Letters to the Editor

Epidemiology deserves better questionnaires

From JOHN GALLACHER

Sir—The call for improved standards in questionnaires design in epidemiology1 is long overdue. The proposed clearing house, committee or working group would likely improve standards. The further proposal of setting up workshops or task forces to address specific issues may also promote better practice. The risk attached to institutionalizing reference points for good practice, however, is that the virtues of validity and comparability become the pressure for conformity.

Easier access to questionnaires is welcome. The benefit of a clearing house, however, is not a straightforward issue of facilitating research. Questionnaires and their findings can have clinical and legal, and increasingly, commercial significance. Unless the contents of a clearing house were limited to research tools of unknown validity and significance it is inconceivable that rules and conditions of access would not be required.

Gathering experts to administer a resource could define a committee. The responsible administration of questionnaires has much to commend it. Professional bodies in psychology have practised it for many years. In epidemiology, however, the potential for questionnaire use is so broad, that quite apart from encroaching on the expertise of other disciplines, the task is unworkable. A way forward might be to include only those questionnaires used for medical diagnosis. Sadly, this would defeat much of the object of a general improvement in questionnaire use throughout the discipline.

We all applaud the standardization of procedures, but can a working group achieve this? In fact, this is the heart of the problem. How can a working group set standards for best practice given the range and variety of questionnaire use in epidemiology? Standardization may be a necessary part of an ongoing process of reducing uncertainty in data, but the real ‘standard’ of course, is comparability of findings. The critical parameters of questionnaire use, those which provide comparable findings, are a matter for research not opinion. Appropriate levels of standardization in any area of research will only come through the comparison of results from a known range of varied practice; not from the deliberations of a working group.

At one level, any vehicle which improves the quality of questionnaire data in epidemiology generally, is welcome. It is worth noting, however, that members of the scientific community have a vested interest in improving standards once the issues are understood and publicly acknowledged. Rather than institutionalize the process of improving questionnaire use, perhaps a more fundamental approach is to stimulate the quality of public debate and understanding, in order to improve the judgement of the researcher.2,3 A task which your leading article attempted admirably.

References


Author’s Response

From JØRN OLSEN

Sir—Our aim was to start a debate and to encourage an appropriate body to act. Gallacher’s letter is therefore most appreciated. The starting point is that much good work in developing questionnaires is wasted. Too often, mistakes are repeated and we lack agreed standards for validating questionnaires. Existing questionnaires are not always readily available and the protocol for their use is not known or observed. In part, these problems relate to the fact that there is no organizational protocol for the development, validation, archiving and licensing of questionnaires. We believe that a committee or group (the name is unimportant) is needed to explore these issues further, and provide advice to the wider academic community with regard to the design, management and re-use of instruments. The amount of work that a single committee can undertake is obviously limited, but discussion around these issues would be a useful first step.

As a way forward, we have suggested the value of a few select questionnaires being subject to scrutiny with regard to their robustness in different settings, and their validity in generating good epidemiological data. We suggest that a peer group might develop guidelines for others to follow in designing and conducting research, for example with respect to adherence to protocol in the use of validated instruments. Third, a body (e.g. a designated university) could set up an archive of questionnaires which have wide acceptance as ‘gold standard’ instruments.

In making these suggestions, we are seeking improved quality of research, not seeking to limit creativity or productivity. Indeed, we would hope that the processes we suggest would stimulate further work rather than stifle criticism. We also hope our paper will encourage funding bodies to give higher priority to this area of research.
Sir—Ananth and Kleinbaum1 provide a lucid and informative review of various analytic methods for ordinal responses. However, it appears as if a discrepancy went unnoticed. On reanalysis of the episiotomy laceration data provided by the authors (Ref. 1, Table 3), I find results that do not accord with those reported for the polytomous logistic, adjacent category, proportional odds and continuation ratio models (Ref. 1, Tables 4 and 6). I was unable to assess the partial-proportional odds model due to a lack of available software.

After replicating the data provided by the authors and fitting the ordinal response models using SAS, I found substantial differences in estimated effects (Table 1). For example, where I find the proportional odds ratio (POR) to be 3.03 ($P = 0.01$) the authors reported a POR of 2.10 ($P = 0.01$). Whether it was a misprint of data in the authors’ Table 3, a problem with model specification, or simple transcription that was erroneous is uncertain. The following point assumes data from the authors’ Table 3 is correct.

A danger of models that summarize ordered responses is that changes in direction, i.e. ‘qualitative’ breaks in the summary assumption may go unnoticed or be ignored. After re-analysis, these data indicate strongly that those with episiotomy are protected against a 1º laceration (odds ratio [OR] = 0.44, $P = 0.02$), while episiotomy was not associated with a 2º laceration (OR = 0.86, $P = 0.70$) and increasingly associated with a 3º (OR = 5.21, $P < 0.01$) or 4º (OR = 7.88, $P < 0.01$) laceration. The score test for the proportional odds assumption ($P < 0.001$) reflects a qualitative difference in response and a summary POR is inappropriate. Under the authors’ analysis1 episiotomy was also observed to be protective for 1º laceration (OR = 0.6), however, this protective effect was non-significant (95% CI : 0.3–1.3) and all further levels demonstrated in increasing association. Therefore, attention to the break in the proportional odds assumption was downplayed. Indeed, where breaks of the proportional odds assumption are a matter of degree, i.e. ‘quantitative’, a summary POR may be appropriate.

Such errors may be found easily by comparing the model output against the $R \times C$ table. The purpose of the paper,1 to provide a synthesis of models for analysing data with ordinal responses and to evaluate their usefulness in epidemiological research, holds despite the errors. Ananth and Kleinbaum’s paper1 and another recent review2 provide the basis for burgeoning widespread use of ordinal regression models in epidemiological research. It is hoped that the lesson gleaned herein will be as widespread.

References

Comment from the Editor
The authors of the paper were invited to respond to this letter but did not do so.

Table 1 Reanalysis of laceration data

<table>
<thead>
<tr>
<th>Degree of Laceration</th>
<th>Episiotomy</th>
<th>Polytomous logistic</th>
<th>Adjacent category</th>
<th>Proportional odds</th>
<th>Continuation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\text{A } \times \text{K}^2$ Re-analysis</td>
<td>$\text{A } \times \text{K}^2$ Re-analysis</td>
<td>$\text{A } \times \text{K}^2$ Re-analysis</td>
<td>$\text{A } \times \text{K}^2$ Re-analysis</td>
</tr>
<tr>
<td>None</td>
<td>9238</td>
<td>1204</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1º</td>
<td>140</td>
<td>8</td>
<td>0.63 (0.20)</td>
<td>0.44 (0.02)</td>
<td>0.63 (0.20)</td>
</tr>
<tr>
<td>2º</td>
<td>71</td>
<td>8</td>
<td>1.09 (0.82)</td>
<td>0.86 (0.70)</td>
<td>1.73 (0.29)</td>
</tr>
<tr>
<td>3º</td>
<td>131</td>
<td>89</td>
<td>2.07 (&lt;0.01)</td>
<td>5.21 (&lt;0.01)</td>
<td>1.90 (0.09)</td>
</tr>
<tr>
<td>4º</td>
<td>37</td>
<td>38</td>
<td>5.39 (&lt;0.01)</td>
<td>7.88 (&lt;0.01)</td>
<td>2.60 (&lt;0.01)</td>
</tr>
<tr>
<td>Total</td>
<td>9617</td>
<td>1347</td>
<td>5.39 (&lt;0.01)</td>
<td>7.88 (&lt;0.01)</td>
<td>2.10 (&lt;0.01)</td>
</tr>
</tbody>
</table>

a Results from Ananth and Kleinbaum1
b $P$ value from score test for proportional odds assumption <0.001.
c $P$ value from score test for equal slopes assumption <0.001.