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

Hallvard Lærum*, Arild Faxvaag

INM, Faculty of Medicine, Norwegian University of Science and Technology,

Trondheim, Norway

emails: hallvard.larum@medisin.ntnu.no , arild.faxvaag@medisin.ntnu.no

*Corresponding author

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 be 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

A task-oriented questionnaire has been developed 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 as in two separate sections, about EMR use and task performance using the EMR, respectively. Using the questionnaire, the evaluator may quickly estimate the potential impact of the EMR system on health care delivery. Problematic areas may be found by identifying clinical tasks for which the EMR system either is not used, or for which performing the task is more difficult when using the system. These results may be compared across time, site or vendor. This paper describes the development of the questionnaire, and shows its performance in two demonstration studies. Its content is validated in an interview study (ten physicians in ten departments), and its reliability investigated in a test-retest study (37 physicians in three hospitals) and a scaling study (30 respondents).

Results

The interviews showed that the physicians found the general clinical tasks in the questionnaire relevant and comprehensible, and that the tasks were interpreted concordant to their definitions. Questions about tasks not explicitly or not completely supported by the EMR systems were, however, found 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. 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, a new questionnaire instrument is presented. 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 either by demonstrating self-reported lack of EMR use or a reduced reported performance of specific tasks. Although clinically oriented task inventories have been published previously, these tasks 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, and a table of minimum functionality requirements for each task is available [8]. In this paper, the development and successful application of the questionnaire in two demonstration surveys are described. 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 merged into higher-level tasks. 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 five principal information needs of physicians defined by Gorman [11] were taken into account, and three new tasks regarding information retrieval were added (table 1, tasks 6, 7 and 8). The refined list of 23 clinical tasks was used 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, video recordings (4.5 h) of two physicians in a rheumatology outpatient clinic attending to nine patients

5

were used in a review of the 24 defined tasks, but the tasks were unchanged. Definitions and examples of all tasks are found in appendix B. Although native English speaking professionals were consulted during translations, all translated material should be regarded as guiding rather than final.

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

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

Development of the questions and the response labels in the questionnaire

The questionnaire 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 B.

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	D1, 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

Table 2 *Questionnaire revisions.* Overall structure of each revision of the questionnaire. Sections not covered in this paper are hidden. For the questionnaires, see appendix B in the thesis summary.

Validation of the questionnaire

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

1. 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. The open questions were analyzed qualitatively by categorizing the responses into themes. The interview guide is provided in appendix C.

2. 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. (see appendix B in the thesis summary)

3. Test-retest study

Test-retest reliability was measured 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 second 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, the second questionnaire was received 4.4 weeks after the first. Cohen's weighted kappa was used to estimate test-retest reliability in the task-oriented questions in the questionnaire. 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. (See appendix B in the thesis summary.)

4: Scaling of response labels

To validate and scale the response labels in the "Frequency of EMR use" scale, thirty respondents were selected by convenience sampling and asked to mark eight response labels' positions 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, the marks on the VAS were measured in millimeters from the "never" end. Mean VAS value and confidence interval were calculated 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 appendix D.

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.

1. 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, this task was partly perceived as an administrative task best performed by other personnel, and partly found 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).
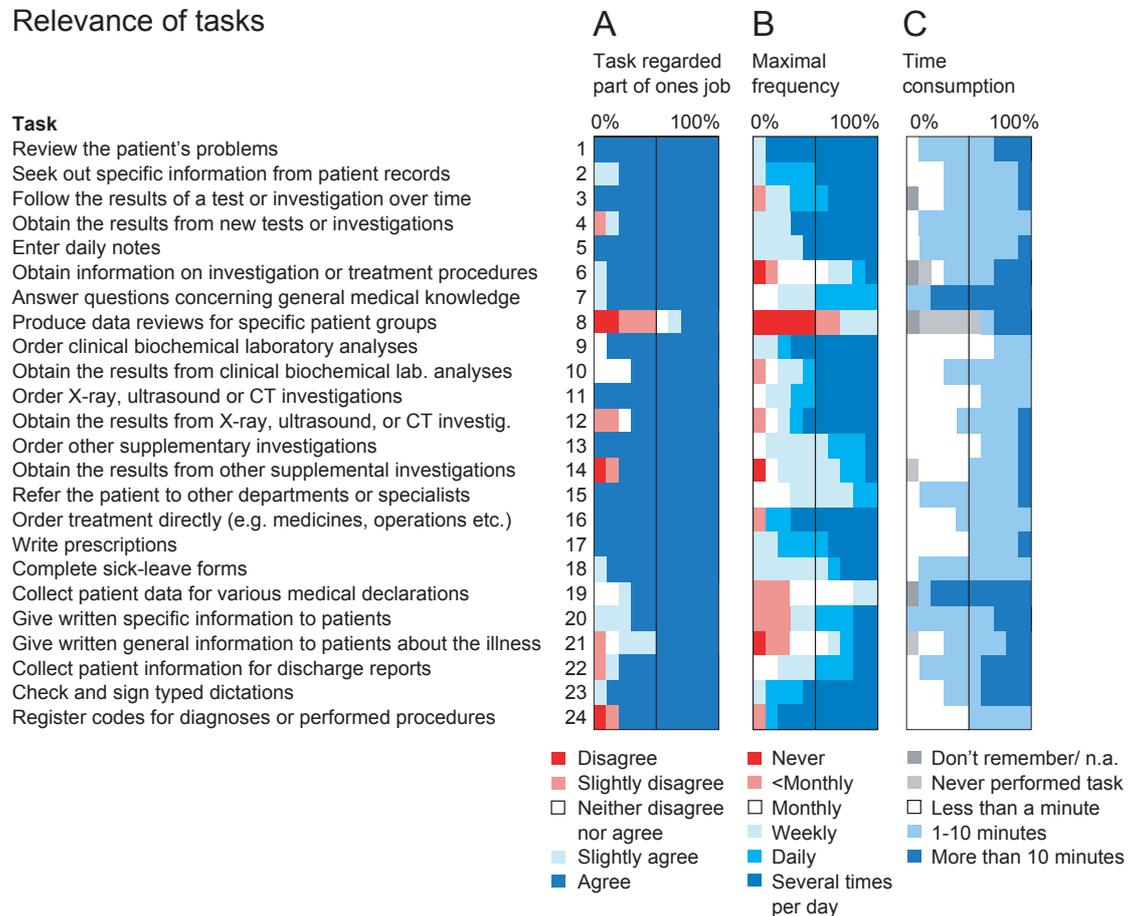
A. "I consider the task to be part of my work as an physician at this hospital"
B. "About how often do you maximally perform this task?"
C. "Try to remember the last time you performed this task. About how much time did it take?"

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.

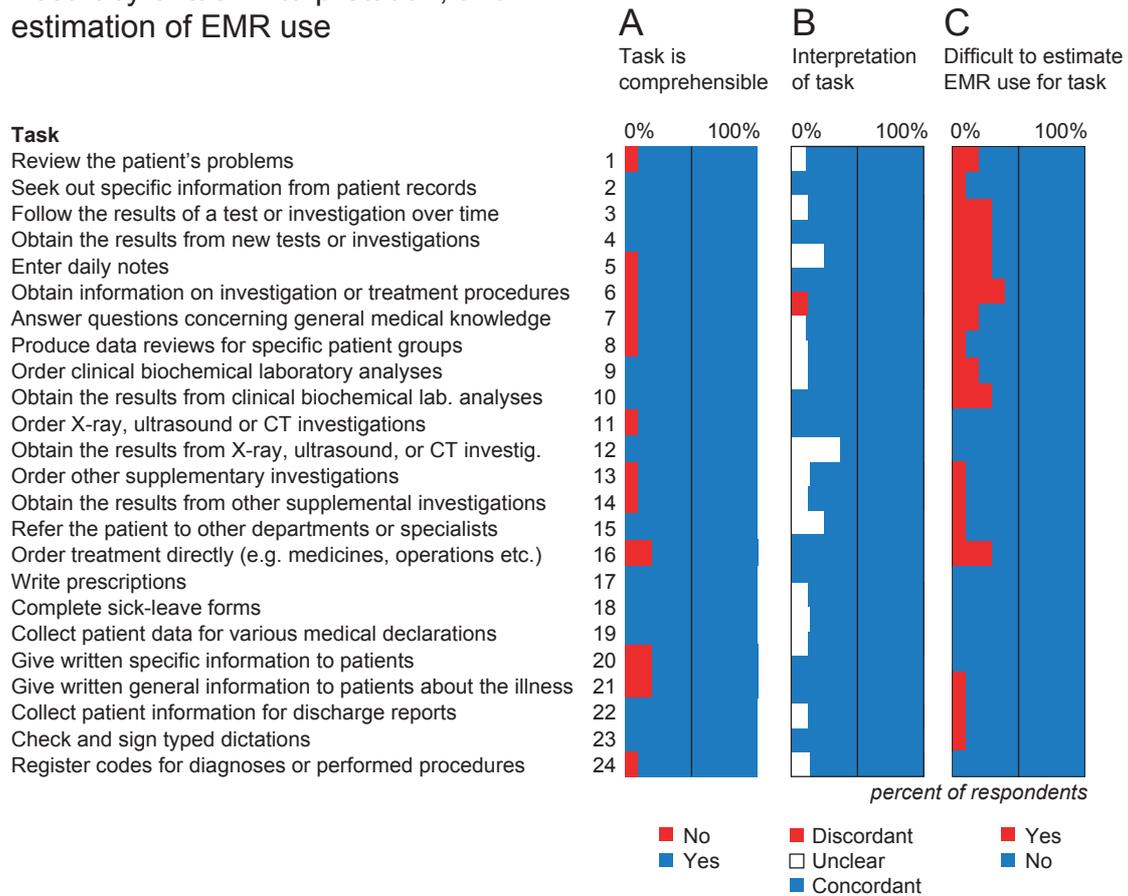To assess the amount of work represented by each task, the physicians were asked to estimate frequency and time consumption of each task. Regarding frequency, most physicians (7-10 of 10) found that all but four tasks were performed frequently, i.e. maximally weekly or daily (median value). Tasks 8, 6 and 19 were all infrequently performed, i.e. maximally less than monthly, but they were relatively time consuming.

11

Regarding the time consumption of each task, most of the defined tasks (17 of 24) took 1-10 minutes, and two tasks took more than 10 minutes (tasks 7 and 19). Some tasks (5 of 24 tasks) took 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 answering the accompanying question about EMR use (figure 2 was difficult or not. Regarding interpretation of the tasks, all of them were found comprehensible by the majority of the physicians (figure 2, part A). As a control, eight of the physicians were requested 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 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).

Accuracy of task interpretation, and estimation of EMR use

Task		A — Task is comprehensible	B — Interpretation of task	C — Difficult to estimate EMR use for task
Review the patient's problems	1			
Seek out specific information from patient records	2			
Follow the results of a test or investigation over time	3			
Obtain the results from new tests or investigations	4			
Enter daily notes	5			
Obtain information on investigation or treatment procedures	6			
Answer questions concerning general medical knowledge	7			
Produce data reviews for specific patient groups	8			
Order clinical biochemical laboratory analyses	9			
Obtain the results from clinical biochemical lab. analyses	10			
Order X-ray, ultrasound or CT investigations	11			
Obtain the results from X-ray, ultrasound, or CT investig.	12			
Order other supplementary investigations	13			
Obtain the results from other supplemental investigations	14			
Refer the patient to other departments or specialists	15			
Order treatment directly (e.g. medicines, operations etc.)	16			
Write prescriptions	17			
Complete sick-leave forms	18			
Collect patient data for various medical declarations	19			
Give written specific information to patients	20			
Give written general information to patients about the illness	21			
Collect patient information for discharge reports	22			
Check and sign typed dictations	23			
Register codes for diagnoses or performed procedures	24			

percent of respondents

Legend: No / Yes | Discordant / Unclear / Concordant | Yes / No

A. "Does this task appear comprehensible to you the way it is worded?"
B. Concordance between task definition and physician's interpretation of the task
C. "Did you find the question about how often you use the EMR for this task difficult to answer?"

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 diffcult or not.

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 explanation was found in the open-ended questions (table 3).

13

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

	Theme	No. of physicians (no. of quotes)	The tasks mentioned in relation to this theme, by number of physicians:				Typical quote (physician's specialty)
			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	*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)
2	Wording problems	7 (21)		16	4, 21	1, 7, 12, 13, 22	*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)
3	Questions regarding use of non-existent functionality	7 (11)			3	6, 9, 14, 15, 18	*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)
4	Distinguishing EMR from other software or media	6 (8)			4	2, 3, 6, 7	*Is [the separate lab system] regarded as a part of [the EMR system]?* (neurology)
5	Task not fully applicable to clinical work	6 (10)			8, 20, 21	3	*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)
6	Functionality missed by the respondent	5 (7)					*..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)
7	Distinguishing other employee's use of the system from one's own	5 (6)			5	15	*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)
8	EMR only partly supports the defined task	5 (12)			1, 3, 19	2, 4, 22	*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)
9	Knowing the EMR functionality	3 (8)					*..is [writing a prescription] available here? Sick leave forms, too?* (neurology)

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.

2. 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 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).

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 %)

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.

In the local study, the task-oriented questions about PC use were simplified, and limited to EMR only. In addition, questions regarding tasks not explicitly supported by the EMR under study were omitted. Here, 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").

3. Criterion validity

Criterion validation was performed 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, general EMR use was assessed by asking how often the EMR was used as an information source in the 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, overall work performance was assessed 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, task performance was correlated separately to 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

16

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)

Criterion validation for task-oriented questions					
	In the...	*...the task-oriented questions about...*	*...correlates to...*	*...in number of questions:*	*Median correlation coefficient (range)*
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 *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 (sum of D1a-D5b)	13 of 19 (68%)	0.483 (0.273 – 0.592)
6	test-retest study	task performance (C1-C24)	the End user Computer Satisfaction measure (sum of D1a-D5b)	15 of 24 (63%)	0.458 (0.328-0.682)
7	local study	task performance (F1-F19)	The Short Global User Satisfaction measure (sum of G1a, G1a and G2)	16 of 19 (84%)	0.512 (0.332 – 0.686)
8	test-retest study	task performance (C1-C24)	The Short Global User Satisfaction measure (sum of E1, E3a and E3c)	20 of 24 (83%)	0.445 (0.348 – 0.711)

Table 5 *Criterion validity.* Significant correlations (Spearmans' 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.

4. Test-retest reliability

In the test-retest study, reliability was measured by Cohen's weighted kappa (quadratic weights) in all task-oriented questions. Generally, the weighted kappa was

high (figure 3), but the questions about EMR use showed better reliability than that of

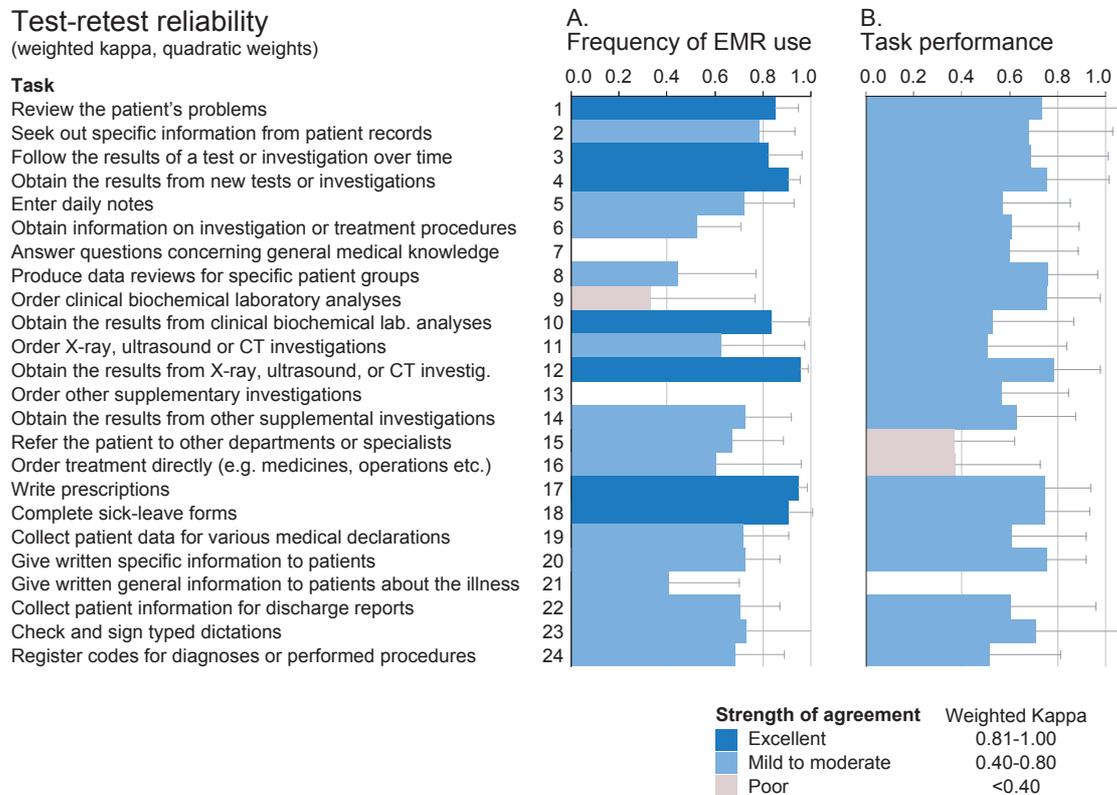task performance (median kappa 0.718 and 0.617, respectively).

Test-retest reliability
(weighted kappa, quadratic weights)

Task
#	Task
1	Review the patient's problems
2	Seek out specific information from patient records
3	Follow the results of a test or investigation over time
4	Obtain the results from new tests or investigations
5	Enter daily notes
6	Obtain information on investigation or treatment procedures
7	Answer questions concerning general medical knowledge
8	Produce data reviews for specific patient groups
9	Order clinical biochemical laboratory analyses
10	Obtain the results from clinical biochemical lab. analyses
11	Order X-ray, ultrasound or CT investigations
12	Obtain the results from X-ray, ultrasound, or CT investig.
13	Order other supplementary investigations
14	Obtain the results from other supplemental investigations
15	Refer the patient to other departments or specialists
16	Order treatment directly (e.g. medicines, operations etc.)
17	Write prescriptions
18	Complete sick-leave forms
19	Collect patient data for various medical declarations
20	Give written specific information to patients
21	Give written general information to patients about the illness
22	Collect patient information for discharge reports
23	Check and sign typed dictations
24	Register codes for diagnoses or performed procedures

A. Frequency of EMR use — B. Task performance

Strength of agreement — Weighted Kappa
Excellent — 0.81-1.00
Mild to moderate — 0.40-0.80
Poor — <0.40

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.

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 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.)

5. Scaling of response labels

In the scaling study, the original labels performed better than the alternative response labels. In the best alternative combination 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.
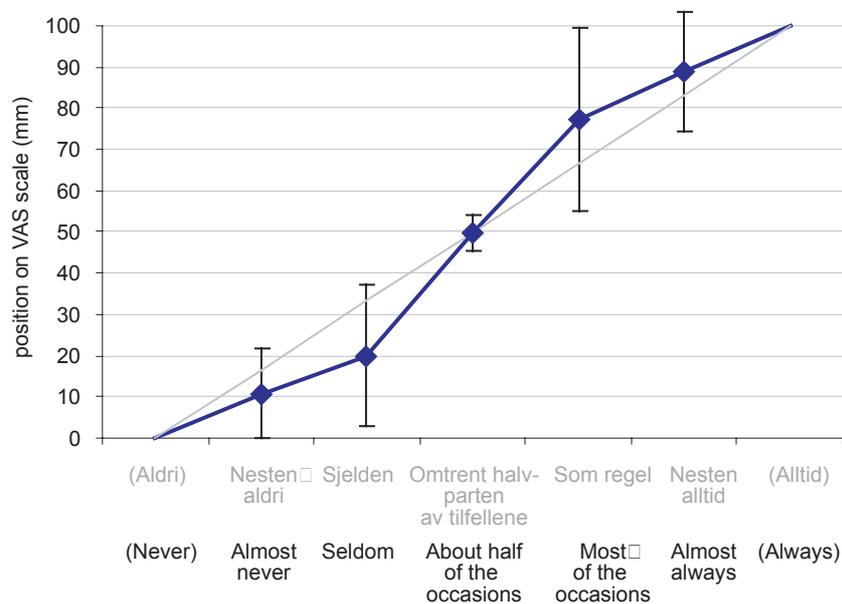
Scaling of response labels

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 interval of the mean. The original Norwegian terms are shown in grey color, the English translations in black.

The response choices in the task performance questions were regarded standard, and were hence not included 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 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

20

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. 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 in 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 was 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 that of 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

22

audit trails [2]. On the other hand, such trails 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 use the EMR mainly for tasks related to retrieval of patient data [12]. For these tasks, 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 signaled 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]. However, 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 that of task performance, in spite of some tasks performing 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. This is due to the uncertainties involved in self-reporting. On the other, 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].

The study of the scale used in the section about EMR use confirms that the order of its response labels coincide with that of the respondent's interpretations expressed on a visual analogue scale (VAS). However, the mean VAS values do not increment linearly, but follows a symmetric s-shaped curve, where the largest increments appear in the middle part of the scale. This means that the data from the scale should not be expected to have a normal distribution, and that differences in frequency of EMR use should be considered slightly larger when involving the central label than when involving labels at each end of the scale.

Limitations of this study

In the demonstration studies, no factor analysis has been performed. In the national study, it was due to the considerable differences in implemented functionality between the various EMR systems. In the local study, it was due to the low sample size relative to the number of questions, i.e. below 10:1 [29]. Also, three different

revisions of the questionnaire appear in this paper. 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 provide interpretable and useful results as part of an evaluation of an EMR system. 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 by providing weighting of the proposed hypotheses.

Conclusions

The task-oriented questionnaire is relevant for clinical work and EMR systems, and provides interpretable and reliable results on its chosen level of detail, as 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

Competing interests: none declared

Author's contributions

AF participated in task formulation, questionnaire design, performance of the demonstration studies and in the writing of this article. HL participated in task formulation, developed the questionnaire, designed and performed the interview, test-retest and scaling studies and wrote this article.

Acknowledgements

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.* England; 2002.

3. Gennip EMSJv, Talmon JL: *Assessment and evaluation of information technologies in medicine / edited by E. M. S. J. van Gennip and J. L. Talmon.* 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 A: *The Survey kit / series editor Arlene Fink.* Thousand Oaks, Calif.: Sage Publications.; 1995.

10. Kirwan B, Ainsworth LK: *A Guide to task analysis.* 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, **M1337:** 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; 1992.

17. **Reliability and Validity: Meaning and Measurement** [http://www.ambpeds.org/research_committee.cfm]

18. Keller SD, Ware JE, Jr., Gandek B, Aaronson NK, Alonso J, Apolone G *et al*.: **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. **OntoLog** [http://www.idi.ntnu.no/~heggland/ontolog/]

20. Sittig DF, Kuperman GJ, Fiskio J: **Evaluating physician satisfaction regarding user interactions with an electronic medical record system[abstract].** *Proc AMIA Symp* 1999, 400-404.

21. Lee F, Teich JM, Spurr CD, Bates DW: **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*. 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 PH, Wright JD, Anderson AB: *Handbook of survey research*. Orlando: Academic Press; 1983.

29. Kerlinger FN: *Foundations of behavioral research : educational and psychological inquiry*. London: Holt, Rinehart & Winston; 1969.

30. Jay SJ, Anderson JG: *Use and impact of computers in clinical medicine*. New York: Springer-Verlag; 1987.

Appendix A

Task list with definitions and examples

This appendix contains a list of the 24 tasks as they appear in the third revision of the questionnaire, including individual definitions and examples. The list also appears in appendix A of the thesis summary.

No	Task	Definition	Examples
1	Review the patient's problems	Gather enough existing information to formulate the patient's main problem(s), in order to perform or order new investigations or make a clinical decision.	1) Assess the patient's history during consultation; 2) "jog the memory" during rounds.
2	Seek out specific information from patient records	Seek out a specified, limited amount of information about the patient in the patient records	1) Find out what blood pressure the patient had five years ago; 2) identify the type of pacemaker used by the patient; 3) seek out the last measured creatinine level; 4) find out what antibiotics the patient actually was taking during a reported spell of allergy-like symptoms three years ago.
3	Follow the results of a test or investigation over time	Get the result of at least to identical investigations performed at different points of time, in order to assess the development of the underlying patterns of disease.	1) Assess how quickly the systolic murmur has changed; 2) identify a suspicious decline in hemoglobin concentration following surgery.
4	Obtain the results from new tests or investigations	Identify and gather results from performed and analyzed investigations not yet assessed by a physician.	Gather the latest lab printout sheet for a patient during rounds.
5	Enter daily notes	Personally type or write updated assessments of the patient's condition like progress notes.	Type a progress note for a patient describing decrease in CRP and improved general condition.
6	Obtain information on investigation or treatment procedures	Gather information about the hospital's consensus about how a certain investigation or treatment procedure should be performed.	Look up a treatment of diabetic ketoacidosis in the Hospital's paper-based collection of endorsed procedures.
7	Answer questions concerning general medical knowledge (e.g.concerning treatment, symptoms, complications etc.)	Gather information answering general medical questions about investigation or treatment of a certain condition.	1) Look up in a medical text book about COPD; 2) Ask an experienced colleague about prognosis of COPD.
8	Produce data reviews for specific patient groups, e.g. complication rate, distribution of diagnoses.	Gather or produce data reviews of a certain patient group relevant to the physician's own work.	1) Gather the frequency of certain diagnoses in the local population; 2) compare it to that of the general population.
9	Order clinical biochemical laboratory analyses	Order one or several clinical biochemical laboratory analyses. The booking of the test may be performed by the physician or other personnel.	1) Tell the nurse to order C-reactive protein and white blood cell count; 2) order white blood cell count by using the computerized physician order entry module in the EMR system.

10	Obtain the results from clinical biochemical laboratory analyses	Obtain any results of clinical biochemical laboratory analyses, new and old.	Look in the lab printout sheet for the latest hemoglobin concentration measured before surgery.
11	Order X-ray, ultrasound or CT investigations	Decide upon and order an X-ray, ultrasound or CT investigation, including providing a clinical summary of the patient if required. The booking of the investigation may be performed by the physician or other personnel.	Order X-ray of the thorax by filling in an investigation form by hand, describing symptoms suggesting pneumonia in the lower left lung.
12	Obtain the results from X-ray, ultrasound, or CT investigations	Obtain any results of X-ray, ultrasound, or computer tomography investigations, new and old.	Gather the results of a previous abdominal CT scan by searching for the results sheet in the patient record
13	Order other supplementary investigations	Order supplementary investigations other than clinical biochemical lab tests, X-ray, ultrasound or CT investigation. The booking of the investigation may be performed by the physician or other personnel.	1) Personally order scintigraphy through the order entry system in the EMR; 2) order a microbiological investigation of specimen from a punctured abscess by completing a form.
14	Obtain the results from other supplemental investigations	Obtain any results of laboratory analyses other than clinical biochemical lab tests, X-ray, ultrasound and computer tomography, new and old.	Look up the results of a scintigraphy in the EMR system.
15	Refer the patient to other departments or specialists	Inquire an assessment by a specialist, leading to advice on further follow-up or transfer to another department. The task includes provision of a clinical summary and the purpose of the referal.	Fill in and send a referral form for an endocrinologist's advice on a patient suffering from a fractured hip and poorly managed diabetes.
16	Order treatment directly (e.g. medicines, operations etc.)	Order treatment (medication, surgery, other) to be performed on the hospital, usually not administered by the patient. Writing a prescription is not included in this task.	1) Order aorto-coronar bypass surgery by filling out a form by hand and contacting the head nurse in surgery unit; 2) order peroral anticoagulant medication by writing in the medication list on a continuous medical treatment form during rounds.
17	Write prescriptions	Order medications (or other types of self-administered treatment) for the patient (or a representative) to buy, collect and administer. The order must include instructions for the patient about how and when the treatment should be applied.	Write a prescription of erythromycin against sinusitis by hand, and give it to the patient.
18	Complete sick-leave forms	Produce standarized recommendation of sick leave, including required clinical information, usually by filling in a form.	Fill in a sick leave form by hand, recommending a week sick leave due to a sprained ankle.
19	Collect patient data for various medical declarations	Gather enough clinical information about the patient to produce a valid medical declaration.	Collect information to provide a declaration about the patien't ability to work after the convalescence from a traffic accident is complete.
20	Give written individual information to patients (such as medication lists, status of the disease, etc)	Provide the patient with written clinical information about the patient's condition relevant for continuing care.	Provide the patient with an updated medication list and a summary of the hospital stay upon discharge
21	Give written general information to patients about the illness	Provide the patient with written information about the patient's condition in general; its cause, prognosis and treatment.	Give a leaflet to the patient describing heart failure in general, its causes, prognosis and treatment
22	Collect patient information for discharge reports	Gather enough information to generate a discharge report, including a summary of the hospital stay, current medications or other treatment and planned follow-up.	Gather the patient record and all lab sheets, read, sort and spread out the relevant papers on a table before dictating the discharge report.
23	Check and sign typed dictations	Verify that the contents of transcribed dictations are correct, and sign it.	Check and sign a printed discharge note.

| 24 | Register codes for diagnoses or performed procedures | Perform selection(s) from various classification system for performed clinical procedures or current diagnoses, and document the selections. | Look up in a classification book, and dictate a relevant code for the performed surgery at the end of the surgery report. |

Appendix B

Details of the development of the questionnaire

Development of the 'Frequency of EMR use' section

In the section covering frequency of use , we chose a relative scale in preference to an absolute frequency scale [30], i.e. pertaining to when the task is performed (Table 1) instead of pertaining to days or weeks. This was done to accommodate for the considerable variation in how often each task is performed by the physicians in different clinical contexts.

The questions in this section were revised in three steps (Table 6) In the first revision, "never" and "always" was added to the first and last response alternative to make the scale complete. Furthermore, the original response choice label "Almost half of the time" was changed to "About half of the time" to make its meaning more symmetrical, and hence more suitable for its middle position. In the second revision of the questionnaire, "time" was replaced by "occasions" to make the questions easier to answer for the physician. Instead of having to estimate the product of the mean duration and the frequency of each task [27] to answer these questions, the physician would estimate just the frequency. The modifications to this scale made a separate validation of the response labels necessary, described in validation study 4. Additionally, in the second revision, the task-related questions regarding use of other

software were removed, to reduce the questionnaire length. In the third revision, used in the test-retest study and in the interviews, two escape response choices were added ("task not supported" and "task does not apply to me").

Revisions	Questions and response labels in section covering frequency of EMR use						
Original, Doll & Torkzadeh [13]	(12 questions about user satisfaction)						
	Almost never	Seldom	Almost half of the time	Most of the time	Almost always		
Rev.1, National survey [8]	How often do you use a personal computer (PC) to assist you with the following tasks: <list of 23 tasks>						
	Never/ Almost never	Seldom	About half of the time	Most of the time	Always/ Almost always		
	What computer program do you use for this task?						
	EMR	Other					
Rev. 2, Local survey [12]	How often do you use the EMR system to assist you with the following tasks: <list of 19 tasks>						
	Never/ Almost never	Seldom	About half of the occasions	Most of the occasions	Always/ Almost always		
Rev. 3, Test-retest study and interviews	How often do you use the EMR system to assist you with the following tasks: <list of 24 tasks>						
	Never/ Almost never	Seldom	About half of the occasions	Most of the occasions	Always/ Almost always	This task is not supported by our EMR	This task does not apply to me

Table 6 Development of questions regarding frequency of EMR use and task performance.

Development of the 'Task performance' section

The questions regarding task performance were revised in two steps (Table 7), starting from the second revision of the questionnaire. Here, the original response labels were replaced with more readable versions, as only the "ease of performance" question was used instead of both the original "ease of performance" and the "quality of work" questions. Each task-oriented question about task performance was modified to fit the "more difficult - easier" form of the response labels. In the third revision of the questionnaire, "compared to previous routines" was added to the main question.

34

Further, one response choice was added ("don't know/not applicable") to the scale, to
provide a response choice suitable for respondents who believed no functionality
regarding this task was available.

Revision	Questions and response labels in section covering task performance						
Original, Aydin & Rice [14]	Overall, to what extent has the system changed these two aspects of your own department? • Ease of performing our department's work. • Quality of our department's work						
	1 Significantly decreased	2 Decreased	3 Slightly decreased	4 No change	5 Slightly increased	6 Increased	7 Significantly increased
Rev. 2, Local survey [12]	In your opinion, how has <the EMR system> changed the performance of the following tasks in your department: 1. To <task 1> become... 2. To <task 2> has become... <similar for rest of task list>						
	Significantly more difficult	More difficult	Slightly more difficult	No change	Slightly easier	Easier	Significantly easier
Rev. 3 Test-retest study and interviews	In your opinion, how has the EMR changed the performance of the following work tasks, compared to previous routines? (Check "Don't know/not applicable if you've never used anything else than the current EMR for this task, or if the EMR in your own department doesn't support this task.) 1. To <task 1> has become... 2. To <task 2> has become... <similar for rest of task list>						
	Significantly more difficult	More difficult	Slightly more difficult	No change	Slightly easier	Easier	Significantly easier	Don't know/ not applicable

Table 7 Revisions of the section covering task performance when using the EMR system

Appendix C: Interview guide for content validation of questionnaire

Interview guide, Norwegian original version

This appendix contains the interview guide used for the content validity study of the third revision of the questionnaire. An English truncated version is available on page 40.

Validerende intervjuer for spørreskjemaet

Hallvard Lærum, April 2003

Kvalifiserende spørsmål (telefon)

1. Arbeider du jevnlig kontakt med pasienter i din stilling ved sykehuset, dvs. minst to ganger i måneden eller hyppigere? (Ja/Nei)

   Hvis nei: *Avslutt*.

2. Har du arbeidet mer enn tre måneder ved dette sykehuset?

   Hvis nei: *Avslutt*

Selve intervjuet
Jeg vil gjerne foreta et digitalt opptak av intervjuet for å sjekke at jeg ikke registrerer feil i mitt eget skjema, og for lettere å få med meg dine kommentarer på en korrekt måte. Kun jeg og min veileder vil ha tilgang til opptaket, og opptaket skal slettes når undersøkelsen er slutt, senest 1.desember 2003. Om du ønsker det, vil opptaket slettes før.

3. Er det i orden for deg at jeg foretar et slikt opptak mens intervjuet foregår? (Ja/Nei)

   Hvis ja: Ofte er det ønskelig å skrive ordrett av det du sier i utvalgte deler av opptaket. Dette vil i så fall gjøres på en måte som ikke gjør det mulig å spore uttalelsen tilbake til deg. Er det i orden for deg om slik avskriving eventuelt blir

foretatt? Da vil den anonymiserte teksten bli bevart etter at opptaket er slettet. (Ja/Nei)

(hvis tillatelse til opptak). Nå starter jeg opptaket, som kommer til å gå i bakgrunnen mens vi arbeider. Jeg kommer til å holde meg strengt til en mal for intervjuet som jeg har på papir, og vil stille deg spørsmålene som står her. Jeg vil registrere svarene dine underveis på PC'en. Du må gjerne komme med kommentarer eller spørsmål til meg underveis når du måtte ønske det. Hvis callingen din går mens vi holder på, avbryter vi intervjuet og opptaket, men da er det fint om vi kan avtale å fortsette på et senere tidspunkt. Du har for øvrig rett til å avbryte intervjuet når du måtte ønske det, uten å måtte begrunne dette nærmere. Er du klar?

Nå får du et skjema, som dreier seg om ditt bruk av elektronisk pasientjournal. Kan du fylle det ut? (ikke B2) *Gi skjema om bruk av elektronisk pasientjournal til respondent.*

4. Observerer utfylling, noterer spørsmål fra respondent. Ta tiden.

Detaljert brukerfornøydhet
Nå vil jeg stille deg noen spørsmål om seksjon D i spørreskjemaet, som er et generelt mål på brukerfornøydhet med informasjonssystemer, i dette tilfellet elektronisk pasientjournal. Kan du slå opp på side 5?

5. Var noen av spørsmålene i seksjon D forvirrende eller vanskelig å svare på?

    Hvis ja: Fortell mer om dette

6. Var noen av spørsmålene i seksjon D irrelevante for deg?

    Hvis ja: Fortell mer om dette

7. Var noen av spørsmålene i seksjon D provoserende eller irriterende?

    Hvis ja: Fortell mer om dette

8. Andre kommentarer til seksjon D?

    Evt. Kommentarer

Global brukerfornøydhet
Så var det seksjon E, på siste side. Dette dreier seg også om brukerfornøydhet, men mer på et overordnet plan enn i forrige seksjon.

9. Var noen av spørsmålene i seksjon E forvirrende eller vanskelig å svare på?

    Hvis ja: Fortell mer om dette

10. Var noen av spørsmålene i seksjon E irrelevante for deg?

    Hvis ja: Fortell mer om dette

11. Var noen av spørsmålene i seksjon E provoserende eller irriterende?

Hvis ja: Fortell mer om dette

12. Andre kommentarer til seksjon E?

Evt. Kommentarer

Bruk av EPJ - spørsmålsform

Nå skal vi gå tilbake til begynnelsen av spørreskjemaet, til seksjon B på side 1. Her spør vi om hvor ofte du bruker elektronisk pasientjournal (EPJ) til å hjelpe deg med bestemte arbeidsoppgaver. Vi skal snakke om disse arbeidsoppgavene til slutt, først må vi se litt på spørsmålsformen. Kan du se på svaralternativene som vises i seksjon B.

13. Var noen av svaralternativene til spørsmålene i denne seksjonen forvirrende eller vanskelig å forstå? (ja/nei)

Hvis ja: Fortell mer om dette

14. Synes du svaralternativene passet til spørsmålene i denne seksjonen?

Hvis nei: Fortell mer om dette. Hopp over de tre neste spørsmålene

15. Var betydningen av svaralternativene vanskelig å skille fra hverandre? (ja/nei)

Hvis ja: Fortell mer om dette

16. Var du i noen spørsmål fristet til å krysse mellom svaralternativene? (ja/nei)

Hvis ja: Fortell mer om dette

17. Var du i noen spørsmål fristet til å krysse av mer enn et svaralternativ? (ja/nei)

Hvis ja: Fortell mer om dette

Gjennomføring av arbeidsoppgaver ved hjelp av EPJ

Så var det spørsmålene i seksjon C, på side 3. Her spør vi om gjennomføringen av arbeidsoppgavene er blitt lettere eller vanskeligere i forhold til tidligere rutiner.

18. Er denne måten å spørre på forståelig? (ja/nei)

Hvis nei: Fortell mer om dette

19. Er det mulig å sammenligne på denne måten? (ja/nei)

Hvis nei: Fortell mer om dette

20. Var noen av svaralternativene forvirrende eller vanskelig å forstå? (ja/nei)

Hvis ja: Fortell mer om dette

21. Synes du svaralternativene passet til spørsmålene i denne seksjonen? (ja/nei)

    <u>Hvis nei</u>: Fortell mer om dette. Hopp over de fire neste spørsmålene.

22. Var noen av svaralternativene vanskelig å skille fra hverandre? (ja/nei)

    <u>Hvis ja</u>: Fortell mer om dette

23. Var du i noen spørsmål fristet til å krysse mellom svaralternativene? (ja/nei)

    <u>Hvis ja</u>: Fortell mer om dette

24. Var du i noen spørsmål fristet til å krysse av mer enn et svaralternativ? (ja/nei)

    <u>Hvis ja</u>: Fortell mer om dette

Nå vil jeg vil gjerne stille deg noen spørsmål om hver enkelt av de 24 kliniske arbeidsoppgavene som er vist i seksjon B og C. Her får du et ark hvor spørsmålene med alternativene står. Jeg vil gjerne at du besvarer hvert spørsmål for seg, i hvert fall til å begynne med.

Spørsmål for hver arbeidsoppgave (24)

25. [A] Er denne arbeidsoppgaven forståelig for deg slik den er formulert? (ja/nei)

    <u>Hvis nei</u>: Forklar

26. Hvordan ville DU ha formulert en slik arbeidsoppgave? (Fritekst)

27. [B] Hvordan stiller du deg til følgende utsagn:

    Jeg anser arbeidsoppgaven som en del av mitt arbeide som lege ved sykehuset.

Uenig	Litt uenig	Hverken enig eller uenig	Litt enig	Enig

    <u>Hvis uenig</u> : Fortell mer om dette

    <u>Hvis uenig</u>: Tror du at denne arbeidsoppgaven kan bli aktuell for deg i fremtiden, f.eks. med støtte av ny datateknologi? (Ja/Nei)

    <u>Hvis uenig</u>: Tror du dette er en arbeidsoppgave som leger idet hele tatt bør utføre? (Ja/Nei)

    <u>Hvis uenig</u>: Gå til neste arbeidsoppgave

28. [C] Omtrent hvor ofte utfører du maksimalt denne arbeidsoppgaven?

Aldri	Sjeldnere enn månedlig	Månedlig	Ukentlig	Daglig	Flere ganger daglig	Vet ikke

29. [D] Tenk på forrige gang du utførte denne arbeidsoppgaven. Omtrent hvor lang tid tok det?

Mindre enn 1 min	1-10 min	Mer enn 10 min	Husker ikke	Klarer ikke å anslå

30. [E]Var spørsmålet om bruk av EPJ til denne arbeidsoppgaven vanskelig å svare på? (Ja/Nei)

Hvis ja: Fortell mer om dette (Fritekst)

Spørsmål for hele spørreskjemaet

31. Kan du komme på andre arbeidsoppgaver innen klinisk arbeid som er relevante for deg, som vi ikke har snakket om ennå?

Helt til slutt

32. Har du noen kommentarer om layout og utseende av spørreskjemaet? (fritekst)

33. Var det kjedelig å fylle det ut? (Ja/Nei)

Hvis ja: Fortell mer om dette

Interview guide, English truncated version

This part of appendix C contains the English truncated version of the interview guide.

The interviews covered content validation of the questionnaire described in this paper.

Interview guide for content validation of questionnaire

Hallvard Lærum, April 2003

The third revision of the questionnaire (the version for the test-retest study) is used during this interview study. Normal text is for the evaluator to say out loud, text in parenthesis are response choices, underlined text are conditions for question branching, and text in italics are instructions to evaluator. Text shown in bold face represent titles for navigation, and should not be read out loud.

In this translation, omitted passages of text are indicated by "[...]"

Inclusion:

1. Do you regularly have patient consultations in your position at the hospital, i.e. at least twice a month or more frequently? (Yes/No)

   If no: Terminate interview

2. Have you worked more than three months in this hospital in your current position? (Yes/No)

   If no: Terminate interview

The interview:
I would like to record the interview digitally. [...]

3. Do you mind me doing so during the interview? (Yes/No)

   If response yes: (...) May I transcribe parts and publish parts of the interviews, provided it is done in a way that makes it impossible to trace the quotes back to you? (Yes/No)

[...]

I'll now give you a questionnaire about your use of electronic medical records. Would you please complete it? *Give the questionnaire to the respondent.*

4. Observe what the respondent does and says during completion, and note all questions the respondent asks. Measure the time needed for the completion.

Section D: End User Computing Satisfaction (Doll & Torkzadeh)
[...] Please turn to page 5 in the questionnaire.

5. Were any of the questions in section D confusing or difficult to answer? (Yes/No)

    <u>If yes:</u> Please tell me more about this

6. Did any of the questions in section D appear irrelevant to you? (Yes/No)

    <u>If yes:</u> Please tell me more about this

7. Were any of the questions in section D provoking or annoying? (Yes/No)

    <u>If yes:</u> Please tell me more about this

8. Other comments to section D?

    Note other comments

Section E: Global user satisfaction (incl. SGUS, Aydin & Rice)
[...] Please turn to page 6 in the questionnaire, and consider section E.

9. Were any of the questions in section E confusing or difficult to answer? (Yes/No)

    <u>If yes:</u> Please tell me more about this

10. Did any of the questions in section E appear irrelevant to you? (Yes/No)

    <u>If yes:</u> Please tell me more about this

11. Were any of the questions in section E provoking or annoying? (Yes/No)

    <u>If yes:</u> Please tell me more about this

12. Other comments to section E?

    Note other comments

Section B: Frequency of EMR use, response choices
[...] Please turn to section B on page 1. First, I'd like to consider the response choices in this section.

13. Were any of the response choices for the questions in this section confusing or difficult to understand? (Yes/No)

    If yes: Please tell me more about this

14. In your opinion, do the response choices correspond to the questions in this section? (Yes/no)

    If no: Please tell me more about this. *Note the answer and skip to question 18*

15. Were the meanings of the response choices difficult to tell apart?

    If yes: Please tell me more about this

16. Were you in any of the questions tempted to check between two response choices?

    If yes: Please tell me more about this

17. Were you in any of the questions tempted to check more than one response choice?

    If yes: Please tell me more about this

Section C: Task performance, response choices
[...] Please turn to section C on page 3.

18. Is this way of asking comprehensible? (Yes/No)

    If no: Please tell me more about this

19. Is it possible to make the comparison asked for in the questions? (Yes/No)

    If no: Please tell me more about this

20. Were any of the response choices confusing or difficult to answer? (Yes/No)

    If yes: Please tell me more about this

21. In your opinion, did the response choices correspond to the questions? (Yes/No(

    If no: Please tell me more about this

22. Were the meanings of the response choices difficult to tell apart?

    If yes: Please tell me more about this

23. Were you in any of the questions tempted to check between two response choices?

    If yes: Please tell me more about this

24. Were you in any of the questions tempted to check more than one response choice?

If yes: Please tell me more about this

Questions about each clinical task (repeated 24 times)
Use the task list with definitions and examples (on a separate sheet) for reference during this part of the interview. The respondent may use a separate questionnaire as an aid, containing the tasks and copies of questions 25 to 30.
[...] Please consider task [X], [description of X] *(X=Clinical task 1-24)*

25. Does this task appear comprehensible to you the way it is worded? (Yes/No)

    If no: Please tell me more about this.

26. How would YOU formulate such a clinical task?

27. How much do you agree with the following statement: "I consider the task to be part of my work as an physician at this hospital" (Disagree-Slightly disagree-Neutral-Slightly agree-Agree)

    If disagree or slightly disagree: Please tell me more about this

    If disagree or slightly disagree: Do you think this task could be relevant for you in the future, f.ex. supported by new computer technology (Yes/No)

    If disagree or slightly disagree: Do you think that physicians should do this task at all? (Yes/No).

    If disagree or slightly disagree: *Skip to next clinical task.*

28. About how often do you maximally perform this task? (Never- More seldom than monthly – Monthly – Weekly – Daily – Several times per day – Don't know)

29. (*If more frequently than never*): Try to remember the last time you performed this task. About how much time did it take? (Less than 1 minute – 1 to 10 minutes – More than 10 minutes – Don't remember – Can't estimate the time needed – Not applicable)

30. Did you find the question about how often you use the EMR for this task difficult to answer? (Yes/No)

Questions regarding the whole questionnaire

31. Do you know of any other tasks within clinical work that are relevant to you, which we not yet have discussed?

32. Do you have any comments about layout and design of the questionnaire?

33. Did you find completing it tedious work? (Yes/No)

    If yes: Please tell me more about this.

Appendix D: VAS form used in the scaling study

Studieansvarlig: Hallvard Lærum (epost: hallvard.larum@medisin.ntnu.no)

kvalis

Dato_____ Initialer_____ Yrke ☐ Lege ☐ Sykepleier ☐ Annet

Din tolkning av åtte begreper

I dette skjemaet vil vi gjerne ha din tolkning av et utvalg begreper. Begrepene kan alle plasseres på en skala mellom to ytterpunkter, men vi vet ikke nøyaktig hvor. Her trenger vi hjelp av deg. Se hvordan begrepet "kjølig" er plassert i eksemplet nedenfor:

Eksempel:

Kjølig

Ekstremt kaldt   *Ekstremt varmt*
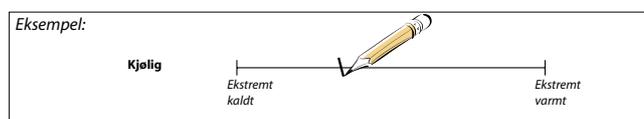
Vi vil gjerne at du plasserer begreper på en tilsvarende måte, ved å tegne et merke på en liggende skala. Prøv selv å plassere begrepene "lunkent" og "iskaldt" på hver sin skala nedenfor:

Lunkent

Ekstremt kaldt   Ekstremt varmt

Iskaldt

Ekstremt kaldt   Ekstremt varmt
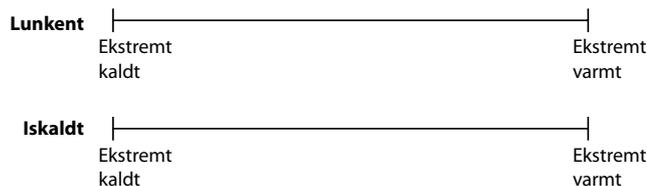
Fint! Nå følger det åtte begreper på hvert sitt ark. Alle begrepene har med hyppighet å gjøre, og kan plasseres et sted mellom ytterpunktene "aldri" og "alltid". Sett et merke mellom disse ytterpunktene der du selv synes hvert begrep passer.

Tusen takk for hjelpen! Resultatene skal brukes til å videreutvikle et spørreskjema for bruk av elektronisk pasientjournal.

Med hilsen

_____
Hallvard Lærum

Scaling study form, front page. The user was asked to interpret words (adverbs) describing frequency by placing a mark on a VAS scale. Each VAS scale ranged from "never" ("aldri") to "always" ("alltid"). The front page describes how to complete the form, by providing examples. The actual data is gathered on the following pages of the form.
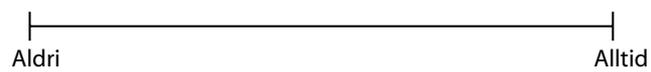
Nesten aldri

├──┤
Aldri   Alltid

Sjelden

├──┤
Aldri   Alltid

Av og til

├──┤
Aldri   Alltid
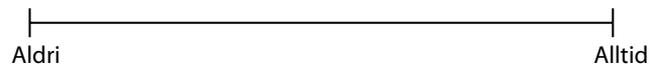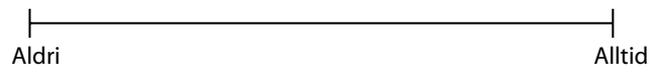
Omtrent halv-parten av tilfellene

├──┤
Aldri   Alltid

Scaling study form, part 2. Each section was separated and presented in random order to the respondent.

Ofte

Aldri   Alltid

Som regel

Aldri   Alltid

Nesten alltid

Aldri   Alltid

Vanligvis

Aldri   Alltid

Scaling study form, part 3. Each section was separated and presented in random order to the respondent.

Additional files

The additional files referred to in this paper will be available on the BioMed Central

Medical Informatics and Decision Making website upon publication of the paper.

(http://www.biomedcentral.com/bmcmedinformdecismak)