Reviewer's report

Title: Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: an experimental study

Version: 1 Date: 5 November 2005

Reviewer: Isabelle Colombet

Reviewer's report:

General comments:
------------------
This paper describes a study which is of high interest for readers who are interested in clinical decision support systems and their evaluation. However, such a complex and rich system as ISABEL raises difficult methodological problems and such a study is difficult to report clearly. I found this paper too long, difficult to read and confusing. The background (especially previous results obtained with ISABEL) and objectives should be more synthetically and clearly stated. Several aspects of methods (especially outcome measures) are also to be clarified, for the reader to be able to appreciate the interest of the results. These are the reasons why there are more questions than constructive suggestions in my report.

Major Compulsory Revisions
-----------------------------

1) p9: the term “the simulated field study” seems a little paradoxical. I would not speak of a field study as the use of the CDSS is not integrated in the clinical workflow of the doctors.

2) Similarly, in the main title of the paper the term "quasi-experimental study" would be more appropriate to the design of the study.

3) The objective of the study is more clearly enunciated in the abstract than in the core paper. The background section reports the overall reasoning and rationale underlying the development and evaluation of the ISABEL system. This may be synthesized by more clearly and directly by stating the main hypotheses underlying the conception of this system and summarizing the results already obtained in the first steps of its evaluation.

4) The objective of the paper could be positioned more clearly in what could be the global strategy of the evaluation of the system. I would imagine that the different steps of this evaluation could be:
   I. Validation of the system:
      a. validation of the interpretation by the system of the case description data entered by users in natural language
      b. evaluation of the quality of the diagnostic suggestions given by the system for a structured case
   II. Evaluation of the potential impact of the system on decision making: simulated cases, quasi-experimental design
   III. Evaluation of the impact of the system, when used in the routine workflow of care: (real case, on the field, quasi experimental design)
      a. On decision making
      b. On different other processes of care (actual prescriptions, hospitalization, length of stay, ..)
   A summary of results obtained in step Ia and Ib should be recalled in background section, and the extent to which the interpretation of the results of the step II described in this paper, depends on previous validation steps should be addressed in Discussion
The report on the “measurement study” (p 7 last § of Background section) could be more appropriately detailed in the Methods section.

5) The measure of outcome (DEO, quality scoring method) is not described in sufficient details so that the work could be replicated.
“The expert panel also provided measurements quality for each of the other clinical decisions”: which dimensions of quality are considered? Are they objectively measured and was the agreement between panel experts evaluated? What was the reliability and validity of the scoring system referred to in ref 50 and what is the validity of its evaluation? (It is mentioned at the end of the Background section that this scoring system was developed and validated on a subset of the same simulated cases as used in the study: what was the gold standard and performance indicator used for this evaluation?)

6) Computation of the DEO: the differential diagnoses are expressed by the user in natural language, sometimes “medical slang”, probably with a lower degree of precision than the diagnoses suggested by the ISABEL system (i.e. “sepsis” proposed in Figure 1).
How is this language interpreted by the panel to be compared to the differential diagnoses suggested by ISABEL and to compute the DEOs? Is this process applied independently by both experts, to which extent did they agree, etc.?
Since answers to these questions are not clear in the paper, it is difficult to appreciate the validity of the main outcome measure. The authors do not give sufficient details on the measurement method, so that the reader be able to appreciate in the results whether the reduction in DEO pre and post ISABEL consultation can be attributed to the impact of the CDSS or to a measurement bias of this outcome.
Describing a complete example of scenario of use and outcome measure may clarify the whole process (for example, how was DEO computed in the example of case and work up described in figure 1 and 2)

7) How many different cases represented 624 case episodes? Should the number of different cases (done by all subjects) be taken into account in any way? Is there a correlation of results obtained across different subjects for one given case?

8) The additional tests and treatment decisions are a valuable outcome (which may be more pertinent than DEO to appreciate the potential concrete impact, and to reflect the operational consequences of diagnosis decision making). Results on this outcome should be more detailed (P19: “some important tests and treatment steps…”) and discussed.

Minor Essential Revisions
----------------------------

9) URL given p 10 (http://trial.isabel.org.uk) is useless since it is not accessible by the reader.
10) Table 1 and Figure 4 present redundant information. Information on participants and cases could be gathered in a unique flow chart or table. The difference between “case” and “case episode” should be explicited in the legend of figure or table (not only in text).
11) Impact on process is enough when the process is well established (evidence-based) as beneficial for clinical outcome. To which extend is it pertinent to consider that the DEO avoided by ISABEL would have had clinical impact…Could this question be discussed more?
12) Conclusion p 20, line 7: as far as I understand, the end of the sentence (“a clinically important alteration in diagnostic decision making resulted from 1 in 8 consultation”) refers to the 12% of cases episodes for which a ‘clinically important diagnosis’ was added. This sentence could be clarified.

Discretionary Revisions
-------------------------
14) Table 1: add percents, add column for total number of cases
15) Table 3 and 4: add some indicator of statistic dispersion, along with the mean

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

**Quality of written English:** Acceptable

**Statistical review:** Yes

**Declaration of competing interests:**

Statistical review
------------------
I would suggest that the manuscript be reviewed by an expert statistician because the choice of the units of analysis and the way to handle the clustering effect (of doctors and case) need to be specifically reviewed.

Declaration of competing interests
-----------------------------------
I declare that I have no competing interests