Evaluation and Implementation: A Call for Action

J. L. Talmon
Department of Medical Informatics, Maastricht University, Maastricht

Summary

Objectives: To raise awareness for actions that are urgently needed to accompany the large scale implementations of ICT in Health Care that are currently taking place in many countries around the world.
Methods: An analysis of a few studies that have recently been described in the literature guided by recent suggestions for research and development of evaluation of health ICT.
Results and Conclusion: Six specific recommendations for action are specified:

- Development of good implementation practice,
- Development of an experience base of implementation of ICT in health care,
- Setting up a surveillance system for unintended effects,
- Build an evidence base of best evaluation practice,
- Developing guidelines for proper reporting of evaluation studies,
- Education of clinicians and decision makers.

Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2006. Methods Inf Med 2006; 45 Suppl 1: S11-5.

Keywords

Medical informatics, medical informatics evaluation, assessement

Introduction: A Brief History

Evaluation has been a key topic within specific domains of Medical Informatics since the seventies. It has its roots in Medical Decision Making: evaluation of the performance/predictive value of (statistical) decision algorithms was done on independent data to achieve credibility. A famous example of such type of studies was the multi-center before-after study by Adams et al of the acute abdominal pain system [1].
Evaluation studies took also place within the domain of signal analysis. The European Study on Common Standards for Quantitative Electrocardiography (CSE project) compared the performance of specific parts of various ECG processing systems using a common set of electrocardiograms [2]. This study was one of the first to apply a Delphi-like technique to obtain a reference library.
These early studies dealt with computer artifacts that were reasonably well defined. The abdominal pain system used Bayes' rule to compute a posteriori probabilities for a small number of diagnostic categories. It used a standardized form as input for patient related data. The evaluation dealt with how accurate the system could predict the true diagnosis. The CSE study dealt also with a rather well described topic. Computer algorithms for the detection of begin- and endpoints of the various wave components in the ECG were compared. Here again the algorithms themselves may have a certain complexity, but the data one has to deal with is well described and limited in extent.
In the early eighties small and large expert systems were developed when Artificial Intelligence techniques became available through various toolkits. The nature of these systems posed a challenge to evaluation. Systems like QMR and Dxplain that have knowledge about several hundred diseases, taking into account a few thousand of signs, symptoms and diagnostic tests are much more difficult to evaluate. It is not easy to have a test set of cases that covers the complete domain of such systems; and even when a large set of cases is used, only a fraction of the underlying knowledge base of the system may have been tested. Early evaluation studies on such systems used only a few dozen cases. In the nineties comparative studies on such systems were performed [3]. But also in these studies, a little more than one hundred cases were used. This is in sharp contrast with the 16.000+ cases that were used in Adams et al's study while the abdominal pain system tested had only a dozen or so different diagnostic categories.
In the early nineties, awareness arose that evaluation of decision support systems in a broad sense needed attention. Guidelines were needed. Various international working groups and EU-funded concerted actions addressed the topic. Gradually the notion emerged that not only systems that support diagnosis should be evaluated but any system that is used in clinical practice [4-6].

In the mean time various stories on both successful and unsuccessful implementations of computer systems in health care surfaced. The books by Lorenzi et al have been instrumental to raise awareness that there is more than the technical artifact that plays a role in successful implementation of ICT in health care [7, 8].

These developments widened the scope of evaluation to a broad activity that dealt with many facets in various phases of the ICT artifact's life-cycle. Jørgensen [9] and later Stoop and Berg [10] identified several domains that could be subject of evaluation: the Technical, Clinical/professional, Organizational, Economic, Ethical and Legal domain.

Also the notion that evaluation was an ongoing activity was becoming more evident. Clarke et al. [11] suggested that evaluation should take place during each phase of the life-cycle of an ICT artifact – in their case a decision support system. This approach has been further elaborated and generalized by Brender et al. [12] for application in system integration.

In recent years there has been a further paradigm shift towards a more holistic view on information systems and their evaluation. Kaplan and Shaw [13] as well as Berg et al. [14] have advocated the socio-technical approach towards evaluation: try to understand why information systems are a success or a failure, taking into account the social context in which the systems are used. This wide scope is also evident in the "declaration of Innsbruck" [15] where an information system is defined as the technical artifact *and* the environment (social, organizational) in which it is used.

What has been sketched before is the development of evaluation as a science, but the question is whether this thinking has lead to a better take up of these issues in evaluation studies as published in the scientific literature. Ammenwerth and de Keizer [16] have analyzed the literature over a period of 20 years. They have noted a steady increase in the number of publications dealing with evaluation studies. In their conclusions they state that over the years the percentage of publications of evaluations in clinical settings has increased: mature systems are being evaluated in practice, but quantitative methods are still used in most of the studies, which is an indication that the socio-technical aspects are not yet well taken up.

In 2005 a wake-up call has been made. Four publications have shown us that evaluation has to be considered serious business. We will discuss these publications in more detail and draw our conclusions for future developments.

Selected Readings

Public Value

The United Kingdom is investing large amounts of money in their National Programme for IT in the NHS (NPfIT). Bend questioned whether clear benefits have been delivered by this programme [17]. He analyzed 40 different evaluation studies for evidence of the value of the NPfIT for the public. He didn't find convincing evidence that the program really contributed to "improving value for money", the satisfaction with the services of the NHS, improvements of the health of the patients and of the public and to an increase in the level of trust in the NHS. Bend was concerned about the potential inadequacy of the evaluation studies performed for two reasons: a) evaluation is a means to learn from experiences and b) building up evidence will enable decision makers, professionals and patients to assess whether spendings on ICT in health care have been – and will be – effective.

He identified four reasons why the evidence was not as solid as hoped for: a) too little time and too few resources allocated for the evaluation, b) unclear objectives of the evaluation, c) weak methodologies, including starting to evaluate when it is already too late to get proper baseline values and d) poor reporting by e.g. mixing personal opinion with conclusions from data and unclear distinction between realized benefits and anticipated benefits.

On the other hand, Bend realized that there are perhaps real problems in the way the technology is being used. Expected benefits may not have been realized at all. He identifies a number of barriers, but here again; there is no solid evidence that these barriers really were the culprit.

Medication Errors

Two reports of the US Institute of Medicine have been instrumental in the increased awareness that ICT might help in reducing medical errors [18, 19]. Various systems have been developed that assist physicians in prescribing medication. Koppel et al [20] studied not the reduction in medication errors due to the use of a computerized physician order entry (CPOE) system, but addressed specifically prescription errors that were due to or aggravated by the use of the system. They identified 22 situations in which the system increased the probability of prescription errors, some of these situations were quite common (observed by more than 50% of the users) and/or occurred frequently (weekly or more often).

There has been a special section in the Journal of Biomedical Informatics were

several authors have taken a stance on the Koppel et al's paper [21-24]. Some of these authors argued that you should see Koppel's study into a proper perspective and that the observations were not necessarily generalizable. There was agreement, however, that the recommendations of Koppel et al where useful and that you should not take an ICT system for granted, even when its proper technical functioning had been determined. You should be aware of the "side effects" and pay continuous attention to its performance as to further improve patient safety.

The Electronic Patient Record (EPR)

Implementation of an information system is not easy. A study of 8000 ICT projects among more than 300 US companies show that more than 50% of ICT projects fail one way or the other [25]. It is unlikely that in health care the situation is different. Scott et al [26] report on Kaiser Permanente's experience with implementing an electronic patient record system. They identify seven key findings that can be used as warning signs for others that are also considering implementation or modification of an EPR system. All issues boil down to the socio-technical context in which the implementation takes place. Poor alignment of the workforce with the decision to implement a specific EPR, technical problems and delays in delivery, reduction in clinician's productivity were all factors that negatively affected the social culture of the organization. All observations of this study are in line with the results of a Delphi-study on success and failure criteria which was conducted after the EFMI special topic conference on The Contribution of ICT to Health Care [27]. Some implementations of ICT in health care will fail; that will be unavoidable. We should though be able to learn from failures as to avoid the old ones and to make new ones in the future.

CPOE Take 2

Another paper that made the headlines of newspapers is the study of Han et al [28]. They expected that the implementation of a commercial CPOE system would increase the quality of care at a neonatal intensive care unit. This expectation was based, among other reports, on the "To err is human" report of the Institute of Medicine [18] and the recommendations of the Leapfrog group (www.leapfroggroup.org).
Their study showed that after implementation of the CPOE system, the mortality rate increased from 2.8% to 6.6%. This seems to be a dramatic development. However, one should keep in mind that while introducing the CPOE system, working procedures were changed as well. The paper has several indications that the increase in mortality is at least partly explainable by these changes in working procedures. This is in line with the observation that introduction of CPOE systems in other organizations did not have such dramatic effects.

Where to Go from here

It is clear from all examples given above that implementing ICT in health care is not a guarantee for improved quality of care. Implementing ICT is a delicate process that may disrupt well-functioning organizations. Without being fully aware that implementing ICT is a change process – and of course acting accordingly – disaster may be around the corner. The disaster may not be as dramatic in as in the neonatal intensive care case, but disruption of service inherently may have negative effects on patient care and hence on the health of the patients.
The examples described above could have been parts of the book by Tenner, named "Why Things Bite Back: Technology and the Revenge of Unintended Consequences." [29]. In this book numerous examples are given of unintended consequences of technology. It covers a wide range of issues related to health care, nature, business etc. The main theme is that whenever you intervene in a system with a certain purpose, the intervention will have consequences – and perhaps even achieve the opposite result – that were not foreseen. Another theme is that whenever we intervene, there will be the need for more vigilance with respect to preconditions for proper working of the intervention. A good example for electronic prescription would be the dosage schemes. With handwritten prescriptions, there is always the pharmacist who has the responsibility of checking the proper dosage and when in doubt he could/should contact the prescribing physician. However, in electronic systems the underlying database is often taken for granted. Hence electronic prescriptions are likely to be believed at face value. Errors in that database may go unnoticed for a long time, specifically when they occur in records of less frequently used drugs. Only when we expect the unexpected and pay close attention to what happens during and after implementation, we are able to safely introduce ICT in health care. Evaluation is an integral part of this process: it documents the (unintended) effects, it informs the decision makers and provides a frame of reference against which the effect of new interventions/developments can be compared.

The following issues should be high on the agenda of those dealing with the implementation of ICT in health care and with the evaluation of the (un)intended effects.

Development of Good Implementation Practice

Although economic arguments have been used to motivate the use of ICT in the Health Care industry, the driving forces in recent years have been the concerns about the quality of care [18, 19]. Hence ICT that supports care processes in health care has to be considered as contributing to continuous health care quality improvement. This requires that the implementation of the ICT itself has to be subject to continuous quality improvement. The development of "good implementation guidelines" as well as the close observation and evaluation of implementation processes become a prerequisite. The socio-technical approach should not only try to determine whether an implementation was successful or not. It should also contribute to developing the theory and good practice for successful implementations.

It seems that at the moment one is not learning from past experience. This does not hold only for the health care industry, it holds for many sectors in our society in relation to the implementation of ICT. We should remember the words of Einstein that "Insanity is doing the same thing over and over again and expecting different results". We have to observe what we are doing and discern what works from that what doesn't. Only then we are able to identify successful approaches and areas where further development and experimentation is necessary.

Development of an Experience Base of Implementation of ICT in Health Care

The development of guidelines for good implementation practice should be based on and accompanied by an experience base, in which various implementations are documented. It should contain both successful implementations as well as (near) failures. Only from failures we can learn. Unfortunately, people are reluctant to report failures. For those responsible for the implementation, it is easy to blame the users or the organization for having poorly organized the implementation. But when these issues are not documented it is likely that others – and even the same organization later on – may make the same mistakes. There is too much at stake to take these risks. Setting up a public database with experiences would significantly contribute to better practice in the future.

Setting up a Surveillance System for Unintended Effects

Not all unintended effects may become evident during implementation of the ICT systems in practice. Only after longer usage, unintended effects may become apparent. Setting up a register of unintended effects of information systems in health care – taken in the sense as defined in the declaration of Innsbruck [15] – would contribute to a safer usage of ICT health care.

One might argue that health care is an industry and that in such environments one does not publish experiences from which the competition can learn. This is a false argument since the health care industry is a risky industry. The medical profession has implemented mechanisms to report errors and to improve procedures. This will improve patient care over time.

Build an Evidence Base of Best Evaluation Practice

By widening the scope of evaluation from the more technical or performance point of view towards a comprehensive approach, potentially dealing with many dimensions, has made evaluation a more daunting task. Not only knowledge and experience from Health Informatics, Computer Science or Medicine – in particular epidemiology and clinical trials – has to be taken into account, but also methods and techniques from the social sciences, economics, operational research and psychology have to be used. Setting up an evidence base of good evaluation practice, including studies with an excellent design and execution and an overview of both quantitative and qualitative methods for evaluation will make it easier to develop and execute proper evaluation studies that address the relevant questions, using the proper methods.

Developing Guidelines for Proper Reporting of Evaluation Studies

The editorial office of our Medical Informatics Journal is often confronted with manuscripts that poorly describe the studies that have been performed. There is a need for improvement in the reporting of evaluation studies as to augment access to study results for meta-analysis with respect to the effects of ICT applications in health care. Much of our evidence is fragmented. The reports of (clinical) studies are in many circumstances of sub-optimal quality [30]. The CONSORT statement, as developed for proper reporting of randomized clinical trials and its subsequent expansions for other types of studies may serve as an example on how to develop guidelines for good reporting of evaluation studies [31].

Education of Clinicians and Decision Makers

Unsupported promises in the past have contributed to the false perception that ICT can solve any problem. The solution has "only to be put in software" and all our troubles are over. Software engineering is still not yet that far developed that developed software can be taken to be error-free. Also implementing an ICT application is more than installing some hardware, installing the software and off-you-go. It is an intervention in a work process and a social organization that has developed over years and that has built-in mechanisms to avoid and correct errors. Implementation of ICT may destabilize such a system. Both decision makers and clinicians have to be aware of these issues. There is still an educational role to be played. Our evidence base can help in this education; not only by demonstrating when things can go wrong but also by showing that when proper measures are taken and implementation processes are executed with care and attention, unintended negative consequences may be tamed and health care can indeed be improved.

References

1. Adams ID, Chan M, Clifford PC, Cooke WM, Dallos V, de Dombal FT, et al. Computer aided diagnosis of acute abdominal pain: a multicentre study. Br Med J (Clin Res Ed) 1986; 293(6550): 800-4.
2. Willems JL. Common standards for quantitative electrocardiography. J Med Eng Technol 1985; 9(5): 209-17.
3. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994; 330(25): 1792-6.
4. Talmon J, Fox J, editors. Knowledge Based Systems in Medicine: Methods, Applications and Evaluation. Berlin: Springer; 1991.
5. van Gennip E, Talmon J, editors. Assessment and Evaluation of Information Technologies in Medicine. Amsterdam: IOS Press; 1995.
6. Talmon J, Enning J, Castaneda G, Eurlings F, Hoyer D, Nykanen P, et al. The VATAM guidelines. Int J Med Inform 1999; 56(1-3): 107-15.
7. Lorenzi NM, Riley RT. Organizational Aspects of Health Informatics: Managing Technological Change. New York: Springer-Verlag; 1995.
8. Lorenzi NM, Riley RT, Ball MJ, Douglas JV. Transforming Health Care through Information: Case Studies. New York: Springer-Verlag; 1995.
9. Jørgensen T. Measuring effect. In: van Gennip E, Talmon J, editors. Assessment and Evaluation of Information Technologies in Meicine. Amsterdam: IOS Press; 1995. p. 99-109.
10. Stoop AP, Berg M. Integrating quantitative and qualitative methods in patient care information system evaluation: guidance for the organizational decision maker. Methods Inf Med 2003; 42(4): 458-62.
11. Clarke K, O'Moore R, Smeets R, Talmon J, Brender J, McNair P, et al. A methodology for evaluation of knowledge-based systems in medicine. Artif Intell Med 1994; 6(2): 107-21.
12. Brender J, Talmon J, Nykanen P, McNair P, Demeester M, Beuscart R. On the Evaluation of System Integration. In: van Gennip E, Talmon J, editors. Assessment and Evaluation of Information Technologies in Medicine. Amsterdam: IOS Press; 1995. p. 189-208.
13. Kaplan B, Shaw NT. Future directions in evaluation research: people, organizational, and social issues. Methods Inf Med 2004; 43(3): 215-31.
14. Berg M, Aarts J, van der Lei J. ICT in health care: sociotechnical approaches. Methods Inf Med 2003; 42(4): 297-301.
15. Ammenwerth E, Brender J, Nykanen P, Prokosch HU, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inform 2004; 73(6): 479-91.
16. Ammenwerth E, de Keizer N. An inventory of evaluation studies of information technology in health care trends in evaluation research 1982-2002. Methods Inf Med 2005; 44(1): 44-56.
17. Bend J. Public Value and e-Health. In: Insitute for Public Policy Research; 2004.
18. Kohn L, Corrigan J, Donaldson M, editors. To err is human: Building a Safer Health System. Washington DC: National Academy Press; 2000.
19. Medicine Io. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academic Press; 2001.
20. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. Jama 2005; 293(10): 1197-203.
21. Shortliffe EH. CPOE and the facilitation of medication errors. J Biomed Inform 2005; 38(4): 257-8.
22. Bates DW. Computerized physician order entry and medication errors: finding a balance. J Biomed Inform 2005; 38(4): 259-61.
23. Nemeth C, Cook R. Hiding in plain sight: what Koppel et al. tell us about healthcare IT. J Biomed Inform 2005; 38(4): 262-3.
24. Horsky J, Zhang J, Patel VL. To err is not entirely human: complex technology and user cognition. J Biomed Inform 2005; 38(4): 264-6.
25. Briggs R, deVreede G-J, Nunamaker J, Sprague R. Special issue: Information systems success. Editorial. J Management Information Systems 2003; 19(4): 5-8.
26. Scott JT, Rundall TG, Vogt TM, Hsu J. Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study. Bmj 2005; 331(7528): 1313-6.
27. Brender J, Ammenwerth E, Nykanen P, Talmon J. Factors influencing success and failure of health informatics systems—a pilot Delphi study. Methods Inf Med 2006; 45(1): 125-36.
28. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics 2005; 116(6): 1506-12.
29. Tenner E. Why things bite back: technology and the revenge of unintended consequences. New York, NY: Vintage Books; 1997.
30. de Keizer N, Ammenwerth E. Trends in Evaluation Research 1982 - 2002: A Study on how the Quality of IT Evaluation Studies Develop. Stud Health Technol Inform 2005; 116: 581-6.
31. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. Jama 1996; 276(8): 637-9.

Correspondence to:
J.L. Talmon
Dept. of Medical Informatics
Maastricht University
PO Box 616
6200MD Maastricht
The Netherlands
E-mail: Talmon@mi.unimaas.nl