Reviewer's report

Title: Systematic review of disability weight studies: comparison of methodological choices and values

Version: 1 Date: 29 September 2013

Reviewer: Marie-Louise Essink-Bot

Reviewer's report:

This is a report of a systematic review of 22 empirical studies to derive disability weights for DALY estimations. The study was well conducted. It is an important piece of work. However, I think the paper may become more valuable if the authors add a discussion of the potential consequences of their findings.

Major compulsory revisions

p.1, second half of the page: the authors suggest that DW studies after 1996 were undertaken because of a need for DW for national purposes or because of practical limitations. I suggest to add that the first ones (such as the DDW study) were also undertaken to test ('validate') the quite novel valuation procedure (condition specific; PTO + interpolation) of the first GBD, and to try to improve on that procedure (e.g., by adding EQ-5D+ descriptions for standardization of the stimulus).

p.4, top half: contains an explanation of the differences between DW and quality weights for QALYs. I suggest to add that Q-weights are scored in the reverse direction, i.e. 1 = perfect health.

p.4, just below half page: as the authors correctly note in the discussion section, the annual profile approach has not been adopted internationally. However, on page 4 this approach is described only very briefly. I suggest to explain it a little more extensively and to consider adding a figure and/or a concrete example.

p.8, last paragraph above Discussion: It seems that studies that provided [ONLY] a short disease specific health state description resulted in slightly higher [suggest to replace this by BETTER] DW compared to ..... where (in Table 3) can the reader find the data on which this statement was based? And how about the standard deviations; I would expect that studies that presented also a description of functional health to standardize the stimulus would result in smaller SDs.

p.9, second half (Panel composition and intercultural differences): I suggest to change 'intercultural' into 'contextual', because I think the differences are now too easily attributed to differences in 'culture'. Although there are many definitions of culture, countries and populations differ in many more characteristics than only culture; the most important ones probably being wealth and development.

Results section & Discussion: A major finding, if not the main result, of this systematic review are the enormous variations in DWs between studies. I wonder if it is possible to make the size of the variations visible in a figure, for example by
showing the DWs from a number of studies for a small series of similar health states? Moreover, I think the authors need to discuss the consequences of this major finding. Can we now conclude that comparisons of DALY estimates across studies that used different sets of weights are not valid at all? Previous studies (e.g. Essink-Bot ML, Pereira J, Packer C, Schwarzinger M, Burstrom K. Cross-national comparability of burden of disease estimates: the European Disability Weights Project. Bull World Health Organ. 2002;80(8):644-52.) found that at least for some diseases the lack of reliable epidemiological estimates was more important than variations in DW. The authors end their paper by stating that 'the global use of the same set of weights is preferable'(page 11) but this could be strengthened. Only if .... DALY estimates will be comparable over time and between regions. Perhaps they might add a recommendation for WHO to invest in deriving a Universal Set of DW, or, at least, a Universal Interpolation Scale to define the endpoints and some marks in-between.

Minor essential revisions
p7, 5th line from above: please change 'lower' into 'worse'.
p 9, top line: suggst to replace 'finely' by 'detailed'.
Reference list: refs 4, 22 and 47 are identical (3 times the same chapter). Refs 39 and 54 relate to the same study; I suggest to mak reference to the DDW study by using the more easily accessible reference to the EJPH (currently ref 54). Ref 39: first author's name is Stouthard MEA (not MEE).

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I have a non-financial interest in relation to this paper, because I am co-author to two of the studies that are included in this systematic review (Dutch Disability Weights study (referenced in current version as no 39 and 54); European disability Weights Project (42)).