

STARE-HI – Statement on Reporting of Evaluation Studies in Health Informatics

Explanation and Elaboration

J. Brender¹; J. Talmon²; N. de Keizer³; P. Nykänen⁴; M. Rigby⁵; E. Ammenwerth⁶

¹ Department of Health Science and Technology, Aalborg University, and V-CHI, Aalborg, Denmark; ² School of Public Health and Primary Care – CAPHRI, Maastricht University, Maastricht, The Netherlands; ³ Department of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands; ⁴ School of Information Sciences, University of Tampere, Tampere, Finland; ⁵ School of Public Policy and Professional Practice, Keele University, Keele, United Kingdom; ⁶ Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

Keywords

Evaluation studie, guidelines, health information systems, informatics

Summary

Background: Improving the quality of reporting of evaluation studies in health informatics is an important requirement towards the vision of evidence-based health informatics. The STARE-HI – Statement on Reporting of Evaluation Studies in health informatics, published in 2009, provides guidelines on the elements to be contained in an evaluation study report.

Objectives: To elaborate on and provide a rationale for the principles of STARE-HI and to guide authors and readers of evaluation studies in health informatics by providing explanatory examples of reporting.

Methods: A group of methodologists, researchers and editors prepared the present elaboration of the STARE-HI statement and selected examples from the literature.

Results: The 35 STARE-HI items to be addressed in evaluation papers describing health informatics interventions are discussed one by one and each is extended with examples and elaborations.

Conclusion: The STARE-HI statement and this elaboration document should be helpful resources to improve reporting of both quantitative and qualitative evaluation studies. Evaluation manuscripts adhering to the principles will enable readers of such papers to better place the studies in a proper context and judge their validity and generalizability, and thus in turn optimize the exploitation of the evidence contained therein.

Limitations: This paper is based on experiences of a group of editors, reviewers, authors of systematic reviews and readers of the scientific literature. The applicability of the details of these principles has to evolve as a function of their use in practice.

Correspondence to:

Elske Ammenwerth
Institute of Medical Informatics,
UMIT – University for Health Sciences,
Medical Informatics and Technology,
Hall in Tyrol, Austria
email: Elske.ammenwerth@umit.at
Phone: +43 50 8648 3809

Appl Clin Inform 2013; 4: 331–358

DOI: 10.4338/ACI-2013-04-RA-0024

received: April 20, 2013

accepted: June 29, 2013

published: July 24, 2013

Citation: Brender J, Talmon J, de Keizer N, Nykänen P, Rigby M, Ammenwerth E. STARE-HI – Statement on Reporting of Evaluation Studies in Health Informatics: explanation and elaboration. *Appl Clin Inf* 2013; 4: 331–358

<http://dx.doi.org/10.4338/ACI-2013-04-RA-0024>

1. Introduction

Health informatics claims its place in supporting clinicians and others delivering health care and support, and in increasing safety and quality of care in its broadest sense including effectiveness, efficacy and efficiency [21]. However, it is also an expensive, and often disruptive, set of technologies [4, 7]. To justify its claims to provide benefits, to optimize use, and to moderate disbenefits, it is an ethical imperative that health informatics must adhere to the principles of justification and good practice as applied to all other interventions, and hence support these claims with evidence [35]. One important aspect of good practice is that health informatics applications should be rigorously evaluated to provide an evidence-base, and that these evaluation studies should be robust, standardized and comparable, and their results should be reported in line with agreed reporting standards [2].

We adopt the definition of evaluation as given in the Declaration of Innsbruck: “Evaluation is the act of measuring or exploring properties of a health information system (in planning, development, implementation, or operation), the result of which informs a decision to be made concerning that system in a specific context.” [2, p. 480]. Following the information systems science approach, a health information system comprises not only the technical artefact (software and hardware) but also the environment (including all actors involved, as well as their interactions) in which it is implemented [43].

As Hansson [19, p. 257] noted: “The role of non-epistemic values in the evaluation of scientific information should, as far as possible, be limited to determining the level of evidence required for various types of practical decisions.” The essence interpreted in our context is that one should carefully consider the level of evidence required when writing a scientific report or journal paper, because that leads to the decision on what information to include and what not. One writes in order to bring a message to an audience. Consequently, an evaluation report should fully inform the intended audience so that they may confidently identify and apply the lessons from the study. It is this perspective that is foundational for the STARE-HI statement [38].

The STARE-HI statement specifies which items should be contained in a publication of a health informatics evaluation study in order to enable others to judge the trustworthiness of a study’s establishment, its design, its execution and line of reasoning, and the validity of the conclusion, as well as its context and thus the potential for generalizability. Table I briefly lists these items and serves as both as a checklist as well as a guide for the remainder of this paper.

The objective of this elaboration paper is to provide elaborations and examples for each principle of the STARE-HI statement. This should support the authors of evaluation studies in applying STARE-HI. Also readers of evaluation papers are supported by these elaborations because they make clear why each item may be of relevance for the interpretation of the study and its results.

Most if not all STARE-HI recommendations hold for either quantitative as well as qualitative evaluations or mixed methods approaches. Furthermore, adherence to the STARE-HI statement does not automatically imply that the quality of the evaluation study is good. Completeness of reporting is no guarantee for study quality, but helps the reader to better assess the quality (and relevance to them) of the research performed. Achievement of good evaluation study quality requires careful planning and design of the evaluation study; this is supported by the Good Evaluation Practice Guideline (GEP-HI, [30]) that complements the STARE-HI Statement.

The structure and aim of this elaboration paper is similar to the elaboration papers of other guidelines on how to report study results, such as CONSORT for RCT studies [1, 28] and STROBE for observational studies in epidemiology [41].

2. Method

2.1 Founding principles and values

Both STARE-HI and this explanatory paper are based on the following principles:

- the reporting must be at a level of detail that enables the (qualified) reader to judge whether or not the design, the outcome and the derived conclusions are valid;

- the reporting must contain adequate information to enable the reader to understand the system in its healthcare context, and thus be able to judge the relevance and applicability of its evaluation to their own situation (section “System Details and System in Use”);
- the reporting must provide the level of detail sufficient for others to exclude uncertainty about the approach, methods, and metrics applied, thereby also enabling them to use the study design in another setting (section “Methods”);
- the reporting must provide the level of detail sufficient for others to follow the interpretation of the data into information and knowledge, the line of reasoning within this process, and the conclusion based on the results (section “Results”);
- the reporting must include a discussion of the pro’s and con’s of the study results to an extent demonstrating the authors’ insight into the literature as well as their understanding of generalizability, strengths and weaknesses of the contribution, thereby putting the information and knowledge gain into broader perspectives including open questions that require future research (section “Discussion”);
- the reporting must summarize the most significant pieces of information and knowledge gained, while indicating their level of trustworthiness, enabling the reader to judge whether the contribution has interest for him/her (section “Conclusion”).

2.3 Approach

The approach for preparing the present elaborated guidelines is identical with that of preparing the core STARE-HI statement as described in [38]. In short, a group of methodologists, researchers and editors developed this STARE-HI elaboration paper based on the literature and on their experience as researchers and reviewers. Examples (good and bad) were collected from the authors’ collection of literature.

For each STARE-HI principle, we repeat the brief description of the principle and provide examples from the literature. We believe that bad examples also present good learning cases, and thus, we present both good and bad examples. Our positive examples from the literature are referenced to give the authors credit for their recommendable style of reporting. After the example, the rationale for the principle is given as well as a further elaboration.

3. STARE-HI principles and their elaboration¹

The following chapter is organized according to the STARE-HI elements with each item and sub-item serving as headings and sub-headings in this section. The major structure is summarized in Table I that may well serve the reader as a check list (► Table 1).

3.1 Title

The title should give a clear indication of the type of evaluated system and the study question as well as the study design.

Examples

- ☉ CPOE: Its effect on Adverse Drug Events, a field study (*constructed example*)
- ☉ A retrospective record analysis to assess the effect of a CPOE system on medication errors (*constructed example*)
- ☉ Assessment of effects of health information systems (*constructed example*)

¹ Our examples – when referenced – may include references and footnotes as they appear in the original text. These references and footnotes will not be included in the reference list or as footnotes in this elaboration paper.

Elaboration

The purpose of the title is to enable readers to judge whether the paper is relevant for him/her. It is the main element for a reader's first screening of the relevancy. Based on the title, readers may decide to look at the abstract or the full paper for more details.

The use of the term "evaluation" (or "assessment" or "study") helps to detect evaluation studies. But also other terms such as "effect" or "impact" (as in the first and second example) can clearly indicate an evaluation study.

The title should then provide key information that enables readers to identify the subject of the study before reading the paper. It should make clear what has been studied about which type of system by which method. The last example does not clearly describe the type of information system evaluated e.g. anesthesia or PACS, nor does it specify the type of effects.

The title also facilitates those who make systematic reviews to quickly identify relevant papers among the often large set of hits from e.g. a PubMed search. A descriptive title will reduce the chance that the paper is erroneously omitted.

3.2 Abstract

The abstract must clearly describe the objective, setting, participants, measures, study design, major results, and conclusions.

Examples

- ⊙ "Objective: To determine use, appreciation and effectiveness of ... Design: A prospective study. ... Setting: Department of a tertiary health care centre in ... Participants: ... cancer patients, ...hospital physicians, ... Main outcome measures: Actual use of the system ... Patients' appreciation of Capability to detect potential problems ... Results: The system was used by 36 H&N cancer patients, 10 hospital physicians ... The total number of patient-sessions was 982. Patients appreciated the system highly, rating it 8.0 on a 10-point scale.... Conclusion: The ... system was used The system enabled the early detection ICT can play an additional role in the management of patients ..." [40, p. 839]
- ⊙ "This paper presents the design and implementation of a clinical picture archiving and communication system (PACS) module ... Preliminary evaluation based on formal survey and usage statistics shows that the system is rapidly being accepted by radiologists and clinicians for the review and processing of digital radiographic images." (anonymized example)

Elaboration

The purpose of an abstract is to provide the reader with the decision-making basis for whether or not to read the article in more detail. A structured abstract with a series of headings as in the first example enables the reader to access the relevant information quickly. A structured abstract is currently required by many journals but is seldom required for conference papers. There is evidence that structured abstracts provide more information than unstructured abstracts [37]. This is possibly due to the fact that a structure forces the author to list all the relevant items. An abstract should never provide additional information compared to the main text of a paper.

In the second example, no details of results are presented ("rapidly accepted" is not defined, and usage statistics may not indicate acceptance so much as compulsion).

3.3 Keywords

Among the keywords should be "evaluation" and keywords describing the type of system evaluated, the setting, outcome measures, and study design.

Examples

- ⊙ Evaluation; medical order entry system; Inpatients; Usability; Cognitive walkthrough; Think-Aloud
- ⊙ Usability evaluation, Information Systems

Elaboration

Author-provided keywords are used directly in literature databases to index papers or as a starting point for the indexing by the database provider. Good keywords hence enhance the retrievability of papers, and consequently also the chances of being cited by others. In this regard, the first example is better than the second, because it gives details of the type of evaluated systems and of the applied evaluation approaches.

MeSH terms [29] are used in many literature databases to index papers. Hence keywords based on MeSH terms list are preferred where available. Unfortunately, only a few MeSH terms are relevant for evaluation papers, and hence it is useful to extend the keyword list with terms that are relevant for indexing the report, but which do not yet appear in MeSH. In the first example, the first three terms are MeSH terms. In the second example “usability evaluation” is a compound of two terms, while “Information Systems” is too broad.

3.4 Introduction

The introduction should provide the reader with the necessary background for the rest of the paper.

3.4.1 Scientific background

Description of what is already known about the (type of) intervention that is the object of study, what are still open research questions, and why there is a need to answer them.

Examples

- ⊙ “... Besides, in these studies, the relation between hypothesized, actual usability problems and the occurrence and frequency of errors that result from suboptimal user interfaces is explored [19,26]. These studies have shown that cognitive analysis methods are not only suited to characterize user-interaction flaws in a system but also to identify (the opportunities for) errors attributable to these usability problems. ...” [33, p. 158]
- ⊙ “At ... University of Technology, research about usability has been performed during several years. Especially usability issues in the medical technology domain have been addressed (e.g. 6-10). To contribute to a deeper understanding of the effects of user expertise on the results of usability tests, the present study has focused on investigating users’ interaction with a simple user interface of an insulin pump.” (anonymized example)

Elaboration

It is important to provide an objective statement of the scientific background. This usually includes critical analysis of existing theories and prior evidence on the topic of the study. Included should be a brief account of the authors’ previous work with respect to the method applied, or the system evaluated.

The first example above provides a good example. It demonstrates what the current knowledge base is and what the power of certain research paradigms is in this area. Still, it is primarily synthesizing the pro’s. So, a way to improve the above example would be to also include arguments against the approach taken, and reasons for still choosing it.

The second example has limitations, because it only specifies the track record of the organization, but not how those studies have informed the current study, and how they built on existing theories and evidence.

3.4.2 Rationale for the study

Short description of the motivation for the study; stakeholders and actors

Examples

- ⊙ “If the OM/Lab pilot implementation would prove to be a success, the system would be implemented at other clinical departments of the AMS. We were asked to evaluate the prototype OM/Lab system.” [33, p. 159]
- ⊙ “However, despite CPOE’s ability to reduce medication errors, a few investigators have begun to question whether CPOE implementation necessarily results in improved outcome and have

raised concerns regarding the Leapfrog Group's CPOE directive.⁵ Some have proposed that under certain circumstances, CPOE may actually foster "unintended consequences"⁶ ..."[18, p. 1506-07].

- © The first author served throughout the purchase project in an action-research case role as (external) consultant ... The authors have full freedom to publish, provided that patient security issues and business secrets are kept confidential. This has been achieved without consequence for the conclusion. (*constructed example*)

Elaboration

To be better able to assess the chosen approach of a study, the selection of measures and the interpretation of the results, it is necessary to know why the study was done. It makes a difference whether the results of a study are primarily to inform a local decision or whether it is to gain insights that have a wider application. In the first case one would not expect to see any statement on the generalizability of the study in the paper, while in the second case, not having such a discussion – including the limitations – would be an omission in the reporting.

In our view, a study that is done just to measure something without a decision to be made based on these measurements is a potential waste of resources and in certain circumstances even unethical [2].

Another important component is the role of the various actors in a study and their relations with stakeholders in respect to the system and its evaluation study. A description should include all interests that may influence the evaluation because of the potential risks of lack of impartiality that the reader ought to know. The discussion and the conclusions of a study may be influenced and hence biased when those who do the evaluation study a) are members of the organizational unit in which the study takes place, b) were decision makers on aspects that will be scrutinized by the evaluation study, c) were actually developers of the system to be evaluated [8, p. 260]. The last example suggests that the researchers were independent of the organization. If in such a publication only the positive aspects are dealt with without a thorough discussion of the limitations of the study and no report of unintended consequences, one may question their freedom to publish their findings.

3.4.3 Objectives of the study

The specific study questions and hypotheses, accompanied by permissions obtained in relation to the study.

Examples

- © "The main objective of this evaluation was to assess whether the OM/Lab system complied with the user requirements and to reveal potential usability flaws in the system. Besides these aims we were also interested in the relation of the usability problems detected in the user testing sessions with end-users' order behaviour in terms of omissions, factual errors in orders and cancelled orders." [33, p. 159]
- © "In light of reemerging uncertainty and discussion regarding the impact that CPOE might have on patient outcome, we examined mortality rates.... before/after CPOE implementation, testing the hypothesis that patient outcome would improve after this intervention." [18, p. 1507]
- © Approval for the study design was obtained from the institutional research board of the participating hospital" [32, p. 803].

Elaboration

An explicit statement of the objectives of the study (as in the first example) or the hypotheses that will be tested (as in the second example) allows the reader to assess to what extent the study aligns with current theories of the domain, to assess how appropriate the study design has been, and to later assess the results in the light of the original research questions.

Many journals demand a statement on approval of the study by an Institutional Review Board (IRB) or Ethics Committee has to be provided (see the third example). It has to be clear that there has been oversight that the participants in the study were given proper protection for their participation. Even though some evaluation studies that e.g. assess user acceptance and implementation processes may not provide potential harm for patients, the users of the system should be free to ex-

press their thoughts on the system without any pressure. In some countries only medical research affecting patients is regulated, while in other countries any research with human subjects is regulated. When IRB approval is legally not required even though human subjects are involved, a statement to that fact should be made..

Note that statements on IRB approval could also be included in the methods section. It is advisable to consult recent issues of the target journal to assess what the most appropriate place is.

3.5 Study context

This section describes the organizational environment of the study. It is sometimes part of the Methods or Methods & Materials section. We have separated it for ease of comprehension, but urge the reader to follow the requirements of the specific journal to which they intend to submit.

Information on study context is important for the later assessment of generalizability of results, and hence for others to assess whether the approach or results are applicable for their purpose as well. So, any conditions or constraints that may influence the study results should be reported. A challenge for authors is to find the right level of detail.

3.5.1 Organizational setting

The name, location and kind of health care facility and involved departments.

Examples

- ☉ “CHP is a 235-bed regional pediatric referral center with ~12000 annual admissions...”[18, p. 1507]
- ☉ “The study was conducted in four clinical units in two Australian metropolitan public teaching hospitals within the same city and under the same administrative structure. The choice of hospitals for the study was based on their long-term use of the same hospital-wide, mandatory CPOE system for all inpatients.” [10, p. 644]

Elaboration

The organizational and health system context determines implementation and usage of ICT systems. Study results in one kind of health system context may not be transferable to other health system contexts. For instance, a tertiary referral hospital is different in organizational and cultural context from a local general hospital. Also, for instance, universal access publicly funded systems will have different requirements and resources than a private or fee-funded setting. Finally, the results of a study in the context of a developing country have different meaning than obtained from a similar study in a better resourced country. Thus, some details of the involved organization (as in the first example) and where needed an indication of the country (as in the second example) should be provided.

3.5.2 System details and system in use

A description that enables the reader to understand how the system works (or is intended to work) and its phase in the system's life cycle.

Examples

- ☉ “Prior to the study, we developed an electronic health information support system for ...cancer patients ...The system was designed to: 1. facilitate communication between all involved health care providers and between health care providers and patients; 2. provide information to health care providers and patients; 3. facilitate contact with fellow sufferers; 4. facilitate the early detection of patient problems by means of monitoring. Access to the functions ‘communication’ and ‘monitoring’ was restricted to authorised users only, whereas the functions ‘information’ and ‘contact with fellow sufferers’ were readily accessible to anyone with access to the Internet. For an extensive description of this system, ... see: [10]” [40, p. 840]
- ☉ “The CPOE system (PowerOrders; Cerner [a member of Leapfrog Group], Kansas City, Missouri) that was purchased ... is a commercially sold ... application... Approximately 3 months before CPOE implementation all hospital health care personnel were trained Hospital wide

implementation of CHP's CPOE system (along with its clinical applications platform) occurred over a 6-day period, reaching full operation by October 29, 2002. ... Physician orders are entered primarily through selecting from various order "menus" and "sub-menus" that require completion of requisite fields before orders are accepted. For example, ... Confirmation of this selection then opens a series of sub-menus that request specific fields to be filled: ..." [18, p. 1507].

Elaboration

The main features should be described as to give the reader an understanding of how much functionality is provided. This information helps the reader to understand how the system may be integrated in clinical practice. Still, the main purpose of an evaluation study is not for description of the system. A good balance has to be found between the level of detail and the space available. Also, the way the system has been introduced in the environment where it was intended to be used and how long it has been used has to be described as they may be important factors influencing the outcome of the study.

The first example illustrates a level of detail on functionality that is needed to judge causal mechanisms leading to the results and conclusions. In the first example it is also clear that the system was developed by the authors' organization. It would have been even better to state whether it was the authors themselves that designed the system. The second example illustrates the reporting of the implementation process (including training). This information may turn out to be important for the interpretation of the results.

In addition to the fairly elaborate recommendations in [38, p. 5], a number of issues should be emphasized:

One or two screenshots may provide the reader with a feeling of the system, though space limitations may prevent the inclusion of these figures. One should assess whether these screenshots are useful for understanding the properties of the system that is evaluated or not. If not it is better to leave them out, whereas if they explain (some of) the findings, they should be included.

A reference should be made to another publication where a detailed system description can be found (see the first example). Since journals may currently allow on-line only supplementary material, a detailed system description may be included in such a format as well.

The system description should explicitly address how the system performs in a real-world setting or how the system is intended to work (see first example). Differences in workflow often explain differences in findings. Without information on the real use, the reader cannot judge comparability and generalizability. It is important to clarify how the system is used. For instance, outlining the clinical process (e.g. the ordering process, see the second example), and by highlighting by whom and how the system is used for what. Even the same type of systems may be used in completely different ways (e.g. a CPOE medication system may transfer data directly to the pharmacy without nurses' involvement, or nurses may check the orders). If necessary, one can describe how the system is assumed to be used, and then describe the actual use in the results section and differences in the discussion section.

Where relevant for the interpretation of the findings, information should be provided on the type of technology used (e.g. availability of bedside terminals for each bed, a laptop on a trolley, or tablets may influence the results in a study on bedside chart review). It would also better support comparison among different studies using different hardware solutions.

It is relevant to describe the phase of the life cycle of the system in which the evaluation took place (such as development/prototyping, implementation in practice/early use, or routine use). If the ICT system was implemented only shortly before the evaluation there has been no time for the use of the system to become mature and stable, and then it may be relevant also to provide information on the training level of the users (see second example).

Finally, any special actions that were taken (or omitted) for the study that deviate from normal procedures should be mentioned, e.g. (additional) training, parallel operation on the ICT system and the old paper-based system, or additional technical equipment or support.

3.6 Methods

This section should contain sufficient information as to allow others to perform the same kind of study in another environment. It not only describes the study design and the study methods, but should also provide the rationale for the chosen study design and methods. Any information that contributes to the understanding of the choice of study design and methods should be given here.

It is important that the presentation is factual and objective, and that perils and pitfalls are taken properly into account.

3.6.1 Study design

The overall study design and the arguments for choosing it.

Examples

- ⊙ “In this prospective cohort study, a comparison was made between the occurrences of errors in the medication ordering process before and after implementation of a CPOE system in the PCCU.” [34, p. 60]
- ⊙ “We performed a quantitative and qualitative study incorporating structured interviews with house staff, pharmacists, nurses, nurse-managers, attending physicians, and information technology managers; real-time observations of house staff writing orders, nurses charting medications, and hospital pharmacists reviewing orders; focus groups with house staff; and written questionnaires administered to house staff. Qualitative research was iterative and interactive (ie, interview responses generated new focus group questions; focus group responses targeted issues for observations).” [25, p. 1198]
- ⊙ “We randomised GP practices rather than individual patients to prevent contamination. In this clustered design at any single practice, all GPs used either teledermatology or standard procedures (control group). Using dedicated randomisation software, practices were assigned to teledermatology or standard care.” [13, p. 559]

Elaboration

This subsection describes the overall approach taken for the study. It should specify the nature of the study design (experimental or observational, qualitative or quantitative) as well as more detailed information. Presenting reasons for any decisions regarding study design (such as resource limitations preventing a randomized controlled trial) is vital for the accurate interpretation of the study outcome (see example). It is permissible to explicitly plan with a limited scope or a reduced level of ambition, but the reader should be informed about the reasons for that choice.

Choosing the right terminology for the description of the design is of importance for the reader and for later systematic reviews. To describe the study designs of quasi-experimental studies, see [20].

An essential part of the description of the study design is the definition of the frame of reference for the interpretation of the results. Assessment of the reference situation can be part of the study design (e.g. measured in a control group, or in a before-after study as in example 1), but benchmark, historical data or results of other studies can be used as well.

The authors must demonstrate their awareness of specific and potential study design biases and how they have taken these into account by the design. For this purpose see the review of biases in [8, pp. 243-323].

3.6.2 Theoretical background

Theories – with appropriate references – on which the study is based and that guided the selection of the measurement instruments used.

Examples

- ⊙ “In this study we sought to identify the determinants of overall CIS user satisfaction. Building on the TAM and the DeLone & McLean Information System Success models [6,7], we focused on the following five dimensions: user characteristics, user satisfaction, use, system quality, perceived usefulness, and service quality.” [31, p. 614]

- © “Data collection and quantitative content analyses were guided by the constant comparative method of grounded theory methodology [27]. We used this methodology because it facilitates the understanding of a phenomenon that has not previously been studied (i.e. the conceptual osteoporosis tool), and it enables the exploration of the ways in which the “reality” of the tool is socially constructed [29], particularly to clarify how the tool might be used at the point of care in real practice settings.” [24, p. 6]

Elaboration

Certain studies build upon theories that have been developed and tested either in health care or in other domains. Whenever a study is based on such theories, it is important to describe that theoretical background (such as TAM model in example 1 or the grounded theory in example 2) and provide the relevant references. Examples of formal theories founding a study method would be the use of socio-technical theory behind e.g. a system’s analysis method, and the Delphi approach for predicting specific aspects of the future. The study results should then also be interpreted against those theories and theoretical models. In both examples above, references are given to where details on the methods and theories can be found.

3.6.3 Participants

Methods of selection of participating users, patients, units, hospitals, etc, including if applicable inclusion and exclusion criteria.

Examples

- © “All persons admitted to the medical, surgical (including subspecialty surgical services), neurology, and obstetrics and gynecology services between September 1997 and April 1998, whose admission and discharge were within the boundaries of 4 consecutive 2-month periods were included in the study. Admission periods did not overlap.” [11, p. 2841]
- © “All physicians at 3 of the sites (6 physicians at each of the 2 sites and 5 at the third) participated in the study; 6 of 27 physicians at the fourth site were randomly selected to participate and agreed to do so. They were not blinded to the study purpose. Once the study began, a seventh physician was added at the fourth site to augment the total number of patients. All physicians were board-certified internists with a mean age range from 39 to 46.” [16, p. 838]

Elaboration

How participants in the study are included should be described in an explicit and comprehensible way, as they may affect study feasibility, interpretation and transferability of the results. Exclusion criteria can relate for instance to patients’ characteristics (such as exclusion of those with co-morbidity), or to staff (for instance, exclusion of temporary staff) – both these examples might make the study smoother but significantly less valuable by excluding important parts of the real world. The results of a power analysis have to be reported where relevant.

When it comes to different levels of units to be included (hospitals, departments, physicians, etc.) inclusion and exclusion criteria should be specified for each level. The second example reports in detail the composition in the various sites, but how the sites were selected is not described – at least not in this extract. In the paper this should be addressed.

Participants can be of many professional categories, and even within one category there may be huge variances in apprenticeship (e.g. from registrar to consultant to chief physician), in specialization (e.g. from pediatrics to geriatrics), and in computer literacy. The report should be clear on how the study has dealt with this kind of differences when one group of professionals is compared with another group of professionals. In the second example, this is described in terms of age and certification.

Any measures taken to get a proper selection of different groups or matching between groups should be specified. The results section should in any case include a description of the relevant characteristics of all participants in the study.

The authors’ awareness of confounding factors in terms of inter- and intra-subject variability has to be demonstrated where relevant. Randomization is generally a good choice for experimental studies, but it is an illusion to think that it can compensate for inequalities within a small population.

Somewhere it should be specified whether compensation is provided (financially or in time) for the effort to participate in the study since it can influence the interpretation of the results. This can be done here, or in relation with the IRB approval statement. It could even be part of the acknowledgement section. It will depend on the guidelines of the target journal what the best place is for such a statement.

3.6.4 Study flow

Details on date of beginning and end of the overall study and any study periods with clear descriptions of intervention.

Examples

- ⊙ “The performance of the decision support application was measured during 4 consecutive 6-week study periods. The first and third study periods were control periods consisting of usual computerized order entry. During control periods, a highlighted dose and frequency was standard for all adult patients, The decision support application was activated for the second and fourth periods. All orders ... on eligible patients were recorded in a log file at the time of ordering.” [32, p. 803]).
- ⊙ “Data were collected before CPOE implementation for a 2-month period from October 4, 2001, to December 4, 2001. There was a 1-month period when no data were collected to allow for CPOE implementation and training of all attendings, fellows, residents, and staff. Post-CPOE data collection then occurred for a 2-month period from January 4, 2002, to March 4, 2002.” [34, p. 60]

Elaboration

The timing of the various phases of a study can have a considerable impact on the outcome of the study. All relevant considerations to take this into account have to be clarified in the study description. Both examples above show a textual description of study flow. In addition, a time-oriented flow chart or a graphical illustration may illustrate the time aspects better than words. Adding specific time instants and/or duration information in such illustrations may further enhance the understanding of the design of the study. Also, precise details should be given on the timing of interventions for every group considered in the study.

Because of the time-wise separation of the baseline and the effect phases, before-after studies may suffer from effects by a number of factors outside the control of the assessment study, simply caused by changes in the dynamic organizational environment of a healthcare facility. Any changes that may affect the outcome should be reported in this section. In the second example, it is made clear that between the two measurement periods, training took place. Hence a change in the outcome measures could also be attributable to the training, where good practice may be discussed rather than the use of the information system.

The second example describes a one-month implementation and training period. Such information needs to be complemented in the results and discussion sections with information on the experience of the users in using the system. For instance, information that a system is in use for more than two years may be of less relevance when the subjects studied are the newly recruited nurses one month after they started to work, compared with a case of a newly installed system that is evaluated one month after installation by nurses that have practiced for more than two years in the same hospital. Further, experience with a previous system may have an effect on how a new system is used. Another aspect that may require attention is possible seasonal effects.

3.6.5 Outcome measures or evaluation criteria

Description of outcome measures used or other evaluation variables of interest together with definitions of key concepts.

Examples

- ⊙ “This study focused on errors that occurred during the medication ordering process. An error was determined to have occurred when an order was found to be incomplete, incorrect, or inap-

appropriate at the time of physician ordering. Errors were classified as potential ADEs, medication prescribing errors (MPEs), or rule violations (RVs).” [34, p. 60]

- ⊙ “Questions included in the survey were designed to measure perceptions of individuals within an organization about the organization. Internal diffusion, the first of the dependent variables, was measured for each separate innovation using a scale developed by Zmud²⁶” [3, p. 105]
- ⊙ “The two data sets were analyzed using a thematic grounded theory approach⁴⁶ to derive themes that would explain the way physicians used the CPOE system.” [10, p. 645]

Elaboration

In quantitative trials, the dependent and independent variables that are measured should be defined clearly (see the second example). Often quite a few outcome measures will be measured, but not all will be reported in a publication. There could be many reasons for doing so, including a multi-faceted study of which only one or a few facets are reported. Still, it is useful to document all outcome measures that have been measured, at least to show that there has been no selectivity in result reporting. In the report, it should be documented why certain measures are not reported. Apart from a listing of the outcome measures, a proper, unambiguous definition for the measures should be given. For example, just stating that one measures medication prescription errors is insufficient. A proper definition of what medication prescription errors are should be included (see the first example).

For qualitative studies, the concept of outcome measures may not be appropriate. For these studies conducting e.g. focus groups and semi-structured interviews, the focus of these group discussions and interviews should be indicated (see the third example).

In any study that covers more than one unit, one has to address the issue of a) how one ensured that all units involved have the same understanding of the outcome measure, and b) that the methods used to measure the outcomes are aligned, such that data coming from those different units can indeed be pooled.

3.6.6 Methods for data acquisition and measurement

Provide sufficient detail on data acquisition and measurement such that others are able to assess the appropriateness and any limitations, as well as to be able to replicate the measurement procedures of the study.

Examples

- ⊙ “The survey instrument utilized the Stronge-Brodts Nurses’ Attitudes Toward Computers Questionnaire (6). This is a validated and reliable instrument that uses a 20-item Likert scale survey to assess attitudes in the areas of job security, legal ramifications, quality of patient care, capabilities of computers, willingness to use computers, and benefit to the institution (6, 7). The survey collected data specific to demographics, general computer experience, attitudes toward computers, and perceptions on time taken in nursing documentation.” [14, p. 121]
- ⊙ “Copies of all prescriptions written by participating physicians were collected daily from each site for a 4-week period. A pharmacist then screened up to 3 prescriptions at random ... per patient to identify possible prescribing errors. If a study pharmacist discovered a possible error on prescription review, 2 physicians independently reviewed the prescription, judged whether an error had occurred ... Interrater agreement for the presence of medication errors and ADEs was high (k=0.92; 95% confidence interval (CI): 0.88–0.96 ...).” [16, p. 838]
- ⊙ “The non-participant observations, relating to physicians’ test ordering and viewing work practices, were undertaken by two researchers over the four sites ... A letter outlining the study, its voluntary nature, the confidentiality of findings and participants, and a consent form, were provided to all participants. A case study protocol (45) was developed which provided a framework for the observations and interviews with general rules to be followed in the field. Pilot observations were undertaken which allowed the researchers to familiarize themselves with the sites and participants.” [10, p. 644]
- ⊙ “Three patient outcomes were measured using administrative data collected in the course of standard patient care. Length of stay was measured by ... Falls were identified through the hospital incident reporting system. Altered mental status was measured as part of ...” [32, p. 804]

Elaboration

The methods section needs to present the applied methods and instrument with a level of detail that - in principle - allows the reader to reproduce the study. The four examples show how authors can present basic information on the quantitative or qualitative methods they employed. The methods section also needs to give information on the validity and the fulfillment of assumptions for the application of the assessment methods, instruments, techniques, and metrics used. This should be made clear, e.g. by stating whether methods have been validated (with their proper references included, see the first example), or by describing how the internal and external validity of not yet validated methods has been assessed (e.g. by means of Cronbach's Alpha for questionnaires).

Nevertheless, one has to be careful in transferring validated methods from one environment to another. As an example, when translating a 'standard' questionnaire from the literature into another language for one's own application purposes, the original verification of the construct validity is no longer valid per se, because of the potential for cultural bias and change of the meaning of key terms. The same is the case when a questionnaire is used in a different social/cultural (or even professional) context, including the risk that the original phrasing is not properly understood.

Cultural assumptions are prevalent in many methodologies and methods and perhaps also in the systems evaluated. Failure to recognize this fact is called "ethnocentricity". As an example, in some cultures it is highly impolite to answer "No", so a questionnaire with the options 'yes', 'no' and 'don't know' will get a biased outcome. When relevant, it has to be clear from the reporting whether cultural issues have been explicitly addressed.

Quantitative outcome measures need to be well defined and verified with respect to their quality characteristics such as objectivity, reliability, or sensitivity (the degree to which the measure is able to detect changes in what is being measured). In the case of multiple investigators, clear definition of measures, standardized instruments, sufficient investigator training and assessment of interobserver reliability is an indication for objective and reliable investigations (see the second example).

Another specific pitfall is that authors sometimes unintentionally 'mis'-use the concept of 'experts' when wanting to certify a high level of quality, but without specifying the area or level of expertise for these 'experts'. For instance, when asking a group of 'experts' to rank a list of Adverse Drug Events according to severity, it is relevant to specify why and in which way this group was considered 'experts'.

3.6.7 Methods for data analysis

For quantitative data, state which statistical techniques were used for analysis. For qualitative data, indicate the analysis methods in detail. For all data analysis methods, indicate any software product used.

Examples

- ☉ "Differences between groups (before vs after ... and survivor vs nonsurvivor) were determined by Mann-Whitney rank sum test for continuous data and by 2 or Fischer's exact tests for categorical data. Data were analyzed using SPSS statistical software program (version 12.0; Chicago, IL)." [18, p. 1507]
- ☉ "The two data sets were analyzed using a thematic grounded theory approach⁴⁶ to derive themes that would explain the way physicians used the CPOE system. Categories were generated from line by line coding and were included in a memoing document that consisted of notes and reflections on the data that were recorded by the principal researcher (JC) during analysis. This constant comparative method of grounded theory analysis meant that the data were repeatedly studied, analyzed and re-analyzed. The analysis was inductive to allow the categories, themes and relationships to come from the data rather than being imposed prior to data collection." [10, p. 645]
- ☉ "Categories (from the used questionnaire) were collapsed according to frequency distributions. Microsoft Survey Pro was used to collate the responses. Analysis was performed using chi-square calculations with a p value of 0.05 regarded as significant" [14, p. 121]

Elaboration

Besides a detailed description of data acquisition, the methods section should also present the applied analysis methods with sufficient detail to allow reproduction, given that the raw data are available. The examples above show how basic information on data analysis in quantitative and qualitative trials can be presented.

The analysis/interpretation of data into information and further on to knowledge has many pitfalls, see [8, p. 243-323], most of them related to the assumptions for the analysis or for the validity of the data material analyzed, and hence important for the trustworthiness of the reported findings and conclusions. Therefore it is important that a clear description is given on how the data have been analyzed. In many cases only the analysis is described, but no account is given of whether and how the validity of analysis has been assessed.

One example of a frequent pitfall is the misuse of statistical techniques for the analysis of data where the data does not fulfill the precondition of the used statistical technique. Another example is the need to verify inter- and intra-person (or case) variability before data of sub-group are merges.

Full confidence may be achieved by the authors demonstrating an awareness of which pitfalls are relevant for the study (and therefore may lead to a bias in the outcome if not taken into account) by making an account of how they are handled and their potential impact on the study outcome (see [8, p. 243-323] for a full description of possible pitfalls).

In addition, a solid frame of reference is essential when authors claim that they are documenting the actual change or benefit. In this respect, it is important that the data and the methods used for data analysis should be similar to those used in the establishment of the frame of reference (e.g. in before-after studies). Some measures have an intangible frame of reference, for instance those of measures of usability and user acceptability. In such cases, the challenge is that of clarifying to the reader which criteria have been used for judging and concluding on the usability and user acceptability.

3.7 Results

All results of the study, including unplanned events and unexpected observations, must be provided with a level of detail sufficient for others to follow the interpretation of the data and the assessment of the validity of the conclusions.

3.7.1 Demographic and other study coverage data

Baseline demographic data and clinical characteristics of study participants (users, patients, and units) and of the study.

Examples

- ☉ “Of the 1202 patients whose prescriptions were collected, 661 (55%) completed the 2-week survey and 600 (50%) completed the 3-month survey. The 541 nonparticipants included 168 who declined to participate when contacted by telephone, 139 who opted-out by postcard, 205 who could not be contacted, 24 who had language or hearing problems, and 5 with other reasons for not participating. ... Of study participants, 65% were women, the mean age was 52 years, 92% spoke English as a primary language, and 81% were white.” [16, p. 838-39]
- ☉ “A total of 104 type 2 diabetes patients participated in the study, 56 patients in the SMS group and 48 patients in the control group (Figure 1). No significant differences in baseline characteristics were found between the two groups (Table 1). Patients in both groups were predominantly male and the average age was 55 years. Co-medication was comparable between the groups. At baseline, the refill adherence (number of days for which oral antidiabetics were dispensed during the eleven months preceding the intervention divided by the total number of days and multiplied by 100%) was 62% in both groups.” [42,]
- ☉ “Since the study compares organizations with differences in epidemiology, metrology and terminology these issues have been investigated and found to have no implication on the study outcome.” (*anonymized example*)

Elaboration

An overview of the demographics of the study population is necessary for the reader to make an assessment of how applicable the results may be for his/her own situation. It also demonstrates the structure of the case-mix, and any drop-outs that may influence the interpretation of outcome.

In addition to the general recommendation to provide demographics like age and sex (see the first example), one should list other characteristics that could have an influence on the outcomes of the study. This could be, for example, co-medications or patterns of behavior (see second example), or resources that participants may have had access to. When a group of users e.g. comprises several levels of apprenticeship (e.g. level of training or professional competence, but also prior computer literacy) information on the composition of the group with respect to these characteristics is necessary.

For studies that compare the results in two or more groups, the demographics should be presented for both groups (see the second example) and tests may be performed to identify any differences on the key parameters. The anonymous negative example above demonstrates awareness, but details are necessary to ensure trustworthiness.

Studies that relate to the use of systems should include facts describing the frequency of use, the kind of use as well as some volume parameters, for instance, the number of orders, number of documented items, number of care plans, etc.

In the case of alternative ways of achieving the objective of the system – e.g. paper and electronic – state the percentage of situations in which the system was used and whether usage was enforced or voluntary.

3.7.2 Unexpected events during the study

Any unforeseen events that may have influenced the study results or outcome.

Examples

- ⊙ “The national e-Medikation pilot project was influenced by strong political concerns from the side of the Chamber of Physicians. Around four months after start of the pilot project, the Chamber of Physicians initiated a two-month boycott of e-Medikation, and started activities to inform the public on possible risks of e-Medikation. During this time of boycott, the usage of e-Medikation by physicians and patients nearly ceased. After end of the boycott, the usage numbers increased again, but did never reach the earlier numbers during the remaining months of the pilot project.” (*constructed example after a real case*)

Elaboration

During a study, events may take place that may affect the outcome of the study. These events can take place within the study context (e.g. a promoter of a system being evaluated leaves the organization and his follow-up is opposing the system) or can be external to the study context (a change in hospital policy may reduce willingness to partake in the study and hence a low recruitment or a high drop-out, or organizational changes may suddenly require that medication can only be ordered for patients, not for local depots). This will have impact on workflow, how the system is used, user satisfaction and all kind of other parameters including complications and patient outcome. Any of such events that could influence the study and its outcome should be reported (see the example above) as to provide a context for the interpretation of the results.

3.7.3 Study findings and outcome data

Presenting the results of the study for each study question, for each outcome variable and evaluation criterion.

Examples

- ⊙ “The mean number of SP visits per participant at baseline was 1.34 visits (SD 5 0.48) and at follow-up it was 1.56 visits (SD 5 0.50). There was no statistically significant difference between intervention or control groups in number of visits“ [6, p. 174]
- ⊙ “Of 1879 prescriptions reviewed, 143 (7.6%; 95% CI 6.4% to 8.8%) contained a prescribing error. Of these, 62 (43%) represented potential ADEs, 3 led to preventable ADEs, and 78 (55%) were er-

rors with no potential for harm. The most frequent errors were incorrect or missing dose (n=77, 54%) or frequency (n=26, 18%; Table 2). “ [16, p. 839]

- © “Our qualitative and quantitative research identified 22 previously unexplored medication-error sources that users report to be facilitated by CPOE. We group these as (1) information errors generated by fragmentation of data and failure to integrate the hospital’s several computer and information systems and (2) human-machine interface flaws reflecting machine rules that do not correspond to work organization or usual behaviors.” [25, p. 1199]

Elaboration

This section should report the results of the study. It includes a display of the observations and the results of any analysis of the data. The three examples above show how basic results for quantitative (first two examples) and qualitative studies (last example), respectively can be presented. The interpretation of the findings should be postponed to the discussion section.

As in the second example, the author should always provide absolute numbers and not only relative numbers, because otherwise the reader cannot properly appreciate the differences found. For instance, the reader may take a different position when a 10% difference is found in two groups of 10 cases each as compared to the same difference in two groups of 250 cases.

It is preferred to present quantitative results in tabular and graphical form and to highlight the most striking findings in the text. Since tables and figures play an important role authors should pay close attention to their design so as to convey the results in an optimal way. In a scientific paper clarity of the presentation is preferred over an artistic approach. Readers have different cognitive styles: Some are better in reading tables, others prefer graphs. Colors should be used with caution; while this can make illustrations more clear, they may lose their meaning if photocopied or printed in monochrome. There are several books that give guidance on the best representation of certain types of data, e.g. [39].

Qualitative results can be presented in tabular form or in plain text (see the third example). It is good style to provide citations to increase credibility of findings.

3.7.4 Unexpected observations

Any unintended (positive or negative) side-effects of the system that were not in focus at the study.

Examples

- © “The proportion of desired responses ... was 57.2% ... in the intervention group and 13.5% ... in the control group ... However, the study was terminated early because of 4 unintended consequences identified among patients in the intervention group: a delay of treatment with trimethoprim-sulfamethoxazole in 2 patients and a delay of treatment with warfarin in another 2 patients.” [36, p. 1578]

Elaboration

It is necessary that the data and the observations presented give account for all observed phenomena. So as to be able to fully appreciate the results of the study, it is necessary to know the context in which the system is used and how it is intended to be used. This information is dealt with in section 3.5. In the current section any observations that are at variance with the context information should be reported in so far as they may have an influence on the outcome of the study.

It is also important to look not only for the intended effects, but also for side effects and unintentional results. The example shows how unintended effects lead to termination of the study.

These effects may be discovered not only in qualitative studies, but may also surface in quantitative studies. When this is found during data analysis, a statement should be made in the results section and details should be provided. The authors should - as part of data analysis - explore the data sets and sources to identify potential reasons behind such unexpected observations, as the causal reason behind such deviations may be common, may skew the outcome, and not least may provide a rich picture of the system. A discussion of the reasons for the deviation as well as of the implications of these results should take place in the discussion section of the manuscript.

3.8 Discussion

The purpose of discussion is that the study should be subjected to a critical analysis of the study results by the authors themselves. Any lessons learned - both with respect to study questions and the methodology - should be reported for others to learn from.

A structured discussion as proposed below is often not required by the journal but it will ease the writing as well as the reading.

3.8.1 Answer to study questions

■ A discussion of the answers identified versus the questions posed for the study.

Examples

- ⊙ “In this study, we performed a usability evaluation of the OM/Lab system, an emerging CPOE Laboratory system. Overall the results of the usability evaluation revealed that the OM/Lab’s usability fell short of the designers expectations.” [33, p. 165]. *This statement was followed by an elaborate discussion of why it is so and how it became so.*
- ⊙ “This study investigated the effect of RTMM² with SMS reminders on the precision with which type 2 diabetes patients adhere to their prescribed oral medication. Our results show that RTMM with SMS reminders has positive effects on patients’ adherence to oral antidiabetic medication. Patients who received reminders for forgotten doses took their oral antidiabetics more regularly than patients who were not reminded. In addition, patients who were reminded tended to miss doses less frequently.” [42, p. 601]

Elaboration

This section addresses the answers to the study questions, based on an interpretation of the data and information as presented in the results section. The study questions should clearly be answered, as shown in both examples.

The key issue is whether the answer to the study questions is convincing for the reader and sufficiently supported by evidence. This includes the question of whether the actual set of outcome measures is likely to have exhausted the study questions. Was the target application the right one for the study? And, was the study design appropriate to answer the study question?

One should avoid presenting new data in the discussion section. One can only make reference to results of secondary analysis of the data that was not part of the original study design, but which were performed to find explanations for certain findings of the study.

The linkage between an outcome and assumed causal effect mechanisms behind this outcome has to be cautiously explained. One should be aware that causal explanations have a number of prerequisites that need to be met. They include, among others, a proper temporal relation (effect after stimulus), not explainable by other events (control group may be needed), a dose-effect relation (although this may be difficult for IT interventions in health care), or be explainable by existing theories (e.g. cognitive or behavioral theories). Just observing an effect is insufficient to state that a causal relation exists.

In the case of unexpected observations, the authors should discuss the potential reasons behind such observed deviations. Often such an analysis will lead to a series of new research questions. If there is a theoretical backing for the observed unexpected observations, the authors should provide sufficient references to those theories (and other studies) that provide the supporting evidence.

3.8.2 Strengths and weaknesses of the study

■ Critical discussion of the methods used.

Examples

- ⊙ “Several limitations of our study should be considered. First and foremost, inherent limitations of study design preclude any statements regarding cause and effect, and ... Second, because we have

² RTMM = Real Time Medication Monitoring

examined a unique patient population admitted through interfacility transport, our findings may not be generalisable to Third, our observation period after CPOE implementation was brief and may simply reflect the adjustment period that commonly follows any major, sweeping change. ... Fourth, in a related consideration, the relative imbalance between our pre- and post-CPOE observation periods raises potential confounding from seasonal variability of illness Fifth, we again consider the possibility that our finding may reflect a clinical applications program implementation and systems integration issue rather than a CPOE per se. Sixth, although we have attempted to control for many important mortality co-variables, it remains possible that our observation that CPOE implementation is associated with increased mortality may have resulted from an unidentified confounding factor. A “regression to the mean” phenomenon cannot be discounted.” [Han et al. 2005, pp. 1511-2].

- ⊙ “Our study has several limitations. It was conducted for a limited duration in only 4 urban primary care practices including many physicians with part-time practices, so the results may not be generalizable. Our study was not powered to detect modest differences in potential ADE rates between computerized and handwritten sites. In addition, physicians were not blinded to the purpose of the study and might have been particularly careful when prescribing or may have excluded patients they knew to be at high risk“ [16, p. 841]
- ⊙ “The limitations of the study were that it was conducted in a single clinic with participants within the same specialty and a small number of cases may limit generalizability. Although formal data on PDA use were not collected on those who chose not to participate, we do know that at least some of them already had PDAs and were not interested in changing to a new one, but it is possible that others were not interested in using the technology. Those who did choose to participate were, by definition, receptive to learning to use the technology. It is possible that this group may have implicitly understood that use was expected, whereas a nonvolunteer sample would not use it as frequently. However, we emphasized that they should use it often as they felt it was needed, and we made the training as minimal as possible to limit this effect.” [5, p. 176]

Elaboration

Any study has strong and weak points. Although the authors may not be the best persons to identify the weaknesses, a reflective investigative author will be well-placed to consider limitations, and a critical self-reflection on the study demonstrates that the authors are cautious about the strength of the evidence that is provided by the study. This is particularly true when controversial results have been obtained. All three examples above show a careful discussion of possible limitations.

In this section, the authors should at minimum report on the following aspects:

- Strengths and weaknesses of the study design, including an account for why a possibly stronger design has not been followed – or in hindsight should have been followed.
- Appropriateness of the selected outcome measures to solidly answer the study question. This is particularly the case when proxy measures have been used (e.g. medication errors as proxy for patient safety)
- Any potential biases not accounted for in the study design or data analysis (e.g. Hawthorne Effect, cultural biases). Even the demonstration of an awareness of pitfalls for one’s study type and setting will increase the credibility of a study.
- Any (potential) violation of the known assumptions for application of the applied methods, techniques and metrics and the potential impact on the outcome and conclusions of the study.
- Any known or observed phenomenon that have a potential impact on the validity of the data collected and hence on the conclusions of the study (beyond biases, for instance organizational pressure)

For quantitative experimental trials, the internal validity of findings needs to be critically assessed. As Coolican et al. [12, p. 20] write: “We should be able to indicate in the discussion ...where our design has weaknesses, where we did not have control and, therefore, how limited we are in assuming that our independent variable *really did* affect our dependent variable.” And, the authors continue with what to look for: “We should be able to point out possible differences between our groups, and differences in their experiences, which might be responsible for any differences in the dependent

variable which we identify, making it difficult to attribute these differences *solely* to the change in the independent variable.”

3.8.3 Results in relation to other studies

■ Make clear what exactly is novel about the obtained results.

Examples

- ⊙ “Although our observation complements the report by Koppel et al (7) that highlights potential problems with CPOE resulting from “systems integration failure” and “human-machine interface flaws,” our finding does not support the overwhelming majority of studies that have reported that CPOE systems reduce potential ADEs (2–4,9–11) Of particular interest and concern, our result seems to conflict with other investigators from our own institution who recently reported in their study ... a significant decrease in harmful ADEs ... after CPOE implementation. (4). Although differences in study method and data source used by Upperman et al (4) can partly explain our vastly divergent observations, we are reminded by Berger and Kichak that “although the literature suggests [CPOE] systems have the potential to improve patient outcomes through decreased adverse drug events, actual improvements in medical outcomes have not been documented.” [18, p. 1510]
- ⊙ “No previous study has looked at CPOE use in different clinical environments. Our study showed that different clinical environments between the EDs and the hematology wards affected who ordered clinical laboratory and radiology tests, how they were ordered ... and where they preferred to order ... A number of previous studies have emphasized the importance of ‘fit’ between the technology and work practices.(1,3,17,25,55–57) Our study has shown that this ‘fit’ between CPOE and work practices will also vary depending on the clinical environment.” [10, p. 649]

Elaboration

Progress is made by scientific inquiry – i.e. doing studies – and integrating the results in the existing body of knowledge. Hence, it is relevant to know how the results of a study relate to the existing body of knowledge. Authors should demonstrate that they know the state of affairs and how their study adds to the existing knowledge base (see both examples). This requires that the results of the study are compared with results of similar studies either in the same domain – in order to strengthen or challenge the knowledge that already exists – or in related domains as to demonstrate the similarity or lack thereof of the studied domains. If similarities exist, other results may be transferable as well.

3.8.4 Meaning and generalizability of the study

■ Implication of the study findings, for the various stakeholders within the study and beyond.

Examples

- ⊙ “Although we anticipated that the CDSS would reduce inappropriate prescribing, the effect was more complicated. Instead of the performance of the intervention group improving significantly over time, their performance remained relatively stable, while overall the control group performance degraded over the time period from baseline to follow-up. Other data have shown that house staff thoroughness, especially for preventive measures, history, or other data not closely related to the acute presenting problem, may decrease as they progress through training.(49) The present study indicated that the CDSS may have minimized that performance degradation in the intervention group.” [6, p. 176]
- ⊙ “Although multidisciplinary settings are common in contemporary health care, no studies have yet evaluated the effect of computerised decision support in such a setting.(7) ... Our results show that computerized decision support can also be an effective instrument in multidisciplinary settings, where such social factors play an important role in decision making. This study shows that, in a multidisciplinary team motivated to adopt a computerised decision support aid that assists in formulating guideline based care plans, such support can be effective in improving multidisciplinary teams’ concordance with guidelines. On the basis of our findings, we encourage the

use of computerised decision support aids in settings where multidisciplinary teams are motivated to use them.” [17, p. 7]

Elaboration

In general a study is conducted in a specific context, among a specific group of participants. The relevancy of a study for a broader audience is not only determined by the actual findings, but also by the generalizability of the results. The authors should report to what extent their findings may have wider implications, taking into account limitations by study design and environmental context (see both examples above). For example, the results and conclusions obtained with a CPOE system in an internal medicine ward may not be comparable to those for an emergency department due to largely different working environments and case-mix. Similarly, results and conclusions obtained in an academic hospital will likely not be relevant to the situation in a rural health care clinic, but may be transferable to a larger teaching hospital. And, a result obtained in one country, culture or health system may not provide transferable evidence appropriate for another country, culture, or health system.

3.8.5 Unanswered and new questions

Future research needs and opportunities.

Examples

- ⊙ “Although the literature has shown that information systems may decrease error in medicine, the idiosyncratic and varied nature of design of health information applications argues for the usability testing of each of these systems and devices to predict aspects of design that may be related to error prior to their widespread dissemination.” [26, p. 525]
- ⊙ “Our results suggest that a computerised decision support system alone is insufficient to improve concordance with guidelines....More research is therefore needed to understand how computerised decision support can improve concordance with guidelines and which additional strategies for change need to be considered to overcome the remaining barriers.” [17, p. 1440]

Elaboration

No study will provide the answer to all questions. It is more likely that the results of the study will raise more new questions than being answered. In this section the authors should report which new questions have arisen from the study, as shown in the examples. These questions need not to be limited to the topic of the study, but can equally well address methodological questions that have arisen. By making the new and open research questions explicit the authors contribute to defining the research agenda for the future. However, the identification of new questions should not be allowed to deflect from the reaching of conclusions on the core study questions – conclusions should be sought and evidence thereby created. The elaboration of new questions should be focused in most cases on new lines of enquiry, not the perpetuation of self-fulfilling research or avoidance of conclusions.

3.9 Conclusion

Summary of the main findings, including the impact of the findings and how they relate back to the big picture provided.

Examples

- ⊙ “CPOE is an important medical information technology that holds great promise as a tool to reduce human error during health care delivery. In this current study, however, we observed an unexpected increase in mortality coincident with CPOE implementation. Our unanticipated finding suggests that when implementing CPOE systems, institutions should continue to evaluate mortality effects, in addition to medication error rates, for children who are dependent on time-sensitive therapies. CPOE technology is still evolving and requires ongoing assessment of “systems integration” and “human-machine interface” effects, both predictable and unpredictable, on patient care and clinical outcomes.” [18, p. 1512]

- ☉ “Prescribing errors are common in ambulatory care, and often have potential to harm patients. Office practices and health systems are beginning to develop or purchase computerized prescribing systems. Basic computerized prescribing did not result in fewer errors compared with hand-written prescribing. Therefore, to achieve a major safety benefit, computerized prescribing with advanced decision support will likely be needed.” [16, p. 841]

Elaboration

Besides the (structured) abstract, the conclusion is an important part of the account of an evaluation study. It helps the reader in quickly assessing what the study has contributed. It should be a concise, balanced and objective summary of the conclusions that have been drawn from the study (see both examples). It summarizes parts of the discussion, in particular the aspects addressed under 3.8.1 and 3.8.4 and possibly parts of 3.8.5. The authors should make sure to choose the right concluding words not to over- or understate the strength of their claims made.

3.10 Authors' contribution

Explicit description of the contributions of the authors to make sure that each author qualifies for authorship.

Examples

- ☉ “The idea for STARE-HI was raised during the ESF workshop in Innsbruck. J.T. took the initiative to develop STARE-HI, he is the guarantor of the study. J.T. and E.A. drafted a first list of issues. E.A. and J.T. drafted the first version of the manuscript. J.B., N.dK., P.N., and M.R. all contributed by critically assessing the items and their descriptions. They have made suggestions for expansion and provided various parts of the text. J.T. and E.A. integrated the various contributions and wrote the final version of the manuscript. All authors have approved this final version.” [38, p. 8]

Elaboration

It is relevant for the scientific record that the contributions of the authors to the study are documented. This is becoming increasingly relevant due to the multidisciplinary nature of the research. Not all authors need to be experts on all details of the study described, but all should have made an explicit contribution. Hence a statement on the contribution of authors makes clear who is responsible for the various aspects of a study, and that all were active contributors at some stage. The International Committee of Medical Journal Editors (ICMJE) has published guidelines on who are entitled to be author on a paper [22]. It is clear from these guidelines that all authors should agree on the final version of the manuscript and hence have joint responsibility for what is published. It is good practice to indicate who of the authors takes responsibility for the whole study.

3.11 Competing interests

A statement of the interest, financial or otherwise, the authors may have with respect to the outcome of the study.

Examples

- ☉ “An evaluation study was designed to discover the benefits of ... Both researchers were from outside the institution; they were asked to perform the evaluation by the system developer.” [23, p. 223]
- ☉ “There is full freedom for the evaluators to publish, provided that patient security issues and business secrets are kept confidential. This is fulfilled for the present study without consequence for the conclusion above.” (fictive example)

Elaboration

We strongly suggest that any relation between the authors and the object of study (including its organizational setting) is declared. It is often observed that the developers of a system also perform an

evaluation. It is not always clear from the text that such relation exists, in particular when the system is already in routine use in an organization. One should be strict and open about the relations and agreements between the parties involved as well as about the type of engagement or role of these parties.

The first example provided above is illustrative in more than one sense. It is perfect that this information is reported. The authors performed a job initiated by the developers, but because of the delimitation to only study benefits it would have been even better, if they had also stated the actual terms of the task, such as potential task directives and delimitations (as opposed to the complete academic freedom to design the study themselves) and financial conditions, including the terms for publishing their findings, whether there was active discouragement from seeking overall effects, and whether any negative findings were put to one side.

The second example touches a dilemma. For legislative reasons one cannot disclose confidential information, and for scientific reasons nor can one suppress information that will bias the conclusion of a study. When there is a contract with the provider of the system that imposes restrictions on the disclosure of a specific type of information, this should be mentioned as to indicate that the authors have made a conscious decision to not mention certain details, and whether this has implications for the conclusion.

There are guidelines developed by the ICMJE [22] on stating competing interests. Unfortunately this covers mainly the situation that is seen in medical research and in particular drug research. Financial interests (both stock holders and payment for services) have to be declared. In the domain of health informatics other relations may exist as well that may potentially influence the reporting of study results and their interpretation – these may range from responsibility for the policy, to implement a system, to a leadership role in the clinical practice of users.

Besides these issues, there are other circumstances that may have an influence on the study design and the conclusions of the study, such as the relation between the evaluators and the sponsors of a study, the purpose of the study (e.g. to take an investment decision) or an employment relation. There are no clear guidelines yet on how to deal with these issues, but it may be advisable to report that one is employed or otherwise paid by the organization in which the implementation of a given ICT system was evaluated. In both cases there is the risk that less favorable outcomes are not reported because of financial or political consequences.

3.12 Acknowledgements

Acknowledgements of any financial or other support.

Examples

- © “We would like to acknowledge the support of the European Science Foundation, enabling us to accomplish the ESF Exploratory Workshop on New Approaches to the Systematic Evaluation of Health Information Systems (HIS-EVAL). A special thanks to Hui Wang for his support in this respect, to Gudrun Hübner-Bloder and Frieda Kaiser for local organization, and to Karl-Peter Pfeiffer for scientific support.” [2, p. 490]

Elaboration

It is essential that credit is given to those that have contributed to the study, but not in a way that would entitle them to authorship. The contributions can vary from development of software, organizational support, data acquisition, writing assistance to financial support, as shown in the example. The ICMJE also provides advice on the distinction between authors and contributors [22].

3.13 References

All references needed for the argumentation

Elaboration

References are an essential part of the scientific discourse. They serve as evidence in the same way as the data material does and should be handled in a similar manner [9]. The references support the ar-

gumentation or provide the detailed background for the statements being made. The readers should be able to retrieve the references so as to be able to study that material in more detail; and if relevant material is unavailable in the public literature then at least the authors must be willing to provide copies upon request from readers.

Although material from sources other than peer-reviewed journals, proceedings or books (for example, reports from the Internet) is generally not recognized as high quality scientific material and hence should in general not be cited since it has less strength as evidence, such material should be referenced when it forms a part of the foundation for the study, since the alternative would be to take credit for work or ideas that are not the authors'.

In the domain of health informatics relevant material (both peer-reviewed and non-peer-reviewed) may only be available on-line. Such material should be properly referenced, preferably by providing also the authors of the document where identified, a full URL and the last date the item was visited by the authors using that URL. The latter helps the reader in determining whether the current version is the version that has been cited.

Details instructions for formatting references are usually given by the respective journal.

3.14 Appendices

Any supporting material, such as detailed descriptions of methods/tools (e.g. a questionnaire), specific data analysis techniques and detailed study results

Elaboration

The main body of the paper should contain the information that is relevant for the general reader of the manuscript. Still there may be relevant aspects of the study to report that are only of interest for specific readers, such as detailed statistical methods, specific questionnaires developed or adapted for the study at hand, detailed results from statistical analyses like factor analysis, structural equation modeling or regression analysis, or detailed results from qualitative studies such as interviews transcripts or concept maps.

There may be restrictions on the amount of information that can be put in an appendix. Furthermore, what can be put in an appendix needs to be printable. Currently, many authors may want to provide other material as well to supplement their paper. Hence, many journals currently offer the possibility of supplementary material of any kind, including software, sound and video clips and 3-d models to be published on the journal's website, accessible through with a link from the article.

4. Discussion

In the CONSORT Explanation and Elaboration paper it is stated that "Sound science encompasses adequate reporting ..." [28]. In analogy to what these authors argue for trials, we would argue that the conduct of valid evaluation studies rests on sound science. Thus, evaluation of health informatics applications and interventions encompasses trustworthy reporting of the study. Consequently, evaluators are requested to enable unbiased insight into study details and to enable others to perform unbiased comparisons and conclusions.

The STARE-HI statement and the elaborations put forward here should assist authors to present the right information in the right way. Implicitly also, this explanatory article may help researchers during the design of evaluation studies and this way also support the GEP-HI initiative [30]. Moreover, it may guide peer reviewers and editors in their assessment of evaluation study reports, and such extended application of this explanatory article will likely improve the clarity and transparency of published evaluation studies.

The STARE-HI statement should be usable for researchers writing for all Western-oriented journals publishing evaluation studies on health informatics applications. In other geographic settings, specific cultural conditions may have to be taken into account; still, the majority of items should be relevant.

There are other special cases, where the above recommendations may have to be adapted by the authors, such as assessment of evaluation methods, formative evaluation as opposed to summative

evaluation, or action-case research. In each and every case, evaluators need to consider which of the guidelines, recommendations and statements are valid in his/her situation and/or adapt according to the situation. In any case, neither STARE-HI nor this elaboration paper is meant to be prescriptive for a special evaluation methodology or approach. It should give guidance, but requires that the user still uses his/her judgement on the applicability of the different items.

We have chosen good and bad examples for our recommendations, striving to find real-life examples in the literature. We have not included citations from our own evaluation papers in the examples. We considered a good example a part of a manuscript that addressed as much as possible, in a clear and succinct way the aspects listed in an item as well as in the elaboration description. Bad examples were considered those that lacked essential information related to the issue at stake. For the bad examples, we didn't look only at the parts cited, but to the whole paper to assure that the issues at stake were not addressed elsewhere in the paper. All chosen examples are considered to have face validity, since all authors have long-year experience in health IT evaluation and have agreed on their inclusion.

This elaboration paper does not intend to replace basic literature on research methodology or evaluation methodology in health informatics. Here, the reader should consult handbooks such as [8, 15, 27].

5. Conclusions

We believe that publishing higher quality evaluation studies is a prerequisite for achieving the vision of evidence-based health informatics. A comprehensive list of principles relevant for properly describing health informatics evaluation studies has been described in detail and with a focus on frequent omissions and pitfalls in writing. Evaluation manuscripts adhering to these principles will enable readers of such papers to better place their studies in a proper context and judge their validity and generalizability. Finally, but not least, it will facilitate also the analysis of whether given papers fit in the scope of systematic reviews and meta-analyses of health informatics interventions.

6. Clinical relevance

The resulting implications of this paper for healthcare practitioners are more complete and mature reports from health IT evaluation studies, and hence, more transparent and reliable information on quality and impact of health information systems or services that are the object of an evaluation study. In a long term perspective, this may facilitate systematic reviews and meta-assessment regarding health information systems or services. It should also yield evidence to stimulate better system design, health informatics policies and investment decisions, and thus health service delivery and patient outcomes.

7. Authors' contributions

The idea of this standard/guideline contribution was launched as a part of the STARE-HI Statement, which again was initiated at the European Science Foundation working conference in 2004. JB is the guarantor and the coordinator of the study. JT and EA have provided extensive critical appraisal and substantial support in the editing. The other co-authors constitute members of the editorial team and all have contributed by critically assessing the items and their descriptions in an on-going fashion, as well as by making substantial input or suggestions for change or expansion.

Protection of human subjects

Human subjects were not included in the project.

Competing Interests

The authors declare that they have no financial conflicts of interest in relation to the work presented here. JB and PN are co-chairs of the EFMI working group on Evaluation (EVAL) that has EA as the chair person. NdK is chairperson of IMIA's working group on Technology Assessment and Quality Improvement. JT and EA are editors of the International Journal of Medical Informatics and of Methods of Information in Medicine respectively.

Acknowledgment

We are indebted to all persons who have provided helpful input to this paper. In particular, the following members of the HIS-EVAL mailing list, and beyond, have provided intellectual contributions of a substantial nature: Emily Campbell, Joshua Richardson, and Dean F. Sittig.

Table 1 STARE-HI elements [38]. The table can be used as a checklist, to check own manuscripts on completeness.

Chapter in this elaboration paper	STARE-HI item	Item comprised in manuscript? (√/–)
3.1	Title	
3.2	Abstract	
3.3	Keywords	
3.4	Introduction	
3.4.1	Scientific background	
3.4.2	Rationale for the study	
3.4.3	Objectives of the study	
3.5	Study context	
3.5.1	Organizational setting	
3.5.2	System details and system in use	
3.6	Methods	
3.6.1	Study design	
3.6.2	Theoretical background	
3.6.3	Participants	
3.6.4	Study flow	
3.6.5	Outcome measures or evaluation criteria	
3.6.6	Methods for data acquisition and measurement	
3.6.7	Methods for data analysis	
3.7	Results	
3.7.1	Demographic and other study coverage data	
3.7.2	Unexpected events during the study	
3.7.3	Study findings and outcome data	
3.7.4	Unexpected observations	
3.8	Discussion	
3.8.1	Answers to study questions	
3.8.2	Strengths and weaknesses of the study	
3.8.3	Results in relation to other studies	
3.8.4	Meaning and generalizability of the study	
3.8.5	Unanswered and new questions	
3.9	Conclusion	
3.10	Authors' contribution	
3.11	Competing interests	
3.12	Acknowledgement	
3.13	References	
3.14	Appendices	

References

1. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T, Consort G. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 8: 663-694.
2. Ammenwerth E, Brender J, Nykänen P, Prokosch H-U, 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; 6: 479-491.
3. Ash J. Organizational factors that influence information technology diffusion in academic health sciences centers. *J Am Med Inform Assoc* 1997; 2: 102-111.
4. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 2004; 2: 104-112.
5. Berner E, Moss J. Informatics Challenges for the Impending Patient Information Explosion. *J Am Med Inform Assoc* 2005; 6: 614-617.
6. Berner ES, Houston TK, Ray MN, Allison JJ, Heudebert GR, Chatham WW, Kennedy JI, Jr., Glandon GL, Norton PA, Crawford MA, Maisiak RS. Improving ambulatory prescribing safety with a handheld decision support system: a randomized controlled trial. *J Am Med Inform Assoc* 2006; 2: 171-179. doi: 10.1197/jamia.M1961.
7. Bloomrosen M, Starren J, Lorenzi NM, Ash JS, Patel VL, Shortliffe EH. Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting. *J Am Med Inform Assoc* 2011; 1: 82-90. doi: 10.1136/jamia.2010.007567.
8. Brender J. *Handbook of evaluation methods for health informatics*. Burlington, MA: Elsevier Academic Press 2006.
9. Brender J, Talmon J. On using references as evidence. *Methods of information in medicine* 2009; 6: 503-507. doi: 10.3414/ME09-03-0001.
10. Callen JL, Westbrook JI, Braithwaite J. The effect of physicians' long-term use of CPOE on their test management work practices. *J Am Med Inform Assoc* 2006; 6: 643-652. doi: 10.1197/jamia.M2152.
11. Chertow GM, Lee J, Kuperman GJ, Burdick E, Horsky J, Seger DL, Lee R, Mekala A, Song J, Komaroff AL, Bates DW. Guided medication dosing for inpatients with renal insufficiency. *Jama* 2001; 22: 2839-2844.
12. Coolican H. *Introduction to Research Methods and Statistics in Psychology*. London: Hodder & Stoughton 1999.
13. Eminovic N, de Keizer NF, Wyatt JC, ter Riet G, Peek N, van Weert HC, Bruijnzeel-Koomen CA, Bindels PJ. Teledermatologic consultation and reduction in referrals to dermatologists: a cluster randomized controlled trial. *Arch Dermatol* 2009; 5: 558-564. doi: 10.1001/archdermatol.2009.44.
14. Fraenkel DJ, Cowie M, Daley P. Quality benefits of an intensive care clinical information system. *Crit Care Med* 2003; 1: 120-125.
15. Friedman C, Wyatt JC. *Evaluation Methods in Medical Informatics*. New York: Springer 2006.
16. Gandhi TK, Weingart SN, Seger AC, Borus J, Burdick E, Poon EG, Leape LL, Bates DW. Outpatient prescribing errors and the impact of computerized prescribing. *J Gen Intern Med* 2005; 9: 837-841.
17. Goud R, de Keizer NF, ter Riet G, Wyatt JC, Hasman A, Hellemans IM, Peek N. Effect of guideline based computerised decision support on decision making of multidisciplinary teams: cluster randomised trial in cardiac rehabilitation. *BMJ* 2009; b1440. doi: 10.1136/bmj.b1440.
18. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, Bayir H, Orr RA. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005; 6: 1506-1512.
19. Hansson S. Values in pure and applied science. *Found Sci* 2007: 257-268.
20. Harris A, McGregor J, Perencevich E, Furuno J, Zhu J, Peterson D, Finkelstein J. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc* 2006; 1: 16-23.
21. Haux R. Medical informatics: past, present, future. *Int J Med Inform* 2010; 9: 599-610. doi: S1386-5056(10)00114-0 [pii] 10.1016/j.ijmedinf.2010.06.003.
22. ICMJE. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publications. 2012, March-28. <http://www.icmje.org/index.html>.
23. Kaplan B, Lundsgaarde HP. Toward an evaluation of a clinical imaging system: identifying benefits. *Methods of Information in Medicine* 1996; 3: 221-229.
24. Kastner M, Li J, Lottridge D, Marquez C, Newton D, Straus SE. Development of a prototype clinical decision support tool for osteoporosis disease management: a qualitative study of focus groups. *BMC Med Inform Decis Mak* 2010; 40. doi: 10.1186/1472-6947-10-40.
25. Koppel R, Metlay J, Cohen A, Abaluck B, Localio A, SE K, Strom B. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005; 10: 1197-2003.

26. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. *Int J Med Inform* 2005; 7-8: 519-526. doi: S1386-5056(05)00011-0 [pii] 10.1016/j.ijmedinf.2005.01.003.
27. Layman E, Watzlaf V. *Health Informatics Research Methods: Principles and Practice*. Chicago, IL: AHIMA 2009.
28. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; c869. doi: 10.1136/bmj.c869.
29. NLM. MeSH. Medical Subject Headings. <http://www.nlm.nih.gov/mesh/MBrowser.html>.
30. Nykanen P, Brender J, Talmon J, de Keizer N, Rigby M, Beuscart-Zephir MC, Ammenwerth E. Guideline for Good Evaluation Practice in Health Informatics (GEP-HI). *Int J Med Inform* 2011; 815-827. doi: S1386-5056(11)00168-7 [pii] 10.1016/j.ijmedinf.2011.08.004.
31. Palm JM, Colombet I, Sicotte C, Degoulet P. Determinants of user satisfaction with a Clinical Information System. *AMIA Annu Symp Proc* 2006: 614-618.
32. Peterson JF, Kuperman GJ, Shek C, Patel M, Avorn J, Bates DW. Guided prescription of psychotropic medications for geriatric inpatients. *Arch Intern Med* 2005; 7: 802-807.
33. Peute LW, Jaspers MW. The significance of a usability evaluation of an emerging laboratory order entry system. *Int J Med Inform* 2007; 2-3: 157-168. doi: 10.1016/j.ijmedinf.2006.06.003.
34. Potts AL, Barr FE, Gregory DF, Wright L, Patel NR. Computerized physician order entry and medication errors in a pediatric critical care unit. *Pediatrics* 2004; 1 Pt 1: 59-63.
35. Rigby M. Evaluation - the cinderella science of health IMIA Yearbook of Medical Informatics 2006: 114-120.
36. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, Pifer E. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. *Arch Intern Med* 2010; 17: 1578-1583. doi: 10.1001/archinternmed.2010.324.
37. Talmon J, Ammenwerth E, Geven T. The quality of reporting of health informatics evaluation studies: a pilot study. *Stud Health Technol Inform* 2007; Pt 1: 193-197.
38. Talmon J, Ammenwerth A, Brender J, de Keizer N, Nykanen P, Rigby M. STARE-HI - Statement on Reporting of Evaluation Studies in Health Informatics. *Int J Med Inform* 2009; 1: 1-9.
39. Tufte E. *The Visual Display of Quantitative Information*. Cheshire (Connecticut): Graphics Press 2001.
40. van den Brink JL, Moorman PW, de Boer MF, Pruyn JF, Verwoerd CD, van Bommel JH. Involving the patient: a prospective study on use, appreciation and effectiveness of an information system in head and neck cancer care. *Int J Med Inform* 2005; 10: 839-849. doi: 10.1016/j.ijmedinf.2005.03.021.
41. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, initiative S. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 8: W163-W194.
42. Vervloet M, van Dijk L, Santen-Reestman J, van Vlijmen B, van Wingerden P, Bouvy ML, de Bakker DH. SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored. *Int J Med Inform* 2012; 9: 594-604. doi: 10.1016/j.ijmedinf.2012.05.005.
43. Winter A, Haux R, Ammenwerth E, Brigl B, Hellrung N, Jahn F. *Health Information Systems: Architectures and Strategies*. New York, Berlin, Heidelberg: Springer-Verlag 2011.