Reviewer's report

**Title:** A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format

**Version:** 4  **Date:** 26 November 2003

**Reviewer:** Marek Brabec

Reviewer's report:

The author addressed questions and remarks raised in the review. Comparison with results of an independent study made in Figure 4 and discussed in the text is certainly valuable. Nevertheless, it is clear that (as the author states correctly in the Abstract): “further validation will be needed in future to establish whether the proposed standard is valuable for general use”. Explanations and description of statistical methods were added as indicated in the author’s response to the review. On the other hand, smoothing of the differences between studies is highly informal and once again, only empirical check can show whether it is justified or not.

Additional minor point is that some of the statistical procedures and tests mentioned in the text could have been done more effectively if some (slightly) more sophisticated techniques were used. Just one example: more efficient estimate (so called L-estimate) of standard deviation can be based on more percentiles than just the 3rd and 97th (which were actually used and mentioned on page 8).

Statement appearing on page 13: “The larger sample sizes used here may provide better confidence in the extreme percentiles.” should be taken cautiously. This is because validity is influenced both by variability and bias. While variability inevitably goes down as the sample size increases, bias (which might be induced by non-representative coverage of the primary studies and other factors) does not need to go away even in very large samples. It remains to be shown in future studies, papers and analyses of the chart’s application to independent datasets whether bias or random variability is more important.

From the statistical point of view, also the phrase quoted on page 14 is somewhat problematic: “While measurement errors are important for accurate growth monitoring of an individual, they might not be as important for population surveys since errors tend to be evenly distributed.”