Author’s response to reviews

Title: Patient safety in Dutch primary care: Study protocol

Authors:

Mirjam Harmsen (M.Harmsen@iq.umcn.nl)
Sander Gaal (S.Gaal@iq.umcn.nl)
Simone van Dulmen (S.vanDulmen@iq.umcn.nl)
Eimert de Feijter (E.deFeijter@iq.umcn.nl)
Paul Giesen (P.Giesen@iq.umcn.nl)
Annelies Jacobs (J.Jacobs@iq.umcn.nl)
Lucie Martijn (L.Martijn@iq.umcn.nl)
Theodorus Mettes (D.Mettes@dent.umcn.nl)
Wim Verstappen (W.Verstappen@iq.umcn.nl)
Ria Nijhuis-van der Sanden (R.Nijhuis@iq.umcn.nl)
Michel Wensing (M.Wensing@iq.umcn.nl)

Version: 4 Date: 15 March 2010

Author’s response to reviews: see over
Cover letter 'Patient safety in Dutch primary care: Study protocol'

We like to thank the reviewers for their very useful comments. Below our reply to each of their remarks.

Reviewer 1:

#1 Thank you for asking me to review this paper. The proposal covers a very important area of primary care which is in its infancy in terms of research. This proposal will need some major modifications to achieve its aims. The protocol in its current state not specific or clear enough on both what it is trying to achieve or how the research is to be carried out. I will make specific comments as to the text, but feel in general that the proposal needs to be a clearer description of what it is trying to achieve, and perhaps use a flow diagram or similar to demonstrate how it will be carried out.

>> We tried to specify the relation between objects, methods, and measures by 1) adjusting the objectives in the ‘Aims and objectives’ section (e.g. adding the objective for the survey), 2) adjusting Table 2 (former figure 1) by adding the objectives), and 3) referring to Table 2 in the ‘Design and methods’ section.

#2 (minor discretionary) P5 last sentence: this is a description of significant event reporting: it would be useful to use this term, rather than just describe it

>> We added the words ‘significant event’ to the sentence ‘When an adverse event has occurred (e.g. the patient died during treatment), a significant event analysis has to be made to determine the preventability of this adverse event’.

#3 P6 first para: this states ‘insight in the frequency and seriousness of potential unsafe situations might improve patient safety’- the paper needs to discuss whether this is just an assertion of the authors, or whether there is any evidence that this is actually the case- and quote evidence where it exists

>> We write ‘might’ to indicate that the evidence is anecdotal. We have rephrased the sentence into ‘Therefore, insight in the frequency and seriousness of potential unsafe situation may be a first step towards improving patient safety’.
#4 (minor discretionary) P6 second para, although interesting only refers to secondary care, so not strictly relevant to this protocol

>> We shortened this section.

#5 P 7 para 1: there is important research not mentioned here. For example,

>> We have included a number of the suggested references, where we felt that these were indeed highly relevant. Although it is important to build on previous research, we do not claim to provide a systematic overview of the literature.

#6 (minor discretionary) P8 last para, really is discussing aims, best to include in the next section. It also needs to follow a logical sequence: e.g. identify under performance (not gaps), examine underlying factors, tailor interventions, test interventions, then set targets for improvement.

>> We replaced ‘gaps’ by ‘under and over performance’ (it is not just under performance), adjusted the sequence, and included the paragraph in the next section.

#7 (minor essential) P9 para 1: objective (3) seems to be a repeat of parts of objectives 1 and 2, unnecessarily complicating the study

>> We rewrote this section, so it should be clearer now what our objectives are.

#8 (minor essential) Definitions of harm will need to be more tightly defined for this study to work- as the authors recognise, but the criteria in box 1 are I think is too vague for reproducibility between different observers, we will need more information to make this definition more precise, or evidence that these criteria actually will work.
These criteria actually did work. The procedures, among them the interpretation of information from the patient records, were pilot-tested, and found to be working. We also measured the percentage of agreement between reviewers. For more details see the revised version of the manuscript.

#9 (minor discretionary) P10 para 1: ‘broadly accepted as valid’: I understand where this idea comes from, but this runs the risk of increasing variation- and disagreement- between observers. For example this definition might have been interpreted by one observer to include the absence of a primary prevention intervention in- for example- aspirin use as was recommended by some guidelines in the UK recently. However in this case this advice was not derived from Grade I/II/III levels of evidence but expert opinion, and later shown to be incorrect in the clinical trial setting. So this ‘error of omission’ in the eyes of guideline followers turns out to be a correct patient care pathway for the evidence based practitioner, and the guideline incorrect. I would suggest that the authors consider sticking to evidence for interventions which are supported by clinical trials of at least grade I/II/III levels of evidence.

#10 (minor discretionary) P14 para 1; selection criteria for interviewers is rather loosely defined, in terms of reputation and experience, and could be more specific.

#11 (major compulsory) P14 second para; it is not clear how many reviewers will review the patient record in the retrospective study- if I understand it correctly I think its only 1 reviewer for 50% of records, and 2 for the remaining 50% for quality assurance. However this is not clear, and a (for example) flow diagram of
the study design (including time scales) would help the reader to understand the method clearly.

>> We rewrote the section.

#12 (major compulsory) P15 para 1; if the review only discusses those cases they are unsure of meeting the criteria, then there is still a risk that incidents are included/excluded inappropriately when one reviewer thought their decision was correct: therefore I think the authors need to consider 2 reviewers looking at all records.

>> We adjusted the section; two reviewers will look at uncertainties and all incidents possibly to be included.

#13 (minor discretionary) P15 para 3: is chart audit the same as patient record? If so please use one term consistently for clarity.

>> We changed the term in all cases into 'patient record study'.

#14 (minor essential) P16: we will need more details about what the survey questionnaire includes, perhaps include a copy of it?

>> The questionnaire itself is too large to be included in the paper. However, we did add more information to the text. Besides, information is stated in Table 2..

#15 (minor essential) P16 para 4, power calculation; I am not familiar with this approach and think it needs a more knowledgeable opinion from a specialist in statistics; however if the assumption is correct of an incidence rate of 30 patient safety issues in the 1000 records reviewed, then it seems to be a very small sample to meet the study objectives 1+2, of determining the incidence/type/impact/causes of these incidents, as there will be many possible different types of incidents (though prescribing error is the most likely from previous research). Pirmohamed et al looked at over 18,000 case records of higher risk patients in their (those admitted from primary care to hospital)

>> The calculation of the accuracy of estimates was focused on the 95% confidence intervals regarding the total number of incidents (not the estimates for each type of incident). We
agree that 1000 medical records is still a fairly small number for accurate data on specific types of incidents.

#16 P17: a significant (but perhaps unavoidable) weakness of reviewing notes looking for evidence of harm is that if this has not been identified by the attending clinician, then its possible that the information included in the case notes might not lead the reviewer to come to the correct conclusion of harm, leading to under detection of cases.

>> Remark added to the discussion section.

#17 Clustering of practices skewing the results is discussed by the authors, but quite possibly will have the effect of decreasing the power of the study.

>> The power is adjusted for the effect of clustering by applying the intra cluster correlation (ICC), as mentioned in the section ‘Accuracy of figures’.

#18 (minor essential) P19 para 1: need to give more information on how level of harm is defined into the different groups mentioned

>> We explained the different classifications in more detail in the ‘data processing and data analysis’ section by adding Table 4.

#19 (minor essential) P 19 para 3: please can the authors clearly state what analyses they will be performing: I assume this will include regression, but if so we need to know which dependant and independent variables, dummy variables to identify practice clusters etc. Are the retrospective reviews and prospective incident reporting study going to be linked? What precisely will happen/ what data will be collected in the prospective study?

>> To be honest, we did not specify the regression analyses in such detail as suggested, although we mostly use the same type of approach. It seems most appropriate to label this as ‘explorative analyses’, and describe it in general terms. Please note: We have started to do these analyses, and it would be inappropriate to suggest that we had planned these a priori in such detail.

We removed hypothesis 2.
#20 (minor essential) P 20 ethics; an area where there may be greatest risk is when a reviewer detects probable harm to a patient that previously wasn’t identified, and what precisely will happen in this case (? Just discussion with internal ethics committee, will the patient be told, if so what is the potential risk, will this impede recruitment for the study?)

>> We rewrote the section.

#21 (major compulsory) The abstract at the moment isn’t clear, but will need to be revised when the above points are addressed.

>> We revised the abstract.

#22 The 2 attachments I couldn’t read, as they were in Dutch. It is possible that they may answer some of the points mentioned above

>> The two attachments were 1) a statement of the CCMO that ethical approval is not necessary for this study, and 2) a letter by the Dutch Ministry of Health, Welfare and Sport (VWS) about financing the study. So, none of the points mentioned by the reviewer were answered by these documents.

#23 (major compulsory) So in summary, this subject area is important; this protocol needs to be written much more clearly, so the reader can easily understand the methods used, how the prospective and retrospective components inter relate (if indeed they are meant to), details of the proposed analyses and how the authors will deal with previously unrecognised but serious safety issues that this study might uncover. The literature review needs to be widened. The power calculations need to be examined by someone more experience in this area than me, to ensure this study will achieve its aims and objectives.

>> See reactions to the specific remarks.

**Reviewer 2:**
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.

#1 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.

>> We rewrote the protocol so it describes the actual situation. However, the differences between planned and actual procedures were rather small.

#2 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.

>> We added a short description of the terms ‘incident’ and ‘adverse event’ in the ‘Background’ section. Throughout the text we tried to use these two terms. But because of the different terms used by different other studies, and because we want to reproduce the results of these studies as much as possible, it is unavoidable to use some other terms as well. For our own study we used the term ‘incident’.

#3 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?
We wanted to explore what went wrong, in a very broad context. That's why we used the term 'incident'. As the definition states, incidents might not have resulted yet in harm, but it might be possible it still results in harm.

#4 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?

We rewrote the section by adding more information.

#5 Have the authors used a clear distinction between the event with unnecessary harm and possible preventability?

Implicitly we felt that unnecessary harm implies possible preventability, at least in an ideal world. We reflected on this remark, but do not see how harm can be unnecessary and yet not preventable.

#6 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 very box. A solution could be the stratification of records.

We agree with the reviewer, and changed it into only the percentage of agreement.

#7 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?

This experience is indeed experience-based and not evidence-based. An example of a study with about 20 practices per country is the EPA study, which was awarded in 2009, and led to several publications. Here are some examples of the publications:

#8 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?

>> We accept that this is not a fully representative sample for Dutch primary care or the Dutch population. However, we tried to reflect the national situation as good as possible by the stratification scheme for practices, and the random sampling of patients. The geographical areas were chosen for pragmatic reasons, but also with a view on stratification by urbanisation level, in order to reflect the national situation. However, 20 practices is still a fairly low number, they were all volunteers, and we used a clustered sampling procedure (clusters defined by regions). These are all limitations for the representativeness to Dutch primary care.

The representativeness for the Dutch population may be better. Note that practices have listed a larger number of patients. For instance, the 20 practices had listed approximately 50,000 patients. Of these, approximately 37,500 have visited the practice in the recent 12 months. These 37,500 individuals can be seen as the sampling frame, from which our 1,000 patients were randomly selected.

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

>> We used a standardised rate within the project. This is added to the text.

#10 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.

>> Reviewers were researchers with clinical experience or health professionals. We adjusted the text to make this more clear.
#11 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.

>> The study protocol was not clear enough about the use of this method. We did not actually use the whole PRISMA method, but just the classification into the Eindhoven Classification Model, which is part of the PRISMA method. This ECM has different levels; we only used the first levels.

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

>> No, they will not be reviewed one year after the selection date. Please note that the selection date is not the date an incident occurred. The selection date was the date two months before the researchers visited the practice to select records. A slightly different method was used in the GP cooperatives, in which the patient record was prospectively reviewed in the GP practice. To make this more clear, Table 3 is added.

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

>> The judgement of the impact/severity was based on the information found in the patient records or written down on the incident reporting forms. These results were classified according to the International Taxonomy of medical errors in primary care (see reference list).

One limitation that has not been mentioned is Hindsight bias.

>> We added a sentence to the Discussion section.

Nijmegen, March 15, 2010

Mirjam Harmsen, Sander Gaal, Simone van Dulmen, Eimert de Feijter, Paul Giesen, Annelies Jacobs, Lucie Martijn, Theodorus Mettes, Wim Verstappen, Ria Nijhuis-van der Sanden, and Michel Wensing