

Research article

Open Access

Task-oriented evaluation of electronic medical records systems: development and validation of a questionnaire for physicians

Hallvard Lærum* and Arild Faxvaag

Address: INM, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

Email: Hallvard Lærum* - hallvard.lerum@rikshospitalet.no; Arild Faxvaag - arild.faxvaag@medisin.ntnu.no

* Corresponding author

Published: 09 February 2004

Received: 17 July 2003

BMC Medical Informatics and Decision Making 2004, 4:1

Accepted: 09 February 2004

This article is available from: <http://www.biomedcentral.com/1472-6947/4/1>

© 2004 Lærum and Faxvaag; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Evaluation is a challenging but necessary part of the development cycle of clinical information systems like the electronic medical records (EMR) system. It is believed that such evaluations should include multiple perspectives, be comparative and employ both qualitative and quantitative methods. Self-administered questionnaires are frequently used as a quantitative evaluation method in medical informatics, but very few validated questionnaires address clinical use of EMR systems.

Methods: We have developed a task-oriented questionnaire for evaluating EMR systems from the clinician's perspective. The key feature of the questionnaire is a list of 24 general clinical tasks. It is applicable to physicians of most specialties and covers essential parts of their information-oriented work. The task list appears in two separate sections, about EMR use and task performance using the EMR, respectively. By combining these sections, the evaluator may estimate the potential impact of the EMR system on health care delivery. The results may also be compared across time, site or vendor. This paper describes the development, performance and validation of the questionnaire. Its performance is shown in two demonstration studies ($n = 219$ and 80). Its content is validated in an interview study ($n = 10$), and its reliability is investigated in a test-retest study ($n = 37$) and a scaling study ($n = 31$).

Results: In the interviews, the physicians found the general clinical tasks in the questionnaire relevant and comprehensible. The tasks were interpreted concordant to their definitions. However, the physicians found questions about tasks not explicitly or only partially supported by the EMR systems difficult to answer. The two demonstration studies provided unambiguous results and low percentages of missing responses. In addition, criterion validity was demonstrated for a majority of task-oriented questions. Their test-retest reliability was generally high, and the non-standard scale was found symmetric and ordinal.

Conclusion: This questionnaire is relevant for clinical work and EMR systems, provides reliable and interpretable results, and may be used as part of any evaluation effort involving the clinician's perspective of an EMR system.

Background

Evaluation is a challenging but necessary part of the development cycle of clinical information systems like the electronic medical records (EMR) systems in hospitals. EMR systems handle the storage, distribution and processing of information needed for health care delivery of each patient. Such systems have been described as "complex systems used in complex organizations", and their evaluation seems to follow that logic. It is generally believed that multiple perspectives need to be considered, and that qualitative and quantitative methods should be integrated when evaluating EMR systems [1]. In addition, the evaluation should include a comparative element [2] and rely heavily on how humans react to the system [3]. Since the multi-perspective, multi-methodical approach easily exceeds any perceivable amount of allocated resources, methods that require modest resources should be considered whenever possible. Task-oriented self-reporting of EMR use and task performance is one such quantitative method.

In this paper, we present a new questionnaire instrument. The questionnaire may be used to survey and compare the physicians' use of and performance with a given EMR system at various points of time. Furthermore, it may be used to compare general patterns in use and performance to that of EMR systems in other hospitals and from other vendors. EMR use is not necessarily a quality indicator by itself, but an indicator of potential impact of the system. Specific problem areas may be identified by demonstrating a self-reported lack of EMR use or a reduced reported performance of specific tasks. Although clinically oriented task inventories have been published previously, these task inventories have been found either too broad [4,5], or too detailed [6] for the questionnaire's intended purpose. Also, very few of them have been tested in several sites or with various EMR systems. Bürkle et al [7] states that questionnaires should be specified depending on the functions of the observed computer system. The design of the questionnaire makes this specification possible, as the tasks generally follow the boundaries of common EMR functionality. In addition, a table of minimum functionality requirements for each task is publicly available [8]. In this paper, we describe the development and successful application of the questionnaire in two demonstration surveys. Support for the validity of its content is demonstrated in an interview study, and that of the questions' reliability by a test-retest study [9]. In addition, a modified response choice scale is investigated in a scaling study.

Methods

Development of the task list for the questionnaire

The questionnaire is task-oriented, i.e. it builds upon 24 general tasks essential to physicians' work. These tasks have been formulated by a work group comprised of two

computer scientists and two physicians, including the author. The group based their work on observations of 40 hours of clinical activity in five departments in two university teaching hospitals, performed January-February 2000 by two of the members of the group. Parts of the observations (7 hours observation time, five physicians from two departments, 27 patients) were transcribed verbatim and categorized by hierarchical task analysis [10]. However, the resulting hierarchy of low-level tasks was too large (104 tasks) for use in questionnaires. Thus, the tasks were transformed and merged into higher-level tasks. In the process, they were aimed at being easy to understand, relevant for clinical work in all specialties and attributable to the functionality found in present EMR systems. Tasks regarded as rarely performed, representing negligible time consumption or not likely to be supported by an EMR system in the near future were deleted. Further, the principal information needs of physicians defined by Gorman [11] were taken into account by adding three new tasks (table 1, tasks 6, 7 and 8). We used the refined list of 23 clinical tasks in a national survey, the first demonstration study in this paper [8]. Preceding the second demonstration study, a local survey [12], the questionnaire was reviewed in Aust-Agder hospital by six internists in two focus group sessions, and one new task (table 1, task 24) was added to the list. In November 2002, we used video recordings (4.5 h) of two physicians in a rheumatology outpatient clinic attending to nine patients to review the 24 defined tasks, but the tasks were unchanged. Definitions and examples of all tasks are found in additional file 1. Although native English speaking professionals were consulted during translations, all translated material should be regarded as guiding rather than final.

Development of the questions and the response labels in the questionnaire

The questionnaire principally consists of two sections; one covering self-reported frequency of use of a given EMR system, the other covering perceived ease of performing them using the system. The first section appeared in the national survey, and both sections in the local survey. The questions and response labels were adapted from validated questionnaires, Doll & Torkzadeh [13] and Aydin & Rice [14], both appearing in Anderson et al [15]. Within each section, the questions are equally worded for every task. For details on the incremental changes of each revision of the questionnaire, see appendix A in additional file 17.

Validation of the questionnaire

The validation of the questionnaire was performed in four separate studies.

Table 1: List of tasks. Tasks used in the various revisions of the questionnaire.

No.	Task	Rev. 1 National study	Rev. 2 Local study	Rev. 3 Test-retest study and interviews
1	Review the patient's problems	x	x	x
2	Seek out specific information from patient records	x	x	x
3	Follow results of a test or investigation over time	x	x	x
4	Obtain results from new tests or investigations	x	x	x
5	Enter daily notes	x	x	x
6	Obtain information on investigation or treatment procedures	x		x
7	Answer questions concerning general medical knowledge (e.g. concerning treatment, symptoms, complications etc.)	x		x
8	Produce data reviews for specific patient groups	x	x	x
9	Order clinical biochemical laboratory analyses	x	x	x
10	Obtain results from clinical biochemical laboratory analyses	x	x	x
11	Order X-ray, ultrasound or CT investigations	x		x
12	Obtain results from x ray, ultrasound, or CT investigations	x	x	x
13	Order other supplementary investigations	x		x
14	Obtain results from other supplemental investigations	x	x	x
15	Refer patient to other departments or specialists	x	x	x
16	Order treatment directly (e.g. medicines, operations etc.)	x		x
17	Write prescriptions	x	x	x
18	Write sick leave notes	x	x	x
19	Collect patient data for various medical declarations	x	x	x
20	Give written specific information to patients (e.g. about medications, disease status.)	x	x	x
21	Give written general information to patients	x	x	x
22	Collect patient information for discharge reports	x	x	x
23	Check and sign typed dictations	x	x	x
24	Register codes for diagnoses or performed procedures		x	x

Structured interviews with physicians

Content validity of the questionnaire was addressed by a structured interview study of physicians from ten selected departments in a university teaching hospital. The two senior residents and eight consultants were named by the head of each department. Three physicians refused to be interviewed, and were substituted by others from the same department. Each one-hour interview was recorded digitally, initiated by the physician filling out the questionnaire whilst being observed. A fixed set of 153 open and closed questions were asked [9,16] mostly about the defined tasks in the questionnaire. During the interviews, answers to the open questions were transcribed and that of the closed questions were registered directly in a database. Unclear or incomplete transcriptions were revised and completed using the recordings of the interviews. We analyzed the open questions qualitatively by categorizing the responses into themes. The interview guide is provided in additional file 11 and 12.

Post hoc analysis of two demonstration studies

The data from two published demonstration studies were used for missing response analysis and criterion validation. The first, a national survey, comprised of responses

from 219 of 307 physicians (72%) in 17 hospitals [8]. The survey included task-oriented EMR use and two translated user satisfaction measures; the Doll & Torkzadeh's "End User Satisfaction scale" [13] and Aydin & Rice's "Short global user satisfaction measure" [14]. The second demonstration study, a local survey, comprised of responses from 70 of 80 physicians (88%) in Aust-Agder Hospital [12]. The questionnaire contained all of the questions from the national survey, except those regarding five tasks not supported in this hospital (table 1). In addition, the section covering task performance was added in this second revision of the questionnaire (table 2). The questionnaires used in these studies are provided in Norwegian original and English translated versions in additional files 2, 3 and 5, 6.

Test-retest study

We measured test-retest reliability in a postal survey of physicians from three hospitals having EMR systems from separate vendors. Within each hospital, equal groups of physicians were randomly selected from surgical, medical and other wards. The first questionnaire was sent to the 96 included physicians, and a reminder was sent to 57 non-responders two weeks later. Three weeks after this, the sec-

Table 2: Questionnaire revisions. Overall structure of the revisions of the questionnaires. Sections not covered in this paper are hidden. For the questionnaires, see additional files 3, 6 and 9.

Questionnaire revision	No. of questions	Section in questionnaire
Rev. 1 National study		
Frequency of PC use for each task, use of EMR or other program	23 + 23	D
End User Computing Satisfaction[13]	12	F
Short Global User Satisfaction[14]	5	G
Rev. 2 Local Study		
Frequency of EMR use for each task	19	DI, D2
Task performance using the EMR, compared to previous routines	19	F
End User Computing Satisfaction[13]	12	E1, E2
Short Global User Satisfaction[14]	5	G
Rev. 3, Test-Retest study and Interviews		
Frequency of EMR use for each task	24	B1, B2
Task performance using the EMR, compared to previous routines	24	C
End User Computing Satisfaction[13]	12	D
Short Global User Satisfaction[14]	5	E

ond questionnaire was sent to the 52 responders along with a music compact disc as inducement. The response rate of the first and second questionnaire was 55.2% (52/96) and 71% (37/52), respectively. On average, we received the second questionnaire 4.4 weeks after the first. To estimate test-retest reliability in the task-oriented questions, we used Cohen's weighted kappa. The kappa values were interpreted according to Lewis' guidelines [17]. The questionnaire used in this study is provided in Norwegian original and English translated version in additional files 8 and 9.

Scaling of response labels

To validate and scale the response labels in the "Frequency of EMR use" scale, we selected 31 respondents by convenience sampling and asked them to interpret a set of response labels by placing marks on a visual analogue scale (VAS). The VAS ranged from "never" to "always", and the eight Norwegian labels (five original response labels and three alternatives) appeared on separate sheets in random order. Using a standard ruler, we measured the marks on the VAS in millimeters from the "never" end, and calculated the mean VAS value and confidence interval for each response label, as well as the number of disordinal label pairs [18]. The combination of labels providing the lowest number of disordinal pairs was selected for the final frequency scale. The VAS form used in this study is provided in additional file 15.

Computer programs used

Teleform™ 8 was used for data acquisition of postal surveys, Microsoft Access 2002™ for data management and data acquisition during interviews, OntoLog [19] 1.4 for

indexing and analysis of video and audio material, StatExact™ 5.0 for calculating the kappa statistic and SPSS™ 11.0 (Windows) for all other statistical analysis.

Results

The studies provided evaluation of the questionnaire in terms of 1) content validity, 2) compliance, 3) criterion validity, 4) test-retest reliability and 5) scaling of response labels.

Content validity

Relevance of tasks

The interviews included structured questions about task relevancy, frequency and time consumption. The majority of the physicians (7-10 of 10) found each of the 24 tasks part of their work, except task 8 (figure 1, section A). In the open-ended questions, they perceived this task partly as an administrative task best performed by other personnel, and partly as not fully applicable to medical work (table 3, themes 1 and 5). However, four of five physicians who did not consider this task a part of their job agreed that it could be a part of it in the future, provided new technology was implemented. The comments transcribed during the interviews suggested that tasks otherwise considered appropriate for other staff could be done by physicians (e.g. gather and present data to the physicians, mediate orders to other instances), if computer support would make the tasks less time consuming (theme 1).

To broadly assess the amount of work represented by each task, the physicians were asked to estimate frequency and time consumption of each task. Regarding frequency, most physicians (7-10 of 10) found that all but four tasks

Questions about relevance of tasks

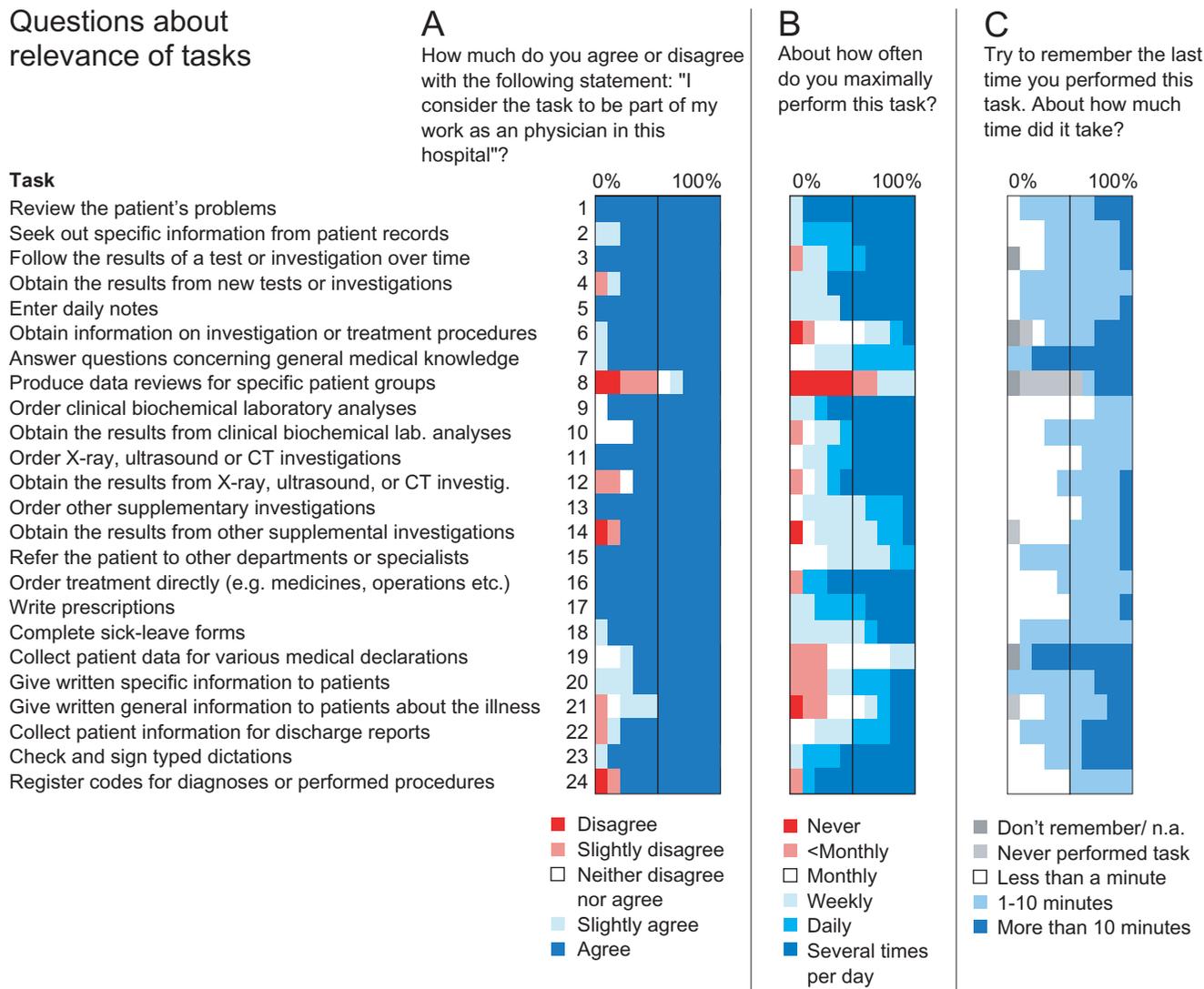


Figure 1
Relevance of tasks Responses in the interview study about A) task relevance, B) how frequently they maximally are performed, and C) how much time the physicians estimate that they take.

were performed frequently, i.e. maximally weekly or daily (median value). Tasks 8, 6 and 19 were all infrequently performed, i.e. maximally less than monthly, but they were relatively time consuming. Regarding the time consumption of each task, most of the tasks (17 of 24) were estimated to 1–10 minutes, and two tasks to more than 10 minutes (tasks 7 and 19). Some tasks (5 of 24 tasks) were estimated to take less than a minute using current paper-based routines (e.g. order lab tests, write prescriptions, register codes), but these tasks were performed frequently (figure 1, part B).

Accuracy of task interpretation, and estimation of EMR use
 The interviews included structured questions about how the physicians interpreted each task, and whether they found answering the accompanying question about EMR use (figure 2) difficult or not. The majority of the physicians found all tasks comprehensible (figure 2, part A). As a control, we asked eight of the physicians to formulate their interpretation of each task in their own words. All respondents who chose the identical wording to that of the defined task were requested to name an example. The answers, either formulations or examples, were compared

Table 3: Themes from the interviews. The themes, typically appearing in open-ended questions, are sorted in descending order by the number of physicians providing answers attributable to the given theme. In the "Tasks" column, the tasks to which each answer is attributed are sorted in descending order by number of physicians commenting the task. In the "Typical quote" column, the quotes are followed by the physician's specialty in parentheses.

Theme	No. of physicians (no. of quotes)	The tasks mentioned in relation to this theme, by number of physicians:				Typical quote
		4	3	2	1	
1 Work role issues	8 (34)	10	19	6, 9, 8, 24	2, 4, 5, 7, 11, 12, 13, 14, 18, 20, 21, 22	<i>The third method would be the "ask-the-nurse" method. This is convenient, though, then I may do other things. [In the future] It could be that it will be so easy to do it, that I could do it myself...if it's really easy, a completely negligible task. But if it takes some time..if I have to wait or something..then I feel that it should be a medical secretary's task, at least in a hospital. (respiratory diseases)</i>
2 Wording problems	7 (21)		16	4, 21	1, 7, 12, 13, 22	<i>I don't understand what you mean with "directly"...write orders on the [order entry form], request or order an operation...one other [example] is requesting treatment by physiotherapist (orthopedy)</i>
3 Questions regarding use of non-existent functionality	7 (11)			3	6, 9, 14, 15, 18	<i>Some questions are difficult to answer, as we can't log on [to the EMR system] and find results from X-ray investigations (plastic surgery)</i>
4 Distinguishing EMR from other software or media	6 (8)			4	2, 3, 6, 7	<i>Is [the separate lab system] regarded as a part of [the EMR system]? (neurology)</i>
5 Task not fully applicable to clinical work	6 (10)			8, 20, 21	3	<i>I've hardly ever been there. I spend a lot of my time providing information [to the patient] verbally. Written information is rarely demanded [by the patient]. I'm sceptical towards providing it in writing...because it must be individualized, and that's much harder in writing than verbally...and if I do, it will usually be copies of notes from the medical record. (oncology)</i>
6 Functionality missed by the respondent	5 (7)					<i>..well, this is about everyday work, after all. You don't ask about what [in the EMR system] might be improved...This is all only about what's already there. (orthopedy)</i>
7 Distinguishing other employee's use of the system from one's own	5 (6)			5	15	<i>Here I was wondering whether you mean the notes I write myself, or the dictation and [the text] typed by others. I'd recon that it would include dictation. (neurology)</i>
8 EMR only partly supports the defined task	5 (12)			1, 3, 19	2, 4, 22	<i>Well, you use [the EMR system], too, but you may never write any of those things without having the rest of the medical record available. You sort of get "black-and-white" alternatives, without being permitted to comment anything. It's not a simple yes or no type of question. After all, you can't found your work on [the EMR system] only. (oncology)</i>
9 Knowing the EMR functionality	3 (8)					<i>..is [writing a prescription] available here? Sick leave forms, too? (neurology)</i>

Questions about accuracy of task interpretation and estimation of EMR use

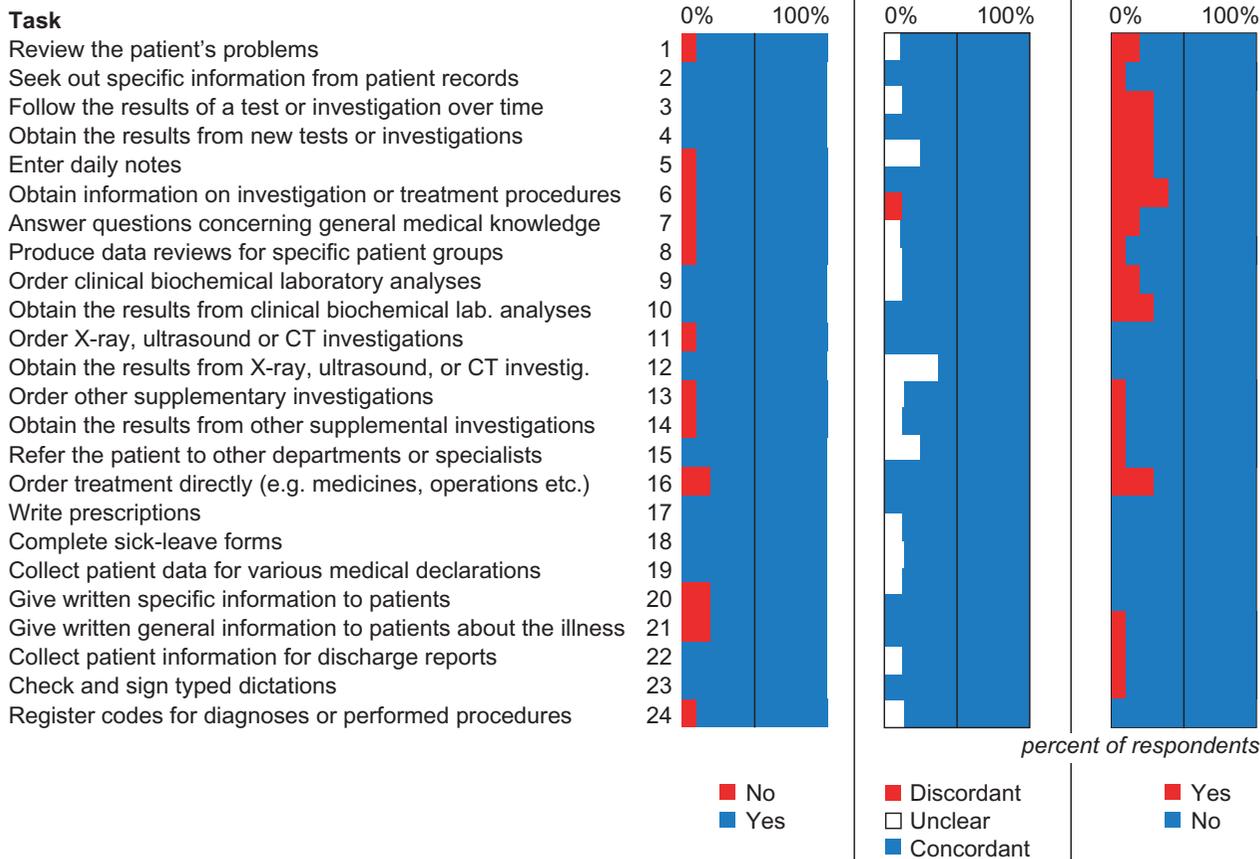


Figure 2
Accuracy of task interpretation, and estimation of EMR use Responses in the interview study about A) whether a task is comprehensible or not, B) whether the physicians' interpretation of each task fitted the actual definition or not, and C) whether estimation of own EMR use for given task was found difficult or not.

to the original task definitions. Answers that complied to whole or essential parts of the task definitions were categorized as concordant, and those that did not comply as discordant. Unclear, incomplete or ambiguous answers were categorized as unclear. All of the tasks had a majority of concordant answers, despite some unclear answers (figure 2, part B). Only tasks 7 had a small proportion of discordant interpretations (1 of 8 respondents).

Nine of the 24 task-oriented questions about EMR use were found difficult to answer by 2-4 of 10 physicians

(figure 2, part C). Five of these addressed functionality not specifically supported by the EMR. An escape choice ("Task not supported by EMR") had been provided, but the physicians never the less found answering these questions confusing. Further explanations were found in the open-ended questions (table 3).

Themes appearing in open-ended questions

The answers to the open-ended questions and the spontaneous comments were categorized into themes. Those mentioned by at least two physicians are shown in table

Table 4: Missing responses in the demonstration studies. The median proportions of missing responses to task-oriented questions in the national and local demonstration study are shown in this table.

Demonstration study	Task-oriented questions	Median missing responses (range)
National study	Frequency of PC use	1.8% (1.4% – 3.2%)
	Use EMR / use other program	21.0% (5.9% – 51.1%)
Local study	Frequency of EMR use	0.0% (0.0% – 1.4%)
	Task performance	2.9% (1.4% – 14.3%)

3. The quantitative and qualitative data from the interview study are provided in additional files 13 and 14, respectively.

Compliance

Overall, the task-oriented questions had a low percentage of missing responses both in the national and in the local demonstration study. However, the questionnaire design in the former was slightly problematic. In the national study, each question about frequency of PC use for a given task was followed by a question about type of computer program used (i.e. "EMR" and/or "other program"). The percentage of missing responses was low in the former, but quite high in the latter (table 4). As a consequence, a number of respondents reported that they were using a computer without telling whether they were using the EMR or not. This subgroup needed to be presented along with explicitly reported EMR use, making interpretation and presentation of the results challenging. The subgroup was particularly large in tasks 10 [Obtain results from clinical biochemical laboratory analyses] and 4 [Obtain results from new tests or investigations] (27.4% and 24.7%, respectively).

In the local demonstration study, we simplified the task-oriented questions about PC use by limiting them to EMR only. In addition, we omitted questions about tasks not explicitly supported by the EMR under study. In this study, the percentages of missing responses were low, both in the questions about EMR use and in those about task performance. In the latter, the question for task 8 [Produce data reviews for specific patient groups] had the highest proportion of missing responses (14.3%). However, the reported EMR use for this task was very low in this study (91% of the physicians answered "seldom" or "never/almost never").

Criterion validity

Criterion validation was assessed in three ways, by correlating task-oriented EMR use to general EMR use, task performance to overall work performance, and task performance to user satisfaction. As the first criterion, we assessed general EMR use by asking the physicians about

how often they used the EMR as an information source in their daily clinical work (table 5, row 1). This question correlated to nine of the 12 tasks about information retrieval, and to 12 of all 24 tasks. This suggests that a considerable proportion of the tasks are regarded essential to EMR's function of information retrieval. Of the remaining three tasks of this kind (tasks 6–8), explicit functionality was available only for task 8 [Produce data reviews for specific patient groups] in this study. As a second criterion, we assessed overall work performance by asking whether performance of the department's work, and that of the respondent's work, had become easier or more difficult using the EMR system (table 5, row 2–4). A high proportion of the questions about task performance correlated to both forms of overall work performance, which suggests that these tasks are regarded important elements of clinical work. As a third criterion for validation of the tasks, we calculated correlations between task performance and two standard measures of user satisfaction (table 5, row 5–8). Both measures correlated to high proportions of the tasks, but the Short Global user Satisfaction measure correlated to more tasks than that of End User Computing Satisfaction measure. The EMR was seldom or never used for the tasks for which no correlation between task performance and user satisfaction was found (notwithstanding tasks 19 [Collect patient data for various medical declarations] in the local study and task 15 [Refer patients to other departments or specialists] in the test-retest study). The data from the demonstration studies are provided in additional files 4 and 7.

Test-retest reliability

In the test-retest study, we measured reliability by calculating Cohen's weighted kappa (quadratic weights) for all task-oriented questions. Generally, the weighted kappa was high (figure 3), but the questions about EMR use showed better reliability than that of task performance (median kappa 0.718 and 0.617, respectively).

In the questions about EMR use, kappa values indicating excellent test-retest agreement was found in seven tasks (figure 3). On the other hand, a low or non-significant kappa was found in tasks 7, 9, 13, and in the questions

Table 5: Criterion validity. Significant correlations (Spearman's rho) between task-oriented and overall questions about frequency of EMR use, work performance and user satisfaction. In the test-retest study, data from its first part was used for this analysis (61 physicians from three hospitals). *Tasks related to information retrieval.

Criterion validation for task-oriented questions					
	<i>In the...</i>	<i>...the task-oriented questions about...</i>	<i>...correlates to...</i>	<i>...in number of questions:</i>	<i>Median correlation coefficient (range)</i>
Frequency of EMR use: individual tasks vs. general information retrieval					
1	test-retest study	frequency of EMR use (B1-1 to B1-24)	question B2-2: "All considered, how often do you use the EMR as an information source in the daily clinical work? (never-always)"	12 of 24 (50%) and 9 of 12* (75%)	0.516 (0.308 – 0.675)
Task performance vs. overall work performance					
2	local study	task performance (F1-F19)	question G1a "The performance of our department's work has become... (significantly more difficult - significantly easier)"	17 of 19 (89%)	0.513 (0.286 – 0.684)
3	test-retest study	task performance (C1-C24)	question E3a: "The performance of our department's work has become... (significantly more difficult - significantly easier)"	20 of 24 (83%)	0.427 (0.329 – 0.662)
4	test-retest study	task performance (C1-C24)	question E3b: "The performance of my own tasks has become... (significantly more difficult - significantly easier)"	21 of 24 (88%)	0.435 (0.291 – 0.689)
Task performance vs. user satisfaction					
5	local study	task performance (F1-F19)	the End user Computer Satisfaction measure	13 of 19 (68%)	0.483 (0.273 – 0.592)
6	test-retest study	task performance (C1-C24)	the End user Computer Satisfaction measure	15 of 24 (63%)	0.458 (0.328–0.682)
7	local study	task performance (F1-F19)	The Short Global User Satisfaction measure	16 of 19 (84%)	0.512 (0.332 – 0.686)
8	test-retest study	task performance (C1-C24)	The Short Global User Satisfaction measure	20 of 24 (83%)	0.445 (0.348 – 0.711)

about task performance in tasks 15, 16 and 21. No tasks performed poorly in both EMR use and task performance. (The data from the test-retest study is provided in additional file 10).

Scaling of response labels

In the scaling study, the original set of labels performed better than the alternatives. In the best alternative set of labels, the number of disordinal pairs was 5%, but the original combination of labels remained the better choice at 4%. The mean positions of the original labels (figure 4) constituted a symmetrical, s-shaped curve. The confidence intervals of the sample show some overlap between adjacent labels (figure 4), whereas the confidence intervals of the mean do not (data not shown, ANOVA $p < 0.001$, LSD $p < 0.001$ between all labels).

We regarded the response choices in the task performance questions as standard, and hence did not include them in this study. (The data from the scaling study is provided in additional file 16.)

Discussion

The results suggest that this questionnaire may provide valid and reliable information about how an implemented EMR system is utilized on an overall level in clinical practice, and how well the system supports clinical tasks.

The tasks-oriented questions are relevant for clinical work, but some are difficult to answer

During development, the tasks have been based on observations of clinical activity, and further refined to suit their purpose as a common denominator for assessments of

Test-retest reliability

(weighted kappa, quadratic weights)

Task

- Review the patient's problems
- Seek out specific information from patient records
- Follow the results of a test or investigation over time
- Obtain the results from new tests or investigations
- Enter daily notes
- Obtain information on investigation or treatment procedures
- Answer questions concerning general medical knowledge
- Produce data reviews for specific patient groups
- Order clinical biochemical laboratory analyses
- Obtain the results from clinical biochemical lab. analyses
- Order X-ray, ultrasound or CT investigations
- Obtain the results from X-ray, ultrasound, or CT investig.
- Order other supplementary investigations
- Obtain the results from other supplemental investigations
- Refer the patient to other departments or specialists
- Order treatment directly (e.g. medicines, operations etc.)
- Write prescriptions
- Complete sick-leave forms
- Collect patient data for various medical declarations
- Give written specific information to patients
- Give written general information to patients about the illness
- Collect patient information for discharge reports
- Check and sign typed dictations
- Register codes for diagnoses or performed procedures

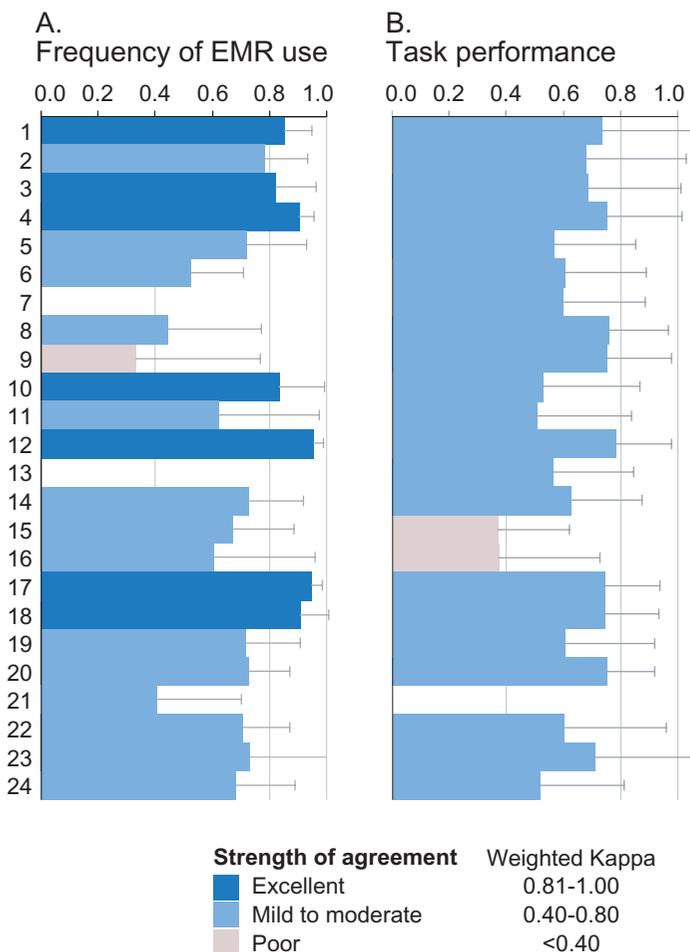


Figure 3

Test-retest reliability Reliability (weighted kappa, quadratic weights) is shown for task-oriented questions about A) frequency of EMR use and B) task performance. Error bars show confidence intervals of kappa values. Non-significant tests ($p > 0.05$) are hidden.

various EMR systems. In the interviews, the tasks were recognized and correctly interpreted (figure 2) by a wide range of physicians. However, some of the task-oriented questions about EMR use were found difficult to answer, particularly for the higher-level tasks. Four themes appearing in the interviews provided reasons for these problems. First, the respondents were confused when asked about use of EMR for tasks for which no explicit functionality was offered (table 3; theme 3), despite the presence of relevant 'escape' response choices. This confusion may partly explain the contradictory responses in the national survey, where a minor proportion of respondents reported use of the EMR system for tasks it did not explicitly support (tasks 6 and 7)[8], and the low reliability of three questions about EMR use in the test-retest study (tasks 7, 9 and 13). It may also explain the few missing responses in the

local study, where unsupported tasks were omitted. As a second problem in describing EMR use, distinguishing EMR from other software or media appeared as a problem in the interviews (theme 4). This problem may explain the many missing responses in parts of the national study (table 4). The reduction of missing responses in the local study suggests that just considering EMR use (and not use of other software) is easier for the respondent. However, the problem will remain for respondents who are using other software than the EMR during clinical work, making reviews of all software available to the physicians necessary. As a third problem, questions about tasks which were not completely supported by the EMR system were found hard to answer, despite the fact that the wording of the questions only implied a supportive role. This problem was in particular attributed to general tasks.

Scaling of response labels

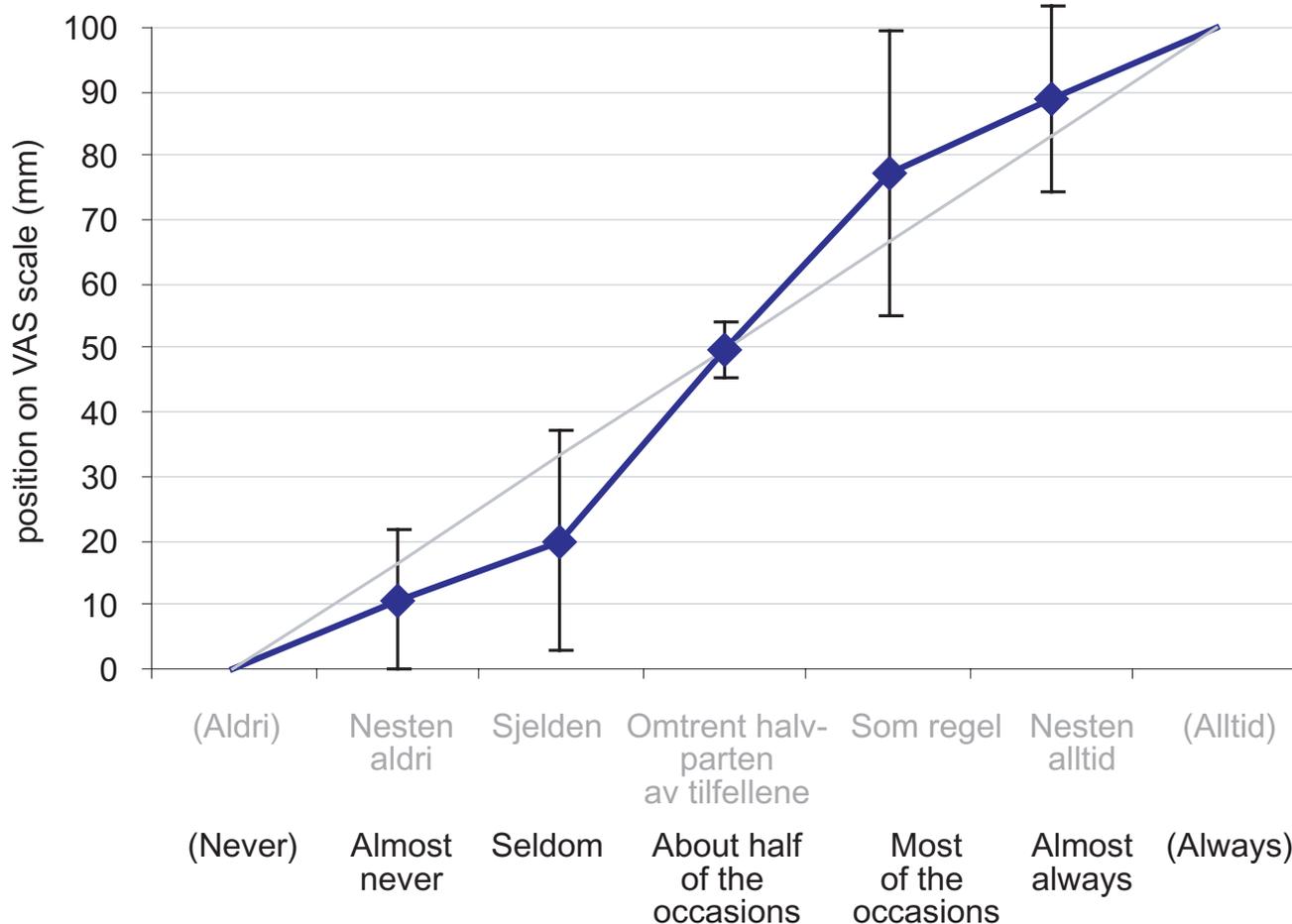


Figure 4
Scaling of response labels The labels comprise the scale used in the questions about frequency of EMR use. The data points represent measured position on the visual analog scale (mm), and the error bars represent confidence intervals of the sample. The original Norwegian terms are shown in grey color, the English translations in black.

However, the test-retest reliability was relatively high in these questions, suggesting a limited negative effect. Fourth and final, distinguishing other employee's use of the system from one's own appeared as a problem in the interviews (theme 7) in tasks 5 and 15. Regarding task 5 [Enter daily notes], the explanation was confusion about whose use of the EMR should be stated, the physician's or the transcriptionist's. This problem is probably amendable by revising the instructions to the respondent in the questionnaire.

In addition to providing explanations to the findings of the closed questions, the results from the open-ended questions addressed a number of themes on their own. First, wording problems (table 3, theme 2) were expressed particularly for tasks 16, 4 and 21. However, the respondents' interpretations of these tasks (figure 1) were all concordant with and covering essential parts of the task definition. Another important theme involved functionality missed by the respondent (table 3, theme 6), i.e. that the questionnaire did not allow them to express what functionality they were missing in the EMR system. This in particular made it difficult to answer the questions about

user satisfaction, as the respondent had problems deciding whether to provide answers based on the functionality actually available in the EMR system, or on the functionality that should have been in the system. The problem is closely related to the problems regarding EMR only supporting parts of a given defined task (table 3, theme 8).

The tasks are relevant for EMR systems

Moderately high correlations were consistently found between a majority of task-oriented questions and overall questions on EMR use, task performance and user satisfaction. The correlations to self-reported overall EMR use suggest that the tasks are regarded essential to EMR systems as such, and the correlations to work performance suggest that the tasks are regarded important to clinical work. The correlations to user satisfaction agree with the results of both Sittig et al [20] and Lee et al [21], who found significant correlations between user satisfaction and questions about how easily the work was done. In combination, this means that high reported EMR use for individual tasks equals high reported use of the EMR on the whole, and that improved performance of individual tasks equals improved overall work performance and high satisfaction with the system as a whole. Although not proving the validity of each task, it is highly suggestive. Furthermore, the correlations were limited to tasks for which clear functionality existed in the EMR systems. For the uncorrelated tasks, further clarification must await completion of the functionality of current EMR systems.

This way of correlating a set of lower-level task-oriented questions to higher-level questions is commonly used as criterion validation [22]. However, higher-level questions regarding EMR use are difficult to answer, as physicians' work consists of a complex mix of tasks that are suited for computer support and tasks that are not. A more direct form of criterion validation could have been achieved by studying system audit trails [2]. Such trails are readily available, but they must be validated themselves, and they cannot be more detailed than the structure of the EMR system itself. In Norway, the EMR systems are document-based in structure [12]. This limits the interpretation of such trails, particularly when considering information-seeking behavior.

The questionnaire produces interpretable results

The demonstration studies provided readily interpretable results. In the national study, the physicians generally reported a much lower frequency of EMR use than what was expected by the functionality implemented in each hospital [8]. In the local study, the physicians reported a very high frequency of EMR use, mainly for tasks related to retrieval of patient data [12]. In this study, the physicians generally had little choice of information sources, as the paper-based medical records were obliterated in this

hospital. The use of the EMR system for other tasks was however much lower. The results from both the national and the local study indicate that the physicians are able to report overall patterns in their use of EMR that is not in line with the implicit expectations signalled by this questionnaire. These results should not be too surprising. The physicians' traditional autonomous position may allow them to withstand instructions from the hospital administration, e.g. regarding ordering of clinical biochemical investigations [23]. Also, in most hospitals having EMR systems, the physicians may freely choose source of patient data. This is due to the fact that both the paper-based and electronic medical record generally are updated concurrently [12], and they are only two of many information sources available in clinical practice (e.g. asking the patient, calling the primary care physician, etc.).

Compared to the 400–600 tasks commonly found in full task inventories [6], the number of tasks in the questionnaire is moderate (24). The high response rates suggest that the number of questions is manageable to the respondents. Compared to that of similar questionnaires [4,21], the task list provides the evaluator with more details about areas for improvement, and it is not designed with one particular EMR system in mind [21]. In addition, more emphasis is placed on clinical use of the EMR system, since the tasks are limited to information-related instead of both practical and information-related tasks [24], and to clinical instead of both clinical and academic work [4]. On the other hand, questionnaires describing self-reported usage patterns have previously been criticized for lack of precision and accountability [25,26]. However, the critics often seem to actually consider poorly validated questionnaires or too optimistic interpretations of them [27], rather than the very principle of self-reporting. When interpreting the results from a survey describing self-reported work patterns, the inherent limitations of self-reporting must be taken into account. Respondents remember recent and extraordinary events much more easily than distant or everyday events, suggesting in our case an over-estimation by those who use the EMR infrequently. Also, in even a systematically validated questionnaire, a considerable degree of bias should be expected towards answers that the respondents believe are expected from them. However, when the responses both fit with the structural premises (i.e. the marked EMR use in the local study, where the paper-based medical record was missing), and defy the implicit expectations (i.e. the lack of EMR use in the national study), the degree of bias seem to be manageable.

Reliability and scaling

The test-retest reliability study generally showed high kappa values both in the section about EMR use and in that of task performance, in spite of some tasks perform-

ing poorly in either section. The poorly performing tasks in the EMR use section addressed functionality that was available to few respondents, while those performing excellently addressed functionality supported by all EMR systems. This means that changes demonstrated for well supported tasks are more likely to reflect real changes in the underlying processes than they are likely to happen by chance. On the one hand, small differences should be interpreted with caution when using the questionnaire, e.g. when significant differences are found in rank values but not in median response values. On the other hand, the evaluator should be careful not to disregard non-significant differences in small samples in the tasks having reliability less than 0.6, as the most likely effect of reliability issues are attenuation of real differences [28].

In the study of the frequency scale (appearing in the questionnaire section about EMR use), the order of the response labels coincide with that of the respondent's visual analogue scale (VAS) markings. In addition, the confidence intervals of the means are clearly separated in this relatively small sample. This suggests that response labels are considered separate steps on an ordinal scale by the respondent. However, the mean VAS values do not increment linearly, but follows a symmetric s-shaped curve, in which the largest increments appear at the middle part of the scale. This suggests that differences in frequency of EMR use might be considered slightly larger when involving or spanning the central label than when involving the labels at each end of the scale. In sum, the scale is ordinal but not linear, making non-parametric methods the best choice for statistical analysis.

Comparing development and evaluation of this questionnaire to that of other questionnaire

When developing questionnaires, existing literature[22,29] and expert groups[30,31] are commonly used to produce the initial items. For our questionnaire, the literature search was mostly unfruitful, and we had to rely on expert groups and observational work. A common way of structuring the initial collection of items is by identifying latent (and possibly unrelated) variables by performing exploratory factor analysis[22]. For our questionnaire, no factor analysis has been performed. In the national demonstration study, it was due to the considerable differences in implemented functionality between the various EMR systems. In the local demonstration study, it was due to the low sample size relative to the number of questions, i.e. below 10:1 [32]. Although consistent patterns of use (e.g. "the notes reader", "the super-user", "the lab test aficionado", etc.) might be identified by factor analysis, it is unlikely that completely unrelated variables would be extracted from a set of work tasks all designed for the same profession. Work tasks found irrelevant by the physicians could have been identified by analyses of internal consist-

ency among the task-oriented questions, e.g. Cronbach's alpha[22]. However, such investigations should ask about the work tasks per se, not about tasks for which the EMR system is used, rendering our demonstration studies of little value in this respect. Instead of performing another survey, we chose to explore the tasks as well as the task-oriented questions in a structured interview study. This way, we had an opportunity of explaining why some of the tasks were performing better than the others in the demonstration studies.

When evaluating questionnaires, criterion and content validation is frequently used[29,33]. As the list of tasks in our questionnaire is rather heterogeneous and covers a considerable field of clinical activity, a single global criterion is hard to find. Instead, we used either criteria explaining parts of the task list (e.g. the tasks regarding information retrieval) or indirect criteria based on well-documented relations (e.g. overall user satisfaction vs. task performance).

Limitations of this study

The questionnaire described in this study applies to physicians only, missing the contribution of other types of health personnel. Further, the list of tasks does not cover communication or planning, suggesting that the list could be augmented in future versions of the questionnaire. Finally, three different revisions of the questionnaire appear in this paper, which might appear confusing. The revisions are however incremental, and should be considered consequences of lessons learned during the demonstration studies.

Application of the questionnaire

The questionnaire described here may be used as an important part of an EMR system evaluation. Instead of a simple summed score, the questionnaire's task list provides a framework by which EMR systems may be described and compared in an informative way. Since the questionnaire does not provide reasons or hypotheses for the results it produces, surveys involving it should always be accompanied by a qualitative study. The combination of methods will, however, provide more than the sum of its parts. Qualitative studies like in-depth interviews may be probing deeper when the results of the preceding survey are presented to the informant, and observational studies may focus on phenomena explaining the survey results. Conversely, the interpretation of a qualitative study may be aided by the results of a following quantitative study, as it provides a way of weighting the proposed hypotheses.

Conclusions

The task-oriented questionnaire is relevant for clinical work and EMR systems. It provides interpretable and reli-

able results on its chosen level of detail, as a part of any evaluation effort involving the hospital physician's perspective. However, development of a questionnaire should be considered a continuous process, in which each revision is guided by further validation studies.

List of abbreviations

EMR Electronic Medical Records

VAS Visual Analogue Scale

Competing interests

None declared.

Author's contributions

AF participated in formulating the tasks, designing the questionnaire, performing the demonstration studies and writing this article. HL participated in formulation the tasks, designing the questionnaire, designing and performing the interviews, performing the test-retest and scaling studies and writing this article.

Additional material

Additional File 1

Task list. List of the 24 tasks as they appear in the third revision of the questionnaire, including individual definitions and examples.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S1.xls>]

Additional File 2

Questionnaire revision 1, Norwegian original version. First revision of the questionnaire, used in the national study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S2.pdf>]

Additional File 3

Questionnaire revision 1, English translated version. First revision of the questionnaire, used in the national demonstration study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S3.pdf>]

Additional File 4

Data from national demonstration study. Data from the national demonstration study, performed in 2001. The results are published in Lærum H, Ellingsen G, Faxvaag A: *Doctors' use of electronic medical records systems in hospitals: cross sectional survey*. *BMJ* 2001, **323**: 1344–1348.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S4.xls>]

Additional File 5

Questionnaire revision 2, Norwegian original version. Second revision of the questionnaire, used in the local demonstration study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S5.pdf>]

Additional File 6

Questionnaire revision 2, English translated version. Second revision of the questionnaire, used in the local demonstration study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S6.pdf>]

Additional File 7

Data from local demonstration study. Data from the local demonstration study, performed in 2002. Results are published in: H Lærum, TH Karlsen, A Faxvaag: *Impacts of scanning and eliminating paper-based medical records on hospital physicians' clinical work practice*. *J Am Med Inform Assoc* 2003, **10**: 588–595

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S7.xls>]

Additional File 8

Questionnaire revision 2, Norwegian original version. Third revision of the questionnaire, used in the test-retest and the interview study in 2003.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S8.pdf>]

Additional File 9

Questionnaire revision 3, English translated version. Third revision of the questionnaire, used in the test-retest and the interview study in 2003.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S9.pdf>]

Additional File 10

Data from the test-retest study. Data from the test-retest study, used for the weighted kappa statistic and criterion validity.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S10.xls>]

Additional File 11

Interview guide, Norwegian Original version. Original interview guide used for the content validation of the questionnaire.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S11.doc>]

Additional File 12

Interview guide, English Translated version. English, truncated version of the Norwegian interview guide used for content validation of the questionnaire.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S12.doc>]

Additional File 13

Quantitative data from interview study. Results from closed questions in the interview study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S13.xls>]

Additional File 14

Qualitative data from interview study. Norwegian quotes from the interview study categorized into English themes (Norwegian only).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S14.xls>]

Additional File 15

Form for scaling study, containing the Visual Analogue Scales. Form used in the scaling study of the Norwegian, modified "frequency of EMR use" Norwegian original version

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S15.pdf>]

Additional File 16

Data form scaling study. Data from the scaling study of the Norwegian modified "frequency of EMR use" scale.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S16.xls>]

Additional File 17

Details of development of the questionnaire. The incremental changes in the three revisions of the questionnaire are described here, along with the intentions of the changes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-4-1-S17.doc>]

Acknowledgements

We thank Peter Fayers for statistical advice, and linguistic and professional support in writing this article.

References

1. Heathfield HA, Pitty D, Hanka R: **Evaluating information technology in health care: barriers and challenges.** *BMJ* 1998, **316**:1959-1961.
2. Heathfield HA, Felton D, Clamp S: **Evaluation of Electronic Patient and Health Record Projects. Volume 2001-IA-533.** Edited by: Heathfield HA. England, ERDIP Programme, NHS Information Authority; 2002:1-80.
3. Gennip E.M.S.J., van, Talmon JL: *Assessment and evaluation of information technologies in medicine / edited by E. M. S. J. van Gennip and J. L. Talmon* Edited by: Gennip EMSJvan and Talmon JL. Amsterdam., IOS Press.; 1995.
4. Cork RD, Detmer WM, Friedman CP: **Development and initial validation of an instrument to measure physicians' use of, knowledge about, and attitudes toward computers.** *J Am Med Inform Assoc* 1998, **5**:164-176.
5. Wirth P, Kahn L, Perkoff GT: **Comparability of two methods of time and motion study used in a clinical setting: work sampling and continuous observation.** *Med Care* 1977, **15**:953-960.
6. Jacoby I, Kindig D: **Task analysis in National Health Service Corps field stations: a methodological evaluation.** *Med Care* 1975, **13**:308-317.
7. Burkle T, Ammenwerth E, Prokosch HU, Dudeck J: **Evaluation of clinical information systems. What can be evaluated and what cannot?** *J Eval Clin Pract* 2001, **7**:373-385.
8. Lærum H, Ellingsen G, Faxvaag A: **Doctors' use of electronic medical records systems in hospitals: cross sectional survey.** *BMJ* 2001, **323**:1344-1348.
9. Fink Arlene: *The Survey kit / series editor Arlene Fink* Edited by: Fink Arlene. Thousand Oaks, Calif., Sage Publications.; 1995.
10. Kirwan B, Ainsworth LK: *A Guide to task analysis* Edited by: Kirwan Barry and Ainsworth LK. London., Taylor & Francis.; 1992.
11. Gorman PN: **Information Needs of Physicians.** *Journal of the American Society for Information Science* 1995, **46**:729-736.
12. Lærum H, Karlsen TH, Faxvaag A: **Impacts of scanning and eliminating paper-based medical records on hospital physicians' clinical work practice.** *J Am Med Inform Assoc* 2003, **MI337**:XX-YY.
13. Doll WJ, Torkzadeh G: **The measurement of end-user computing satisfaction - theoretical and Methodological issues.** *MIS Quarterly* 1991, **15**:5-10.
14. Aydin CE, Rice RE: **Social worlds, individual-differences, and implementation - predicting attitudes toward a medical information-system.** *Information & Management* 1991, **20**:119-136.
15. Anderson JG, Aydin CE, Jay SJ: *Evaluating Health Care Information Systems* SAGE; 1994.
16. Sprangers M, Cull A: **Guidelines for moduel development.** Amsterdam, The Netherlands Cancer Institute; 1992.
17. Lewis RJ: **Reliability and Validity: Meaning and Measurement.** Ambulatory Pediatric Association 1999.
18. Keller SD, Ware J.E., Jr., Gandek B, Aaronson NK, Alonso J, Apolone G, Bjorner JB, Brazier J, Bullinger M, Fukuhara S, Kaasa S, Lepage A, Sanson-Fisher RW, Sullivan M, Wood-Dauphinee S: **Testing the equivalence of translations of widely used response choice labels: results from the IQOLA Project. International Quality of Life Assessment.** *J Clin Epidemiol* 1998, **51**:933-944.
19. Heggland J: **OntoLog.** Faculty of Information Technology, Mathematics and Electrical Engineering, NTNU NTNU; 2003.
20. Sittig DF, Kuperman GJ, Fiskio J: **Evaluating physician satisfaction regarding user interactions with an electronic medical record system.** *Proc AMIA Symp Bethesda, Maryland USA, American Medical Informatics Association;* 1999:400-404.
21. Lee F, Teich JM, Spurr CD, Bates DV: **Implementation of physician order entry: user satisfaction and self-reported usage patterns.** *J Am Med Inform Assoc* 1996, **3**:42-55.
22. Doll WJ, Torkzadeh G: **The measurement of end-user computing satisfaction.** *MIS Quarterly* 1988, **12**:259-274.
23. Massaro TA: **Introducing physician order entry at a major academic medical center: I. Impact on organizational culture and behavior.** *Acad Med* 1993, **68**:20-25.
24. Mendenhall RC, Lloyd JS, Repicky PA, Monson JR, Girard RA, Abrahamson S: **A national study of medical and surgical specialties. II. Description of the survey instrument.** *JAMA* 1978, **240**:1160-1168.
25. Kushniruk AW, Patel VL: **Cognitive computer-based video analysis: its application in assessing the usability of medical systems.** *Medinfo* 1995, **8 Pt 2**:1566-1569.
26. Brender J: **Methodological and methodical perils and pitfalls within assessment studies performed on IT-based solutions in healthcare.** *Virtual Centre for Health Informatics, Aalborg University;* 2002.
27. Nelson EC, Jacobs AR, Breer PE: **A study of the validity of the task inventory method of job analysis.** *Med Care* 1975, **13**:104-113.
28. Rossi Peter H., Wright James D., Anderson Andy B.: *Handbook of survey research Orlando, Academic Press;* 1983.
29. Joshi K, Bostrom RP, Perkins WC: **Some new factors influencing user information satisfaction: implications for systems professionals.** *Calgary, Canada, Special Interest Group on Computer Personnel Research Annual Conference;* 1986:27-42.
30. Doll WJ, Xia W, Torkzadeh G: **A Confirmatory Factor Analysis of the End-User Computing Satisfaction Instrument.** *MIS Quarterly* 1994:453-461.
31. Chin JP, Diehl VA, Norman KL: **Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface.** *NY Association for Computing Machinery;* 1988:213-218.
32. Kerlinger Fred N.: *Foundations of behavioral research : educational and psychological inquiry* London, Holt, Rinehart & Winston; 1969.
33. Friedman Charles P., Wyatt JC: *Evaluation methods in medical informatics* New York, Springer; 1997.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6947/4/1/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

