Reviewer's report

Title: A Tutorial on Pilot Studies: The What, Why and How

Version: 2 Date: 25 September 2009

Reviewer: George Borm

Reviewer's report:

The authors provide an examination and description of the key aspects of pilot studies. This would be interesting, and of interest, because there is a lot of confusion: what is a pilot study and what qualities (with respect to sample size, thoroughness of methods and reporting, etc.) it should have.

Major compulsory revisions

1. However, the article rather makes me more confused, mainly because it is not clear to me what they mean by a pilot study.
2. According to the definition in the abstract, all studies, apart from phase III and phase IV seem to be pilot studies. So phase I and phase II studies are special cases of pilot studies? But then there is a lot in the article that I do not understand, because it deviates from what is usually considered to be appropriate for phase I and II.

In my view, phase II and I trials are not pilot studies. They do not (yet) have the aim to investigate the feasibility of large studies (in fact they may be large themselves). They do not fulfill the criteria in table 1. Their aim is to identify a safe and possibly efficacious treatment. How this treatment then has to be tested in a larger trial is another matter.

In any case the authors should clarify the difference (if any) between pilot, phase I and phase II studies.

3. The authors explicitly discuss the relationship between proof of concept (poc) studies and pilot studies. I am not sure why, because (in their view) a poc study simply is a special case of pilot study.

4. The authors even consider studies that evaluate marker data in cohorts as pilot studies. So they not only discuss clinical trials, but also observational studies? However, further in the article they create the impression that they are talking about trials only, for example, they propose the CONSORT format for reporting. In fact, throughout the paper I have the impression that the authors discuss trials, but I may be wrong.

Minor revisions (in no way exhaustive or conclusive, because I am not sure what the authors consider a pilot study and they would depend on that):

1. I do not think that dose finding studies get no attention in the literature.
2. I cannot see how there can be confusion between a pilot study and a study with an adaptive design. There can be doubt as to how to choose between them, to decide what is the best approach in a particular situation, but once the decision is taken, it is obvious what is what. Whether or not a study is a pilot, depends on its objectives. Whether or not an adaptive method is used is a strategy to reach that objective and is of a completely different nature. Both a pilot and a non-pilot could be adaptive.

3. On page 11, the authors claim that data from the pilot can be combined with data from the main study. In general, this is definitely not the case. For example, multiple testing issues may arise or other opportunistic actions may create bias: the investigator may be tempted to include the results of the pilot only when they are 'suitable'. If any pooling is considered, this should be planned beforehand and the protocol should describe in advance how and discuss the statistical consequences and methods.

4. On page 16, the authors state that the primary outcome of the main study should be defined in the pilot. But is it not the case that pilots may be conducted with the aim to find out what is a suitable (practical or surrogate) endpoint?

5. I would be interested in the opinion of the authors about randomization and blinding in pilot studies. Should randomization (if it is done) not be clearly described in the report?

The authors could provide a useful contribution when they first could provide a clear, logical and consistent framework to classify studies. A framework that is an extension of existing frameworks, such as the Phase I-IV classification. Maybe they could provide an organogram of the different types and subtypes of studies (or maybe just trials)? I would advice them to choose a rather narrow definition of pilot trials. If it is too wide, it is difficult to provide guidance. Once this is clear what they are talking about, guidance in the way they now propose would be useful.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I have no competing interests