Reviewer's report

Title: Patient safety in Dutch primary care: Study protocol

Version: 3 Date: 15 January 2010

Reviewer: Cordula Wagner

Reviewer's report:

The article describes the study protocol of a patient safety study in Dutch primary care. A combination of methods will be used to get insight into the incidence, type, impact and causes of these incidents.

In general the article is well written and follows the standard approach for these kind of studies that has been described earlier by other authors.

Nevertheless, I have a couple of questions and remarks.

Does the study protocol describe the actual situation of the data gathering or does it describe the way researchers want to handle the data gathering? It is important to describe the reality, because in a later stage in method sections of result articles authors will refer to this study protocol.

It is confusing that the authors use various terms: are they looking for unintended injuries, unintended harm, preventable adverse events or incidents. Example: “.. which types of adverse events have to be recognized as incidents and which are not.” All adverse events (at least based on the definition of international adverse event studies) fall in the broad definition of an incident.

These terms differ in their definition and focus. In most existing studies where record review had been used the focus lies on adverse events, judged by three criteria. Studies on incident reporting have a broader scope; they look for incidents (this includes near misses and process deviations). In most cases incident reporting can not be used to establish the incidence of safety problems. It can of course establish the type, impact and causes of safety problems. Authors should state more clearly which information they will use for the investigation of incidence rates and what for the other objectives.

The central definition for this study is the definition of a patient safety incident. This is a very broad definition and therefore very difficult for reviewers to use. Have near misses be included? If yes, was the whole range included? If no, why not using the definitions of adverse events?

The context and content of the training of reviewers is very important. How extended was the training, what was the content of the training, was there a learning curve and pilot reviews at the beginning? How many meetings have been organised?

Have the authors used a clear distinction between the event with unnecessary harm and possible preventability?
The authors want to investigate the inter-rater-reliability based on 50 records. If it is true what the authors assume, e.g. that in 3% of the contacts an incident occurs, we can expect to find 1.5 incidents in the sample. I am not sure whether it will be possible to calculate a kappa value. In nearly none of these records an incident can be found. So, the agreement will be high, but for a kappa calculation you need enough cases in every box. A solution could be the stratification of records.

The authors state that experience showed that a sample size of 20 practices will be large enough to give reliable results. On which study (experience) is this assumption based and what are the numbers of that study?

The authors state that for logistical reasons it is acceptable to sample in one or a few geographical areas. I understand the logistical reasons, but is it also acceptable for the study objectives. Lies the focus on Dutch primary care or the study population?

What is the standardised rate to compensate practices for their activities?

The records have been used by a team of researchers and health professionals. Later on criteria for reviewers have been given. Did the researchers fulfil the same criteria (e.g. clinical experience, reputation)? In one of our pilot studies we found a difference in judgement between researchers and professionals.

How have reviewers been trained to use the PRISMA method? It is a complex method and very hard to use with the limited information in patient records.

Records will be reviewed one year prior to the selection date. Will they also be reviewed one year after selection date?

How have reviewers judged the impact/severity of an incident?

One limitation that has not been mentioned is Hindsight bias.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests.