MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

An inventory of evaluation studies of information technology in health care: Trends in evaluation research 1982 - 2002

Elske Ammenwerth[a], Nicolette de Keizer[b]

[a] *UMIT - University for Health Informatics and Technology Tyrol, Research Group Assessment of Health Information Systems, Innsbruck, Austria*

[b] *Department of Medical Informatics, Academic Medical Center, University of Amsterdam, The Netherlands*

Abstract

During the last years the significance of evaluation studies as well as the interest in adequate methods and approaches for evaluation has grown in medical informatics. In order to put this discussion into a historical context, we conducted a systematic overview on trends in evaluation research of information technology in health care 1982 – 2002.

The inventory is based on a systematic literature search in PubMed. Each of the found 1.035 papers from 1982 – 2002 was indexed based on a taxonomy coding type of information system, clinical domain, research strategy, evaluation methods, setting and evaluation criteria.

We found interesting developments in evaluation research in the last 20 years. For example, there has been a shift from medical journals to medical informatics journals. With regard to methods, explanative research and quantitative methods have dominated evaluation studies in the last 20 years. From 1982 to today, the number of lab studies and technical aspects has declined, while the number of studies focusing on process quality or outcome quality of patient care has increased. Based on our results, we are inclined to talk of a kind of maturation of evaluation studies in medical informatics research.

Keywords

Evaluation studies, technology assessment, information technology, health care, health information systems, inventory

Introduction

During the last years the significance of evaluation studies has grown in medical informatics. Evaluation studies are more and more considered as integral part of the planning, introduction and operation of information technology in health care.

Evaluation can be defined as the act of measuring or exploring some property of a system, the result of which informs a decision concerning that system in a specific context [1]. Evaluation of health information systems has to deal with the actors (the people), the artifacts (the technology), the environment in which it is implemented as well as with their interaction.

Some discussion is going on in medical informatics about the best methods and approaches to be used, the need for an evaluation framework, and the often considered insufficient quality of evaluation studies. Some of those topics have already been discussed for some years now (see e.g. in [2], [3] or [4]).

In order to put this discussion into the historical context of evaluation research, we conducted a systematic overview on evaluation studies of information technology in health care from 1982 – 2002. The aim of this overview was to get insight into trends of evaluation research in the last 20 years.

In this paper, we will present some selected results of this inventory of studies. We will answer the following questions:

1. How does the number of published studies develop?
2. From which countries do the studies come from?
3. In which journals have they been published?
4. Which types of information systems have been evaluated?
5. In which clinical domains did the studies take place?
6. Which research strategy has been chosen?
7. Which evaluation methods have been used?
8. In which setting did the studies take place?
9. Which aspects have been evaluated?

The inventory of evaluation studies described in this paper is freely available on the internet, to help people planning evaluations and to find comparable evaluations.

Methods

The inventory is based on a systematic literature search in PubMed. This PubMed search was done on April 25th 2003; an update was done on July 28th 2003.

PubMed Search Strategy

We searched for evaluation studies of health information systems. We defined an evaluation study as the systematic, empirical assessment of a component of a health information system. We defined a health information system as comprising all computer-based components which are used by health care professionals or the patient themselves in the context of inpatient or outpatient patient care.

We decided to exclude medical-technical components (such as robotics or virtual reality systems) and all systems which are only used to analyze images or signals (e.g. automatic image analysis system). We also excluded all computer-based training and education systems for health care professionals since these

are not part of direct patient care. For the same reason we did not include administrative systems. Telemedical systems were only included when there was a clear indication of computer-based data transmission (not just use of videoconferencing tools).

The first step in the search strategy in PubMed comprised a selection of papers from 1982 to 2002 based on the following three queries which were combined by "AND":

1. Search for health information systems by searching in title words (e.g. computer, documentation, program), in Major Mesh Heading (medical informatics), and in Minor Mesh Heading (e.g. computers, decision-support systems, hospital information systems).
2. Search for evaluation studies by searching in title words (e.g. impact, effect, evaluation), publication type (e.g. clinical trial, evaluation studies), or Minor Mesh Heading (e.g. feasibility study, costs and cost analysis).
3. Search only for papers with an abstract.

The amount of selected papers was then reduced by conducting the following steps:

1. Automatically exclude all papers which had a certain MeSH heading such as e.g. animals, DNA, plants, robotics, neurosurgery, epidemiology etc. Resulting papers were stored in a database.
2. Manually exclude papers by checking title and abstract for inclusion and exclusion criteria. This was done by the two authors. Any differences were solved by discussion.

The complete and rather complex query is available upon request. We checked the precision of our query by checking whether the papers cited in other review papers were found by our query.

Taxonomy

The following taxonomy was used to classify each evaluation study:

- Type of information system which was evaluated: patient information system (PIS), order entry system (CPOE), radiology information system (RIS), PACS, lab information system (LIS), pharmaceutical information system (PHARM), operation management system (OP), anesthesia documentation system (ANAST), patient data management system (PDMS), nursing information system (NIS), GP information system (GP), teleconsultation system (TC), telemonitoring system (TM), expert system (XPS), other or general clinical information system (CIS).
- Clinical domain where study took place: at the patient's home, at a general practitioner, in an oupatient unit, in an intensive care unit, in a normal inpatient care unit, in a lab or pharmacy, in a radiology department, transinstitutional, others or unclear.
- Research strategy: explorative, explanative, review, mixed or unclear.
- Methods used: more quantitative, more qualitative, mixed or unclear.
- Setting: field study, lab study, mixed or unclear.
- Evaluation aspects: software or hardware quality, computer knowledge/attitudes, information quality, costs of information system, user satisfaction, usage patterns; efficiency of care, appropriateness of care, organizational aspects; quality of care, costs of care, patient satisfaction with care, patient-related behavior. More than one aspects could be chosen.

Results

The first step in the PubMed search resulted in about 45.000 hits between 1982 and 2002. After excluding all irrelevant papers (step 2 and 3), we found 1.035 papers which met our inclusion criteria.

While checking the quality of our query, we found precision rates of 80% and higher. For example, from the 107 papers cited in [5], 40 were eligible for our study (the others e.g. had no abstracts, or did not meet our inclusion criteria). From those 40 papers, 32 (= 80%) were found in our selection. Similar rates (> 80%) were found by checking the papers cited in [6], [7] and [8].

We will now present selected results of the analysis.

Number of published studies

The overall number of published studies is steadily increasing (Fig. 1). The percentage of evaluation studies compared to the overall number of papers with the Major Mesh Term "medical informatics" nearly doubled from 0.6% in 1982-1984 to 1.0% in 2000-2002.

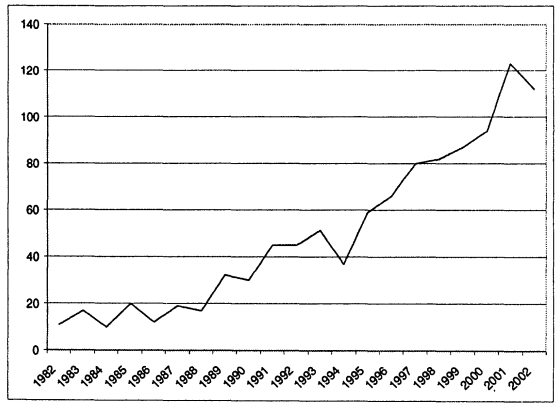

Figure 1 - Overall number of published studies 1982-2002.

Origin of first author

More than half of the first authors (53%) came from North America (mostly USA), about one third (34%) from Europe (Fig. 2). The dominance of papers from the USA was stable over the last 20 years. 93% of all papers were in English, being the dominant language for the last 20 years (the lowest percentage of English papers was 80% in 1992).

Distribution of journals

Half of the studies (n=509) have been published in 21 different journals in the last 20 years (each of them containing at least 9 publications). The number of papers in those journals is given in Table 1.

Not all journals exist since 1982. Some of the 'journals' are in fact proceedings. A detailed analysis of the journals showed that some journals were dominant in the 1980th but lost their influence later (e.g. Am J Health Syst Pharm and Med Care), while others only came up in the 1990th (e.g. J Am Med Inform Assoc and J Telemed Telecare).

Table 1: Journal distribution or evaluation studies 1982-2002.

Journal	No. paper	Perc.
Proc AMIA Symp	98	9,5%
J Telemed Telecare	75	7,2%
J Am Med Inform Assoc	35	3,4%
J Digit Imaging	35	3,4%
Am J Health Syst Pharm	26	2,5%
Bmj	26	2,5%
Telemed J E Health	25	2,4%
Int J Med Inf	23	2,2%
Comput Inform Nurs	21	2,0%
Methods Inf Med	20	1,9%
Medinfo	17	1,6%
Comput Methods Programs Biomed	14	1,4%
J Med Syst	12	1,2%
Jama	12	1,2%
J Biomed Inform	11	1,1%
Med Care	11	1,1%
Hum Pathol	10	1,0%
Med Inform Internet Med	10	1,0%
Stud Health Technol Inform	10	1,0%
Diabetes Care	9	0,9%
Radiology	9	0,9%
Other journals	526	51,5%

An interesting development can be found when classifying the journals either as „medical informatics" or as „medical or other" journals. The number of studies which are published in medical informatics journals has risen from 10% in the early 1980th to about 60% in mid-1990th and is around 50% nowadays (Fig. 3)

Evaluated information systems

Mostly, expert systems, teleconsultation systems and general clinical information systems (e.g. EPR) have been evaluated; they cover about 60% of all studies (Fig. 4).

An analysis over the last 20 years shows mostly stable percentages for the different types with two exceptions: In the 1980th, the evaluation of expert systems (XPS) counted for around 50% of all evaluation studies. Since then, their percentage has fallen to less then 20% today. Secondly, in the 1990th, teleconsultation systems (e.g. teleradiology) first came up. Their percentage today is around 30% of all evaluation studies.

Clinical domain

Most studies (around 29%) took place in normal inpatient care (e.g. on a ward). Around 19% were transinstitutional studies (mostly evaluations of telemedical applications). 13% took place in an intensive care or emergency unit, 10% at a general practitioner, and 8% in an outpatient unit. The other studies (around 20%) took place in other areas such as the patient's home or in a laboratory. The number of transinstitutional studies has risen in the last years (up to 25% today), that of studies in normal inpatient care units fallen (but it is still around 23% of all studies). .

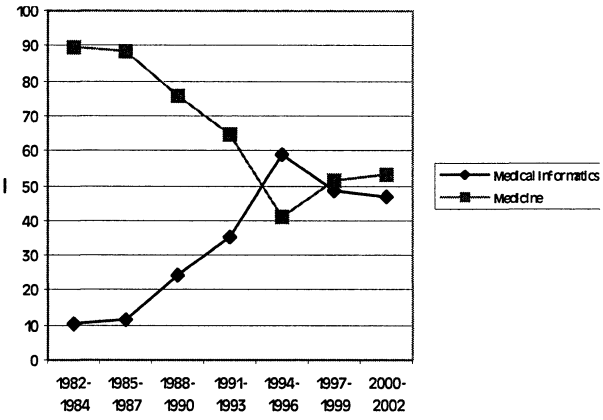

Figure 2 - Publication in medical informatics and in medical or other journals 1982-2002.

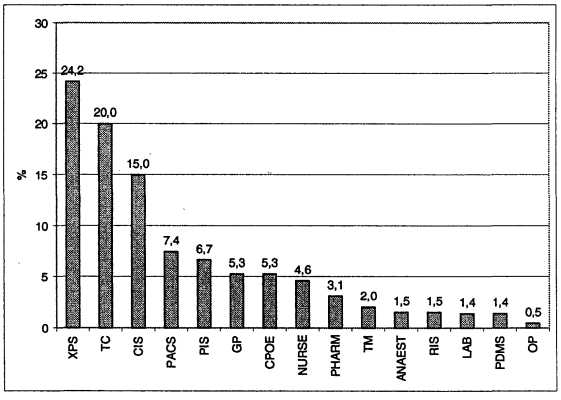

Figure 4 - Evaluated information systems 1982-2002.

Research Strategy

Most of the studies were explanative studies, i.e. they test a predefined hypothesis (e.g. "IT increases the quality or efficiency of care"). Only less then 3% were explorative studies, i.e. they explore an unknown area and try to generate hypotheses (e.g. "Which effects occurred after the introduction of IT?"). The number of reviews is slowly rising and is today at almost 9% of all publications (Fig. 5).

Evaluation methods

Throughout the last 20 years, quantitative evaluation methods (e.g. time measurements, quantitative user acceptance measurements, length of stay measurements, error rate scores) dominated evaluation studies (Fig. 6). Qualitative methods were mostly used in explorative studies (about 60% of all explorative studies applied qualitative methods), but seldom in explanative studies (about 1%). A detailed analysis showed that qualitative methods were often applied when the organizational and social impacts of IT were evaluated (35% of those studies used qualitative methods).

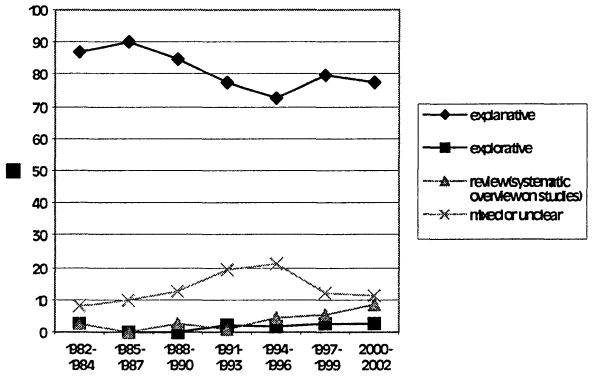

Figure 5 - Evaluation research strategy 1982-2002.

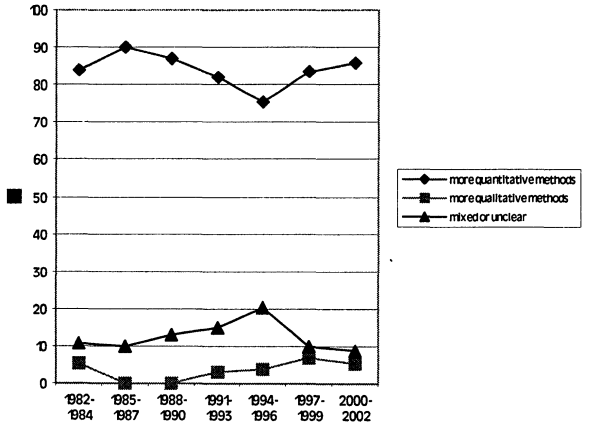

Figure 6 - Evaluation methods 1982-2002.

Study settings

About 75% of all evaluation studies were field studies, about 22% were lab studies. The number of field studies has slowly been rising in the last years.

Lab studies were rather often used for the evaluation of expert systems (about 50% of such studies were lab studies). Lab studies were seldom used in explorative studies (< 5%), but more often in explanative studies (25%).

Evaluation aspects

The most chosen evaluation aspects in the last 20 years were appropriateness of care (31% of all studies), efficiency of work processes (24%), user satisfaction (22%), and software quality (20%). Effects on quality of patient care were only considered in 7% of all studies.

A detailed analysis shows that the percentage of software quality fell from 44% in the late 1980th to 12% today. During this time, hardware quality rose up to 25% of all studies, mostly as part of studies of telemedical systems. Since the early 1990th, appropriateness of care staid continuously high and is now the most evaluated criteria (36% of all studies).

When we add up all studies which evaluate the outcome quality of patient care (that is, quality and costs of patient care, patient satisfaction and patient behavior), we can see that their percentage rose from around 15% in the 1980th to 35% today.

Criteria for outcome quality of patient care were rather often used during the evaluation of patient information systems (around 22% of those studies) and for telemonitoring systems (29%). Hardware and software quality were often evaluated in lab studies (36 and 62% of those studies), while outcome quality of patient care were nearly never evaluated in lab studies.

Discussion

There have been a few comparable literature reviews in the last years. For example, Moorman [9] analyzed publications on Electronic Patient Records between 1991 – 2002 with regard to journal, origin and MeSH term. Sawyer [10] analyzed publications on information systems between 1990 and 2001 with regard to evaluation construct (information, technology, or people), evaluation level (team or institution), and evaluation research strategy (experimental/explanative, intensive/explorative, or computational). However, this research did not concentrate on IT in health care, but on IS literature in general.

Our systematic overview is better comparable to the work of van der Loo [11]. He analyzed 108 evaluation studies with regard to type of system, study design, data collection, economic evaluation type, and type of effect measure. Our analysis differs in the following ways: We used a more detailed list of information system type. We used an extended list of evaluation criteria, in order to allow a more detailed analysis. We added information on research strategy and methods used. We did not analyze study design, data collection methods, or quality of the study, as the information in the abstracts is often insufficient of this. We found a larger number of studies then van der Loo, as we used broader inclusion criteria (e.g. including also lab studies), And we focused more strongly on the trends in evaluation research in the last 20 years than on absolute numbers.

We cannot be sure that we found all available studies in our search. The possible variation of terms in title, abstract or MeSH-headings is much too diverse to cover all possible combinations. We tried to check the precision of our search by taking other review articles as gold standard. Here we found satisfying rates of 80% and more. However, the limit to PubMed may have produced a bias with regard to U.S. publications. As a next step, other databases will have to be included to extend the inventory.

Our inventory concentrated on some attributes of evaluation studies which can be described only based on the abstract. We did not try to describe study design, study methods or results in detail, or to check the quality of studies, as this was not the aim, and as this would have required the analysis of the full paper.

As evaluation and medical informatics terminology has changed over the years, we found it difficult to retrieve older evaluation studies from PubMed. We tried to take this into account by using a rather broad number of terms to search for evaluation studies. However, this may nevertheless have lead to an underestimation of studies especially in the 1980th. The drop in the number of evaluation studies we found in 2002 may be caused by a still incomplete documentation of 2002-studies in PubMed.

The classification of information systems we used was developed while doing the inventory. As definitions and terminology in medical informatics may have changed over the last 20 years, we invested some energy to develop and properly define and test the categories.

We found interesting developments in evaluation research in the last 20 years. For example, there has been a shift from medical or other journals to medical informatics journals. Then, the number of systematic reviews is steadily rising. In addition, the evaluation of expert systems lost significance, while that of telemedical systems rose. The number of field studies is also steadily rising.

On the other side, we also found rather stable trends. For example, explanative research and quantitative methods dominated evaluation studies in the last 20 years. Explorative studies with more qualitative methods were mostly used for the evaluation of organizational and social impacts of IT.

However, it must also be taken into account that we only searched for studies in PubMed which is dominated by journals from the positivistic or objectivistic research tradition. Thus, we may find a larger number of explorative or qualitative studies when we include other sources such as literature databases from social science such as the Social Science Index. Our next step will thus be to extent our review to further databases and to check the trends we have detected so far.

We must also take into account that the taxonomy we used does not always allow a clear classification of studies using only information from abstracts. For example, qualitative and quantitative methods are more extremes on a continuum of used methods than clear-cut classification criteria.

Rigby [12] wrote that the focus of evaluation of an information system is changing during his life cycle: While during the implementation phase, the focus of evaluation is more on technical aspects, it shifts to impacts on patient care and then to impact on the overall organization during routine use of an information system. We found similar developments in medical informatics: From 1982 to today, the number of lab studies and technical aspects has declined, while studies focusing on process quality or outcome quality of patient care have increased. Thus we are inclined to talk of a kind of maturation of evaluation research in medical informatics.

Conclusion

There have been some shifts as well as some stable trends in evaluated information systems and publication organs in the last 20 years. There seems to be a slow maturation in evaluation research of IT in health care, with rising numbers of field studies on process or outcome of patient care.

In order to help other researchers to search for evaluation studies, and to reproduce our review, we developed a web-based interface to allow searching our database. It is available for free at:

http://evaldb.umit.at.

Acknowledgement

We thank Frieda Kaiser and Megan Stepushyn for their valuable contribution to the website and the analysis.

References

[1] Declaration of Innsbruck, Results from the ESF Exploratory Workshop on Systematic Evaluation of Health Information Systems, 4. - 6. April 2003, Innsbruck. In preparation.

[2] Anderson JG, Aydin CE, and Jay SJ, Eds.: Evaluating Health Care Information Systems - *Methods and Applications.* London, New Delhi, Sage Publications. 1994.

[3] van Gennip E, and Talmon J, Eds.: Assessment and evaluation of information technologies in medicine vol. 17. *Studies in Health Technology and Informatics.* Amsterdam, IOS Press. 1995.

[4] Friedman C, and Wyatt JC. *Evaluation Methods in Medical Informatics*, New York: Springer. 1997.

[5] Balas EA, Austin SM, Mitchell JA, Ewigman BG, Bopp KD, and Brown GD. The clinical value of computerized information services. A review of 98 randomized clinical trials. *Arch Fam Med* 1996: 5: 271-8.

[6] Kaplan B. Evaluating informatics applications - clinical decision support systems literature review. *Int J Med Inf* 2001: 64: 15-37.Roine R, Ohinmaa A, and Hailey D. Assessing telemedicine: a systematic review of the literature. *Canad Med Assoc J* 2001: 165(6): 765-71.

[7] Friedman C. Is Medical Informatics a mature science? *Int J Med Inform* 2003: 69: 261-272.

[8] Moormann P, and van der Lei J. An Inventory of Publications on computer-based medical records: an update. *Method Inf Med* 2003: 42: 199-202.

[9] [Sawyer S, and Chen T. Conceptualizing Information Technology in the Study of Information Systems: Trends and Issues. In: Myers M, Whitley E, Wynn E, and DeGross J, eds. *Global and Organizational Discourse about Information Technology.* London: Kluwer, 2002; pp. 1-23.

[10]van der Loo R. Overview of Published Assessment and Evaluation Studies. In: van Gennip EMSJ, and Talmon JS, eds. *Assessment and evaluation of information technologies.* Amsterdam: IOS Press, 1995; pp. 261-82.

[11]Rigby M. Health informatics as a tool to improve quality in non-acute care - new opportunities and a matching need for

a new evaluation paradigm. *Int J Med Inform* 1999: 56: 141-150.

Address for correspondence

Elske Ammenwerth,

Research Group Assessment of Health Information Systems, UMIT - University for Health Informatics and Technology Tyrol, Innrain 98, 6020 Innsbruck, Austria,

elske.ammenwerth@umit.at