 Mike Clarke. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Prof Archibald Cochrane, CBE
(1909 - 1988)

The Cochrane Collaboration is named in honour of Archie Cochrane, a British researcher.

In 1979 he wrote, “It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomized controlled trials”

Source: http://www.cochrane.org/cochrane/archieco.htm
The Cochrane Collaboration

Archie Cochrane’s challenge led to the establishment during the 1980s of an international collaboration to develop the *Oxford Database of Perinatal Trials*. His encouragement, and the endorsement of his views by others, led to the opening of the first Cochrane centre (in Oxford, UK) in 1992 and the founding of The *Cochrane Collaboration* in 1993.

Source: http://www.cochrane.org/cochrane/archieco.htm
2500 SRs per year, of which about 20% are Cochrane reviews (estimate by Moher et al. PLoS Med 2007)
Centre for Reviews and Dissemination databases

CRD was established in January 1994, and produces and promotes the use of research based knowledge in health and social care.

DARE – (Database of Abstracts of Reviews of Effects) contains over 4000 abstracts of quality assessed and critically appraised systematic reviews. The database focuses on the effects of interventions used in health and social care.

NHS Economic Evaluation Database (NHS EED) contains over 6000 abstracts of quality assessed economic evaluations. The database aims to assist decision-makers by systematically identifying and describing economic evaluations, appraising their quality and highlighting their relative strengths and weaknesses.

Both DARE and NHS EED include details of abstracts in the process of being written and these can be ‘fast-tracked’ on request.

The HTA database brings together details of completed and ongoing health technology assessments from around the world. The abstracts in the database are descriptive rather than analytical and do not form critical appraisals of the reports. The database is produced in collaboration with the INAHTA Secretariat, based at SBU, Sweden.
Systematic reviews are done in different domains

Tuberculosis among Health-Care Workers in Low- and Middle-Income Countries: A Systematic Review

Rajnish Joshi\textsuperscript{1,2}, Arthur L. Reingold\textsuperscript{1}, Dick Menzies\textsuperscript{3}, Madhukar Pai\textsuperscript{3*}

\textsuperscript{1} Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, California, United States of America, \textsuperscript{2} Department of Medicine, Mahatma Gandhi Institute of Medical Sciences, Sevagram, Maharashtra, India, \textsuperscript{3} Montreal Chest Institute, McGill University, Montreal, Canada

Meta-analysis of “rates”
Systematic reviews are done in different domains

**Annals of Internal Medicine**

**Systematic Review: T-Cell–based Assays for the Diagnosis of Latent Tuberculosis Infection: An Update**

Madhukar Pai, MD, PhD; Alice Zwerling, MSc; and Dick Menzies, MD, MSc

GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis

D.I. Ling*, A.A. Zwerling* and M. Pai*#
Systematic reviews are done in different domains

Chloroquine or amodiaquine combined with sulfadoxine–pyrimethamine for uncomplicated malaria: a systematic review

Jimee Hwang¹, Edward Bitarakwate², Madhukar Pai³, Arthur Reingold³, Philip J. Rosenthal⁴ and Grant Dorsey⁴

1 Department of Internal Medicine, University of California San Francisco, San Francisco, CA, USA
2 Elizabeth Glaser Pediatric AIDS Foundation, Kampala, Uganda
3 Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, CA, USA
4 Department of Infectious Diseases, University of California San Francisco, San Francisco, CA, USA

Effect of Duration and Intermittency of Rifampin on Tuberculosis Treatment Outcomes: A Systematic Review and Meta-Analysis

Dick Menzies¹, Andrea Benedetti¹, Anita Paydar¹, Ian Martin¹, Sarah Royce², Madhukar Pai¹, Andrew Vernon³, Christian Lienhardt⁴, William Burman⁵

¹ Respiratory and Epidemiology Clinical Research Unit, Montreal Chest Institute & Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Canada, 2 University of California at San Francisco, San Francisco, California, United States of America, 3 Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, 4 International Union against Tuberculosis and Lung Diseases, and Institut de Recherche pour le Développement, Paris, France, 5 Denver Public Health, Denver, Colorado, United States of America

Meta-analysis of “RCTs [therapy]”
Systematic reviews are done in different domains

Risk of Tuberculosis From Exposure to Tobacco Smoke

A Systematic Review and Meta-analysis

Michael N. Bates, PhD; Asheena Khalakdina, PhD; Madhukar Pai, MD, PhD; Lisa Chang, MPH; Fernanda Lessa, MD, MPH; Kirk R. Smith, PhD

Arch Intern Med. 2007;167:335-342

Meta-analysis of “observational studies [etiology]”
Are these the same or different?

- Traditional, narrative review
- Systematic review
- Overview
- Meta-analysis
- Pooled analysis
Types of review articles

All reviews (also called overviews)

- Meta-analyses
- Systematic reviews
- Individual patient data (IPD) meta-analyses
- Reviews that are not systematic (traditional, narrative reviews)

In practice, not all meta-analyses are conducted as part of systematic reviews.

All reviews (also called overviews)

- Meta-analyses
- Systematic reviews
- Reviews that are not systematic (traditional, narrative reviews)
- Individual patient data (IPD) meta-analyses
Some definitions

Traditional, narrative reviews, usually written by experts in the field, are qualitative, narrative summaries of evidence on a given topic. Typically, they involve informal and subjective methods to collect and interpret information.

“A systematic review is a review in which there is a comprehensive search for relevant studies on a specific topic, and those identified are then appraised and synthesized according to a predetermined and explicit method.”*

Some definitions

“A meta-analysis is the statistical combination of at least 2 studies to produce a single estimate of the effect of the healthcare intervention under consideration.”*

Individual patient data meta-analyses (pooled analyses) involves obtaining raw data on all patients from each of the trials directly and then re-analyzing them.

# Narrative vs. Systematic Reviews

<table>
<thead>
<tr>
<th>Components of a review</th>
<th>Traditional, narrative reviews</th>
<th>Systematic reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation of the question</td>
<td>Usually address broad questions</td>
<td>Usually address focused questions</td>
</tr>
<tr>
<td>Methods section</td>
<td>Usually not present, or not well-described</td>
<td>Clearly described with pre-stated criteria about participants, interventions and outcomes</td>
</tr>
<tr>
<td>Search strategy to identify studies</td>
<td>Usually not described; mostly limited by reviewers’ abilities to retrieve relevant studies; usually not reproducible and prone to selective citation</td>
<td>Clearly described and usually exhaustive; transparent, reproducible and less prone to selective citation</td>
</tr>
<tr>
<td>Quality assessment of identified studies</td>
<td>Usually all identified studies are included without explicit quality assessment</td>
<td>Only high-quality studies are included using pre-stated criteria; if lower-quality studies included, the effects of this are tested in subgroup analyses</td>
</tr>
<tr>
<td>Data extraction</td>
<td>Methods usually not described</td>
<td>Usually undertaken by more than one reviewer onto pre-tested data forms; attempts often made to obtain missing data from authors of primary studies</td>
</tr>
<tr>
<td>Data synthesis</td>
<td>Qualitative description employing the ‘vote counting’ approach, where each included study is given equal weight, irrespective of study size and quality</td>
<td>Meta-analysis assigns higher weights to effect measures from more precise studies; pooled, weighted effect measures with confidence limits provide power and precision to results</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Usually dealt with in a narrative fashion</td>
<td>Heterogeneity dealt with by graphical and statistical methods; attempts are often made to identify sources of heterogeneity</td>
</tr>
<tr>
<td>Interpreting results</td>
<td>Prone to cumulative systematic biases and personal opinion</td>
<td>Less prone to systematic biases and personal opinion</td>
</tr>
</tbody>
</table>

Elements of a Systematic Review

- Formulate the review question & write a protocol
- Search for and include primary studies
- Assess study quality
- Extract data
- Analyze data
- Interpret results & write a report

Define a focused 4-part review question (Patient, Intervention, Comparison and Outcome)

- Review protocols on systematic reviews, and prepare a protocol
- Identify appropriate databases and sources of studies
- Run searches on all relevant databases and sources
- Save all citations (titles/abstracts) in a reference manager

Reviewer 1 screens all titles/abstracts and makes selections for second screen

- Reviewers meet and resolve disagreements on citations they do not agree on
- The final number (N) selected after this process is ready for second screen (review of full-text articles)

Get full texts of all articles identified for second screen (N)

- Articles considered eligible after full-text review (by two reviewers) is the final set of studies for inclusion (n)
- Studies included in the final analysis (n)

Excluded after second screen

- Keeping a log of excluded studies with reasons for exclusion
- Paper data extraction forms (after pilot test)

Reviewer 1 extracts data (including quality assessment) from the final selected articles

- Reviewers meet and resolve disagreements on data
- Compute inter-rater reliability (e.g. Kappa statistic)
- The final data after this process is ready for data entry

Enter data into database manager software

- Import data and analyze using software
- Tabulate study characteristics
- Create forest plots of effect measures
- Check for heterogeneity

Check for heterogeneity: Chi-squared or I-squared tests; those tests have low power; consider a conservative p value of <0.10 for significance

- Using QUOROM or MOOSE as guides for report writing
- Discuss applicability of results and limitations of the review
- Make recommendations for practice or policy, and research

You made it! Celebrate!!
Meta-analysis options

Should Data Be Combined Statistically?

Yes
- Type of Data
  - Discrete
    - 1. Peto Method
    - 2. Mantel-Haenszel
    - 3. Woolf Method
    - 4. DerSimonian-Laird
  - Continuous
    - Same Units of Measurement Used Across Trials?
      - Yes
        - 1. Weighted Mean Difference
        - 2. Standardized Mean Difference
      - No
        - 1. Standardized Mean Difference

No
- Complete Qualitative Systematic Review

Algorithm of statistical choices available to systematic reviewers.

Fixed effect models

- In a fixed effect model, we assume the studies come from the same hypothetical population of studies.
- We assume a single, ‘fixed’, parameter.
- Study weights are a function of within-study variances.
- Confidence interval relatively narrow.

Courtesy: Leon Bax, http://www.mix-for-meta-analysis.info/index.html#
Random effects models

- There is no longer the assumption of a single, homogeneous source population
- Studies are allowed to come from different distributions
- Study weights not just based on within-study variances; a random effects constant is used to distribute the weights more evenly
- Results in a wider confidence interval

Courtesy: Leon Bax, http://www.mix-for-meta-analysis.info/index.html#
**Forest Plot: when outcomes are dichotomous**

**Review:** Supplementation with M in condition X
**Comparison:** 01 Supplement M versus placebo
**Outcome:** 01 Adverse effects

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention group</th>
<th>Control group</th>
<th>Relative risk (fixed)</th>
<th>Weight (%)</th>
<th>Relative risk (fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/N</td>
<td>n/N</td>
<td>95% CI</td>
<td></td>
<td>95% CI</td>
</tr>
<tr>
<td>Study A</td>
<td>1/141</td>
<td>2/142</td>
<td></td>
<td>17.8</td>
<td>0.50 [0.05, 5.49]</td>
</tr>
<tr>
<td>Study B</td>
<td>7/27</td>
<td>9/29</td>
<td></td>
<td>77.7</td>
<td>0.84 [0.36, 1.93]</td>
</tr>
<tr>
<td>Study C</td>
<td>1/100</td>
<td>0/100</td>
<td></td>
<td>4.5</td>
<td>3.00 [0.12, 72.77]</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>268</td>
<td>271</td>
<td></td>
<td>100.0</td>
<td>0.87 [0.41, 1.87]</td>
</tr>
</tbody>
</table>

Total events: 9 (supplement M), 11 (control)
Test for heterogeneity Chi-square=0.79 df=2, p=0.67, I²=0.9%
Test for overall effect z=0.35, p=0.7

**p value indicating level of statistical significance**

**Heterogeneity (I²) = diversity between studies**

**Line of no effect**

**Scale of treatment effect**

**Overall effect**

**Outcome effect measure**
- Shown graphically and numerically
- Fixed effect model used for meta-analysis

**Influence of studies on overall meta-analysis**

**Study IDs**

**Details of review**

N = total number in group
n = number in group with the outcome

---

Figure 1. Meta-analysis of binary outcome measure

Ried K. Aus Fam Phys 2006
Forest Plot: when outcomes are continuous

### Study
<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention group</th>
<th>Control group</th>
<th>Weighted mean difference (fixed) 95% CI</th>
<th>Weight (%)</th>
<th>WMD (fixed) 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study A</td>
<td>34 9.77 (2.93)</td>
<td>34 10.29 (3.43)</td>
<td></td>
<td>275</td>
<td>-0.52 [-2.04, 1.00]</td>
</tr>
<tr>
<td>Study B</td>
<td>36 8.40 (1.90)</td>
<td>36 8.90 (3.00)</td>
<td></td>
<td>46.9</td>
<td>-0.50 [-1.66, 0.66]</td>
</tr>
<tr>
<td>Study C</td>
<td>30 10.26 (2.96)</td>
<td>30 12.09 (3.24)</td>
<td></td>
<td>25.6</td>
<td>-1.83 [-3.40, -0.26]</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100.0</td>
<td>-0.85 [-1.64, -0.05]</td>
</tr>
</tbody>
</table>

Test for heterogeneity: Chi-square = 2.03 df = 2, p = 0.36, I² = 14%
Test for overall effect: z = 2.09, p = 0.04

**p value indicating level of statistical significance**

-4.0 -2.0 0 2.0 4.0
Favours intervention Favours control

### Details of review
- Review: Medicines for condition X
- Comparison: 01 Medicine Z versus placebo
- Outcome: 01 Fasting blood glucose levels (mmol/L)

### N = total number in group
- Mean (standard deviation) of outcome

### Outcome effect measure
- Shown graphically and numerically
- Fixed effect model used for meta-analysis

### Influence of studies on overall meta-analysis
- Overall effect
- Line of no effect
- Scale of treatment effect

---

Figure 2. Meta-analysis of continuous outcome measures

Ried K. Aus Fam Phys 2006
Forest Plot: Cumulative Meta-analysis

Beta-blockers after acute myocardial infarction
Passive smoking and lung cancer review

Meta-analysis Software

- Free
  - RevMan 5 [Review Manager]
  - Meta-Analyst
  - Epi Meta
  - Easy MA
  - Meta-DiSc
  - Meta-Stat
  - MIX

- Commercial
  - Comprehensive Meta-analysis
  - Meta-Win
  - WEasy MA

- General stats packages (commercial)
  - Stata
  - SAS
  - S-Plus
Evaluation of echinacea for the prevention and treatment of the common cold: a meta-analysis

Sachin A Shah, Stephen Sander, C Michael White, Mike Rinaldi, Craig I Coleman

Echinacea is one of the most commonly used herbal products, but controversy exists about its benefit in the prevention and treatment of the common cold. Thus, we did a meta-analysis evaluating the effect of echinacea on the incidence and duration of the common cold. 14 unique studies were included in the meta-analysis. Incidence of the common cold was reported as an odds ratio (OR) with 95% CI, and duration of the common cold was reported as the weighted mean difference (WMD) with 95% CI. Weighted averages and mean differences were calculated by a random-effects model (DerSimonian-Laird methodology). Heterogeneity was assessed by the Q statistic and review of L'Abbe plots, and publication bias was assessed through the Egger weighted regression statistic and visual inspection of funnel plots. Echinacea decreased the odds of developing the common cold by 58% (OR 0.42; 95% CI 0.25–0.71; Q statistic p<0.001) and the duration of a cold by 1.4 days (WMD −1.44, −2.24 to −0.64; p=0.03). Similarly, significant reductions were maintained in subgroup analyses limited to Echinaguard/Echinacin use, concomitant supplement use, method of cold exposure, Jadad scores less than 3, or use of a fixed-effects model. Published evidence supports echinacea’s benefit in decreasing the incidence and duration of the common cold.

Lancet Infect Dis 2007;7:473-80

University of Connecticut School of Pharmacy, Storrs, CT, USA (S A Shah PharmD), S Sander PharmD, C M White PharmD, M Rinaldi PharmD, C I Coleman PharmD); University of Connecticut Health Center Division of Infectious Diseases, Farmington, CT, USA (M Rinaldi); and Division of Drug Information, Hartford
Table 2: Individual study characteristics

<table>
<thead>
<tr>
<th>Analyses included in study</th>
<th>Incidence in echinacea group</th>
<th>Incidence in control group</th>
<th>Number of patients with cold in echinacea group</th>
<th>Number of patients with cold in control group</th>
<th>Mean duration in echinacea group (SD)</th>
<th>Mean duration in control group (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turner et al (2005)</td>
<td>73/149</td>
<td>58/160</td>
<td>NA</td>
<td>NA</td>
<td>1.60 (1.50)</td>
<td>2.90 (1.50)</td>
</tr>
<tr>
<td>Cohen et al (2004)</td>
<td>85/160</td>
<td>150/168</td>
<td>138</td>
<td>308</td>
<td>1.10 (1.30)</td>
<td>2.40 (1.50)</td>
</tr>
<tr>
<td>Sperber et al (2004)</td>
<td>14/24</td>
<td>18/22</td>
<td>NA</td>
<td>NA</td>
<td>9.00 (9.37)</td>
<td>9.00 (9.81)</td>
</tr>
<tr>
<td>Taylor et al (2003)</td>
<td>Duration of cold</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>6.27 (1.50)</td>
<td>5.75 (1.50)</td>
</tr>
<tr>
<td>Barrett et al (2002)</td>
<td>Duration of cold</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>3.34 (0.93)</td>
<td>3.60 (0.93)</td>
</tr>
<tr>
<td>Schulten et al (2001)</td>
<td>Duration of cold</td>
<td>NA</td>
<td>35</td>
<td>39</td>
<td>3.34 (1.08)</td>
<td>3.33 (0.93)</td>
</tr>
<tr>
<td>Turner et al (2000)</td>
<td>Incidence of cold</td>
<td>11/50</td>
<td>NA</td>
<td>NA</td>
<td>7.34 (0.83)</td>
<td>7.50 (0.83)</td>
</tr>
<tr>
<td>Lindemuth and Lindemuth (2000)</td>
<td>Incidence of cold</td>
<td>48</td>
<td>NA</td>
<td>NA</td>
<td>4.74 (1.08)</td>
<td>3.60 (0.93)</td>
</tr>
<tr>
<td>Grimm and Muller (1999)</td>
<td>Incidence of cold</td>
<td>35/54</td>
<td>40/54</td>
<td>NA</td>
<td>8.00 (5.10)</td>
<td>8.7 (3.60)</td>
</tr>
<tr>
<td>Berg (1998)</td>
<td>Incidence of cold</td>
<td>0/14</td>
<td>7/26</td>
<td>NA</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Melchart et al (1999)</td>
<td>Incidence of cold, duration of cold</td>
<td>60/199</td>
<td>33/90</td>
<td>60</td>
<td>3.33 (1.50)</td>
<td>3.37 (1.57)</td>
</tr>
<tr>
<td>Hoheisel et al (1997)</td>
<td>Incidence of cold, duration of cold</td>
<td>24/60</td>
<td>36/60</td>
<td>NA</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Scaglione and Lund (1995)</td>
<td>Duration of cold</td>
<td>16</td>
<td>16</td>
<td>NA</td>
<td>3.37 (1.52)</td>
<td>3.37 (1.57)</td>
</tr>
<tr>
<td>Braunig and Knick (1993)</td>
<td>Duration of cold</td>
<td>70</td>
<td>45</td>
<td>NA</td>
<td>1.00 (0.00)</td>
<td>1.00 (0.00)</td>
</tr>
</tbody>
</table>

NA—not applicable. *Data shown as number of events/total population; †Reported data is number of cold episodes, not number of patients with cold; ‡Reported data as difference of -0.52 days, 95% CI: -1.09 to 0.22.

Figure 3: The effect of echinacea on incidence of common cold
The squares represent individual studies and the size of the square represents the weight given to each study in the meta-analysis. Error bars represent 95% CIs. The diamond represents the combined result. The solid vertical line extending upwards from 1.0 is the null value.
Meta-analysis using STATA

NOTE: Weights are from random effects analysis

Overall (I-squared = 70.7%, p = 0.001)

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turner (2005)</td>
<td>0.75 (0.45, 1.23)</td>
<td>15.82</td>
</tr>
<tr>
<td>Cohen (2004)</td>
<td>0.14 (0.08, 0.24)</td>
<td>15.12</td>
</tr>
<tr>
<td>Sperber (2004)</td>
<td>0.31 (0.08, 1.20)</td>
<td>8.27</td>
</tr>
<tr>
<td>Schulten (2001)</td>
<td>0.15 (0.02, 1.34)</td>
<td>4.43</td>
</tr>
<tr>
<td>Turner (2000)</td>
<td>0.56 (0.22, 1.43)</td>
<td>11.72</td>
</tr>
<tr>
<td>Grimm &amp; Muller (1999)</td>
<td>0.64 (0.28, 1.47)</td>
<td>12.68</td>
</tr>
<tr>
<td>Berg (1998)</td>
<td>0.09 (0.00, 1.70)</td>
<td>2.70</td>
</tr>
<tr>
<td>Melchart (1998)</td>
<td>0.75 (0.44, 1.26)</td>
<td>15.64</td>
</tr>
<tr>
<td>Hoheisel (1997)</td>
<td>0.44 (0.21, 0.92)</td>
<td>13.82</td>
</tr>
<tr>
<td>Overall (I-squared = 70.7%, p = 0.001)</td>
<td>0.42 (0.25, 0.71)</td>
<td>100.00</td>
</tr>
</tbody>
</table>
All systematic reviews are not meta-analyses!

“...it is always appropriate and desirable to systematically review a body of data, but it may sometimes be inappropriate, or even misleading, to statistically pool results from separate studies. Indeed, it is our impression that reviewers often find it hard to resist the temptation of combining studies even when such meta-analysis is questionable or clearly inappropriate.”

All systematic reviews are not systematic!

Many Reviews Are Systematic but Some Are More Transparent and Completely Reported than Others

The PLoS Medicine Editors

Epidemiology and Reporting Characteristics of Systematic Reviews

David Moher\textsuperscript{1,2,3*}, Jennifer Tetzlaff\textsuperscript{1}, Andrea C. Tricco\textsuperscript{1,4}, Margaret Sampson\textsuperscript{1}, Douglas G. Altman\textsuperscript{5}

1 Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada, 2 Department of Paediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Canada, 3 Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada, 4 Institute of Population Health, University of Ottawa, Ottawa, Canada, 5 Centre for Statistics in Medicine, Oxford, United Kingdom

All systematic reviews are not systematic!

- 300 SRs were identified (one month)
- Majority (272 [90.7%]) reported in specialty journals
- Most reviews (213 [71.0%]) were categorized as therapeutic, and included a median of 16 studies
- Reviews typically searched a median of three electronic databases and two other sources
- Most (197/295 [66.8%]) reviews reported information about quality assessment, while few (68/294 [23.1%]) reported assessing for publication bias.
- A little over half (161/300 [53.7%]) reported combining their results statistically, of which most (147/161 [91.3%]) assessed for consistency across studies.
- There were large differences between Cochrane reviews and non-Cochrane reviews in the quality of reporting.
When can meta-analyses mislead?

- When a meta-analysis is done outside of a systematic review
- When poor quality studies are included or when quality issues are ignored
- When small and inconclusive studies are included
- When inadequate attention is given to heterogeneity
  - Indiscriminate data aggregation can lead to inaccurate conclusions
- When reporting biases are a problem
  - Publication bias
  - Time lag bias
  - Duplicate publication bias
  - Language bias
  - Outcome reporting bias

Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S.,
Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D.

ABSTRACT

BACKGROUND
Evidence-based medicine is valuable to the extent that the evidence base is complete and unbiased. Selective publication of clinical trials — and the outcomes within those trials — can lead to unrealistic estimates of drug effectiveness and alter the apparent risk–benefit ratio.

METHODS
We obtained reviews from the Food and Drug Administration (FDA) for studies of 12 antidepressant agents involving 12,564 patients. We conducted a systematic literature search to identify matching publications. For trials that were reported in the literature, we compared the published outcomes with the FDA outcomes. We also compared the effect size derived from the published reports with the effect size derived from the entire FDA data set.

RESULTS
Among 74 FDA-registered studies, 31%, accounting for 3,449 study participants, were not published. Whether and how the studies were published were associated with the study outcome. A total of 37 studies viewed by the FDA as having positive results were published; 1 study viewed as positive was not published. Studies viewed by the FDA as having negative or questionable results were, with 3 exceptions, either not published (22 studies) or published in a way that, in our opinion, conveyed a positive outcome (11 studies). According to the published literature, it appeared that 94% of the trials conducted were positive. By contrast, the FDA analysis showed that 51% were positive. Separate meta-analyses of the FDA and journal data sets showed that the increase in effect size ranged from 11 to 69% for individual drugs and was 32% overall.
If exposure and disease are not associated

False positive study

100 studies will be designed

If $\alpha = 0.05$

5 studies show false positive results

Publication Bias

Hot topic Bias

Positive results bias

5 studies will be published

Likely to be meta-analyzed

THE FALSE POSITIVE RESEARCH CYCLE

(Choi, 1998)

Courtesy: Bernard Choi, PHAC
Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials

Comparison of Protocols to Published Articles

An-Wen Chan, MD, DPhil
Ashjorn Hjortdalsen, MD, PhD
Mette T. Haahr, BSc
Peter C. Gotzsche, MD, DrMedSci
Douglas G. Altman, DSc

Selective publication of studies with statistically significant results has received widespread recognition. In contrast, selective reporting of favorable outcomes within published studies has not undergone comparable empirical investigation. The existence of outcome reporting bias has been widely suspected for years, but direct evidence is limited to case reports that have low generalizability and may themselves be subject to publication bias. Our study had 3 goals: (1) to determine the prevalence of incomplete outcomes and statistical significance among published reports of randomized trials, (2) to assess the association between outcome reporting and statistical significance, and (3) to evaluate the consistency between primary outcomes described in the study protocols and those described in the published articles.

Methods

Comparison of Protocols to Published Articles

JAMA 2004


Selection in Reported Epidemiological Risks: An Empirical Assessment

Fotini K. Karvouna1, George Liberopoulos1, John P. A. Ioannidis1,2*

1 Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. 2 Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, United States of America

ABSTRACT

Epidemiological studies may be subject to selective reporting, but empirical evidence thereof is limited. We empirically evaluated the extent of selection of significant results and large effect sizes in a large sample of recent articles.

Methods and Findings

We evaluated 389 articles of epidemiological studies that reported, in their respective abstracts, at least one relative risk for a continuous risk factor in contrasts based on median, tertile, quartile, or quintile categorizations. We examined the proportion and correlates of reporting statistically significant results in the abstract and whether the magnitude of the relative risks presented (coined to be consistently >1.00) differs depending on the type of contrast used for the risk factor. In 342 articles (87.9%), ≥1 statistically significant relative risks was reported in the abstract, while only 160 articles (43.4%) reported ≥1 statistically nonsignificant relative risk in the abstract. Reporting of statistically significant results was more common with structured abstracts, was less common in US-based studies, and in cancer outcomes. Among 50 randomly selected articles in which the full-text was examined, a median of nine (interquartile range 5–16) statistically significant and six (interquartile range 3–16) statistically nonsignificant relative risks were presented (p = 0.25). Periocularly, the smallest presented relative risk was based on the contrasts of extreme quintiles; on average, the relative risk magnitude was 1.14–1.14, and 1.36-fold larger in a contrasts of extreme quintiles, extreme tertiles, and above versus below median values, respectively (p < 0.0001).

Conclusions

Published epidemiological investigations almost universally highlight significant associations between risk factors and outcomes. For continuous risk factors, investigators selectively present contrasts between more extreme groups, when relative risks are inherently lower.
Figure 1. Publication bias. A, The black circle represents the underlying truth. The white square represents the pooled estimate from a systematic review of all the evidence (small shaded circles). B, The white circles represent evidence that was not identified by the reviewers because it was not published. Note the error in the pooled estimate (publication bias).
Funnel plot to detect publication bias

http://www.cochrane-net.org/openlearning/index.htm
Funnel plot to detect publication bias

http://www.cochrane-net.org/openlearning/index.htm
Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

Clinical research on important questions about the efficacy of medical interventions is sometimes followed by subsequent studies that either reach opposite conclusions or suggest that the original claims were too strong. Such disagreements may upset clinical practice and acquire publicity in both scientific circles and in the lay press. Several empirical investigations have tried to address whether specific types of studies are more likely to be contradicted and to explain observed controversies. For example, evidence exists that small studies may sometimes be refuted by larger ones.\(^1,2\)

Similarly, there is some evidence on disagreements between epidemiological studies and randomized trials.\(^3-5\) Prior investigations have focused on a variety of studies without any particular attention to their relative importance and scientific impact. Yet, most research publications have little impact while a small minority receives

Context Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

Objectives To understand how frequently highly cited studies are contradicted or find effects that are stronger than in other similar studies and to discern whether specific characteristics are associated with such refutation over time.

Design All original clinical research studies published in 3 major general clinical journals or high-impact-factor specialty journals in 1990-2003 and cited more than 1000 times in the literature were examined.

Main Outcome Measure The results of highly cited articles were compared against subsequent studies of comparable or larger sample size and similar or better controlled designs. The same analysis was also performed comparatively for matched studies that were not so highly cited.

Results Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials \((P=.008)\). Among randomized trials, studies with contradicted or stronger effects were smaller \((P = .009)\) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with “negative” results.

Conclusions Contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes. The extent to which high citations may provoke contradictions and vice versa needs more study. Controversies are most common with highly cited nonrandomized studies, but even the most highly cited randomized trials may be challenged and refuted over time, especially small ones.

JAMA. 2005;294:218-228
Discrepancies between meta-analyses

A guide to interpreting discordant systematic reviews

Alejandro R. Jadad, MD, DPhil; Deborah J. Cook, MD, MSc; George P. Browman, MD, MSc

Abstract

SYSTEMATIC REVIEWS ARE BECOMING prominent tools to guide health care decisions. As the number of published systematic reviews increases, it is common to find more than 1 systematic review addressing the same or a very similar therapeutic question. Despite the promise for systematic reviews to resolve conflicting results of primary studies, conflicts among reviews are now emerging. Such conflicts produce difficulties for decision-makers (including clinicians, policy-makers, researchers and patients) who rely on these reviews to help them make choices among alternative interventions when experts and the results of trials disagree. The authors provide an adjunct decision tool — a decision algorithm — to help decision-makers select from among discordant reviews.

Résumé

LES EXAMENS CRITIQUES SYSTÉMATIQUES DEVIENNENT des outils importants pour guider les décisions relatives aux soins de santé. Comme le nombre des examens critiques...
DISCREPANCIES BETWEEN META-ANALYSES AND SUBSEQUENT LARGE RANDOMIZED, CONTROLLED TRIALS

JACQUES LE LORIER, M.D., PH.D., GENEVIEVE GREGOIRE, M.D., ABDELTIF BENHADDAD, M.D., JULIE LAPIERRE, M.D., AND FRANÇOIS DERDERIAN, M.SC.

Available online at www.sciencedirect.com

Contemporary Clinical Trials 28 (2007) 324–328

Discussion

Mega-trials vs. meta-analysis: Precision vs. heterogeneity?

Ian Shrier a,*, Robert W. Platt b, Russell J. Steele c
Yes, there are problems, but meta-analysis has made and continues to make major contributions to medical research, clinical decision making, and standards of research reportage. However, it is no panacea. Readers need to examine any meta-analyses critically to see whether researchers have overlooked important sources of clinical heterogeneity among the included trials. They should demand evidence that the authors undertook a comprehensive search, avoiding covert duplicate data and unearthing unpublished trials and data. Lastly, readers and researchers alike need to appreciate that not every systematic review should lead to an actual meta-analysis...

David Naylor. BMJ 1997;315:617-619
EPIB672: Systematic Reviews and Meta-analyses
Special Topic in Epidemiology and Biostatistics

Academic credits: 2
Dates: May 3 to 14, 2010
Class times: 2:00 – 4:30 PM, Monday through Friday
Faculty: Dr. Mehlukar Pai, MD, PhD (mehlukar.pai@mcgill.ca)
Enrollment limit: 25

Systematic reviews and meta-analyses are critical for evidence-based clinical and public health practice. The widespread and growing application of systematic reviews to synthesize evidence on key research and clinical questions makes it useful for health professionals to be able to understand and critique this research design. This course will provide a detailed description of the systematic review process, discuss the strengths and limitations of the method, and provide step-by-step guidance on how to actually perform a systematic review.

Specific topics to be covered include:
- Formulation of the review question
- Searching of literature
- Quality assessment of studies, data extraction, meta-analytic methods, and report writing
- The course will also cover statistical issues such as selection of statistical models for meta-analyses, practical examples of fixed and random effects models as well as examples of methods to evaluate heterogeneity and publication bias; graphical and tabular templates for the presentation of meta-analysis data. STATA software package will be used, along with computer lab tutorials on how to effectively use tools such as STATA and Excel for conducting reviews. This course will feature invited speakers who will provide overviews of special topics.

Prerequisites: A background level in statistics (EPI 400) and biostatistics (EPI 360), or permission of the instructor. Students who have not done prior coursework in introductory epidemiology/statistics may contact the instructor.

Recommended textbook: Figure 1, Smith GD, Albrizio LG, editors. Systematic reviews in health care: meta-analysis in context. Second Edition. London: BMJ Publishing Group 2001. URL www.systat.com. All participants will receive a CD that contains useful resources for systematic reviews, such as free software, guidelines, sample data extraction forms, quality checklists, permission slides, etc.

Note: The language of instruction is English, and students are advised that fluency in English is essential to benefit from the course. However, students may submit their course assignments and examinations in French. Courses may be taken for Academic Credit, Continuing Medical Education (CME) Credit, or for a Professional Internship Certificate.