International Journal of Medical Informatics (2003) xxx, xxx–xxx

ELSEVIER

International Journal of Medical Informatics

www.elsevier.com/locate/ijmedinf

Evaluation of health information systems—problems and challenges

Elske Ammenwerth[a,*], Stefan Gräber[b], Gabriele Herrmann[c], Thomas Bürkle[d], Jochem König[b]

[a] *Research Group Assessment of Health Information Systems, University for Health Informatics and Technology Tyrol (UMIT), Innrain 98, 6020 Innsbruck, Austria*
[b] *Institute of Medical Biometrics, Epidemiology and Medical Informatics, University Hospital of Saarland, Homburg, Germany*
[c] *Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany*
[d] *Department of Medical Informatics and Biomathematics, University of Münster, Münster, Germany*

Received 7 June 2002; received in revised form 22 October 2002; accepted 9 July 2003

KEYWORDS
Technology assessment;
Information systems;
Information technology;
Evaluation studies;
Health care

Summary *Objectives:* Information technology (IT) is emerging in health care. A rigorous evaluation of this technology is recommended and of high importance for decision makers and users. However, many authors report problems during the evaluation of information technology in health care. In this paper, we discuss some of these problems, and propose possible solutions for these problems. *Methods:* Based on own experience and backed up by a literature review, some important problems during IT evaluation in health care together with their reasons, consequences and possible solutions are presented and structured. *Results and conclusions:* We define three main problem areas—the complexity of the evaluation object, the complexity of an evaluation project, and the motivation for evaluation. Many evaluation problems can be subsumed under those three problem areas. A broadly accepted framework for evaluation of IT in healthcare seems desirable to address those problems. Such a framework should help to formulate relevant questions, to find adequate methods and tools, and to apply them in a sensible way.

1. Introduction

Information technology (IT) is emerging in health care. For example, decision support systems are introduced, knowledge servers allow direct access to state-of-the-art clinical knowledge, and health care professional workstations offer a vast amount of functionality (such as order entry, workflow management, report writing) to support health care professionals in inpatient and outpatient units.

It is evident that the use of modern information technology offers tremendous opportunities to reduce clinical errors (e.g. medication errors, diagnostic errors), to support health care professionals (e.g. availability of timely, up-to-date patient information), to increase the efficiency of care (e.g. less waiting times for patients), or even to improve the quality of patient care [1].

However, there are also hazards associated with information technology in health care: modern information systems are costly (according to [2], about 4.6% of the budget of health care enterprises

*Corresponding author. Tel.: +43-512-586734809; fax: +43-512-586734850.
E-mail address: elske.ammenwerth@umit.at (E. Ammenwerth).

1386-5056/$ — see front matter © 2003 Published by Elsevier Ireland Ltd.
doi:10.1016/S1386-5056(03)00131-X

is spend on information and communication technology), and their failures may cause negative effects on patients and staff.

Therefore, a rigorous evaluation of IT in health care is recommended [3] and of great importance for decision makers and users [4]. Evaluation can be defined as the decisive assessment of defined objects, based on a set of criteria, to solve a given problem [5]. Monitoring and evaluation of clinical software even may become a must in the future, when software programs become considered as medical devices [6], and when recommendations become regulations backed up by the United States Food and Drug Administration (FDA) or other national legal bodies.

When evaluating IT in health care, we must take into account that IT is only one part of the information system of an organization. Information systems can be defined as the overall information processing in an organization, including the involved human players and the information technology used [7]. Thus, during IT evaluation, not only the technology itself, but also the interaction between IT and human players in their information processing role must be taken into account. Evaluation thus has to consider the environment in which IT is used.

Evaluation should accompany the whole life cycle of information technology [8]. Technical verification and validation are most important during system development. Pilot studies and feasibility studies can be conducted after implementation and are followed by cost-benefit or cost-effectiveness studies. Finally, monitoring studies can be used during routine use to track the functioning of information technology in a given environment over a longer period of time.

Evaluation studies can be formative or summative. Formative evaluation strives to improve the information technology under evaluation by providing the developers (and implementers) with feedback. Summative evaluation tries to demonstrate the outcome of an information technology in clinical routine [9].

There are various phase models for evaluation studies in the literature. For example, Holle and Zahlmann [10] proposes four phases (technical pilot study, feasibility study, controlled effectiveness study, cost-effectiveness study), while VATAM [8] is oriented on a eight-phase life cycle of information technology (conception, design, development, integration, early use, exploitation, routine use, end of life cycle). Some more approaches are presented for example in Brender [11].

Many different questions can lead the evaluation of information technology. Typical evaluation questions are (some study example in brackets):

- Which information technology should be selected and installed (e.g. [12–14])?
- What is the usability of the information technology (e.g. [15–17])?
- What are the technical and system features (e.g. performance, software quality) of the information technology that affect its use (e.g. [18])?
- Do the users accept the information technology and use it as intended (e.g. [19–22])? If not, why not?
- How does the information technology affect structural or process quality (time saving, data quality, clinical workflow, patient administration) with regard to different users (physicians, nurses, administrative staff)? Does it work effectively? If not, why not (e.g. [23–27])?
- What are effects of an information technology on the quality of care (e.g. [28–31])?
- Are the patients satisfied with the information technology (e.g. [29,32,33])?
- What are the investment and operational costs of information technology (e.g. [34–36])? Is it cost-effective (e.g. [37–39])?

Despite a large amount of published evaluation studies (e.g. van der Loo [40] found over 1500 citations on evaluation of healthcare IT between 1967 and 1995), many authors report problems during evaluation, such as:

- Unclear, conflicting or changing evaluation goals during the study [41].
- Large efforts needed for the preparation and execution of the study [42].
- Complex and sometimes contradictory results [28].
- Dependence of the evaluation results on the motivation and expectations of the users [43].
- Uncertainty whether results can be generalized to other environments [32,44].

Where do these problems stem from? Are they merely the result of a non-systematic study design and inexperienced evaluation management? Or are there deeper reasons for those problems, perhaps based on the special structure and processes in health care?

Several authors state that IT evaluation in health care can learn from clinical trials (e.g. [45]) and from the systematic study designs which are standard for all major clinical trials in health care. Strong recommendations for the design, execution and publication of clinical studies exist (e.g. [46,47]).

However, some experts are of the opinion that there are inherent problems in information technology evaluation which do not allow the simple transfer of the objectivist study approach of clinical trials to IT evaluation studies [44,48]. Traditional clinical trials strive to objectively measure effects of a diagnostic or therapeutic intervention on an individual patient, while IT evaluation studies attempt to measure the quality as well as the effects of a new information technology on structure, process and outcome of patient care. The question of whether problems in the evaluation of health care IT are due to methodological insufficiencies, or rather due to more complicated circumstances than in clinical trials still seems to be unanswered.

In this paper, we want to review some of the underlying reasons which make evaluation of health care IT so difficult. As many failed evaluation studies may not be published [49], it is difficult to do a real comprehensive review of such reasons. Thus, we will mainly rely on our own experiences with evaluation studies. We will support our experiences with literature.

We will structure the problems into three main problem areas: the complexity of the object of evaluation, the complexity of the evaluation project, and the motivation to perform evaluation. We will end the paper with some general recommendations for evaluation studies in health care, and discuss the need for a comprehensive evaluation framework.

2. Typical problems in evaluation of IT in health care

2.1. First problem area: complexity of the evaluation object

2.1.1. Problem

When understanding IT as part of the information system of an organization, it is clear that an evaluation will often not concentrate only on hardware and software, but on the information processing, i.e. on the interaction between information technology and users in a given environment. That means it is often more a situation or process which is evaluated than a single product. Thus, evaluation requires not only an understanding of computer technology, but also of the social and behavioral processes that affect and are affected by the technology. The success of IT heavily depends, e.g. on how it matches with clinical workflow, on how the technology is introduced in the organization, on the quality of information it offers, on training and support, on the depth of usage, and on the motivation of the users and on their use of the system [50,51]. The evaluation object is therefore usually broader and more complex than, e.g. a drug or a new medical diagnostic procedure.

2.1.2. Consequences

The complexity of the evaluation objects has some important consequences.

- The *introduction of IT takes time*. It is not enough to implement the technology and then to immediately measure the effects. Users and workflow need a lot of time to get used to new tools and to completely exploit the new possibilities [51,52]. Hardware or software modifications done to improve, e.g. IT usability or functionality may also change the use and the effects of the technology. Thus, evaluation results can change during this first period of use. For example, an evaluation study of the quality of nursing documentation after IT introduction found significant changes of several quality indicators after 3 and 9 months of use [53]. For a summative evaluation, the evaluator thus may have to wait much longer than the typical wash-in period needed in clinical trials.
- Even after an introduction period, the evaluation object often steadily changes [48] (*moving evaluation target*). For example, the use of information technology may be affected by changes in work organization, or in staff [52]. Freezing the environment during the study period is often neither useful nor possible [41,48]. Thus, after finishing a study, the environment may already have markedly changed compared to the beginning of the study, making the results obsolete. It is nearly impossible to reach a stable situation in a flexible health care environment which makes evaluation results dependent of the point in time where the evaluation took place. For example, during the 3-year evaluation of a nursing documentation system [54], changes in documentation workflow and in the use of the technology took place, making it difficult to directly compare, e.g. user acceptance scores from different measurement times. In clinical trials, the evaluation object is usually more stable.
- Each information system in our definition is *quite unique*. While the information technology may be similar in various departments, workflow, users and used functionality may be different. In addition, the organization of its introduction [55] as well as the overall user motivation [56] is an important factor. Thus, even when the same information technology is introduced, its effects may

be varying [4]. The influence of such factors on the results of an evaluation study is often hard to disentangle [57]. The influencing factors can usually only partly be controlled, posing the problem of *external validity* [48]: many evaluation studies may be valid only for the particular institutions with their specific information system. For example, the evaluation of documentation quality and user acceptance after introduction of a nursing documentation system found significant different results on the various study wards [58], e.g. due to differences in workflow, in computer knowledge of users, or in organization of user support. In many cases, those factors cannot as easily be controlled through adequate study design as in controlled clinical trials.

2.1.3. Possible solutions

The complexity of the evaluation object is an inherent attribute in health care IT evaluation and cannot be reduced. However, there are some ways to handle this problem in evaluation studies.

- To address the problem of external validity, the information technology and its environment which is going to be evaluated should be *defined in detail* before the beginning of the study. Not only the software and hardware which is used should be described, but also the number of users and their experience and motivation, the way information technology is introduced and used, the general technical infrastructure (e.g. networks) and any further aspects which may influence the usage of an information technology and its effects. Of special importance should also be the functionality and the way it is really used. Only this information may allow interpretation of the study results and comparison of different locations.
- To address the problem of the moving evaluation target, *all changes in the information technology* and its interaction with the users should be carefully documented during the study [52]. For example, changes in workflow, in staffing, or in hardware or software should be documented with reference to the ongoing evaluation. This permits the explanation of changes and differences in effects measured during the study period.
- Another approach to address the problem of the moving evaluation target may be to define smaller *evaluation modules*. This would allow the evaluation design or evaluation questions to be adapted to changes in the environment. For example, the evaluation of a nursing documentation system started with an assessment of the time needed for documentation during the introduction phase [23], continued with an evaluation module on changes in documentation quality over the first year [54], and then focused on the question of workflow support and user acceptance after longer use [58]. Each module answered a question related to a defined phase of the introduction of the information technology.
- An evaluation must be planned in a *long-term perspective* in order to allow the users and the environment to integrate the new information technology. Hence enough resources for long-term evaluation (e.g. over several months or even years) should be available. For example, the already described stepwise evaluation (and consecutive improvement) of a nursing documentation system spanned 3 years. This long term evaluation takes into account the learning curve in the introduction phase as well as subsequent changes of the moving evaluation target.
- With regard to the complexity of the evaluation object, special attention should also be paid to unexpected, *adverse effects*, e.g. with regard to the quality of patient care, such as an increase in patients' stay, or high drop-out of participants. This may indicate that something is 'going wrong'. They may then demand deeper analysis and further interventions (e.g. expand study questions, call off the project, or modify the information technology). For example, during the evaluation of user acceptance, unexpected disturbances in nursing workflow were found after the introduction of a nursing documentation system [58]. These problems were addressed in a subsequent study.

2.2. Second problem area: complexity of the evaluation project

2.2.1. Problem

Evaluation of information technology is usually performed in the real and complex health care environment, with its different professional groups (such as physicians, nurses, patients, administration, IT staff hospital management, funding bodies), and its high dependency on external influences such as legislation, economic constraints, or patient clientele.

This poses problems to the evaluation projects. For example, the different stakeholders often have different conceptions and views of successful information technology [59]. The medical user may want an easy-to-use system releasing him or her

from documentation tasks, while the administrative user wants a system enforcing complete documentation. The different stakeholder requirements can serve as a frame of reference for evaluation during the early phases of the IT life cycle, but also guide evaluations during later phases. In each case, multiple stakeholder view may lead to a multitude of (possibly conflicting) evaluation questions [41]. Evaluation can, for example, be done from an economic, sociologic, psychological, organizational, technical, information logistical, or clinical point of view [60].

Depending on the point of view adopted, the evaluation will require different study designs and evaluation methods. The evaluation researcher must decide, e.g. on the evaluation approach (e.g. objectivistic versus subjectivistic [9]), on the adequate evaluation methods (e.g. quantitative versus qualitative), and on the study design (e.g. RCT versus observational study). Each has its own advantages and drawbacks [48,61], making their selection a rather challenging endeavor. This multitude of possible evaluation questions and available evaluation methods makes the planning of an evaluation study quite complex.

2.2.2. Consequences

The complexity of the evaluation project has several consequences.

- The overall success of information technology is elusive to define [51], and it is therefore often difficult to establish *clear-cut evaluation criteria* to be addressed in a study [57]. Each stakeholder group may have individual questions [8], and a universal evaluation in terms of absolute or relative benefits is usually not feasible (or, from a more subjectivistic point of view, even not possible). It is also unrealistic to expect that the information technology itself will have a direct and easy to measure effect on the outcome quality of patient care like a drug. Thus, indirect measures are often used such as user satisfaction or changes of clinical processes, which, however, do not give a really complete picture of the benefit of information technology. Many studies thus tend to integrate different views and questions, leading to very complex questions that are difficult to manage with limited resources in a given period of time.
- However, when *the multitude of possible evaluation criteria* are reduced to an appropriate number, the study runs the risk of being of limited value for certain stakeholders, or of presenting a distorted picture of reality. For example, a study found an increased time effort for nursing documentation, but did not analyze in parallel changes in quality and completeness of documentation, thus leading to a potentially unbalanced picture of the effects of a nursing documentation system [23].
- Often, *changes in the evaluation questions* may occur during the study, e.g. based on intermediate evaluation results, new insights, changes in stakeholders' opinions, or changes of the information technology (scope creep [62]). For example, after the evaluation of a nursing documentation system found increased time efforts [23], the question arose whether the completeness of documentation had changed. Changes in study questions, however, may be difficult to balance with study resources.
- The selection of *adequate evaluation design and evaluation methods* is often regarded as a problem during evaluation studies. Evaluators may not be sufficiently aware of the broadness of available approaches, or be too deeply embedded in either the qualitative or the quantitative paradigm, neglecting the possible contributions of the complementary approach [48]. Thus, inadequate methods or study designs may be chosen which may not be able to answer the original study questions [44].

2.2.3. Possible solutions

The following suggestions may be useful in order to deal with the complexity of the evaluation project.

- It is recommended that the *general intention of the evaluation* and the starting point should be decided early on. Evaluation should be started early in the life cycle of information technology [8]. In formative studies striving to support the further development of IT in health care, evaluation is often carried out too late, making it difficult to modify the information technology. In principle, evaluation should start before the new information technology is implemented, in order to allow for early gathering of comparative data, and then continue during all phases of its life cycle.
- The *areas of evaluation should be restricted* to aspects which can be measured with the available resources. A complete evaluation of all aspects of a system (such as economics, effectiveness, and acceptance) is usually not feasible. Nevertheless, the evaluation should be able to answer the most important questions. For example, during the evaluation of a nursing documentation system, the evaluation mostly focussed mainly on the nurses' point of view, however the

medical staff was also shortly questioned on the usefulness of the system [63]. A balance between the resources of a study and the inclusion of the most relevant aspects has to be found. Aspects which are not essential for the evaluation should also be formulated.

- Sufficient time should be invested into the definition of *relevant study questions*. All involved stakeholder groups should discuss and agree on the goals of evaluation [8]. The selected study questions should be relevant for decision-making with regard to introduction, operation or justification of information technology. Conflicting goals should be discussed and solved, as they are not only problematic for an evaluation, but for the overall management of new information technology.
- In quantitative studies, the *evaluation criteria* such as outcome variables (e.g. documentation quality), explanatory variables (e.g. users' documentation skills) and possible cofounders (e.g. number of patients) have to be strictly defined before the start of the study. In qualitative studies, the constructs and variables which are going to be investigated should also be stated in detail. This allows researchers to better analyze the results and to compare different studies.
- When *new evaluation questions* emerge during the study, they should only be included in the study design when it is possible without creating problems. Otherwise, they should be tackled in consecutive studies. Strict quantitative evaluation methodologies based on accepting or rejecting of a small number of predefined hypotheses may not permit the introduction of new questions during the study. In any case, each shift in evaluation questions must thoroughly be documented. For example, during the evaluation of the nursing documentation system, the issue of its effect on nursing workflow which arose during the study of user acceptance was addressed in a subsequent study [58].
- For each study question, *adequate methods* must be chosen. A triangulation of methods may sometimes help to best answer the study questions [64]. For example, to address the effects of a nursing documentation system, both quantitative methods (time measurement, user acceptance scales, documentation quality measurement) as well as qualitative methods (focus group interviews) were used [58]. It may also he useful to take into account both subjective data, addressing the perceptions of different user groups, as well as objective data [4,65].

2.3. Third problem area: motivation for evaluation

2.3.1. Problem

An evaluation study can normally only be conducted when there is sufficient funding, and a sufficient number of participants (e.g. staff members, wards). Both these variables depend on the motivation of stakeholders (e.g. hospital management) to perform an evaluation. Sometimes, this motivation is not very high, because, for example, of fear for negative outcome [58], or of fear for revealing deficiencies of already implemented technology. In addition, the introduction of IT in an organization is a deep intervention that may have large consequences. It is thus often very difficult to organize IT evaluation in the form of an experiment, and to easily remove the system again at the end of the study in case the evaluation was too negative.

Even with a motivated management, it may be difficult to find suitable participants. Participating in a study usually requires some effort from the involved staff, such as filling out questionnaires, or being involved in time measurements [66]. In addition, while the users have to make large efforts to learn and use a new, innovative system, the benefit of joining a pilot study is usually not obvious (the study is conducted in order to investigate possible effects), but participation may even include some risks for the involved staff. For example, disturbances in workflow occurred after the introduction of a nursing documentation system [58]. Such efforts and risks which may be part of the evaluation of new information technology are usually not paid for by the organizers of studies, as opposed to clinical trials.

In summary, due to the given reasons, the hospital management, as well as involved staff members are often reluctant to participate in IT evaluation studies.

2.3.2. Consequences

The described problem has consequences for the study.

- Without the support and motivation of the stakeholders to conduct an evaluation study, it will be difficult to get *sufficient resources* for an evaluation and sufficient participants willing to participate.
- Due to the given problems, the study organizer is often happy to recruit any participant who volunteers to participate. However, those participants may be more motivated to participate than the 'normal' user. This leads to the well-known *volunteer effect*, where results are better when

participants are motivated. This in turn reduces the external validity of the study [9].
- Evaluation results are not only important for the involved units, but also for the overall organization or for similar units in other organizations. To allow transfer of results, the pilot wards or pilot users must be *sufficiently representative* for other wards or users. But, as each information technology within its environment is quite unique (see problem area 1), it is difficult to find comparable or representative participants. In addition, a small number of participants (or the complexity of workflow) does not allow randomization which is normally used in clinical trials to obtain comparability of study groups.

2.3.3. Possible solutions

Some suggestions may show how the number of available participants can be increased.

- To increase the number of participants, two approaches should be combined. First, the responsible management should be *informed and motivated* to support the study. The result of an evaluation study may be important to decide on new information technology, and to support its continuous improvement. Second, the possible participants could be directly addressed. It should be made clear that the study provides the opportunity to influence not only the future development of IT in health care but also the own working environment. It is their chance to participate. For the staff, personal experience with new information technology may be a motivating factor, and the unit gains reputation by participation in innovative projects. For units where the technology is already operational, the participation in a study may allow to better influence its further development. User feedback of study results may act as an important driving force for users to participate in the study. However, both possible benefits as well as required efforts for the study and potential risks of innovative technology must be honestly explained.
- Offering *financial compensation* or additional staff for the study period may help to gain support from participants and from management. Clinical trials are often funded by the pharmaceutical and medical-technical industry with interest in proving the safety and usefulness of their new drugs and medical procedures. In the future, when legal regulations may demand such proof for health care IT (certification), the IT industry may be willing to finance larger studies [6,45].
- As in clinical trials, *multi-centric studies* should be considered [25,66]. This would largely increase the number of available participants. This means however, that study management requires much more effort. A multi-centric study design is difficult when the environment is completely different. In addition, the variation between study participants will be bigger in multi-centric trials than in single-centre ones. This may render interpretation and comparison of results even more difficult (cp. discussion in problem area 1).

2.4. Summary of general recommendations

The above discussed problems and approaches will now be summarized in a list of 12 general recommendations for IT evaluation in healthcare:

1. Evaluation takes time, thus take your time for thorough planning and execution.
2. Document all of your decisions and steps in a detailed study protocol. Adhere to this protocol; it is your main tool for a systematic evaluation.
3. Strive for management support, and try to organize long-term financial support.
4. Clarify the goals of the evaluation. Take into account the different stakeholder groups. Dissolve conflicting goals.
5. Reduce your evaluation questions to an appropriate number of the most important questions which you can handle within the available time and budget. If new questions emerge during the study, which cannot easily be integrated, postpone them for a new evaluation study.
6. Clarify and thoroughly describe the information technology which is the object of your evaluation, and the environment. Take note of any changes of the information technology and its environment during the study that may affect results.
7. Select an adequate study design. Think of a stepwise study design.
8. Select adequate methods to answer your study questions. Neither objectivist nor subjectivist approaches can answer all questions. Take into account the available methods. Consider being multi-methodic and multi-disciplinary, and consider triangulation of methods, data sources, investigators, and theories. Strive for methodical (e.g. biometrics) advice.
9. Motivate a sufficient number of users to participate. Consider multi-centric trials and financial or other compensation.
10. Use validated evaluation instruments wherever possible.

11. Be open to unwanted and unexpected effects.
12. Publish your results and what you learned to allow others to learn from your work.

3. Discussion

Medical informatics is an academic discipline, and thus evaluation is an important part of any system development and implementation activity [67,68]. However, many problems with regard to health care IT evaluation have been reported. Wyatt and Spiegelhalter in 1992 [66] as well as Grémy and Degoulet in 1993 [69] already discussed the complexity of the field, the motivation issue, and methodological barriers to evaluation. First examples of meta-analysis of IT evaluation studies confirmed those barriers (e.g. [70–73]). The problems they addressed still seem to be valid today.

In this paper, we elaborated a number of important problems to some more detail, and structured them into three areas: the complexity of the evaluation object, the complexity of the evaluation project with its multitude of stakeholders, and the motivation for evaluation.

Others may find further evaluation problems, or may structure them differently. We did not address the validity of those areas, but found them helpful to structure the problems which we have experienced personally, or which are often mentioned in the literature. Addressing those main problem areas in future evaluation research may help to design and conduct better evaluation studies and produce more valid evaluation results.

It is interesting so see that some of the problems addressed in health care can also be found in other evaluation fields. For example, Palvia et al. [51] discuss the different perspectives of the stakeholders when dealing with expert systems in insurance companies. Chan [74] presents results from a literature review which show a ''schism'' between the use of quantitative and qualitative measures in IT-value research. Changing evaluation objectives in the context of e-commerce investments are reported by Doherty and Mcaulay [75]. Grundén [76] discusses the need for organizational and social evaluation of computer-supported cooperative work (CSCW) systems supporting complex informal communication between different user groups and proposes an evaluation model for this. These few examples show that it may be promising to look into other evaluation fields outside health information systems as to learn from each other.

A kind of framework to support evaluation studies of information systems may be useful to address the problem areas discussed in this paper. In fact, many authors have formulated the necessity for such a framework (e.g. [4,77,78]). Based on our analysis in this paper, such a framework could be helpful when it supports the evaluator in answering, e.g. the following questions:

1. Which stakeholders should I take into account when planning an evaluation study?
2. How can I motivate stakeholders and get funding for an evaluation study?
3. How can I come to a consensus on the aims of an evaluation study?
4. How detailed should the description of the evaluation object (the information system) be?
5. What are the steps for planning, executing and analyzing an evaluation study?
6. Which evaluation questions are possible, and which may be best suited to reach my evaluation aims?
7. How can I derive clear evaluation criteria from the evaluation questions?
8. Which methods (and tools) for data acquisition, data analysis, and data presentation are available, and which would be optimal to answer my evaluation questions?
9. What should a study protocol and study report contain?
10. How can I address the moving target problems during the evaluation study, as well as other frequent problems encountered during the evaluation of health information systems?

Some authors have already addressed a certain number of these questions. For example, the VATAM guidelines [8] focus on the aims of an evaluation based on the views of the stakeholders, on possible evaluation areas (such as IT development, IT quality, user requirements, HTA, or marketing), and on the description of the information system (application type and life cycle). These guidelines offer a classification to describe methods and tools for evaluation. Some authors concentrate more on the content of an evaluation protocol (e.g. [79] or [80]). Many papers deal with possible areas of evaluation: Shaw [77] structured them into clinical, human and organizational, educational, administrative, technical, and social aspects, Sapirie [81] distinguishes data entry, data analysis, information use, IT resources, and IT management, Hebert [82] discussed criteria of structure, process or outcome quality of IT, Grant et al. [78] focused on a strategic, tactical or operational evaluation level, Palvia et al. [51] discussed task, technology, people, or organization criteria, Kaplan et al. [83] presented the structure of communication, care, control, or context, and Jorgensen [60] distinguishes need for the resource, development process, resource's intrin-

sic structure, resource's functions, and resource's impact. Some of the authors combined the axis of the measured effects with other axis such as phase of life cycle of the information system, or involved stakeholders (e.g. [78]).

While all those works can be regarded as important steps towards a framework, it is obvious that it only addresses some of the above stated questions. In addition, many of those approaches have not yet been fully empirically validated. The quest for an evaluation framework seems to be an important future task for medical informatics.

4. Conclusion

Evaluation studies in health care IT take a lot of time, resources, and know-how. Research in the area of health care IT evaluation is just beginning. It is still mostly unclear how 'good' information systems should look like. Clearly defined methodological guidelines which take the difficulties of information system evaluation in health care into account may help to conduct better evaluation studies. We have classified some of the problems encountered in healthcare IT evaluation under the three main problem areas complexity of the evaluation object, complexity of the evaluation project and limited motivation for evaluation. We suggested a list of essential recommendations to support the evaluation of information systems. A broadly accepted framework for IT evaluation in healthcare which goes more into details seems desirable, supporting the evaluator during planing and executing of an evaluation study. Such a framework should help to formulate adequate evaluation questions, to find adequate methods and tools, and to apply them in a sensible way. Many problems may be reduced or avoided following such a framework. Some research seems still necessary in this area.

Acknowledgements

We want to thank the anonymous reviewers for their very helpful comments on first versions of this paper, and Jytte Brender for information on her recent work on perils and pitfalls of assessment studies. Thanks also to Frieda Kaiser for her help with the English language. Special thanks to Jan Talmon for intensive discussions and comments.

References

[1] D.W. Bates, M. Cohen, L.L. Leape, J.M. Overhage, M.M. Shabot, T. Sheridan, Reducing the frequency of errors in medicine using information technology, JAMIA 8 (4) (2001) 299—308.

[2] HIMSS, The 11th Annual HIMSS Leadership Survey Sponsored by IBM: Trends in Healthcare Information and Technology—Final Report, Healthcare Information and Management Systems Society, http://www2.himss.org/survey/2000/survey2000.html (last access: January 2003).

[3] M. Rigby, Evaluation: 16 powerful reasons why not to do it—and 6 over-riding imperatives, in: V. Patel, R. Rogers, R. Haux (Eds.), Proceedings of the 10th World Congress on Medical Informatics (MedInfo 2001), Studies in Health Technology and Informatics, vol. 84, IOS Press, Amsterdam, 2001, pp. 1198–1202.

[4] B. Kaplan, N. Shaw, People, organizational and social issues: evaluation as an exemplar, in: R. Haux, C. Kulikowski (Eds.), Yearbook of Medical Informatics 2002, Schattauer, Stuttgart, 2002, pp. 91–102.

[5] L. Heinrich, Informationsmanagement, Oldenbourg, München Wien, 1999.

[6] R. Miller, R. Gardner, Recommendations for responsible monitoring and regulation of clinical software systems, JAMIA 4 (6) (1997) 442—457.

[7] A. Winter, E. Ammenwerth, O. Bott, B. Brigl, A. Buchauer, S. Gräber, A. Grant, A. Häber, W. Hasselbring, R. Haux, A. Heinrich, H. Janssen, I. Kock, O.-S. Penger, H.-U. Prokosch, A. Terstappen, A. Winter, Strategic information management plans: the basis for systematic information management in hospitals, Int. J. Med. Inf. 64 (2–3) (2001) 99—109.

[8] VATAM, VATAM guidelines, Validation of Health Telematics Applications (VATAM), http://www-vatam.unimaas.nl (last access: July 2003).

[9] C. Friedman, J.C. Wyatt, Evaluation Methods in Medical Informatics, Springer, New York, 1997.

[10] R. Holle, G. Zahlmann, Evaluation of telemedical services, IEEE Trans. Inf. Technol. Biomed. 3 (2) (1999) 84—91.

[11] J. Brender, Trends in assessment of IT-based solutions in healthcare and recommendations for the future, Int. J. Med. Inf. 52 (1–3) (1998) 217—227.

[12] J. Chocholik, S. Bouchard, J. Tan, D. Ostrow, The determination of relevant goals and criteria used to select an automated patient care information system: a Delphi approach, JAMIA 6 (3) (1999) 219—233.

[13] L. Einbinder, J. Remz, D. Cochran, Mapping clinical scenarios to functional requirements: a tool for evaluating clinical information systems, in: J. Cimino (Ed.), Proceedings of the AMIA Annual Fall Symposium, Hanley & Belfus, Philadelphia, 1996, pp. 747–751.

[14] R. Stiefel, E. Rizkalla, The elements of a complete product evaluation, Biomed. Instrum. Technol. November-December (1995) 482–488.

[15] M. Beuscart-Zéphir, P. Sockeel, B. Bossard, R. Beuscart, Activity modelling for assessing the usability of telematics applications in healthcare, in: B. Cesnik, A. McCray, J. Scherrer (Eds.), Proceedings of the 9th World Congress on Medical Informatics (MedInfo 1998), Studies in Health Technology and Informatics, vol. 52, IOS Press, Amsterdam, 1998, pp. 832–836.

[16] S. Gräber, Application of clinical workstations: functionality and usability, Clin. Perform. Qual. Health 5 (2) (1997) 71—75.

[17] A.W. Kushniruk, C. Patel, V.L. Patel, J.J. Cimino, 'Televaluation' of clinical information systems: an integrative approach to assessing Web-based systems, Int. J. Med. Inf. 61 (1) (2001) 45—70.

[18] G. Braccini, F. Fabbrini, M. Fusani, Software quality assessment for health care systems, in: C. Pappas (Ed.), Proceedings of the Medical Informatics Europe 1997, Studies

in Health Technology and Informatics, vol. 43, IOS Press, Amsterdam, 1997, pp. 746–750.
[19] J. Anderson, Increasing the acceptance of clinical information systems, MD Comput. January–February (1999) 62–65.
[20] C. May, L. Gask, T. Atkonson, N. Ellis, F. Mair, A. Esmail, Resisting and promoting new technologies in clinical practice: the case of telepsychiatry, Soc. Sci. Med. 52 (2001) 1889—1901.
[21] D. Goodhue, Understanding user evaluations of information systems, Manage. Sci. 41 (12) (1995) 1827—1844.
[22] D. Sittig, G. Kuperman, J. Fiskio, Evaluating physician satisfaction regarding user interactions with an electronic medical record system, in: Proceedings of the Annual Symposium on Computer Application in Medical Care, Hanley & Belfus, Philadelphia, 1999, pp. 400–404.
[23] E. Ammenwerth, R. Eichstädter, R. Haux, U. Pohl, S. Rebel, S. Ziegler, A randomized evaluation of a computer-based nursing documentation system, Methods Inf. Med. 40 (2) (2001) 61—68.
[24] T. Bürkle, R. Kuch, H. Prokosch, J. Dudeck, Stepwise evaluation of information systems in an university hospital, Methods Inf. Med. 38 (1) (1999) 9—15.
[25] K. Herbst, P. Littlejohns, J. Rawlinson, M. Collinson, J. Wyatt, Evaluating computerized health information systems: hardware, software and human ware: experiences from the Northern Province, South Africa, J. Public Health Med. 21 (3) (1999) 305—310.
[26] S. Shea, R. Sideli, W. DuMouchel, G. Pulver, R. Arons, P. Clayton, Computer-generated informational messages directed to physicians: effect on length of hospital stay, JAMIA 2 (1) (1995) 58—64.
[27] J. Watkins, G. Weatherburn, S. Bryan, The impact of a picture archiving and communication system (PACS) upon an intensive care unit, Eur. J. Radiol. 34 (1) (2000) 3—8.
[28] R. Currell, C. Urquhart, P. Wainwright, R. Lewis, Telemedicine versus face to face patient care: effects on professional practice and health care outcomes (Cochrane review), in: The Cochrane Library, Issue 1, Update Software, Oxford, 2000.
[29] J.E. Gray, C. Safran, R.B. Davis, G. Pompilio-Weitzner, J.E. Stewart, L. Zaccagnini, D. Pursley, Baby CareLink: using the internet and telemedicine to improve care for high-risk infants, Pediatrics 106 (6) (2000) 1318—1324.
[30] C. McCowan, R.G. Neville, I.W. Ricketts, F.C. Warner, G. Hoskins, G.E. Thomas, Lessons from a randomized controlled trial designed to evaluate computer decision support software to improve the management of asthma, Med. Inf. Internet. Med. 26 (3) (2001) 191—201.
[31] P.R. Dexter, S. Perkins, J.M. Overhage, K. Maharry, R.B. Kohler, C.J. McDonald, A computerized reminder system to increase the use of preventive care for hospitalized patients, N. Engl. J. Med. 345 (13) (2001) 965—970.
[32] F. Mair, P. Whitten, Systematic review of studies of patient satisfaction with telemedicine, BMJ 320 (7248) (2000) 1517—1520.
[33] R. Nahm, I. Poston, Measurement of the effects of an integrated, point-of-care computer system on quality of nursing documentation and patient satisfaction, Comput. Nurs. 18 (5) (2000) 220—229.
[34] J. Enning, A. Bakker, Analysis of costs of information systems, in: E. van Gennip, J. Talmon (Eds.), Assessment and Evaluation Technologies in Medicine, Studies in Health Technology and Informatics, vol. 17, IOS Press, Amsterdam, 1995, pp. 87–97.
[35] E. van Gennip, A. Bakker, Assessment of effects and costs of information systems, Int. J. Biomed. Comput. 39 (1995) 67—72.
[36] S. Bryan, G. Weatherburn, M. Buxton, J. Watkins, J. Keen, N. Muris, Evaluation of a hospital picture archiving and communication system, J. Health Serv. Res. Policy 4 (4) (1999) 204—209.
[37] C. Lock, What values do computers provide to NHS hospitals? BMJ 312 (1996) 1407—1410.
[38] F.S. Mair, A. Haycox, C. May, T. Williams, A review of telemedicine cost-effectiveness studies, J. Telemed. Telecare 6 (Suppl. 1) (2000) S38—S40.
[39] S. Bryan, M. Buxton, E. Brenna, Estimating the impact of a diffuse technology on the running costs of a hospital. A case study of a picture archiving and communication system, Int. J. Technol. Assess. Health Care 16 (3) (2000) 787—798.
[40] R. van der Loo, Overview of published assessment and evaluation studies, in: E.M.S.J. van Gennip, J.S. Talmon (Eds.), Assessment and Evaluation of Information Technologies, Studies in Health Technology and Informatics, vol. 17, IOS Press, Amsterdam, 1995, pp. 261–282.
[41] H. Heathfield, V. Peel, P. Hudson, S. Kay, L. Mackay, T. Marley, L. Nicholson, R. Roberts, J. Williams, Evaluating large scale health information systems: from practice towards theory, in: D. Masys (Ed.) Proceedings of the AMIA Annual Fall Symposium, Hanley & Belfus, Philadelphia, 1997, pp. 116–120.
[42] T. Bürkle, M. Schmitz, H.U. Prokosch, J. Dudeck, A systematic approach for evaluation of nursing work in an university hospital, in: P. Barahona, M. Veloso, J. Bryant (Eds.), Medical Informatics Europe 1994, European Federation for Medical Informatics (EFMI), Lissabon, 1995, pp. 1321–1325.
[43] R. Henderson, F. Deane, User expectations and perceptions of a patient management information system, Comput. Nurs. 14 (3) (1996) 188—193.
[44] H. Heathfield, D. Pitty, R. Hanka, Evaluating information technology in health care: barriers and challenges, BMJ 316 (1998) 1959—1961.
[45] W. Tierney, J. Overhage, C. McDonald, A plea for controlled trials in medical informatics, JAMIA 1 (4) (1994) 353—355.
[46] ICH-GCP, Good Clinical Practice (CPMP/ICH/135/95), International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, 1996 (http://www.ifpma.org/ichl.html).
[47] H. Schäfer, J. Berger, K.-E. Biebler, U. Feldmann, E. Greiser, K.-H. Jöckel, J. Michaelis, A. Neiss, H.H. Rapse, B.-P. Robra, M. Schumacher, J.-J. Trampisch, N. Victor, J. Windeler, Empfehlungen für die Erstellung von Studienprotokollen (Studienplänen) für klinische Studien, Informatik Biometrie und Epidemiologie in Medizin und Biologie 30 (3) (1999) 141—154.
[48] J.R. Moehr, Evaluation: salvation or nemesis of medical informatics? Comput. Biol. Med. 32 (3) (2002) 113—125.
[49] W. Tierney, C. McDonald, Testing informatics innovations: the value of negative trials, JAMIA 3 (5) (1996) 358—359.
[50] M. Berg, Patient care information systems and healthcare work: a sociotechnical approach, Int. J. Med. Inf. 55 (1999) 87—101.
[51] S. Palvia, R. Sharma, D. Conrath, A sociotechnical framework for quality assessment of computer information systems, Ind. Manage. Data Syst. 101 (5) (2001) 237—251.
[52] M. Butler, A. Bender, Intensive care unit bedside documentation systems—realizing cost savings and quality improvements, Comput. Nurs. 17 (1) (1999) 32—38.
[53] C. Mahler, E. Ammenwerth, A. Wagner, A. Tautz, T. Happek, B. Hoppe, R. Eichstädter, Effects of a computer-based nursing documentation system on the quality of nursing documentation, J. Med. Syst., 2003, in press.
[54] E. Ammenwerth, U. Mansmann, C. Iller, R. Eichstädter, Factors affecting and affected by user acceptance of

computer-based nursing documentation: results of a two-year study, JAMIA 10 (1) (2003) 69—84.
[55] N.M. Lorenzi, R.T. Riley, Managing change: an overview, JAMIA 7 (2) (2000) 116—124.
[56] D.E. Forsythe, B.G. Buchanan, Broadening our approach to evaluating medical information systems, in: P. Clayton (Ed.), Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care, McGraw-Hill, New York, 1992, pp. 8–12.
[57] J. Wyatt, Clinical data systems. Part 3. Development and evaluation, Lancet 344 (1994) 1682—1688.
[58] E. Ammenwerth, U. Mansmann, C. Mahler, M. Kandert, R. Eichstädter, Are quantitative methods sufficient to show why wards react differently to computer-based nursing documentation? in: G. Surjan, R. Engelbrecht, P. McNair (Eds.), Proceedings of the 17th International Congress of the European Federation for Medical Informatics (Medical Informatics Europe 2002—Health Data in the Information Society), 25–29 August 2002, Budapest, Studies in Health Technology and Informatics, vol. 90, IOS Press, Amsterdam, 2002, pp. 377–381.
[59] H. Heathfield, P. Hudson, S. Kay, L. Mackay, T. Marley, L. Nicholson, V. Peel, R. Roberts, J. Williams, Issues in the multi-disciplinary assessment of healthcare information systems, Ass. Healthcare Inf. Technol. People 12 (3) (1999) 253—275.
[60] T. Jorgensen, Measuring effects, in: E.M.S.J. van Gennip, J.L. Talmon (Eds.), Assessment and Evaluation of Information Technologies, Studies in Health Technology and Informatics, vol. 17, IOS Press, Amsterdam, 1995, pp. 99–109.
[61] J. Frechtling, User-Friendly Handbook for Mixed Method Evaluation, http://www.ehr.nsf.gov/ehr/rec/pubs/nsf97-153/start.htm (last access: July 2003).
[62] N. Dewan, N. Lorenzi, Behavioral health information systems: evaluating readiness and user acceptance, MD Comput. 17 (4) (2000) 50—52.
[63] E. Ammenwerth, R. Eichstädter, T. Happek, R. Haux, B. Hoppe, M. Kandert, A. Kutscha, G. Luther, C. Mahler, U. Mansmann, U. Pohl, Evaluation of computer-based documentation on four wards—final report (in German), Department of Medical Informatics, Heidelberg, 2001.
[64] N. Denzin, Strategies of multiple triangulation, in: N. Denzing (Ed.), The Research Act, McGraw-Hill, New York, 1970, pp. 297–331.
[65] F. Grémy, J. Fessler, M. Bonnin, Information systems evaluation and subjectivity, Int. J. Med. Inf. 56 (1999) 13—23.
[66] J. Wyatt, D. Spiegelhalter, Field trials of medical decision-aids: potential problems and solutions, in: P. Clayton (Ed.), Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care, McGraw-Hill, New York, 1992, pp. 3–7.
[67] J. Talmon, A. Hasmann, Medical informatics as a discipline at the beginning of the 21st century, Methods Inf. Med. 41 (2002) 4—7.
[68] Y. Shahar, Medical informatics: between science and engineering, between academia and industry, Methods Inf. Med. 41 (2002) 8—11.
[69] F. Grémy, P. Degoulet, Assessment of health information technology: which questions for which systems? Proposal for a taxonomy, Med. Inf. (Lond.) 18 (3) (1993) 185—193.
[70] M. Johnston, K. Langton, R. Haynes, A. Mathieu, Effects of computer-based clinical decision support systems on clinician performance and patient outcome—a critical appraisal of research, Ann. Intern. Med. 120 (1994) 135—142.
[71] B. Kaplan, Evaluating informatics applications—clinical decision support systems literature review, Int. J. Med. Inf. 64 (2001) 15—37.
[72] E.A. Balas, S.M. Austin, J.A. Mitchell, B.G. Ewigman, K.D. Bopp, G.D. Brown, The clinical value of computerized information services. A review of 98 randomized clinical trials, Arch. Fam. Med. 5 (1996) 27—278.
[73] R. Walton, S. Dovey, E. Harvey, N. Frreemantle, Computer support for determining drug dose: systematic review and meta-analysis, BMJ 318 (1999) 984—990.
[74] Y. Chan, IT value: the great divide between qualitative and quantitative and individual and organizational measures, J. Manage. Inf. Syst. 16 (4) (2000) 225—261.
[75] N. Doherty, L. Mcaulay, The Relationship Between the ex ante and ex post information systems evaluation: reflections from the literature and the practice of evaluating e-commerce investments, in: A. Brown, D. Remenyi (Eds.), Proceedings of the Ninth European Conference on Information Technology Evaluation (ECITE 2002), Université Paris-Dauphine, 15–16 July 2002, MCIL, Reading, 2002, pp. 129–136.
[76] K. Grundén, An evaluation model for CSCW systems, in: A. Brown, D. Remenyi (Eds.), Proceedings of the Ninth European Conference on Information Technology Evaluation (ECITE 2002), Université Paris-Dauphine, 15–16 July 2002, MCIL, Reading, 2002, pp. 181–187.
[77] N. Shaw, 'CHEATS': a generic information communication technology (ICT) evaluation framework, Comput. Biol. Med. 32 (2002) 200—209.
[78] A. Grant, I. Plante, F. Leblanc, The TEAM methodology for the evaluation of information systems in biomedicine, Comput. Biol. Med. 32 (3) (2002) 195—207.
[79] C. Ohmann, G. Belenky. Leitfaden zur Evaluierung von Wissensbasen des MEDWIS-Arbeitskreises ''Evaluation'', http://www.uni-duesseldorf.de/www/medfak/theochir/ak_eval/lf_main.htm (last access: July 2003).
[80] F. Eurlings, A. van Asten, H. Cozijn, K. Klaassen, R. Stokman, R. van Valkenburg, E. van Gennip, Effects of a nursing information system in 5 Dutch hospitals, in: U. Gerdin, M. Tallberg, P. Wainwright (Eds.), Nursing Informatics—The Impact of Nursing Knowledge on Health Care Informatics, Studies in Health Technology and Informatics, vol. 46, IOS Press, Amsterdam, 1997, pp. 50–55.
[81] S. Sapirie, Assessing health information systems, in: T. Lippeveld, R. Sauerborn, C. Bodard (Eds.), Design and Implementation of Health Information Systems, World Health Organization, Geneva, 2000.
[82] M. Hebert, Telehealth success: evaluation framework development, in: V. Patel, R. Rogers, R. Haux (Eds.), Proceedings of the 10th World Congress on Medical Informatics (MedInfo 2001), Studies in Health Technology and Informatics, vol. 84, IOS Press, Amsterdam, 2001, pp. 1145–1149.
[83] B. Kaplan, P. Brennan, A. Dowling, C. Friedman, V. Peel, Toward an informatics research agenda: key people and organizational issues, JAMIA 8 (3) (2001) 235—241.