The challenges of a systematic review and meta-analysis of prognosis studies

Richard D. Riley ¹
Willi Sauerbrei ², Douglas G. Altman ³

1. Centre for Medical Statistics & Health Evaluation, University of Liverpool
2. Institute of Medical Biometry and Medical Informatics, University Hospital of Freiburg
3. Centre for Statistics in Medicine, University of Oxford

Funding sources
UK HTA Programme
DH NCCRCD
Aims of this Talk

- Examine if a systematic review and meta-analysis is feasible for prognostic marker studies

- Highlight major problems; including:
  - Poor reporting of results in primary studies
  - Heterogeneity across studies
  - Selective reporting / publication bias

- Consider guidelines and approaches to limit these problems

- Encourage the availability of individual patient data

- Consider reasons to be optimistic
Prognostic Markers

- Also called *prognostic variables or factors*

- Identify different risk groups
  - help to stratify patients for treatment
  - ensure balanced groups within RCTs
  - aid patient counselling

- Include biological, clinical, genetic, histological, pathological and demographic features.

- Example: CEA in colorectal cancer
  MYCN in neuroblastoma
  Age in traumatic brain injury
Evidence-Based Prognostic Markers

- Primary studies of prognostic markers important

- Clinical use of markers ideally based on **overall evidence**

  **This is difficult for clinicians** because:

  - Large number of primary studies
  - Conflicting results
  - Small patient numbers

- Formal evidence-based reviews and synthesis of prognostic marker studies needed
Systematic Reviews
- common approach (e.g. Cochrane)
- identifying, evaluating & combining evidence-base
- systematic & transparent framework

Meta-analysis
- statistical analysis
- combines quantitative results across studies
- produces overall summary of effect of interest
- increase power, reduce uncertainty
- can examine impact of study-level covariates
Meta-analysis using aggregate data

- Traditional meta-analysis uses aggregate data
- Obtainable from publications or study authors
- Meta-analysis of prognostic marker studies usually requires from each study:
  - an estimate of the relationship between the marker and outcome;
    e.g. hazard ratio for overall survival
  - the standard error of this estimate;
    e.g. standard error of log hazard ratio
- Meta-analysis synthesises the results
  e.g. each study weighted by inverse of the variance
Example of a meta-analysis

Is VMA a prognostic marker for overall survival in neuroblastoma?
Is a Systematic Review and Meta-Analysis of Prognostic Markers in Neuroblastoma Possible?

- **Neuroblastoma**
  - most common solid tumour of childhood
  - active research area for prognostic markers

- **Prognostic Tumour Markers**
  - Measurable parameter in the blood, urine or body tissue e.g. CEA (protein), Chromosome 1p (gene).

- **Systematic Review** of primary studies reporting results for a potential prognostic tumour marker neuroblastoma

- **‘A Prognosis Paper’**: one presenting aggregate data or individual patient data (IPD) relating marker levels at baseline to survival
Identifying the Prognostic Marker Literature

Search strategy → Medline/Embase/Cancerlit (1966 to 2000) → 3415 papers identified

Inclusion/Exclusion → 260 prognosis papers

131 different prognostic markers studied in the 260 papers identified

This emphasises the need for evidence-based research
13 markers most commonly reported were selected for further study …
13 markers most commonly reported were selected for further study ...

Median no. of papers per marker = 1
What summary statistic to extract?

Need a statistic that compares time to death and/or recurrence of disease in different risk groups

- **Hazard ratio** desirable because:

  1) *Relative risk* for survival data
  2) estimate (with standard error) of the difference in outcome between groups of patients defined by the marker
  3) Takes into account the whole follow-up period, not just one specific time-point
  4) Incorporates those patients censored (lost to follow-up)
  5) log(HR) approx Normal, aiding meta-analysis models
Extracting the Hazard Ratio …. 

Unadjusted or adjusted Hazard Ratio? 
- **Unadjusted** hazard ratios were sought from each paper 
- If not possible, adjusted results were then sought

Extract the estimates required from each study 
- Papers commonly reported > 1 result, 
  - e.g. for **more than one marker** 
    - for both **overall (OS)** and **disease-free survival (DFS)** 
- Estimates for both OS and DFS desired

What is the overall evidence for each of the 13 markers? 
→ **260 different published prognosis studies** 
→ **575 results or IPD** from which an OS or DFS hazard ratio desired for one of the 13 markers
Extracting the Hazard Ratio and Variance

1) Easy if they have presented the hazard ratio & variance directly:

3 out of 575 (0.52%)
(all from just 1 paper out of 260)

**Table 2. Multivariate risk factors (clinical and molecular) in 149 patients with neuroblastoma stages 1–3 (29 events)**

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\beta/SE(\beta)$</th>
<th>$\exp(\beta)$</th>
<th>Unfavourable</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYCN</td>
<td>2.53</td>
<td>4.26</td>
<td>amplified</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>2.06</td>
<td>5.09</td>
<td>&gt;1 year</td>
</tr>
</tbody>
</table>

hazard ratio (HR)

variance of log(HR)
2) **Indirect** estimation needed (Parmar et al, 1998):

(i) Hazard ratio & CI, or (ii) Hazard ratio & p-value

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CATEGORIES COMPARED*</th>
<th>HAZARD RATIO (95% CONFIDENCE INTERVAL)</th>
<th>P VALUE†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td>III or IV vs. I, II, or IVS</td>
<td>5.6 (2.3–13.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>≥1 vs. &lt;1 yr</td>
<td>3.7 (1.7–8.0)</td>
<td>0.001</td>
</tr>
<tr>
<td>Ferritin</td>
<td>&gt;142 vs. ≤142 μg/liter</td>
<td>6.4 (3.0 – 13.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LDH</td>
<td>&gt;1500 vs. ≤1500 U/liter</td>
<td>4.6 (2.1–9.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Genetic factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-myc</td>
<td>&gt;1 copy vs. 1 copy</td>
<td>6.8 (3.5–13.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chromosome 1p</td>
<td>Loss vs. no loss</td>
<td>6.7 (3.4–13.3)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

52 out of remaining 572 (9.0%)
Cumulative: 9.6% of the 575
(3) Use P-value, group sizes and group events

104 out of remaining 520 (20%)
Cumulative: 27.7% of the 575

<table>
<thead>
<tr>
<th>Prognostic Variable</th>
<th>No. of Patients</th>
<th>Deaths</th>
<th>Expected</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracic site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>227</td>
<td>39</td>
<td>120</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>No</td>
<td>1,108</td>
<td>523</td>
<td>442</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1 yr</td>
<td>490</td>
<td>76</td>
<td>247</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>&gt; 1 yr</td>
<td>845</td>
<td>486</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>211</td>
<td>7</td>
<td>119</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>B</td>
<td>118</td>
<td>15</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>248</td>
<td>61</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>675</td>
<td>465</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>83</td>
<td>14</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>DNA index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>228</td>
<td>129</td>
<td>72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>&gt; 1</td>
<td>426</td>
<td>120</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>N-myc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonamplified</td>
<td>396</td>
<td>94</td>
<td>147</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Amplified</td>
<td>96</td>
<td>73</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

P-value
No. patients
No. events
Use Individual Patient Data to fit a Cox regression model

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Sex</th>
<th>Age (months)</th>
<th>Stage</th>
<th>N-myc</th>
<th>Primary Site</th>
<th>Survival (months from diagnosis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>36</td>
<td>IV</td>
<td>1</td>
<td>Abdomen</td>
<td>44, dead</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>36</td>
<td>III</td>
<td>1</td>
<td>Abdomen</td>
<td>54, alive</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>84</td>
<td>II</td>
<td>1</td>
<td>Thorax</td>
<td>43, alive</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>24</td>
<td>I</td>
<td>1</td>
<td>Adrenal</td>
<td>108, alive</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>168</td>
<td>IV</td>
<td>1</td>
<td>Abdomen</td>
<td>34, dead</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>24</td>
<td>II</td>
<td>1</td>
<td>Abdomen</td>
<td>52, alive</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>108</td>
<td>II</td>
<td>1</td>
<td>Paraspinal</td>
<td>32, alive</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>36</td>
<td>IV</td>
<td>1</td>
<td>Abdomen</td>
<td>22, dead</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>11</td>
<td>IV</td>
<td>&gt;10</td>
<td>Adrenal</td>
<td>27, alive</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>108</td>
<td>IV</td>
<td>1</td>
<td>Adrenal</td>
<td>37, dead</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>12</td>
<td>II</td>
<td>1</td>
<td>Cervical</td>
<td>27, alive</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>18</td>
<td>II</td>
<td>1</td>
<td>Adrenal</td>
<td>41, alive</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>NA</td>
<td>III</td>
<td>15th</td>
<td>Abdomen</td>
<td>24, dead</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>NA</td>
<td>II</td>
<td>1</td>
<td>Thorax</td>
<td>94, alive</td>
</tr>
</tbody>
</table>

41 of remaining 416 (9.9%)

Cumulative: 35.3% of the 575
Survival curve extraction

4 of remaining 375 (1.1%)
Cumulative: 35.5% of 575
Summary of the Overall Number Obtained

- 204 estimates obtained out of 575 desired

- For the other 371 (64.5%) we could not extract a hazard ratio by any of the above methods
Could we have done anything else?

- e.g. use % survival at n years for:
  (i) estimating hazard ratio, or
  (ii) as the statistic in the meta-analysis

- The benefit of this is marginal

- % survival is equally poorly reported

e.g. in 26 prognosis papers for marker LDH:
  - 12 gave actuarial estimates of % survival
  - only 6 of these gave a confidence interval or standard error
  - 4 different time-points used: 2, 3, 4, 5 years
Problem for Meta-analysis No. 1

Poor Reporting of Primary Studies

- Prevents a reliable meta-analysis
- Can not include all the evidence
- Only 35.5% of estimates obtained
- Two thirds of the evidence not available
- May introduce bias

What about more recent studies?

- Reporting has improved
  - e.g. the 26 papers giving a hazard ratio published > 1990
  - Yet, still represents only 17% of the total literature since 1990 assessed
Key Reporting Problems

- No appropriate statistical analysis performed or reported
- **Hazard ratio not calculated** or not reported
- Just p-value provided and not confidence intervals
- **Inexact p-values** provided, e.g. p<0.05 or ‘significant’
- Group numbers and group events not given
Why is this happening?

- Lack of statisticians involved
- Lack of statistical knowledge, understanding and ability
- Lack of guidelines on how to do things better
- Unaware of why improved reporting is needed
- Focus on obtaining publications from primary studies
- No understanding of evidence-based research
- Biased and selective reporting of results?
Evidence of small study effects  
(publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies
Evidence of small study effects (publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies
Evidence of small study effects (publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies
Other evidence of reporting & bias problems

- Kyzas et al. (2005) review 331 cancer prognostic studies
  - conclude that the reporting of study design and assay information was often suboptimal

- Kyzas et al. (2007) review 1915 prognostic marker articles
  - nearly all articles present significant findings
  - < 1.5% were fully ‘negative’ in that they did not present statistically significant prognostic results and did not elaborate on non-significant trends.
Other evidence of reporting & bias problems

- A systematic review of studies of Bcl2 in non-small cell lung cancer (Martin et al., 2003)

Small studies
  - all show a statistically significant relationship between Bcl2 and death

Large studies
  - all 3 are non-significant & show a smaller effect
Other evidence of reporting & bias problems

- Simon (2001) comments that the prognostic literature:
  
  “is probably cluttered with false-positive studies that would not have been submitted or published if the results had come out differently.”

- Rifai et al. (2008) believe that is time to take action against reporting biases in prognostic studies.

- Guidelines have been proposed ...
REMARC guidelines (McShane et al., 2005)

- Aim to improve reporting standards
- Suggest the key information to be reported from a prognostic marker study
- Considers the whole study process from pre-defined hypotheses and patients included... to the statistical methods used and results identified... to the study limitations and implications for practice
- Journal editors are encouraged to enforce REMARK
Other guidelines for improved reporting

- Riley et al. (2003); Altman et al. (1995)
  - how aggregate data should be reported
  - effect estimates & confidence intervals
  - clear presentation of survival curves
  - details of adjustment factors
## An Example of Better Reporting

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alive</th>
<th>Dead</th>
<th>Total</th>
<th>Hazard Ratio (HR)</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ve</td>
<td>4</td>
<td>12</td>
<td>16</td>
<td>2.40</td>
<td>1.19 to 4.84</td>
<td>0.014</td>
</tr>
<tr>
<td>+ve</td>
<td>4</td>
<td>29</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NSE status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>2</td>
<td>9</td>
<td>11</td>
<td>1.45</td>
<td>0.64 to 3.28</td>
<td>0.38</td>
</tr>
<tr>
<td>&gt;=100</td>
<td>4</td>
<td>16</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>16</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LDH Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1500</td>
<td>4</td>
<td>9</td>
<td>13</td>
<td>11.11</td>
<td>3.30 to 37.4</td>
<td>0.0001</td>
</tr>
<tr>
<td>&gt;=1500</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>22</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ferritin Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;150</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>1.92</td>
<td>0.65 to 5.66</td>
<td>0.24</td>
</tr>
<tr>
<td>&gt;=150</td>
<td>3</td>
<td>21</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>16</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>3</td>
<td>10</td>
<td>13</td>
<td></td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>4.09</td>
<td>1.58 to 10.62</td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td>4</td>
<td>12</td>
<td>16</td>
<td>0.80</td>
<td>0.34 to 1.87</td>
<td></td>
</tr>
<tr>
<td>&gt;5</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>0.88</td>
<td>0.33 to 2.32</td>
<td></td>
</tr>
<tr>
<td><strong>NMYC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ve</td>
<td>5</td>
<td>21</td>
<td>26</td>
<td>1.38</td>
<td>0.65 to 2.93</td>
<td>0.41</td>
</tr>
<tr>
<td>+ve</td>
<td>1</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>10</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>8</td>
<td>41</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Survival Curve for MYCN

Log-rank = 6.33
p = 0.012

Time (years)

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ve: no. risk</td>
<td>33</td>
<td>17</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Events</td>
<td>0</td>
<td>16</td>
<td>24</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>-ve: no. risk</td>
<td>16</td>
<td>15</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Events</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
Other guidelines for improved reporting

- Riley et al. (2003); Altman et al. (1995)
  - how aggregate data should be reported
  - effect estimates & confidence intervals
  - clear presentation of survival curves
  - details of adjustment factors

- Burton et al. (2004)
  - encourage clearer reporting of missing data

- Numerous authors encourage availability of IPD
Problem for Meta-analysis No. 2

Heterogeneity of clinical and statistical factors

In the neuroblastoma review, of the 204 estimates obtained there was great variability in:

CLINICAL & REPORTING FACTORS: e.g.
- Cut-off level used to dichotomise the continuous markers
- Method of measuring the marker
- Stage of disease
- Age of Patients
- Type of treatment received
- Outcome – overall or disease-free survival
Problem for Meta-analysis No. 2

Heterogeneity of clinical and statistical factors

In the neuroblastoma review, of the 204 estimates obtained there was great variability in:

**STATISTICAL FACTORS:**
- Type of estimate, e.g. unadjusted and adjusted; indirect and direct
- Adjustment factors
- Analysis method

**DESIGN FACTORS**
- Study design (e.g. Method of marker measurement)
- Purpose of the study
- Study quality
### Example of heterogeneity in the 94 estimates obtained for marker MYCN

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>Cut-off</th>
<th>Point</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFS</td>
<td>46</td>
<td>1 copy</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>48</td>
<td>2 copies</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 copies</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Result Type</td>
<td></td>
<td>4 copies</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>77</td>
<td>5 copies</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>adjusted</td>
<td>17</td>
<td>10 copies</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Stage groups</td>
<td></td>
<td>Mean gene expression</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>68</td>
<td>Positive vs negative protein</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(or staining vs no staining)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>unknown</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2, 3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2, 3, 4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 3, 4, 4S</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3, 4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1 year</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1 year</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Dilemma for meta-analysis

- This heterogeneity exists in addition to the incomplete set of evidence from the poor reporting.
- Is it right to pool the estimates available?
- Clinical and statistical interpretation of meta-analysis results difficult.
- **Could we make strong clinical recommendations?**
  - e.g. clear results for specific stages of disease?
  - e.g. marker X is better than marker Y?
  - e.g. marker X should be used in addition to marker Y?
- Compounded by issue of publication/reporting bias.
<table>
<thead>
<tr>
<th>Paper</th>
<th>N</th>
<th>U/A</th>
<th>Age</th>
<th>Stage</th>
<th>Cut-off level</th>
</tr>
</thead>
<tbody>
<tr>
<td>193</td>
<td>40</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>4</td>
</tr>
<tr>
<td>194</td>
<td>12</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>neg. pos</td>
</tr>
<tr>
<td>199</td>
<td>492</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>200</td>
<td>41</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>1</td>
</tr>
<tr>
<td>214</td>
<td>26</td>
<td>u</td>
<td>all</td>
<td>3,4</td>
<td>1</td>
</tr>
<tr>
<td>216</td>
<td>31</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>1</td>
</tr>
<tr>
<td>246</td>
<td>38</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>1</td>
</tr>
<tr>
<td>247</td>
<td>26</td>
<td>u</td>
<td>all</td>
<td>2,3,4,5</td>
<td>10</td>
</tr>
<tr>
<td>256</td>
<td>32</td>
<td>u</td>
<td>all</td>
<td>3,4</td>
<td>10</td>
</tr>
<tr>
<td>260</td>
<td>28</td>
<td>u</td>
<td>all</td>
<td>1,2,3,4</td>
<td>1</td>
</tr>
<tr>
<td>277</td>
<td>42</td>
<td>u</td>
<td>all</td>
<td>1,2,3,4</td>
<td>10</td>
</tr>
<tr>
<td>285</td>
<td>50</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>3</td>
</tr>
<tr>
<td>288</td>
<td>29</td>
<td>u</td>
<td>all</td>
<td>3,4</td>
<td>10</td>
</tr>
<tr>
<td>297</td>
<td>295</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>3</td>
</tr>
<tr>
<td>306</td>
<td>60</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>3</td>
</tr>
<tr>
<td>316</td>
<td>43</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>3</td>
</tr>
<tr>
<td>337</td>
<td>57</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>neg. pos</td>
</tr>
<tr>
<td>356</td>
<td>58</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>10</td>
</tr>
<tr>
<td>373</td>
<td>68</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>376</td>
<td>237</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>387</td>
<td>60</td>
<td>u</td>
<td>all</td>
<td>unknown</td>
<td>1</td>
</tr>
<tr>
<td>393</td>
<td>45</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>396</td>
<td>110</td>
<td>u</td>
<td>unknown</td>
<td>all</td>
<td>10</td>
</tr>
<tr>
<td>469</td>
<td>37</td>
<td>u &gt; 1 ye</td>
<td>unknown</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>12</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>3</td>
</tr>
<tr>
<td>506</td>
<td>237</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>540</td>
<td>32</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>544</td>
<td>31</td>
<td>u</td>
<td>all</td>
<td>all</td>
<td>unknown</td>
</tr>
<tr>
<td>548</td>
<td>313</td>
<td>u</td>
<td>all</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

**Pooled Result (95% CI)**

\[ 5.48 (4.3, 6.97) \]
Overcoming the Problem of Heterogeneity

- **Collaboration** of research groups required

- Seek **consistency** in cut-offs, adjustment factors, outcomes, analysis, measurement methods etc.

- **Multi-disciplinary** teams

- Improve study design standards (Altman and Lyman, 1998) – e.g. **protocol** driven

- Design **large prospective studies** to answer prespecified questions of clinical interest

- Promote better reporting
Overcoming the Problem of Heterogeneity

- Large, prospective multi-centre studies
- Facilitate access to tumour banks, containing detailed patient-level information
- Collaborate across research groups & pool IPD (e.g. breast cancer - Look et al., 2002)
- Prospectively planned pooled analyses
  - seek common aims
  - agree common design and clinical factors
  - agree to pool IPD at the end
The Benefit of Having IPD From Each Study

- IPD would limit poor reporting by allowing:
  - data checking
  - consistent statistical analysis in each study
  - model assumptions to be verified
  - estimates of interest to be calculated
  - proper handling of continuous variables

- IPD would limit heterogeneity in:
  - type of estimates (adjusted/unadjusted)
  - outcome
  - adjustment factors
  - cut-off level (use continuous level?)

- IPD facilitates
  - analysis of subgroups (e.g. age < 1)
  - analysis of combinations of markers
What to include in the IPD?

For **all markers considered** (not just those ‘significant’), include:

- Relevant patient details (e.g. age, stage)
- exact initial marker level and how marker was measured
- time of disease recurrence (if appropriate)
- follow-up time
- final disease status
- important adjustment factors
- treatment received
## An example IPD

<table>
<thead>
<tr>
<th>no.</th>
<th>Marker levels</th>
<th>Adjustment factors</th>
<th>Survival and disease status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TH</td>
<td>LDH</td>
<td>MYCN</td>
</tr>
<tr>
<td>1</td>
<td>Pos</td>
<td>200</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Neg</td>
<td>350</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Neg</td>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Neg</td>
<td>320</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Difference between within-trial and across-trial relationships

-20
-15
-10
-5

Proportion male

HR

Dotted line: within-study relationship
Red dashed line: across-study relationship
IPD – am I being realistic?

- Researchers protective over their own data
- Worried about Data Protection Act – no need to put in ID number.
- **Cost, time** – when does it become worthwhile?
- To identify the best prognostic markers we need to be prepared to collaborate and share data.
- **Try to make IPD available** - in paper, on Web, on request
- Be involved in prospectively planned pooled analyses
Generalisations to other diseases

- Altman (1995) shows a general standard of poor reporting in survival studies

- Other systematic reviews of prognostic markers limited e.g. Lung cancer (Brundage et al., 2002)
  - median no. of papers per marker = 1

  Brain damage (Zanbergen et al., 2001)
  – small samples & different laboratory techniques

  Prostate cancer (Parker et al, 2001)
  – incomplete & heterogeneous nature of reports

- Increasing evidence of reporting biases (e.g. Kyzas work)

- Lack of consensus regarding design standards
Generalisations to other diseases

- Schmitz-Dräger et al. (2000) review 43 trials regarding p53 immunohistochemistry as a prognostic marker in bladder cancer

- Conclusion:

  “From this analysis it becomes evident that further retrospective investigations will not contribute to the solution of the problem and thus are obsolete.

  There is an obvious need for standardization of the assay procedure and the assessment of the specimens as well as for the initiation of a prospective multi-centre trial to provide definite answers.”
Reasons to be optimistic

- IPD can be obtained, although may be a long process (Altman et al., 2006)
- Meta-analyses have been facilitated when IPD available e.g. in determining a consistent cut-off level (Sakamoto et al., 1996; Look et al., 2003)
- Awareness of reporting biases (e.g. Kyzas work)
- Design guidelines & identification of ‘Phases’ of prognosis research (e.g. Altman and Lyman, 1998; Hayden et al., 2008)
- Reporting guidelines (e.g. REMARK)
Reasons to be optimistic

- The initiation of tumour banks

- Hayes et al. (2008) state that the exciting potential of prognostic markers highlights the “importance of prospective collection, processing, and storage of biospecimens”

- e.g. Goebell et al. (2004): establishing a multi-institutional bladder cancer database & virtual tumour bank to evaluate the prognostic significance of potential markers.

- Many others too; e.g. Confederation of Cancer Biobanks
Reasons to be optimistic

- This meeting!
- Cochrane Prognosis Methods Group
  - Aims to facilitate evidence-based prognosis research
  - Improve design, quality & reporting of primary studies
  - Facilitate systematic reviews & meta-analysis in long-run
  - Bring together prognosis researchers
  - Please join!
Summary

- Evidence-based use of prognostic markers essential

- Systematic reviews & meta-analysis limited
  - Poor reporting
  - Publication bias & selective reporting
  - Small, poorly designed primary studies
  - Statistical, clinical & methodological heterogeneity

- Guidelines for improvement
- Availability of IPD necessary
- Work together – multiple disciplines (involve editors)
- Multi-centre studies
- Prospective meta-analysis


