Reviewer's report

Title: An evaluation of classification systems for stillbirth

Version: 2 Date: 3 October 2008

Reviewer: Jan Jaap H.M. Erwich

Reviewer's report:

Major compulsory revisions

Re: MS: An evaluation of classification systems for stillbirth.
Authors: Flenady et al.

Before going into my comments on this manuscript I want to address a potential conflict of interest to make things absolutely clear regarding this review process. Since I am one of the developers of one of the used systems as described in this paper, I am of course biased towards the use of my own system (TULIP). However, I am familiar with this study already since the authors and myself have discussed it several times during meetings. In addition, the editor of BMC is probably aware of my involvement in this field. I feel comfortable to give my comments and I support the process of open peer review, since the authors can comment back on my comments, knowing the person involved.

I can not agree more with their general conclusions, and it is mandatory that an international effort is made to bring this field further and propose a workable classification for both developed and developing countries in order to bring stillbirth numbers down globally. The nine teams did a lot of work to try to classify their cases. InfoKeep and Ease parameters, although subjective, may give some sort of impression on how systems work in real life.

However, some questions remain. Although the general conclusions are valid, whether the detailed analysis on the use, the performance and the comparison of the systems is valid to such an extent is questionable.

Main outcome parameters were 1. the ability to retain important information (using InfoKeep), 2. ease of use (using Ease). 3. inter-observer agreement (Kappa) and 4. proportion of unexplained stillbirths. The proportion of unexplained is a way to analyse one of systems main goal to identify the underlying cause of death (see pg4). However, all 4 items are used equal and a large variation in proportion unexplained exists. The issue of unexplained is hardly elaborated on in the discussion.

Classification for the use in perinatal audit is important to analyse whether substandard factors (if present) are direct responsible for the death. A chronological report with time-scale of the events taken place during pregnancy and delivery is essential for the proper interpretation. No information is given whether systems seems to keep this important information, which I think they
probably do not, some comments on this issue are needed. Systems which focussed on examining suboptimal care were excluded (why was the Nordic/Baltic excluded in this respect, ref 34) but that does not mean that the remaining systems do not need to keep important information for audit purposes.

Why were automated computer systems excluded (pg 7)?

The use of InfoKeep and Ease parameters is sometimes comparing apples with pears. For instance, Fetal growth restriction as a category will always score low in TULIP, since this clinical condition is not a cause of death, but is classified as a contributing factor. The way InfoKeep is explained on pag 8/9 of the methods is not clear, was it only used when the category scored yes? It is much focussed on retaining the clinical conditions, which is OK, but not in a system to classify cause of death. What was the rationale to perform subgroup analyses using these specific categories (IUGR, placenta PA, congen abn etc) ? (pg12), this choice will influence the results to a great extent.

One of the main problems in classification is the mix of clinical conditions with causes/pathology which are or are not the underlying cause which leads to non-equivocal classification illustrated with suboptimal agreement such as this paper shows as well.

If TULIP scored low on ease, how is the best kappa explained?

Although the teams were experienced, wide variation in the proportion of unexplained was shown, reflecting differences in interpretation of the systems (pg14). Why should these differences in interpretation be not a cause of variation of the InfoKeep and Ease parameters?

The authors state correctly that they were involved in the development of two systems and they describe the measurements undertaken to try to avoid bias, on which they should be congratulated. The teams were experienced, however no details on the instruction or prior knowledge of systems is given apart from the remark that no training was provided. It is known that the direct use of a system from obligatory brief publications can be difficult and some short training improves the quality of classification. For instance in TULIP multiple important factors (see Background) are stored as well in a standardized way, but this has not been done in this study. No prior publications on CODAC are available in the peer-reviewed public domain (I have not yet access to the link mentioned in this paper BMC article Froen et al 2008) and although bias is avoided, it cannot be excluded. I am aware of the contents of CODAC by other routes and think it has potential to be widely used, however no details are given on this system in this paper where other systems are detailed present in the supplementary files. Readers which are unfamiliar to this field may find this a problem.

In conclusion, this paper is very worthwhile in addressing the importance and problems in the classification of stillbirths (or perinatal deaths). This is an important issue and warrants publication of this paper, however there is too much emphasis on details of the scoring, which suggests an accuracy which is not necessary true and is more apparent than real. If this is addressed by the authors, publication and many citations will be no problem.
Minor essential revisions
I have no comments regarding the text, language, style or structure of the paper. On pg 20 it is stated that CODAC has a user friendly interface, although this is probably true, no data are shown to underline this statement. A period is lacking after: varies. Keeling (on pg 20)

**Level of interest:** An article of importance 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:**

See my comments above.