

*Viewpoint Paper* ■

# A Viewpoint on Evidence-based Health Informatics, Based on a Pilot Survey on Evaluation Studies in Health Care Informatics

ELSKE AMMENWERTH, PhD, NICOLETTE DE KEIZER, PhD

**Abstract** Concerned about evidence-based health informatics, the authors conducted a limited pilot survey attempting to determine how many IT evaluation studies in health care are never published, and why. A survey distributed to 722 academics had a low response rate, with 136 respondents giving instructive comments on 217 evaluation studies. Of those studies, half were published in international journals, and more than one-third were never published. Reasons for not publishing (with multiple reasons per study possible) included: “results not of interest for others” (1/3 of all studies), “publication in preparation” (1/3), “no time for publication” (1/5), “limited scientific quality of study” (1/6), “political or legal reasons” (1/7), and “study only conducted for internal use” (1/8). Those reasons for non-publication in health informatics resembled those reported in other fields. Publication bias (preference for positive studies) did not appear to be a major issue. The authors believe that widespread application of guidelines in conducting health informatics evaluation studies and utilization of a registry for evaluation study results could improve the evidence base of the field.

■ *J Am Med Inform Assoc.* 2007;14:368–371. DOI 10.1197/jamia.M2276.

## The Problem of Publication Bias

Healthcare IT systems have been shown to increase quality and efficiency of health care.<sup>1–3</sup> However, there are also examples where IT systems failed to provide the expected benefits or even seem to have negative effects on patient care.<sup>4–6</sup> Systematic evaluation is thus needed and even seen as an ethical imperative for health informaticians.<sup>7</sup> As a whole, published IT evaluation studies contribute to the emergence of evidence-based health informatics<sup>7,8</sup> which can be defined as the conscientious, explicit, and judicious use of current best evidence to support a decision with regard to IT use in health care (based on EBM-definition by Sackett).<sup>9</sup>

There have been quite a lot of IT evaluation papers in the last 25 years, as shown in a recent inventory.<sup>10</sup> However, we do not know how representative and complete those IT evaluation publications are. One problem frequently discussed in this context is publication bias. The most common type of publication bias is the one in which well-executed studies

with null, negative, or disappointing results do not find their way into the archival literature.<sup>11,12</sup>

In health care, publication bias is an issue discussed for more than 100 years.<sup>13</sup> Since then, hundreds of publications have discussed the problem of publication bias in clinical research (e.g.,<sup>14–16</sup>), analyzing reasons as well as methods for detection and prevention.

However, despite the fact that publication bias can pose larger threats to the evidence-base of health informatics, publication bias in health informatics has not yet been systematically studied, even when the problem has already been discussed by Tierney<sup>17</sup> and Friedman and Wyatt.<sup>11</sup> Consequently, the authors conducted a limited pilot study to determine:

1. What percentage of IT evaluation studies are not published in international journals or proceedings?
2. What are typical reasons for not publishing the results of an IT evaluation study?

## Publication Bias in Health Informatics: Results of a Pilot Survey

To answer those questions, the authors in Spring 2006 conducted a written, e-mail-based survey of academics. The survey sample included members of the mailing lists of the AMIA working group on Evaluation ( $n=341$ ), the EFMI working group on Assessment of Health Information Systems ( $n=224$ ), and the IMIA working group on Technology Assessment and Quality Development in Health Informatics ( $n=220$ ), and first authors of IT evaluation papers that were published between 2001 and 2006 and Medline Indexed ( $n=204$ ). Overall, after removing duplicate names, 722 academics were included.

Affiliations of the authors: Institute for Health Information Systems (EA), UMIT—University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria; Department of Medical Informatics (NdK), Academic Medical Center, Universiteit van Amsterdam, Amsterdam, The Netherlands.

Parts of the results have been presented at an MIE 2006 workshop and at a AMIA 2006 workshop. We thank all respondents for answering and providing valuable information and comments in this survey.

Correspondence and reprints: Dr. Elske Ammenwerth, Institute for Health Information Systems, UMIT—University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria; e-mail: <Elske.Ammenwerth@umit.at>.

Received for review: 9/14/2006; accepted for publication: 1/26/2007.

The survey consisted of three questions:

1. Which information systems did you evaluate in the last three years?
2. Where did you publish results?
3. If you did not publish: What were reasons for this? (Here, typical reasons could be selected, or free-text entered).

For each study, authors analysed where it had been published, and classified the responses as internationally published (e.g., peer-reviewed journal, proceeding, or book); only local publication (e.g., local conferences, masters thesis); or non-publication (no publication available, just internal reports).

Only 136 academics responded (return rate: 18.8%). The preliminary results are reported herein as indicators of possible trends, with the hope that others might confirm them in larger studies. The 118 individuals who reported completing studies provided information on 217 evaluation studies. From those 118 individuals, 33 came from the mailing list of first authors (they reported on 77 studies), 37 from the EFMI mailing list (reporting on 53 studies), 31 from the AMIA mailing list (reporting on 56 studies), and 17 from the IMIA mailing list (reporting on 31 studies). In addition, 18 of the 136 respondents said they conducted no evaluation studies. The 118 respondents that provided valid information came from the USA ( $n=45$ ), UK ( $n=11$ ), Netherlands ( $n=10$ ), Canada ( $n=8$ ), Germany ( $n=6$ ), Australia ( $n=6$ ), and from 18 other countries ( $n=32$ ). Please refer to the Appendix, available as an electronic data supplement at [www.jamia.org](http://www.jamia.org), for the response rate per country. Based on the e-mail signature, the e-mail address, and the comment of the respondent, we grouped the respondents by background. Most respondents ( $n=92$ ) came from an academic environment, 8 from IT management, 6 from industry, and 5 from governmental institutions, with 7 backgrounds unknown.

Overall, the 118 respondents reported on 217 evaluation studies. The most often evaluated type of IT system was EPR/EHR systems ( $n=28$ ) and CPOE and medication systems ( $n=23$ ) (see Appendix, available as an electronic data supplement at [www.jamia.org](http://www.jamia.org) for details on type of IT systems). For the 217 evaluation studies conducted by respondents, 213 publications were reported. Of those 213 publications, 77 were included in this survey based on pre-selection of authors with prior evaluation study publications. About half of the 217 evaluation studies were published in peer-reviewed international journals, proceedings, or books. Slightly more than 1/3 of studies have not yet been published anywhere. One-tenth of other studies had only internal project reports and 1/16 had only local publications (e.g., local conferences, thesis). From those studies published internationally, more than half appeared in health informatics journals or proceedings, 1/3 appeared in medical or nursing journals, and 1/10 appeared in other journals. See the Appendix (available as an on-line data supplement at [www.jamia.org](http://www.jamia.org)) for details on what journals the evaluations appeared in.

For 107 evaluation studies that were unpublished or only published in internal project reports or local publications, respondents gave reasons for non-publication. These rea-

sons were grouped into the following categories (with multiple reasons per study possible):

1. "Planned or in preparation": Publication is planned or already in progress. Quote: "May publish following validation" (around 1/3 of all non-published studies).
2. "Not of interest for others": The generalizability of the results seemed too limited, or the results seemed not to be of interest for others. Quote: "Constellation of internal social factors, adoption factors, staff training/experience, etc. seemed too unique to make it general enough" (around 1/3 of all non-published studies).
3. "No time for writing": No time found yet for publication, as e.g., making the IT system operational took too much time, funding ran out, or new projects started. Quote: "Too busy implementing CPOE to publish" (around 1/5 of all non-published studies).
4. "Limited scientific quality": The methods used seemed not adequate, or the paper was submitted, but not accepted as the editor said it had insufficient quality. Quote: "The setup (e.g., amount of interviews) was not robust enough" (around 1/6 of all non-published studies).
5. "Political and legal reasons": The organization the author work(ed) in prohibited publication, or the results were too negative to be published (both categories from the original questionnaire). Quote: "Government was unwilling to publicly share negative content of initial responses" (around 1/7 of all non-published studies).
6. "Only meant for internal use": Results were only meant for internal use; there was no academic or scientific interest to publish. Quote: "The evaluation was only meant for the own organization; academic output is not necessary" (around 1/7 of all non-published studies).

See Appendix (available as an on-line data supplement at [www.jamia.org](http://www.jamia.org)) for a more complete list of comments from the respondents.

## Discussion

Based on a limited pilot e-mail survey, the authors found that about half of current study respondents' IT evaluation studies were reported in international publications. Stated reasons for non-publication were diverse, including unclear generalizability of results obtained in a local context, missing time or budget to write down the evaluation results, doubts on scientific quality of the study, political and legal reasons (including "publication bias," i.e., non-publication due to negative results), and studies conducted only for internal use without any academic research interest. With regard to the last point, the authors believe that results of reasonably conducted evaluation studies, independent of the results, should be made publicly available, even though such studies may lack innovative methods or new results.

This pilot survey was an early attempt to quantify rates of non-publication of informatics evaluation results and to explore reasons for this. Limitations of the current study include its lack of information about the respondents (e.g., how well were academics who publish regularly represented, versus non-academic who rarely publish), lack of analysis of non-respondents, and the low survey response rate, with less than 20% of those contacted providing information. As a result, it is unclear whether the results are representative for the overall informatics community—the

scientific validity of the pilot study should be therefore judged carefully.

Most respondents came from academia, and 1/4 of them were selected because they already had published evaluation studies—so survey results have a large academic bias, with possible overestimation of publication rates. We do not know what motivated respondents to participate in this survey, and thus its results are prone to many potential forms of sampling bias—for example, the number of evaluation studies, academic position, professional background, national background, language, and other factors characterizing intended subjects. This means that the numbers given in the results can only be illustrative, and not considered representative for the IT evaluation community in health informatics.

While its validity may be discussed, the current study's results are at least consistent with studies from other domains. For example, Dickerson<sup>12</sup> analysed 204 RCTs in health care and found that 50% had not been published. Reasons for non-publication reported in the latter study included negative results ( $n=35$ ), lack of interest ( $n=16$ ), article planned or in progress ( $n=15$ ), or methodological problems ( $n=5$ ). However, there are two aspects that seem to differ. First, negative publication bias seems to be a larger issue in clinical fields. In our survey, only three respondents explicitly indicated lack of publication due to negative results. A 2001 JAMIA Editorial exploring issues related to negative publication bias in healthcare informatics<sup>11</sup> also came to the conclusion that publication bias in health informatics is not a major reason for non-publication. Study registers being developed for clinical trials such as the AMIA Global Trial Bank and also being promoted for IT evaluation studies (e.g.,<sup>18</sup>) seem to the authors to not be taken in general as a high priority in health informatics. Second, in our survey, a rather high number of authors mentioned "limited scientific quality," pointing to methodological problems within the evaluation study that prevented publication. Reasons for this can comprise the complexity of IT evaluation studies which is often not optimally planned beforehand, or the feeling that the results are not easily generalizable to other settings.

In the authors' opinion, a broad problem exists, which is to establish a foundation for Evidence-Based Health Informatics (EBHI) through providing a means to access results of all systematically-conducted IT evaluation studies. We believe that several steps can be taken in this direction:<sup>7</sup>

- a) Increase the number of IT evaluation studies by providing higher academic or monetary rewards, and by reserving a fixed amount of the budget of each IT project for evaluation.
- b) Develop Guidelines for Evaluation Practice in Health Informatics to increase the methodological and scientific quality of IT evaluation studies (see, as one example of many efforts, ongoing activities on GEP-HI at <http://iig.umat.at/efmi>).
- c) Develop Guidelines for Reporting on Evaluation Studies in Health Informatics to increase the quality of IT evaluation submission (see, as one example of many efforts,

ongoing activities on STARE-HI at <http://iig.umat.at/efmi>).

- d) Increase accessibility of evaluation studies, e.g. by developing open repositories for IT evaluation studies (see as an example the AMIA Global Trial Bank at <http://www.amia.org/gtb> or the IT Evaluation Database at <http://evaldb.umat.at>).

## Conclusion

The study mentioned in this viewpoint paper was a preliminary, and potentially biased, early attempt to explore and quantify the various reasons for non-publication of IT evaluation studies in health informatics. It suggests that possibly half of all IT evaluation studies are never published. The authors believe that further studies are required to better quantify the exact nature of non-publication in our field, and to determine how best to make results of evaluation studies accessible by means other than traditional peer-reviewed publications, such as repositories of evaluation studies such as the Evaluation Database EvalDB or the AMIA Global Trial Bank.

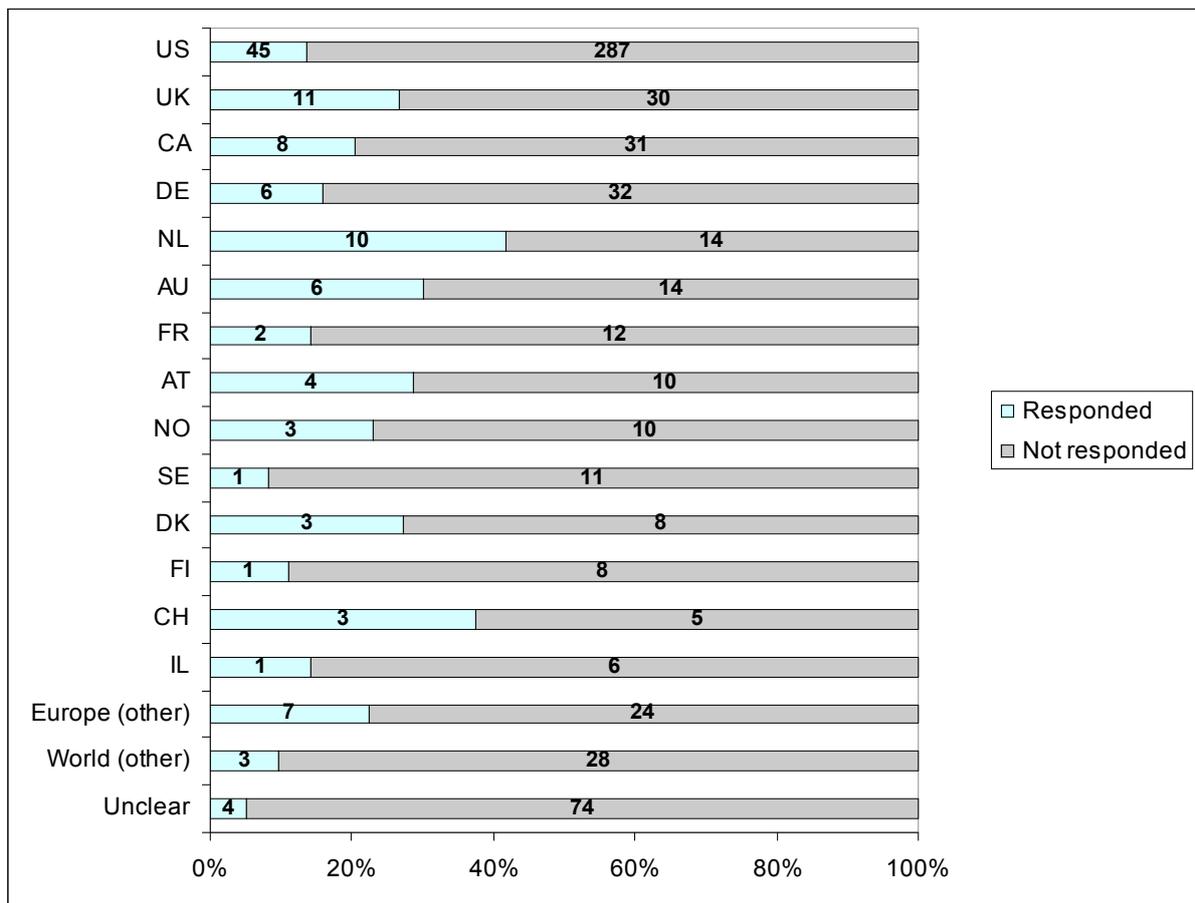
## References ■

1. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;144(10):742–52.
2. Garg A, Adhikari N, McDonald H, Rosas-Arellano M, Deveaux P, Beyene J, et al. Effects of computerised clinical decision support systems on practitioner performance and patient outcomes. A systematic review. *JAMA* 2005;293:1223–38.
3. Rothschild J. Computerized physician order entry in the critical care and general inpatient setting: a narrative review. *J Crit Care* 2004;19(4):271–8.
4. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005;116(6):1506–12.
5. Ammenwerth E, Shaw N. Bad health informatics can kill—evaluation the answer? *Meth Inf Med* 2005;44:1–3.
6. Koppel R, Metlay J, Cohen A, Abaluck B, Localio A, SE K, et al. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005;293(10):1197–2003.
7. 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 Inf* 2004;73(6):479–91.
8. Rigby M. Evaluation: 16 Powerful Reasons Why Not to Do It—And 6 Over-Riding Imperatives. In: Patel V, Rogers R, Haux R, editors. *Proceedings of the 10th World Congress on Medical Informatics (Medinfo 2001)*. Amsterdam: IOS Press; 2001. p. 1198–202.
9. Sackett D, Rosenberg W, Gray J, Haynes R, Richardson S. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–72.
10. Ammenwerth E, de Keizer N. An inventory of evaluation studies of information technology in health care: Trends in evaluation research 1982–2002. *Methods Inf Med* 2005;44:44–56.
11. Friedman C, Wyatt J. Publication bias in Medical Informatics. *J Am Med Inform Assoc* 2001;8(2):189–191.
12. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263(10):1385–9.
13. Heath D. The reporting of unsuccessful cases. *Boston Med Surg J* 1909;August 19;1090:263–4.

14. Olson CM, Rennie D, Cook D, Dickersin K, Flanagan A, Hogan JW, et al. Publication bias in editorial decision making. *JAMA* 2002;287(21):2825–8.
15. Phillips CV. Publication bias in situ. *BMC Med Res Methodol* 2004;4:20.
16. Scholey JM, Harrison JE. Publication bias: raising awareness of a potential problem in dental research. *Br Dent J* 2003;194(5):235–7.
17. Tierney W, McDonald C. Testing informatics innovations: the value of negative trials. *J Am Med Inform Assoc* 1996;3(5):358–9.
18. De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered?—A statement from the International Committee of Medical Journal Editors. *N Engl J Med* 2005;352(23):2436–8.

## Appendix 1

Number of people contacted from each country, and percentage of people that responded. Numbers indicate absolute numbers of respondents resp. and of non-respondents. Overall number of respondents: n=118.



**Appendix 2:**

Type of IT systems evaluated in the 217 reported IT evaluation studies (comprising both published and un-published studies).

<i>Type of evaluated IT system</i>	<i>Number</i>	<i>Percentage</i>
EPR/EHR system	28	13%
CPOE and medication systems	23	12%
General clinical information systems	22	10%
Decision-support systems	19	9%
Telemedical and telemonitoring systems	15	7%
Patient information systems	12	6%
Mobile IT tools	10	5%
Intensive care information systems	9	4%
GP systems	8	4%
Educational systems	8	4%
Nursing systems	7	3%
PACS and RIS	6	3%
Laboratory systems	6	3%
Operation room systems	5	2%
Websites	4	2%
Other/Unclear	35	16%
Sum of reported evaluation studies	217	100%

**Appendix 3:**

Place of publication for the 103 evaluation studies that were internationally published.

<i>Place of publication</i>	<i>Number of studies</i>	<i>Examples of journals/proceedings (as indicated by respondents)</i>
Study is published <i>only</i> in health informatics journals	31 studies (30%)	J Am Med Inform Assoc, Int J Med Inform, Methods Inf Med, Health Inf Manag, Inform Prim Care, Health Inform J, J Med Internet Res, Med Inform Internet Med, MD Comput, Comput Biomed Res, Int J Clin Monit Comput, Comput Nurs, BMC Med Inform Decis Mak, Med Inform
Study is published <i>only</i> in health informatics proceedings	20 studies (19%).	Medinfo conference, AMIA Conferences, MIE Conferences, Nursing Informatics (NI) conference, EFMI STC
Study is published <i>only</i> in medical, nursing or pharmaceutical journals	30 studies (29%).	Acad Med, Acute Care, Am J Health Syst Pharm, Am J Prev Med, Am J Trop Med Hyg, Anesth Analg, Ann Human Biol, AORN J, Arch Dermatol, Arch Pediatr Adolesc Med, Aust Health Rev, BMJ, Can J Anaesth, Cancer Detect Prev, Diabetes Educ, Eur J Anaesthesiol, Intensive Care Med, J Clin Anesth, J Exp Child Psychol, J Telemed Telecare, Med Clin (Barc), Nicotine Tob Res, Pediatrics, Pediatr Crit Care Med, Support Care Cancer, Transplantation.
Study is published <i>only</i> in other types of journals (no health informatics and no medical/nursing/pharmaceutical journal)	11 studies (11%).	IEEE Proc, Patient Educ Couns, J Healthc Qual, Int J Qual Health Care, J Am Med Dir Assoc, Int Libr Rev, Healthc Inf Manage, Health Care Manage Rev, J Educ Multimedia and Hypermedia, Health Policy
Study is published <i>both</i> in health informatics journals <i>and</i>	6 studies (6%).	see above

health informatics proceedings

Study is published *both* in 3 studies see above  
medical journal *and* health (3%)  
informatics journals or

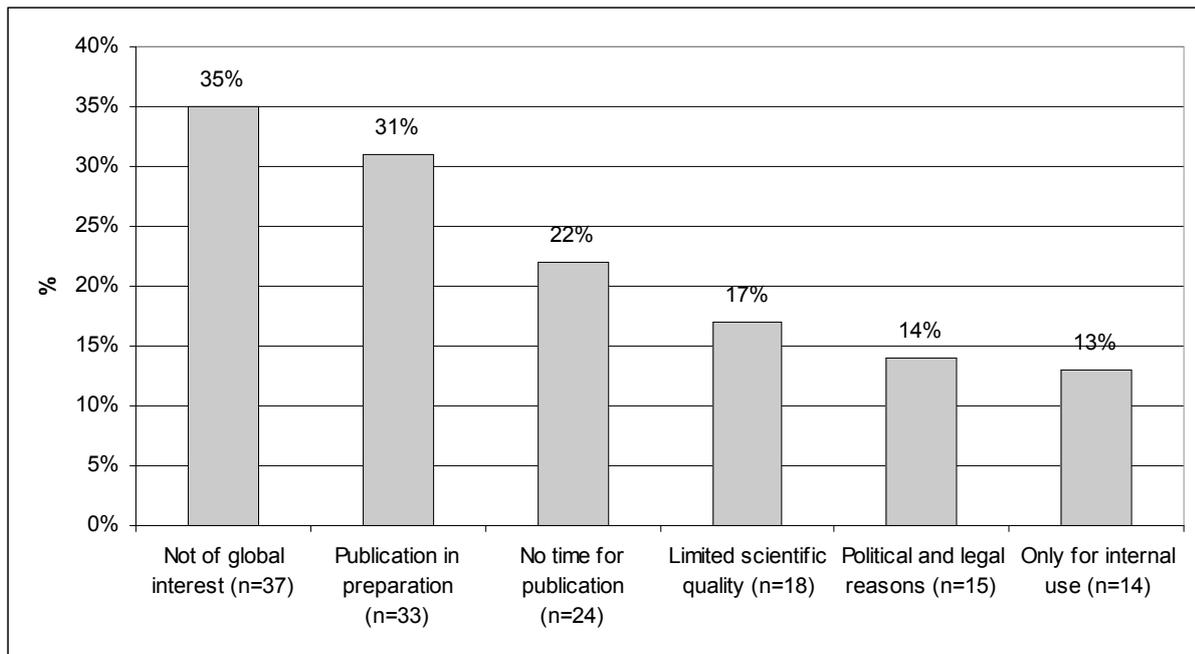
proceedings

Unclear data 2 studies  
(2%)

*Sum:* 103 studies  
(100%)

**Appendix 4:**

Percentages of reasons for not publishing 107 evaluation studies. N indicates the absolute number of studies, the percentage indicates the relative numbers of studies for which each reason was given (multiple nominations for one study were possible).



**Appendix 5:**

Reasons for not publishing evaluation results for 107 studies that have not or only locally been published.

**Comments from respondents (selected examples)**

<b>Not of global interest</b>  (Generalisability of results seemed too limited / Results seemed not to be of interest for others)	<p>“Hard to generalize the results because the evaluated underlying features/functions have been specific to the applications; these features/functions are not exposed in real-life applications”</p> <p>“The population is unique due to employment status (salaried govt employees).”</p> <p>“Constellation of internal social factors, adoption factors, staff training/experience, etc. seemed too unique to make it general enough.”</p> <p>“There was no real innovative characteristic in the software.”</p> <p>“Not sure the paper made a substantive contribution.”</p> <p>“But the main reason is a feeling that, in retrospect, we doubted the generalisability of results.”</p> <p>“The results and systems are often too specific to interest audience in other countries (academic and other international journals).”</p> <p>“When reasons for the lack of efficacy were related to our institution, then of limited interest to others.”</p> <p>“It were constructive evaluations on immature prototypes, implying that there is no scientific news.”</p> <p>“The evaluation (usability inspection and usability tests) was made on a prototype, and was promising, but the Company could not finish the application, then the evaluation was too limited.”</p>
<b>Limited scientific quality</b>  (Methods used seemed not adequate/Submission	<p>“The setup (e.g. amount of interviews) was not robust enough.”</p> <p>“Difficulties in establishing a sampling frame and therefore unable to measure non-response bias.”</p> <p>”The methods used seemed not exhaustive enough.”</p> <p>“Limited data, inconclusive.”</p> <p>“But the main reason is a feeling that, in retrospect, the study design was</p>

was not accepted)

weak.”

“Submission was first not accepted for Proceedings due to insufficient results.”

“Reasons for rejection: Sample size too small (larger sample size could not be achieved).”

“Besides, we faced an evaluation problem there: the new system induced some profound changes in the way Healthcare professionals worked, some of them we did not anticipated on a detailed level. Consequently, some of our “before” quantitative and qualitative indicators could not be compared with equivalent “after” indicators. This made the potential scientific paper complex and difficult to explain: we simply gave up, although we had some interesting results.”

**Political and legal reasons**

“The organization I work(ed) in prohibited publication (well, not expressly, but it was made clear that such a publication would not be welcome).”

(Publication was prohibited/Results were too negative)

“The organisation did not allow me to publish any materials related to the study. I have gone through several rounds of negotiation but it wasn’t successful.”

“Government was unwilling to publicly share negative content of initial responses.”

“Competitive reasons prohibited publication – These results are considered proprietary and have not been published.”

“The organization I work would most likely prohibit publication of what was an ‘internal’ audit and evaluation.”

“Unable to suitably anonymise data due to specific nature of participating clinics and my own employing organization at that time.”

“Details of tender evaluations are confidential.”

“I’m not sure the Company would have allowed the paper if we’d come to write it.”

“Not published because there is business confidential information in them.”

“The results were not all positive. The system administrators and stakeholders of the system would not like to accept the result.”

“The results were too negative, not only regarding the system that the hospital wanted to buy (and actually did), but also regarding the organisation of the emergency department, the work processes, the attitudes of healthcare professionals and so on.”

“At some point in the wait for a decision by the journal, I was informed by my University that EU patent applications would be invalidated if I published the paper.”

“We can only publish scientific data obtained through a research protocol approved by our Institutional Review Board. It is very difficult to do innovative health information technology projects on a formal protocol.”

“Not subjected to our university ethics process, therefore not published.”

**Only for internal use**

(Results were only meant for internal use)

“The study was not actually research-centred “

“In the last 5 years we did no evaluation studies of academic value, the results were only of internal interest.”

“The evaluation was only meant for the own organization, academic output is not necessary.”

“Only for internal business use/quality improvement.”

“Focus was on optimization of IT and alignment with clinical workflows in complex environments.”

“The data was collected for internal use within our government agency to monitor allocation of resources and return on investment.”

“Publication for general audience was not part of the contract and ownership of the results pass to the government.”

“The evaluations were done for a government agency. Once the reports were delivered, no further activities were funded. The majority of authors were consultants, not academics. Publication is not of interest for consultants.”

**No time for publication**

“Overwhelmed with ongoing tasks to implement EMR.”

“Too busy implementing CPOE to publish.”

“Other commitments delayed data analysis to the extent that publication would have been too long after the work was done. This reflects poor time planning on my part, but also the fact that this evaluation work was only a low priority component of the whole project.”

“Funding ended and workload of cooperating physicians was too high.”

“Always a lack of time when funding is coming from outside.”

“The EU-project ended. By the time all the deliverables for the Commissions had been done - new proposals had to be written, new projects started.”

“Funding ran out more than a year ago, analysis was never completed, but the preliminary results did not indicate this would be of sufficient interest to compete with higher-priority publication efforts.”

**Publication in  
preparation**

“We may still pursue publication.”

“May publish following validation. Probably not in [journal name] which seems to be closed shop for those outside of the academic communities.”

(Publication is  
planned/ Study not  
yet finished)

“We are still busy with improving the system before we will perform the described evaluation study.”