Methodological Review

Methods for studying medical device technology and practitioner cognition: The case of user-interface issues with infusion pumps

Jan Maarten Schraagen [a,*], Fenne Verhoeven [b]

[a] TNO Behavioral and Societal Sciences, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
[b] University of Applied Sciences Utrecht, P.O. Box 13102, 3507 LC Utrecht, The Netherlands

ARTICLE INFO

ABSTRACT

Purpose: The aims of this study were to investigate how a variety of research methods is commonly employed to study technology and practitioner cognition. User-interface issues with infusion pumps were selected as a case because of its relevance to patient safety.
Methods: Starting from a Cognitive Systems Engineering perspective, we developed an Impact Flow Diagram showing the relationship of computer technology, cognition, practitioner behavior, and system failure in the area of medical infusion devices. We subsequently conducted a systematic literature review on user-interface issues with infusion pumps, categorized the studies in terms of methods employed, and noted the usability problems found with particular methods. Next, we assigned usability problems and related methods to the levels in the Impact Flow Diagram.
Results: Most study methods used to find user interface issues with infusion pumps focused on observable behavior rather than on how artifacts shape cognition and collaboration. A concerted and theory-driven application of these methods when testing infusion pumps is lacking in the literature. Detailed analysis of one case study provided an illustration of how to apply the Impact Flow Diagram, as well as how the scope of analysis may be broadened to include organizational and regulatory factors.
Conclusion: Research methods to uncover use problems with technology may be used in many ways, with many different foci. We advocate the adoption of an Impact Flow Diagram perspective rather than merely focusing on usability issues in isolation. Truly advancing patient safety requires the systematic adoption of a systems perspective viewing people and technology as an ensemble, also in the design of medical device technology.

Contents

* Corresponding author. Fax: +31 346 353 977.
  E-mail addresses: jan_maarten.schraagen@tno.nl (J.M. Schraagen), fenne.verhoeven@hu.nl (F. Verhoeven).

1. Introduction

Designing for the safe use of medical device technology is an overriding concern for medical device manufacturers, human factors engineers, practitioners, and regulatory bodies alike [1]. Frequently, the design activity takes the perspective of an artifact as an object rather than as a hypothesis about how the artifact shapes cognition and collaboration [2]. As pointed out by Woods [2], standard human factors practice, driven by time and resource constraints, insulates the underlying concepts about how the system will support practitioners from results of usability testing of specific features and choices. In contrast, the Cognitive Systems Engineering contribution to design is not about the artifact as object, but about how the artifact is part of and transforms the distributed cognitive system.

One of the fundamental findings of Cognitive Science is that artifacts shape cognition and collaboration [3]. Technological artifacts impact cognition in context by representing work demands and underlying processes in particular ways. For instance, by showing only current status, some medical devices neither account for events that preceded the current state, nor indicate what to expect in the future [4]. This failure to develop representations that reveal change and highlight events in the monitored process has contributed to incidents where practitioners using such opaque representations miss operationally significant events due to impaired mental models [5]. Hence, there is a direct link from technological choices embodied in artifacts, to deficient cognitive processes in operational contexts, to incident evolution. This is one of the main reasons to go 'behind human error' [5] and develop a more extensive systemic analysis of incidents.

Cognitive Systems Engineering employs a wide variety of methods to study technology and practitioner cognition, ranging from ethnomethodology and conversational analysis, to cognitive work analysis and controlled studies [6–9]. However, there have been few systematic studies to answer the question how one should study the interaction of complex tools, cognition, collaboration and context in the field setting or workplace, in terms of strengths and limitations of various methods. Woods [10] is an exception, even though he did not systematically compare strengths and limitations of various methods. Woods emphasized a family of methods he termed 'staged world studies' in which investigators stage situations of interest through simulations of some type. For instance, by introducing disruptions and contrasting conditions relative to the artefacts, the investigator may observe, by using process-tracing methods, how the distributed cognitive system responds. Our aim in this study was to map methods to the way artifacts shape cognition and collaboration. In particular, we developed an Impact Flow Diagram (adapted from [5]) showing the relationship of computer technology, cognition, practitioner behavior, and system failure in the area of medical infusion devices. We subsequently conducted a systematic literature review on user-interface issues with infusion pumps, categorized the studies in terms of

methods employed, and noted the usability problems found with particular methods. Next, we assigned usability problems and related methods to the levels in the Impact Flow Diagram.

As our approach is primarily descriptive and retrospective, our results will give an indication of the current practice of discovering user-interface issues with a particular medical device. By focusing on user-interface issues, we run the risk of limiting the field of view of our conceptual looking glasses, for instance by ignoring collaborative or organizational aspects. In order to limit the risk of this bias, we will devote a special paragraph to the larger context surrounding user-interface issues, and illustrate this with a selected case study from our literature review.

1.1. Medical infusion devices and use-related hazards

Infusion pumps are medical devices that deliver fluids into a patient's body in controlled amounts. Although infusion pumps have contributed to improvements in patient care, they are not without risks. For instance, from 2005 through 2009, the U.S. Food and Drug Administration (FDA) received approximately 56,000 reports of adverse events associated with the use of infusion pumps, including numerous injuries and deaths [1]. In the UK, at least 700 unsafe incidents with infusion pumps are reported each year [11]. The FDA distinguishes between three types of reported problems: software defects, user interface issues, and mechanical or electrical failures [1]. This systematic review focuses on user interface issues.

That user-interface issues with infusion pumps are widely regarded as a serious issue, is reflected by the FDA's recent initiative to improve pump safety [1]. In order to assure that use-related hazards have been adequately controlled, the FDA [12] states that three central steps are essential:

1. Identify anticipated use-related hazards (derived analytically, for instance by heuristic analysis) and unanticipated use-related hazards (derived through formative evaluations, for instance simulated use testing).
2. Develop and apply strategies to mitigate or control use-related hazards.
3. Demonstrate safe and effective device use through human factors validation testing (either simulated use validation testing or clinical validation testing).

The analytical approaches and formative evaluations are complementary, each having unique strengths and weaknesses with respect to identifying, evaluating, and understanding use-related hazards early in the design process. Formative evaluations can demonstrate sufficient use-safety for an infusion pump. Formative evaluation has its strengths in a focus on critical tasks, challenging or unusual use scenarios and the follow-up to determine the cause of task failures. Potential limitations of formative evaluation include artificial testing conditions and limited range of users and use conditions. Clinical validation testing has its strengths in

realistic testing conditions (e.g., time pressure, distractions, noise, glare), a broader range of users, and unanticipated use conditions, but potential limitations include lack of control over use scenarios and testing conditions.

Although the focus on use-related hazards is important, from a Cognitive Systems Engineering perspective one runs the risk of considering the artifact, in this case the infusion pump, as an object rather than as a hypothesis about how it shapes cognition and collaboration. In practice, this focus implies that specific features of the artifact are iteratively improved by usability testing while one remains blind to how more fundamental, frequently implicit, technological choices impact representations and cognitive processes. Just as the switch to the 'glass cockpit' in airplanes led to certain pilot actions becoming invisible to the co-pilot, the choice for computer technology in developing infusion pumps necessarily implied the implicit adoption and acceptance of certain generalizable characteristics of computer technology, such as the 'keyhole representations' of large data sets [5]. The implication for design methods is that the representations and cognitive processes should be at the core of one's attention, in addition to more traditional outcome measures. One family of methods are the process-tracing techniques, such as verbal protocols, and knowledge elicitation techniques (see [6] for a review). Ultimately, the purpose of these methods is to inform the design of systems for cognitive work from the point of view of people working in fields of practice.

As we expected process-tracing methods to be relatively unknown in the area of medical device technology, the primary aim of this study was to conduct a systematic review on the focus of methods commonly used in discovering user-interface issues with infusion pumps. The rationale for this review and focus is, first, that user-interface issues with infusion pumps have high relevance for patient safety, as human factors are commonly considered to be the leading cause of dosing errors [13], frequently resulting from pump programming errors [14,15]. Second, the case of infusion pumps is highly suitable as numerous studies have been carried out into user-interface issues with infusion pumps, thus providing a potentially large database to draw upon. The current review follows the PRISMA statement for systematic reviews to the extent permitted by the resulting extracted literature [16].

2. Methods

2.1. Eligibility criteria

Literature was sought dealing with the user interface (or usability, human–machine, programming) of infusion pumps (or intravenous pump, infusion device, Patient-Controlled Analgesia [PCA]). This particular focus does not lend itself easily to be formulated in a Population, Intervention, Comparison, Outcome, Study design (PICOS) question, as we did not want to restrict ourselves to a specific population and a specific intervention. By combining human factors or human–machine interface (HMI) issues on the one hand with the particular application area (infusion pumps) on the other hand, we expected to retrieve a manageable number of records. We restricted the reports retrieved to the years 1990–2011 and only included reports written in English.

2.2. Information sources

The search was conducted in the Scopus database, which includes PubMed and all relevant human factors journals. Date last searched was August 1, 2012.

2.3. Search

The search string used was:

(TITLE-ABS-KEY("human factors" OR ergonomics OR interface OR "user-computer" OR "human-machine interaction" OR usability OR hmi OR mmi OR programming))
AND
(TITLE-ABS-KEY("intravenous pump" OR "Patient-Controlled analgesia" OR "Patient controlled analgesia" OR "infusion pumps" OR "infusion pump" OR "intravenous pumps" OR "IV pump" OR "IV pumps" OR "infusion device" OR "infusion devices"))

2.4. Study selection

Screening of records was carried out by the first author based on full abstracts. Articles that evidently addressed only mechanical issues and/or technical issues with the delivery of fluids were excluded. Subsequently, the remaining full-text articles were assessed for eligibility. Articles that did not present empirical data, were insufficiently detailed (e.g., [33,44]) or highly deficient methodologically, review articles, or opinion articles (e.g., Letters to the Editor) were excluded in this step. Studies focusing on highly specific equipment problems [39,43], or specific procedures (e.g., handwritten versus computerized orders [27]) were also excluded.

2.5. Data collection process

As this review was not a quantitative meta-analysis, a qualitative summary was written for each study included during data extraction. In accordance with PRISMA [16], the following items were included in the summaries: sample size, sample characteristics (e.g., experience, clinical area), study period, study location (e.g., size and type of hospital), error-reporting database inspected, type of intervention (e.g., organizational, interface), tasks to be carried out, design issues (e.g., counterbalancing order, within- or between-subjects, repetitions), evaluation criteria, types of pumps used, measures (e.g., dosing errors, critical incidents, acceptance, mode errors, time taken to complete tasks, preference, workload). Due to the different nature of the studies retrieved, not all items were included in each summary.

2.6. Risk of bias in individual studies

Risk of bias at the study level was assessed by comparing several methodologically similar studies and noting differences, assessing these differences, and noting limitations of the studies as reported by the authors themselves. For instance, studies not controlling for order in which different interfaces are evaluated are subject to a higher risk of bias than studies in which order is counterbalanced. Risk of bias at the outcome level was assessed by recording whether outcome measures were based on self-reports, expert judgments, observation of user behavior, or actual readouts from pump databases. For instance, studies that heavily rely on self-reports in error databases are more prone to bias at the outcome level than studies that directly observe programming errors.

2.7. Risk of bias across studies

Chan et al. [17] reported that incomplete outcome reporting is common in randomized trials. Overall, 50% of efficacy and 65% of harm outcomes were incompletely reported (e.g., precise p-values or effect sizes were not reported). Whether this should be called a 'bias' or a reporting convention is a matter of debate. Furthermore,

the demands put upon the researchers in reporting outcomes in terms of levels of statistical detail are subject to change over time. Given that we included studies from 1990 onwards, substantial differences in outcome reporting may be expected. Given, also, that our focus in this review was not on comparing exact outcomes across studies, but rather on the different methodologies used, we decided not to focus on incomplete outcome reporting.

3. Results and discussion

3.1. Study selection

Fig. 1 shows the flow of information through the different phases of the systematic review:

The study selection process (see Fig. 1) shows that no duplicates were removed. In some cases, notably the work by Garmer et al. [18], Lin et al. [19,20], Obradovich and Woods [21], and Wetterneck et al. [22], the same research was presented first at a conference and was later published in a journal. We decided to retain all versions, as slightly different aspects were emphasized in each version. If one were to consider these studies as duplicates, five records would be removed.

3.2. Categorization of studies

The 47 studies included in the final analysis differed widely in terms of methodology used. We decided to categorize the studies in the following categories: experimental comparison ($N = 8$), heuristic evaluations of existing pumps ($N = 4$), medical device evaluation in hospital procurement ($N = 4$), observational studies ($N = 9$), pre–post intervention studies ($N = 9$), retrospective analyses ($N = 6$), and case studies ($N = 7$).

This categorization scheme was informed by established methodological sources such as Shadish, Cook, and Campbell [51]. However, in order to better capture the richness involved in the various studies, we decided to subcategorize studies further. For instance, the categories "Heuristic evaluations of existing pumps" and

"Medical device evaluation in hospital procurement" should theoretically be placed under the main category of "Observational studies". Yet, this would have neglected large methodological differences among these studies as well as potential interesting outcome differences. The final resulting categorization scheme is therefore not so much theoretically valid as well as heuristically valid for this particular domain of research (user interface issues with infusion pumps). Eventually, all 47 studies could be uniquely assigned to one category (when a particular study consisted of more than one methodological approach, the dominant approach was chosen for classification).

In Table 1, the 47 studies are grouped according to study type, and further subdivided into study methods, variables, findings, and methodological limitations. Section 3.3 will relate the study findings to the Impact Flow Diagram, whereas Section 3.4 will map the methods to the Impact Flow Diagram. Section 3.5 will discuss the relative strengths and weaknesses of the methods, based on limitations noted by the authors themselves, and informed by considerations from methodological sources such as [51]. Section 3.6 will describe a case study expanding the analysis to include coordination and organizational aspects.

3.3. Impact Flow Diagram

Table 1 shows a highly diverse list of findings. In order to structure these findings, we first focused on usability issues, our first topic of interest in this review. Second, we made the relationships explicit between the properties of the infusion pump as a medium, the way infusion pumps represent the underlying process for practitioners, and how these representations impact the cognitive and collaborative behavior of practitioners. These relationships are depicted in Fig. 2, in the form of an Impact Flow Diagram [5]. In the following discussion, we will draw upon individual studies to illustrate our general points.

Fig. 2 depicts how infusion pump technology is an instance of computer technology in general, and that there are design shaping properties of the computer medium that make it easy for designers
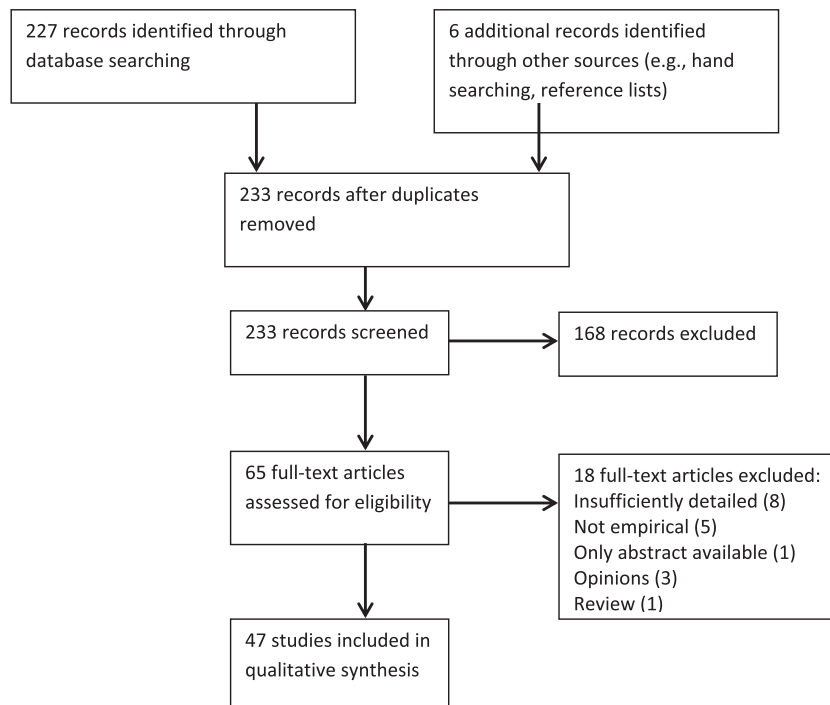

Fig. 1. Selection process used by the authors to identify appropriate published studies about user interface issues with infusion pumps.

Table 1
Study type and methods, variables, findings and methodological limitations for each study included in the final analysis.

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
Lin et al. [19,20]	Experimental	Comparison of existing with redesigned interface	Time, workload, errors, preference	Mean programming time on the New interface was significantly less than with the Old interface. The New interface led to significantly less workload than the Old interface and to significantly more reliable performance. 23 out of 24 participants expressed a preference for the New interface design	Test participants were not experienced nurses; clinical trials not conducted in field settings
Lin et al. [23,24]	Experimental	Comparison of existing with redesigned interface	Time, workload, errors, preference	Nurses made significantly fewer programming errors with the new interface (13 errors) compared to the old (29 errors). Programming time with the new interface was 18% faster. There was a nonsignificant 14% decrease in workload with the new interface over the old. Nine nurses stated they favored the new interface, 1 preferred the old, and 2 were neutral	Clinical trials not conducted in field settings (e.g., impact of interruptions not studied)
Garmer et al. [18,25,26]	Experimental	Comparison of existing with redesigned interface	Time, errors, preference, use of manual	Time to complete test tasks was significantly longer for the existing interface (260 s) as compared to the new interface (188 s). Differences in number of errors between interfaces were not significant. Subjective data (questionnaires) showed that when subjects used the new interface they thought they had better control of operations, were more sure they had set the infusion correctly and that it was easier to correct errors ($p < .05$).The manual was used 29 times for the existing interface but only 8 times for the new interface ($p < .05$)	Relatively small user groups ($N = 6$); clinical trials not conducted in field settings
Trbovich et al. [28]	Experimental	Comparison of three pump types: traditional, smart, barcode	% planted drug errors remedied	Pump type did not significantly impact the ability to remedy "wrong drug" errors. When provided with the flexibility to override limits, nurses often did so, even when clinically inappropriate	Large numbers of planted errors might have influenced participants to behave differently than they would under clinical circumstances where these errors occur less frequently
Zhang et al. [29]	Heuristic	Usability inspection of two volumetric pumps by four evaluators	Number of violations of heuristics; severity of violations	14 heuristics were applied. Consistency and Visibility were the two most frequently violated heuristics. Differences between pumps were found in terms of number and severity of heuristic violations	Heuristic evaluation does not identify major missing functionality; it requires both domain knowledge and usability expertise; it may not identify problems that arise because of the device's use environment, for instance, lighting and noise
Gagnon et al. [30]	Heuristic	User-centered evaluation to determine the effectiveness and usability of three frequently used infusion pumps	Problems, severity, positive and negative features	Both newer devices provided useful dosage calculation assistance, and useful feedback about the current state of the device. A significant shortcoming of both newer devices was the inability to navigate backwards through the infusion set-up screens to correct a previous entry. Neither manufacturer incorporated a Back-button/function	Evaluators lacked both domain experience and experience in heuristic evaluation. This led to considerable differences of opinion and replication of problems across evaluators using the same device
Graham et al. [31]	Heuristic	Four raters conducted a heuristic evaluation of a three-channel infusion pump interface	Violations of 14 usability heuristics	The most severe violations were spread out across at least 8 of the 14 usability heuristics. Two heuristics, "Consistency" and "Language", were found to have the most violations. *Consistency* demands that users should not have to wonder whether different words, situations, or actions mean the same thing. The *Language* heuristic demands that the intended users should always have the language of the system presented in a form understandable to them	Heuristic evaluation does not identify major missing functionality; it requires both domain knowledge and usability expertise; it may not identify problems that arise because of the device's use environment, for instance, lighting and noise
Turley et al. [32]	Heuristic evaluation of manuals instead of devices	Review of five medical device operating manuals	Information contained in the manuals was checked against usability heuristics	On the basis of the number of heuristics violated, the average severity rating, and the affordance violations, one particular pump received the highest recommendation	Method is entirely dependent on the information that the manufacturer provides

Table 1 (continued)

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
Ginsburg [34]	Mixed methods	Heuristic evaluation, user testing	Expert ratings, errors, user preferences	A discrepancy was found between results from user testing and user preference, because of prior experience with particular pumps, and because users rated devices on ease-of-use rather than safety	No novice users available or tested, small sample size in each clinical area, scenarios did not include all tasks but rather a sample of representative tasks, pump order not counterbalanced within each area, testing conditions not strictly controlled across participants or clinical area, some errors may have been missed by the observer as no video recordings were made
Namshirin et al. [35]	Mixed methods	Heuristic evaluation, cognitive walkthrough, usability evaluation, clinical evaluation	Violation of heuristics, user challenges, efficiency, number of errors, user satisfaction	Results of the heuristic evaluation coincided with those of the technical evaluation, and led to the removal of two pumps from the procurement process. The project was subject to stringent time constraints and usability analysis focused on the qualitative metrics rather than the quantitative ones. Largely based on questionnaire responses, one pump was chosen unanimously	Not clear whether the three representative tasks that were chosen for usability evaluation also constituted critical tasks. Behavioral metrics were largely ignored in favor of questionnaire responses, hence final pump chosen, while preferred by users, could be unsafe
Keselman et al. [36]	Mixed methods	Interviews, document analysis	Thematic coding categories, semantic relationships	Participants' conception of safety-relevant device aspects was somewhat narrow and there was no overall collective perception where all perspectives were represented. Administrators equated equipment-related safety with technical accuracy and reliability instead of usability issues	Retrospective bias may have influenced the interpretation of the data
Nemeth et al. [37]	Mixed methods	Expert analysis, usability assessment, adverse event self-reporting, field observation	Subject actions and comments, analysis of programming actions from files	A sample of 19 nurses was recruited for the usability sessions. Results showed no definite advantage for one of the pumps over the others. Subjects regularly ignored dose limiting software. Subjects did not benefit from their previous experience with a particular device. There were some discrepancies between what subjects said they found positive and their actual behavior	Testing was not carried out with instructions for use: nurses only received a brief orientation to pump operation before starting. No quantitative data are reported
Obradovich and Woods [21,45]	Observational (use problems)	Interviews, bench tests, observations of use	Error-prone tasks, device characteristics, context analysis, tailoring strategies	Main categories of use problems: (1) complex and arbitrary sequences of operation (2) different operating modes intended for different contexts (3) ambiguous alarms (4) getting lost: given the arbitrary command sequences and the lack of feedback, users can enter a command and be surprised by the result (5) poor feedback on device state and behavior	Unclear how many users were observed and whether the deficiencies observed were critical and representative of the full set of possible deficiencies
Liljegren et al. [40]	Observational (use problems)	Field studies, evaluation of pump use, incident analysis	Classes of incidents	A total of 13 types of incidents could be found, of which two were connected to the user interface: (1) switching the functions Volume To Be Infused and Flow Rate, which lead to the pump being set to deliver e.g. 27 mL at 350 mL/h instead of 350 mL at 27 mL/h. (2) misreading the numerical display i.e. reading 27.0 mL/h as 270 mL/h or 035 mL/h as 3.5 mL/h	Unclear how many users were observed and what tasks they had to carry out under what circumstances
Nunnally et al. [42]	Observational (use problems)	Video recording of pump programming, finite state analysis	Efficiency, choice of mode and sequence selection	Practitioners (anesthesiologists and ICU nurses) entered 57.1% more keystrokes than necessary to accomplish the tasks. 69.5% of all keystrokes used were goal-directed. More experienced users did not use more goal-directed keystrokes	Pump programming not studied in actual conditions. Tasks not identical across subjects
Ahmad et al. [15]	Observational (prospectively collected incidents)	Critical incident analysis	Types of critical incidents	Over a period of 60 months, 27 Critical Incidents (0.32%) were identified through self-report and investigated. Three main categories of incidents were identified: programming errors, breaches of policy and patient	Reporting of critical incidents partially dependent on self-report

Table 1 (*continued*)

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
				selection. Of the 27 CIs, 18 (66.6%) were due to programming errors and six were breaches in hospital policy. Of these 18, nine were incorrect bolus doses and the other 9 were incorrect drug concentrations	
Husch et al. [47]	Observational (prospectively collected incidents)	Single-day direct observation of every drug administration	Rate deviations and other errors	During the data collection period, 486 patients receiving infusions via PCA and general IV pumps were included in the study. An IV pump was used for 286 of these patients. Of the 389 errors noted overall, 37 were rate deviation errors and three of these were judged to be due to a programming mistake, while errors associated with orders, documentation, labeling and patient identification were more frequent	Data were collected on one particular day only, which may not have been representative for other days or other periods of the year
Taxis and Barber [48]	Observational (drug preparation and administration)	Direct observation of 113 nurses on 76 study days	Number and types of drug errors	265 IV drug errors were identified during observation of 483 drug preparations and 447 administrations. The most common type of error was the deliberate violation of guidelines when injecting bolus doses faster than the recommended speed of 3–5 min	Only one observer recorded drug errors, which may make results less reliable. The observer did not interview nurses in depth, as a result of which some information relevant to prescribing errors may have been missed
Brixey et al. [49]	Observational (legibility)	Observations by two observers of pump use	Ambient light level, photographs, field notes	For the pump used in this study, the only text that was clearly visible from the foot of the patient's bed was for the rate of infusion displayed in the uppermost screens. Legibility for the other screens was reduced because of the font size (3.1–4.7 mm) and background colors (black characters on a yellow background)	Study was limited to a convenience sample of a single model of a dual-channel infusion pump
Johnson et al. [50]	Observational (attitudes)	Questionnaires	Attitudes toward medical device use errors	Traditional view of blaming the operator was still prevalent	Limited sample size ($N = 26$) for an attitude survey
Adachi and Lodolce [14]	Pre–post intervention	Observation of pump-related errors before and after interventions	Pump-related errors as% of dosing errors	In 2003, pump-related errors accounted for 22% (10 of 46 errors) of dosing errors, compared with 41% (24 of 59 errors) in 2002. Although statistical tests were not reported, this is a statistically significant difference ($X^2(1) = 4.234$, $p = 0.04$)	Unclear whether pump-related errors decreased as a result of the interventions. Multiple interventions introduced at the same time, making it impossible to attribute success to one specific process change
Apkon et al. [52]	Pre–post intervention	Observation of resource consumption and staff satisfaction	Purchasing and pharmacy records; questionnaire	The combined effect of prolonging infusion hang times, preparation in the pharmacy, and purchasing premanufactured solutions resulted in 1500 fewer infusions prepared by nurses per year, with fewer opportunities for error. Nursing staff expressed a significant preference for the revised process	Actual failure rates were not measured. Multiple interventions were introduced simultaneously, making it impossible to attribute success to one specific process change
Carayon et al. [53]	Pre–post intervention	Three longitudinal surveys after introduction of smart IV pump	Implementation process; technical performance; usability; user acceptance	The main problems with the Smart IV pump technology reported by nurses included air-in-line alarms, and beeps resulting from a delay. Nurses' perceptions of pump reliability and noise did not improve after 1 year, despite the fact that nurses had been using the Smart IV pump for a significant amount of time. Nurses' perceptions of usability (e.g., learnability, efficiency, error recovery, and satisfaction) tended to improve 1 year after implementation	Actual pump use was not studied, nor how actual pump use influenced safety outcomes
Eade [54]	Pre–post intervention	Evaluation of intervention program to teach nurses how to program pump	Knowledge test, PCA errors	During a 5-month evaluation period, no errors occurred on the unit participating in the study, although PCA errors occurred on other units not participating	Lack of statistical analysis, lack of detail on the educational intervention, and how soon the evaluation period came after the intervention. Most importantly, 17 nurses and a 5-month period are in all likelihood too limited to note any errors, given that on average, every year only 3 errors were reported for the entire hospital under study

Table 1 (continued)

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
Ferguson et al. [55]	Pre–post intervention	Evaluation study of mandatory training for all registered nurses	Number of programming errors	A statistically significant decrease from 8 to 1 programming errors was found after the training program was completed by more than 900 nurses	Possible underreporting of errors, post-study due to awareness of the intervention time, changes in personnel over the course of 1 year, recency of the intervention (there was only a 1-month period between the intervention and post-intervention data collection, so that the effects may not sustain over time)
Moss [56]	Pre–post intervention	Improvement of PCA process by FMEA process	PCA errors	19% decrease in the number of reported PCA errors during 2009 compared with 2003 (21 versus 26). No statistical significance testing reported	Multiple interventions introduced simultaneously, making it impossible to attribute success to one specific process change. PCA errors not measured directly, but dependent on self-reporting
Paul et al. [57]	Pre–post intervention	Review of critical incident reports before and after safety interventions	PCA errors	In more than 25,000 patients using PCA pumps, errors occurred in 0.25% of the cases (62 in total), with negative effects (some harm, e.g., respiratory depression and uncontrolled pain, but no documented deaths) to one-third of these patients. 49 of the PCA errors occurred before the safety intervention and 13 after (odds ratio 0.28; 95% CI = 0.14, 0.53; $p < .001$). The most common causes of PCA errors were programming errors (33.9%). All 21 PCA programming errors occurred before the safety interventions were instituted (odds ratio 0.05; 95% CI = 0.001, 0.30; $p = .001$). For the total errors, 77.4% involved incorrect doses (48 of 62), with 59.6% of such errors being an overdose	This being a pre–post intervention without control study, it is possible that the observed reduction in errors was the result of factors other than the safety interventions. Vicente et al. [58] estimated mortality rates from pump-programming errors between 1 in 33,000 and 1 in 338,800, hence even the large sample size (25,000) in this study was likely too small to estimate mortality risk from PCA misprogramming. Furthermore, it is not clear to what extent the new PCA pumps or any of the other interventions, alone or in combination, contributed to the observed reduction in errors
Rothschild et al. [59]	Pre–post intervention	Non-blinded, prospective time series evaluation of effect of smart pumps on medication errors	Incidence and nature of medication errors and adverse drug events	Smart pumps did not reduce the rate of serious medication errors, in part because the pump setup made it easy for nurses to bypass the drug library (24% bypass rate) and because overrides were frequent	Power for detecting a decrease in the rate of life-threatening events was limited because of their low frequency. The extent of alert overrides and library bypasses may have reduced the effectiveness of the intervention. A randomized controlled trial could not be safely implemented
Wetterneck et al. [60]	Pre–post intervention	Evaluation of smart iv pump by FMEA	Failure modes	No specific before–after data were reported in this study	Impossible to evaluate effectiveness of design changes
Hicks et al. [13]	Retrospective analysis of error records	Analysis of voluntary reports to MEDMARX	% of error records associated with PCA	For a 5-year review period, 9571 (1%) of error records were associated with PCA—624 (6.5%) of which resulted in patient harm (from temporary harm to patient death). The leading type of error was improper dosage or quantity (38%), the majority of which occurred during drug administration	Precise numbers on the frequency of incorrect programming were not reported. Results are highly dependent on the quality of error-reporting, in particular the accuracy and completeness of the reports. Method does not yield detailed insights into usability issues, and may be subject to numerous biases, such as the outcome bias [63,64] and the hindsight bias [65]
Lori Brown et al. [61]	Retrospective analysis of error records	Screening of MDR database for 'use error'	Qualitative description of PCA adverse events	Three categories of PCA pump adverse events are described: product packaging, drug concentration programming, and improper administration set loading	The accuracy and completeness of these reports were not verified, and they varied greatly as to the level of detail. No quantitative results were reported
Thornburg et al. [62]	Retrospective analysis of error records	Categorization of adverse medical events reports	Frequency of occurrence	The three most frequent occurrences were: failure to open intravenous infusion "piggyback" medication bag clamp (23.1%), medication identification failure (13.7%), and pump programming (11.5%)	Event reports limited to three hospitals and a 1-year period. Results are highly dependent on the quality of error-reporting, in particular the accuracy and completeness of the reports
Malashock et al. [63]	Retrospective analysis of device alerts	Data download of smart infusion software	Device alerts	157 (18%) alerts resulted in reprogramming of the device, while users chose to override 696 (82%) alerts during the 8-month study period	Data about over-rides do not indicate how many of these events were true versus false alerts. If true and overridden, then safety issue; if false and frequent and overridden, then potential for future disregard of true alert
Rayo et al. [64]	Retrospective analysis of device alerts	Data download of smart infusion software	Device alerts	In 12% of the cases, the alert caused the clinician to change the input. The alert was overridden in 88% of the cases.	Data about over-rides do not indicate how many of these events were true versus false alerts. If true and

Table 1 (*continued*)

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
				56% of the overridden alerts were not readjusted to within the DERS's recommended limits	overridden, then safety issue; if false and frequent and overridden, then potential for future disregard of true alert
Kingman et al. [65]	Retrospective analysis of administration errors	Interviews and electronic survey	Self-report of prostacyclin administration errors	Serious errors in medication administration were reported by 68% of survey respondents. In separate interviews, 94% reported serious errors. Errors occurred for the following reasons: incorrect cassette placed in pump, due to identical appearance of cassettes; inaccurate pump programming; errant drug dosing; inadvertent cessation of the pump	True error rates could not be reported, as the number of in-patient prostacyclin exposure days for each respondent's center was unknown
Vicente et al. [58]	Case study	Qualitative event reconstruction, MDR database search	Patient records, autopsy report, toxicology results, interviews	A drug cassette containing 1 mg mL^{-1} solution of morphine was unavailable, so the nurse used a cassette that contained a more concentrated solution (5 mg mL^{-1}). The available evidence is consistent with a concentration programming error where morphine 1 mg mL^{-1} was entered instead of 5 mg mL^{-1}	Various factors contributed to the adverse drug event; it is unclear whether a programming error was the major factor. Results may not generalize to other types of pumps or other situations
Draper et al. [66]	Case study	Reproduction of infusion error with pump in question, database search	Syringe size, dosage size	The pump itself worked correctly, hence "no problem found". The problem originated with the "size override" function, which, when enabled, allows the operator to program the pump for a syringe size that differs from the standard program to allow for "nonstandard" infusion scenarios. In this case, the infusion error could be reproduced by setting the syringe size to the same size as the dosage size	"Size override" function may be specific to particular types of pumps. Measures to prevent the error (disabling the override function) may have unanticipated consequences. Reasons behind the infusion error are not uniquely identifiable
Musshoff et al. [67]	Case study	Toxicological analysis, analysis of pump history	Tissue distribution of piritramide	The PCA pump had been changed during a previous servicing from displaying mg/h to mL/h, therefore, the anesthetist had entered "1.5" assuming mg/h, but actually applying 1.5 mL/h (equivalent to 2.25 mg/h). The change of displayed units had been indicated by a red sticker on the backside of the pump	Various factors contributed to the adverse drug event; it is unclear whether a programming error was the major factor. Results may not generalize to other types of pumps or other situations
Perry [68]	Case study	Qualitative case description	None	Users thought they had stopped the infusion but actually had not. A device's operation and status should be apparent to the user	Case described in insufficient detail
Rule et al. [69]	Case study	Root Cause Analysis	Various safety issues	Several safety issues were identified: skipping steps in the checklist; making assumptions about the patient's implicit goals (saving time due to prior knowledge), time and workload pressure on the part of the nurse, the insulin pump not having fail-safe mechanisms to alert the user that the pivotal priming step had not occurred, company's training materials did not contain documentation of the change in priming steps, critical information about additional insulin that was administered in the clinic via syringe was not shared among all parties involved (company's representative, nurse, patient), clouding the subsequent interpretation of the patient's blood sugars	Cause for lack of insulin delivery with the new pump was multifactorial. It is unclear what was the major factor and what recommendations should be made on the basis of the Root Cause Analysis
Syed et al. [70]	Case study	Qualitative event reconstruction	Historical pump data, interviews, chart review	Morphine concentration was incorrectly programmed in an infuser: instead of 5 mg mL^{-1}, it was set at 0.5 mg mL^{-1}. This setting resulted in the administered dose being ten times greater than the prescribed dose (in this case, 20 mg boluses instead of	Retrospective event reconstruction is vulnerable to outcome bias and hindsight bias, particularly in the absence of adequate critical incident reporting

Table 1 (continued)

Study	Study type	Studymethods	Variables	Findings	Methodologicallimitations
Vicente [71]	Case study	Qualitative event reconstruction	Chronology of events	2 mg). Multiple caregivers, insufficient handover, incorrect assumptions, and distributed knowledge contributed to the adverse event The manufacturer initially exhibited the traditional approach to medical error for years, with an emphasis on better nurse training. This long period was followed by a comparatively abrupt shift toward human factors design. The shift was preceded by a 9-month period characterized by new leadership, a perception of poor organizational performance, and a disruption of the operating environment (e.g. pressures from government and public opinion)	Event reconstruction was not based on internal company documents. No valid method available for weighing the relative influence of events on cultural change

to create devices with typical flaws in human–computer operation. For instance, the general property of 'virtuality' means that there is nothing inherent in the computer medium that constrains the relationship between things represented and their representation. What needs to be represented is the larger therapy plan or dose–time relationships [45]. However, contemporary infusion device displays are limited to showing only current status, and offer no evidence of context that drove changes to infusion rate, nor of future implications of infusion rate changes [4,46]. The infusion device has the capability to 'make us smart' [3], yet it 'makes us dumb', as it does not answer the questions in the mind of a clinician.

Further, the 'keyhole' property of the computer medium, shared by infusion devices, means that the size of the available display units is very small relative to the size of the number of data displays that potentially could be examined. This particular property leads to some typical representational properties of the design, such as deep hierarchical levels with a vast number of programming pathways [42], complex and arbitrary sequences of operation [45], and different operating modes intended for different contexts [45]. In turn, these representational properties shape the cognitive systems involved and lead to increased memory demands and impair the development of accurate mental models of the pump, as demonstrated by Nunnally et al.'s [42] failure to find a relationship between level of experience and ability to use the pump. In the end, these cognitive systems have inevitable behavior shaping properties and their impact on operational processes is shown as programming errors [15,57,58,68] or 'mode errors' [40,45].

Finally, the general property of interactivity means that computer technology should make pertinent aspects of its status and intentions obvious, should enable a collaborative approach, and participate in managing attention to the most important signals without overwhelming the user with low-level messages. When not done properly, ambiguous alarms [45,53] and poor feedback on device state and behavior result [45]. Poor feedback and ambiguous alarms shape cognition by complicating situation assessment and enhancing stress on workload management. These properties of cognitive systems shape resultant practitioner behavior, in that alerts are frequently overridden [64,65], drug libraries are bypassed [59] in order to reduce stress on workload, and infusion is inadvertently stopped [66,69,70] because of poor feedback on system status.

In conclusion, the findings present a pattern that is representative for the generally unreflective use of computer technology by designers. The design shaping properties of the computer medium (e.g., virtuality, keyhole effect) stimulate designers to create devices with typical flaws in human–computer cooperation. In the case of infusion devices, typical flaws such as proliferating modes, making the system opaque, and providing poor feedback, create new cognitive demands, such as increased memory demands, impaired mental models, and poor situation assessment. These design deficiencies become problems that possibly contribute to incidents if other factors are present, such as distraction or increase in workload [70]. Although the Impact Flow Diagram may give the impression that the cognition-shaping properties of representations only affect individual caregivers, unreflective use of technology is in fact about miscoordination between the human and machine portion of a single ensemble, with the human portion frequently being distributed across multiple caregivers. Coordination across caregivers is an aspect that has not received sufficient attention in the literature reviewed here, although there are some hints of its importance [21,70]. In paragraph 3.6, we will re-analyze Syed et al.'s case study [70], by paying special attention to coordination and organizational aspects.

3.4. Mapping methods to the Impact Flow Diagram

The findings reported in Table 1, with the associated study type and study methods, were coded for presence of key words listed at the right-hand side of the Impact Flow Diagram (Fig. 2). Next, the associated methods were assigned independently by the two authors to one of the four levels in the Impact Flow Diagram (i.e., Computer Technology; Computer Based Devices; Joint Cognitive Systems; Infusion Pump Technology). A Cohen's unweighted Kappa of .75 showed good agreement between the two coders. Remaining discrepancies were resolved by discussion. This yielded the following mapping (see Table 2):

The results of the mapping process show a number of interesting points. First, the majority of the study methods employed in previous studies uncovering user interface issues with infusion pumps deal with the impact of behavior shaping properties of cognitive systems on operational processes, that is, use errors. Second, none of the methods employed dealt with general properties of computer technology. Apparently, these properties are not the direct focus of most study methods. Third, not surprisingly, the observational studies on use problems excel at determining the impact of the general properties of computer technology on the representational properties of the design. This is not surprising given that these observational studies, in particular [21,42,45], were carried out within a Cognitive Systems Engineering framework that formed the basis for the Impact Flow Diagram. Fourth, the impact of the cognition shaping properties of representations on
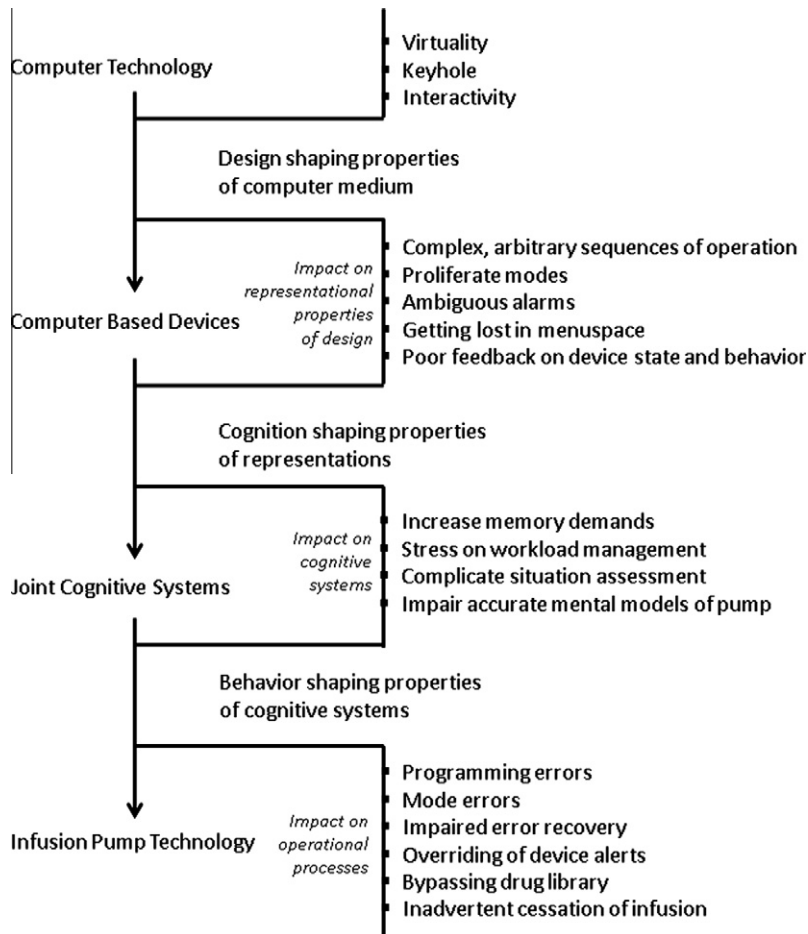
Fig. 2. Impact Flow Diagram defining a failure path for the uncritical use of infusion pump technology.

cognitive systems were dealt with only by case studies. Still, issues such as increased memory demands, complicated situation assessment and inaccurate mental models of pump design are being dealt with only sparingly in the studies retrieved. Finally, although the mapping process yielded some ambiguity regarding heuristic evaluations, inspection of the full list of heuristics in the primary sources, e.g., [29], made it clear that these heuristics focus primarily on representations.

In conclusion, covering all levels in the Impact Flow Diagram requires a combination of methods, in particular observational studies, case studies, heuristic analysis and experimental comparisons. Even then, these methods by themselves do not deal with general properties of computer technology.

3.5. Strengths and limitations of methods

Strengths and limitations of methods were derived from the limitations noted by authors themselves (listed in Table 1), in conjunction with general methodological sources such as [51]. The case studies (1), the heuristic evaluations (2), and the observational studies (3) excel at finding usability issues, ranging from quite specific in some case studies to more general in some observational studies. These usability issues are being dealt with in attempts to design new and improved interfaces for infusion pumps. Comparing these new interfaces with existing interfaces is a relative strength of experimental comparisons (4). These comparisons yield precise and quantitative data on the speed and accuracy with which programming tasks are carried out. Together, these four methods yield information on usability issues that stays closest

to the user interface. When used together in a sensible way, for instance in a mixed-methods study, the methodological limitations of these methods may be mitigated as they are complementary in some cases. It should be noted that some observational studies on use problems ([21,42,45]) and some experimental comparison studies ([19,20]) provided a wealth of information on mental representations and cognitive processes that went beyond observable behaviors. For instance, Lin et al. [19,20] carried out an extensive cognitive task analysis that served as a foundation for their newly developed interface design.

The other methods, retrospective analysis (5) and pre–post intervention (6), although broadening the scope of issues that may go wrong during the infusion process, suffer from a number of limitations. Retrospective analysis of medication error records is highly dependent on the quality of error-reporting, in particular the accuracy and completeness of the reports. Due to the fact that the researcher using these reports is dependent upon a third party for providing these reports, there is ultimately no control over data collection procedures and, hence, quality of data outcome. Retrospective analysis may give a very broad indication of the incidence of PCA-related errors, relative to other types of errors, its effects on patient harm, and its occurrence during particular phases of the medication-use process. However, in comparison with other methods, this method does not yield detailed insights into usability issues, and may be subject to numerous biases, such as the outcome bias [72,73] and the hindsight bias [74]. And, like all retrospective methods, it may prematurely attribute failure to "human error", it may overly simplify the dilemmas and difficulties practitioners face, and may not explain failure at all, but merely

Table 2
Mapping of study type/methods to Impact Flow Diagram levels.

Level	Number of studies	Study type/method
Computer technology	0	None
Computer based devices	10	Observational studies (use problems); Heuristic Evaluation
Joint cognitive systems	3	Case study
Infusion pump technology	34	User testing; Observational studies (prospectively collected incidents; drug preparation and administration); Pre–post Intervention; Experimental comparisons; Retrospective analysis; Qualitative event reconstruction; Root Cause Analysis

represents a primary reaction to failure [5]. Pre–post intervention is quite different in this respect, as it observes reductions in medication errors after a particular suite of interventions has been introduced. These interventions frequently go beyond relatively isolated changes in interface design, but rather involve training programs, changes in work procedures, and the introduction of smart pumps. Frequently lacking a control group and introducing multiple intervention measures simultaneously, these methods do not allow one to draw inferences about causality (see [51]).

Finally, medical device evaluation in hospital procurement (7) constitutes a retrospective reflection on the usefulness of various methods employed during the acquisition of new medical devices. It has yielded interesting issues to take into account during a procurement process. In particular, as already noted by Woods [2], there may be a tendency, due to time and resource constraints, to narrowly focus on user preferences rather than user behavior, and to equate safety with technical accuracy rather than usability. Underlying concepts about how the system will support practitioners are hardly ever being dealt with during hospital procurement processes.

In conclusion, from the perspective of design of cognitive work from the point of view of people working in fields of practice, there is a scarcity of methods that focus on tracing cognitive processes. Combining several methods (in particular, observational studies, heuristic analysis, and experimental comparison) may yield a broader picture, but only when the focus when using these methods is on uncovering representations and cognitive processes. There is nothing inherent in the methods themselves that prevent a researcher from narrowly focusing on observable behavior alone, nor in focusing on individual determinants alone. The next paragraph illustrates how to go beyond individual determinants, as well as providing an illustration of applying the Impact Flow Diagram in a case study.

3.6. Case study: how medical device technology and organizational policy shape cognition and collaboration

In order to prevent a narrow focus on individual cognition shaping properties to the exclusion of collaboration shaping properties, we will discuss a particular case study, [70], in somewhat more detail. It should be mentioned that this particular case study was not carried out from a Cognitive Systems Engineering perspective. However, since it used some typical Cognitive Systems Engineering methods (e.g., qualitative event reconstruction using interviews), it may be reinterpreted in terms of our Impact Flow Diagram.

In a hospital setting, morphine concentration was incorrectly programmed in a PCA infuser by two nurses: instead of 5 mg mL^{-1}, it was set at 0.5 mg mL^{-1}. The concentration programming error with this pump has been reported previously [58] and results from a low default setting as the initial choice. The most common programming error is to enter the default concentration. This setting resulted in the administered dose being ten times greater than the prescribed dose (in this case, 20 mg boluses instead of 2 mg). Because the PCA was incorrectly attached to the patient, the patient initially did not receive morphine. The incorrect concentration setting was discovered by a third nurse and the pump was reprogrammed by the second nurse. Still, the patient reported back pain. A fourth nurse later in the afternoon discovered the incorrect attachment and corrected the position of the back check valve. Shortly after this, the anesthesiologist visited the patient during routine pain rounds and found her to be cyanosed, somnolent and apneic. The patient made a full recovery after resuscitative measures were taken. The pump was replaced by more up to date technology. Neither of the two nurses involved in the initial programming of the pump was familiar with the programming.

Compounding the programming error was the misplacement of the back-check valve, which allowed a large reservoir of morphine to accumulate, most likely in the empty antibiotic bag, which was piggybacked into the main iv line earlier in the day. Nurses were not alerted to potential problems with the system, even after 153 mg of morphine had been delivered from the pump over a period of 90 min and the patient was still complaining of pain. The third nurse suspected a programming error and alerted the second nurse, who reprogrammed the pump to its desired setting. The fourth nurse later recognized that the back check valve was incorrectly attached; however, she assumed that the antibiotic bag contained cefazolin instead of the accumulated morphine. When she flushed the iv line and allowed its contents to be administered, a massive dose of morphine was delivered. The patient's rapid change in level of pain and the onset of drowsiness were taken for an appropriate response to morphine. According to the authors of this case study, multiple caregivers, insufficient handover, incorrect assumptions, and distributed knowledge contributed to the adverse event.

In terms of our Impact Flow Diagram, it is clear that the programming error resulted from the general property of virtuality (freedom from physical constraints), which enabled the pump to return to a low default setting as the initial choice. This, combined with poor feedback on device state and behavior led to an inaccurate mental model of the pump behavior, which resulted in a programming error. In addition to this programming error, there also was an incorrect attachment of the PCA tubing to the patient. This has nothing to do with computer technology, but it represents a design flaw in that there was no 'forcing function' [3] to constrain the sequence of user actions: nurses could misplace the back-check valve without feedback that the morphine would accumulate in an empty antibiotic bag. The impact on the nurses' cognitive system was again that an inaccurate mental model was developed, this time of the joint patient-pump relationship. The impact on operational processes was to inadvertently cease infusion, as the patient did not receive any morphine while it was being redirected to the antibiotic bag.

It is important to think of people and technology, not as independent components, but rather as a single ensemble where breakdowns in coordination may occur [5]. The Joint Cognitive Systems in the Impact Flow Diagram are clearly apparent in this case, as there were four different nurses involved over a time span of a little over five hours, an anesthesiologist, the postanesthesia care

unit, and the patient herself (lack of space permits us to discuss the larger hospital and regulatory environment). This ensemble of people was grouped around the PCA, each with a different perspective, due to differences in experience and training. Systems become brittle rather than resilient when organizations allow for multiple caregivers to operate independently of each other. Just as artifacts constrain cognition, organizational policies and regulations constrain collaboration.

It is also important to emphasize that the patient herself is part of the system to be monitored and controlled by the hospital staff. The fact that the patient for a long time had no complaints other than that she still experienced pain, was correctly explained, first as a programming error, and later as due to an incorrect attachment. Unfortunately, these explanations were limited, partially correct, and symptom-oriented. Her symptoms of drowsiness once the morphine was delivered in a massive dose were quite understandably normalized as representing a to-be-expected reaction to the morphine taking effect. Different nurses taking care of the same patient for brief periods of time leads to the equivalent of the 'keyhole effect' in computer technology: the nurses merely get a glimpse of a more dynamic and continuous process to which the patient herself and her daughter had more privileged access (it is noteworthy that it was the daughter who reported that her mother was "feeling weird" after the massive injection of morphine). Poor feedback on the patient's state and behavior then again led to an inaccurate mental model of her state on the part of the nurses. This inaccurate model was only corrected by the anesthesiologist when the patient's symptoms were so clear that they could not be ignored any more. The fact that the staff anesthesiologist visited the patient during routine pain rounds just after she had received the massive injection of morphine again emphasizes the importance of the 'keyhole effect', this time with a fortunate ending (one wonders what would have happened if the anesthesiologist was held up on his way to the patient).

This case study, using methods such as interviews, chart review and inspection of historical pump data, went a long way in uncovering the sequence of events leading up to the incident. The authors adopted the 'latent failure model' [75] in that they considered the various nurses involved as layers of defense, with the accident progression being stopped by the anesthesiologist who happened to be at the right time at the right place. Although this example clearly indicates the importance of inadequate communications, deficient training, design failures, and unsuitable materials, in the end it fails to adopt systems thinking, because it views 'human error' as the only explanatory construct (the authors list 17 potential human errors associated with PCA opioid administration, of which they claim 6 were present in the current incident). Putting on different conceptual glasses, we have tried to maintain a systems perspective, albeit a limited one within the scope of this paper, by showing how the Impact Flow Diagram may be helpful in analyzing this case. In particular, organizational policies and regulations enforce multiple caregivers to obtain piecemeal information and impede collaboration. This organizational 'keyhole effect', combined with poor feedback on the patient's state, led to only partially correct mental models on the part of the nurses. Note that this explanation goes beyond the traditional human factors explanation of accidents in terms of "loss of situation awareness" and does not blame the nurses in any way [76].

4. Conclusions, limitations, and recommendations

4.1. Conclusions

Our systematic literature review has shown that most study methods used to find user interface issues with infusion pumps fo-

cus on observable behavior rather than on how artifacts shape cognition and collaboration. In terms of the Impact Flow Diagram developed, most methods deals with operational processes rather than with representational properties of design or with joint cognitive systems. This is unfortunate, as it limits our deeper understanding of the multiple constraints involved in this domain. Detailed analysis of one case study showed that a deeper understanding is limited not so much by the methods employed but by the conceptual looking glasses that were put on, in other words, by the, frequently implicit, model of accident causation adopted.

Our results also clearly show that, although several techniques for usability evaluation testing exist in isolation, the concerted application of these techniques when testing infusion pumps is lacking in the literature. There are some advancements reported in the hospital procurement process where human factors engineering has been taken into account from the outset [35]. Even though this process may include multiple forms of usability testing and evaluation, it is still subject to issues such as underrepresentation of stakeholders, vagueness of criteria by which pumps are judged, confirmation bias, and time pressure. There are still a lot of lessons to be learned from detailed descriptions of procurement processes, yet incorporating human factors principles in hospital procurement decision-making is, in the long run, essential to identifying pumps that are difficult to use and that pose potential dangers to patient safety.

Our study showed the importance of going beyond the simple application of methods to solve particular user-interface problems. One way of 'going beyond' is the adoption of a theoretical perspective such as the Cognitive Systems Engineering perspective with the associated Impact Flow Diagram. Assuming that a broad coverage of all levels within the Impact Flow Diagram is desirable, we would recommend combining heuristic analysis, observational studies of use problems and experimental comparisons, as these are the methods that were found to yield the broadest range of issues while at the same time suffering from the fewest methodological problems, if applied sensibly, that is, with a clear view toward describing the cognitive processes and representations from a practitioner's point of view.

4.2. Limitations

Taken as a whole, the 47 studies retrieved differed widely in methodology. We therefore grouped them in seven categories. Of course, this impeded a quantitative meta-analysis, and, due to lack of detail on statistical measures, it turned out to be impossible to perform a quantitative synthesis.

Limitations at the review level may have been the incomplete retrieval of identified research. However, out of the 232 records identified in total, only 5 (2.1%) were identified through other sources than database searching [30,38,41,48,56]. This does not necessarily imply that we identified all studies, but the percentage identified through other sources is small enough to be confident that not many studies were missed.

4.3. Recommendations

Every piece of medical device technology is part of and shapes the human-technology ensemble. Designing for the safe use of medical technology requires us to recognize this fact. Narrowly focusing on the improvement of the technology part is bound to overlook the broader implications of the design process. Research methods to uncover use problems with technology may be used in many ways, with many different foci. Our research has shown that most methods employed in uncovering user-interface problems focus on observable behavior, to the relative neglect of the shaping forces of computer technology on representations

and processes. We advocate the adoption of an Impact Flow Diagram perspective rather than merely focusing on usability issues in isolation. Truly advancing patient safety requires the systematic adoption of a systems perspective, also in the design of medical device technology.

Acknowledgments

References

[1] Center for Devices and Radiological Health, U.S. Food and Drug Administration. Infusion pump improvement initiative; April 2010. <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/GeneralHospitalDevicesandSupplies/InfusionPumps/ucm205424.htm>.

[2] Woods DD. Designs are hypotheses about how artifacts shape cognition and collaboration. Ergonomics 1998;41(2):168–73.

[3] Norman DA. Things that make us smart. Reading, MA: Addison-Wesley; 1993.

[4] Nemeth CP, O'Connor M, Cook RI. The infusion device as a source of healthcare resilience. In: Nemeth CP, Hollnagel E, Dekker S, editors. Resilience engineering perspectives. Preparation and restoration, vol. 2, Farnham; 2009. p. 235–53.

[5] Woods DD, Dekker S, Cook R, Johannesen L, Sarter N. Behind human error. 2nd ed. Farnham, England: Ashgate Publishing Limited; 2010.

[6] Bisantz A, Roth E. Analysis of cognitive work. Rev Hum Factors Ergon 2007;3(1):1–43.

[7] Hutchins E. Cognition in the wild. Cambridge, MA: MIT Press; 1995.

[8] Rogers ML, Patterson ES, Render ML. Cognitive work analysis in health care. In: Carayon P, editor. Handbook of human factors and ergonomics in health care and patient safety. Boca Raton, FL: CRC Press; 2012. p. 465–74.

[9] Rasmussen J, Pejtersen A, Goodstein L. Cognitive systems engineering. New York: Wiley; 1994.

[10] Woods DD. Discovering how distributed cognitive systems work. In: Hollnagel E, editor. Handbook of cognitive task design. Mahwah, NJ: Lawrence Erlbaum Associates; 2003. p. 37–53.

[11] National Patient Safety Agency, National Reporting and Learning Service. Design for patient safety: a guide to the design of electronic infusion devices. Edition 1; 2010.

[12] Center for Devices and Radiological Health, Office of Device Evaluation, U.S. Food and Drug Administration. Applying human factors and usability engineering to optimize medical device design. Draft guidance; June 22, 2011. <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM259760.pdf>.

[13] Hicks RW, Sikirica V, Nelson W, Schein JR, Cousins DD. Medication errors involving patient-controlled analgesia. Am J Health Syst Pharm 2008;65(5):429–40.

[14] Adachi W, Lodolce AE. Use of failure mode and effects analysis in improving the safety of i.v. drug administration. Am J Health Syst Pharm 2005;62(9):917–20.

[15] Ahmad I, Thompson A, Frawley M, Hu P, Heffernan A, Power C. Five-year experience of critical incidents associated with patient-controlled analgesia in an Irish university hospital. Ir J Med Sci 2010;179(3):393–7.

[16] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.

[17] Chan AW, Hrobjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published studies. JAMA 2004;291:2457–65.

[18] Garmer K, Liljegren E, Osvalder A, Dahlman S. In: Usability evaluation of a new user interface for an infusion pump developed with a human factors approach. Proceedings of the XIVth triennial congress of the international ergonomics association and 44th annual meeting of the human factors and ergonomics association, 'ergonomics for the new millennnium', 29 July 2000 through 4 August 2000. San Diego, CA, United States: Human Factors and Ergonomics Society, Inc.; 2000. p. 128–31.

[19] Lin L, Isla R, Doniz K, Harkness H, Vicente KJ, Doyle DJ. In: Anon, editor. Analysis, redesign, and evaluation of a patient-controlled analgesia machine interface. Proceedings of the 39th annual meeting of the human factors and ergonomics society. Part 2 (of 2), 9 October 1995 through 13 October 1995, San Diego, CA, USA. Santa Monica, CA, United States: Human Factors and Ergonomics Society, Inc.; 1995. p. 738–41.

[20] Lin L, Isla R, Doniz K, Harkness H, Vicente KJ, Doyle DJ. Applying human factors to the design of medical equipment: patient-controlled analgesia. J Clin Monit Comput 1998;14(4):253–63.

[21] Obradovich JH, Woods DD. In: Users as designers: how people cope with poor HCI design in computer-based medical devices. Proceedings of the 38th annual meeting of the human factors and ergonomics society. Part 2 (of 2); 24 October 1994 through 28 October 1994, Nashville, TN, USA. Santa Monica, CA, United States: Human Factors and Ergonomics Society, Inc.; 1994. p. 710–4.

[22] Wetterneck TB, Carayon P, Sobande F, Hundt AS. In: Technology characteristics predicting end user acceptance of smart intravenous infusion pumps. 49th Annual meeting of the human factors and ergonomics society, HFES 2005, 26 September 2005 through 30 September 2005. Orlando, FL, United States: Human Factors and Ergonomics Society, Inc.; 2005. p. 1038–41.

[23] Lin L. Human error in patient-controlled analgesia: incident reports and experimental evaluation. Proc Hum Factors Ergon Soc 1998;2:1043–7 [5 October 1998 through 9 October 1998].

[24] Lin L, Vicente KJ, Doyle DJ. Patient safety, potential adverse drug events, and medical device design: a human factors engineering approach. J Biomed Inform 2001;34(4):274–84.

[25] Garmer K, Liljegren E, Osvalder A, Dahlman S. Arguing for the need of triangulation and iteration when designing medical equipment: nature of requirements on medical equipment derived from a human factors approach and from usability tests: methodological aspects. J Clin Monit Comput 2002;17(2):105–14.

[26] Garmer K, Liljegren E, Osvalder A, Dahlman S. Application of usability testing to the development of medical equipment. Usability testing of a frequently used infusion pump and a new user interface for an infusion pump developed with a human factors approach. Int J Ind Ergon 2002;29(3):145–59.

[27] Sowan AK, Gaffoor MI, Soeken K, Johantgen ME, Vaidya VU. Impact of computerized orders for pediatric continuous drug infusions on detecting infusion pump programming errors: a simulated study. J Pediatr Nurs 2010;25(2):108–18.

[28] Trbovich PL, Pinkney S, Cafazzo JA, Easty AC. The impact of traditional and smart pump infusion technology on nurse medication administration performance in a simulated inpatient unit. Qual Saf Health Care 2010;19(5):430–4.

[29] Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1–2):23–30.

[30] Gagnon R, Laberge J, Lamsdale A, Histon J, Hudson C, Davies J, et al. In: A user-centered evaluation of three intravenous infusion pumps. 48th annual meeting of the human factors and ergonomics society, HFES 2004, 20 September through 24 September 2004. New Orleans, Louisiana, United States: Human Factors and Ergonomics Society, Inc.; 2004. p. 1773–7.

[31] Graham MJ, Kubose TK, Jordan D, Zhang J, Johnson TR, Patel VL. Heuristic evaluation of infusion pumps: implications for patient safety in intensive care units. Int J Med Inform 2004;73(11–12):771–9.

[32] Turley JP, Johnson TR, Smith DP, Zhang J, Brixey JJ. Operating manual-based usability evaluation of medical devices: an effective patient safety screening method. Jt Comm J Qual Patient Saf. 2006;32(4):214–20.

[33] Silva APS, Almeida RMA, Ferreira JA, Gibertoni A. In: Human factors engineering applied to risk management in the use of medical equipment. 12th mediterranean conference on medical and biological engineering and computing, MEDICON 2010, 27 May 2010 through 30 May 2010, Chalkidiki; 2010. p. 987–90.

[34] Ginsburg G. Human factors engineering: a tool for medical device evaluation in hospital procurement decision-making. J Biomed Inform 2005;38(3):213–9.

[35] Namshirin P, Ibey A, Lamsdale A. Applying a multidisciplinary approach to the selection, evaluation, and acquisition of smart infusion pumps. J Med Biol Eng 2011;31(2):93–8.

[36] Keselman A, Patel VL, Johnson TR, Zhang J. Institutional decision-making to select patient care devices: identifying venues to promote patient safety. J Biomed Inform 2003;36(1–2):31–44.

[37] Nemeth C, Nunnally M, Bitan Y, Nunnally S, Cook RI. Between choice and chance: the role of human factors in acute care equipment decisions. J Patient Saf 2009;5(2):114–21.

[38] Cook RI, Woods DD, Howie MB, Horrow JC, Gaba DM. Unintentional delivery of vasoactive drugs with an electromechanical infusion device. J Cardiothorac Vasc Anesth 1992;6(2):238–44.

[39] Lamsdale A, Chisholm S, Gagnon R, Davies J, Caird J. In: A usability evaluation of an infusion pump by nurses using a patient simulator. 49th annual meeting of the human factors and ergonomics society, HFES 2005, 26 September 2005 through 30 September 2005. Orlando, FL, United States: Human Factors and Ergonomics Society, Inc.; 2005. p. 1024–8.

[40] Liljegren E, Osvalder A, Dahlman S. In: Setting the requirements for a user-friendly infusion pump. Proceedings of the XIVth triennial congress of the international ergonomics association and 44th annual meeting of the human factors and ergonomics association, 'ergonomics for the new millennnium', 29 July 2000 through 4 August 2000. San Diego, CA, United States: Human Factors and Ergonomics Society, Inc.; 2000. p. 132–5.

[41] Moll van Charante E, Cook RI, Woods DD, Yue L, Howie MB. Human–computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. In: Stassen HG, editor. Analysis, design and evaluation of man–machine systems, London; 1993. p. 263–74.

[42] Nunnally M, Nemeth CP, Brunetti V, Cook RI. Lost in menuspace: user interactions with complex medical devices. IEEE Trans Syst Man Cybern A Syst Humans 2004;34(6):736–42.

[43] Nunnally ME, Bitan Y. Time to get off this pig's back? the human factors aspects of the mismatch between device and real-world knowledge in the health care environment. J Patient Saf 2006;2(3):124–31.

[44] Nyssen A-S. Integrating cognitive and collective aspects of work in evaluating technology. IEEE Trans Syst Man Cybern A Syst Humans 2004;34(6):743–8.

[45] Obradovich JH, Woods DD. Users as designers: how people cope with poor HCI design in computer-based medical devices. Hum Factors 1996;38(4):574–92.

[46] Nemeth CP, Nunnally M, O'Connor M, Klock PA, Cook R. Getting to the point: developing IT for the sharp end of healthcare. J Biomed Inform 2005;38(1):18–25.

[47] Husch M, Sullivan C, Rooney D, Barnard C, Fotis M, Clarke J, et al. Insights from the sharp end of intravenous medication errors: implications for infusion pump technology. Qual Saf Health Care 2005;14(2):80–6.

[48] Taxis K, Barber N. Causes of intravenous medication errors: an ethnographic study. Qual Saf Health Care 2003;12:343–8.

[49] Brixey JJ, Zhang J, Johnson TR, Turley JP. Legibility of a volumetric infusion pump in a shock trauma ICU. Jt Comm J Qual Patient Saf 2009;35(4):229–35.

[50] Johnson TR, Tang X, Graham MJ, Brixey J, Turley JP, Zhang J, et al. Attitudes toward medical device use errors and the prevention of adverse events. Jt Comm J Qual Patient Saf 2007;33(11):689–94.

[51] Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton-Mifflin; 2002.

[52] Apkon M, Leonard J, Probst L, DeLizio L, Vitale R. Design of a safer approach to intravenous drug infusions: failure mode effects analysis. Qual Saf Health Care 2004;13(4):265–71.

[53] Carayon P, Hundt AS, Wetterneck TB. Nurses' acceptance of smart IV pump technology. Int J Med Inform 2010;79(6):401–11.

[54] Eade DM. Patient-controlled analgesia–eliminating errors. Nurs Manage 1997;28(6):38–40.

[55] Ferguson R, Williams ML, Beard B. Combining quality improvement and staff development efforts to decrease patient-controlled analgesia pump errors. J Nurses Staff Dev 2010;26(5):E1–4.

[56] Moss J. Reducing errors during patient-controlled analgesia therapy through failure mode and effects analysis. Jt Comm J Qual Patient Saf 2010;36(8):359–64.

[57] Paul JE, Bertram B, Antoni K, Kampf M, Kitowski T, Morgan A, et al. Impact of a comprehensive safety initiative on patient-controlled analgesia errors. Anesthesiology 2010;113(6):1427–32.

[58] Vicente KJ, Kada-Bekhaled K, Hillel G, Cassano A, Orser BA. Programming errors contribute to death from patient-controlled analgesia: case report and estimate of probability. Can J Anesth 2003;50(4):328–32.

[59] Rothschild JM, Keohane CA, Cook EF, Orav EJ, Burdick E, Thompson S, et al. A controlled trial of smart infusion pumps to improve medication safety in critically ill patients. Crit Care Med 2005;33(3):533–40.

[60] Wetterneck TB, Skibinski KA, Roberts TL, Kleppin SM, Schroeder ME, Enloe M, et al. Using failure mode and effects analysis to plan implementation of smart i.v. pump technology. Am J Health Syst Pharm 2006;63(16):1528–38.

[61] Lori Brown S, Bogner MS, Parmentier CM, Taylor JB. Human error and patient-controlled analgesia pumps. J Intraven Nurs 1997;20(6):311–6.

[62] Thornburg K, Thomas G, Draper S. In: Categorizing adverse medical device and medication event frequency. 50th annual meeting of the human factors and ergonomics society, HFES 2006, 16 October 2006 through 20 October 2006. San Francisco, CA, United States: Human Factors and Ergonomics Society, Inc.; 2006. p. 1009–13.

[63] Malashock CM, Shull SS, Gould DA. Effect of smart infusion pumps on medication errors related to infusion device programming. Hosp Pharm 2004;39(5):460–9.

[64] Rayo M, Smith P, Weinger MB, Slagle J, Dresselhaus T. In: Assessing medication safety technology in the intensive care unit. 51st annual meeting of the human factors and ergonomics society, HFES 2007, 1 October 2007 through 5 October 2007. Baltimore, MD, United States: Human Factors and Ergonomics Society, Inc.; 2007. p. 692–6.

[65] Kingman MS, Tankersley MA, Lombardi S, Spence S, Torres F, Chin KS. Prostacyclin administration errors in pulmonary arterial hypertension patients admitted to hospitals in the United States: a national survey. J Heart Lung Transplant 2010;29:841–6.

[66] Draper S, Nielsen GA, Noland M. Using "no problem found" in infusion pump programing as a springboard for learning about human factors engineering. Jt Comm J Qual Saf 2004;30(9):515–20.

[67] Musshoff F, Padosch SA, Madea B. Death during patient-controlled analgesia: piritramide overdose and tissue distribution of the drug. Forensic Sci Int 2005;154(2–3):247–51.

[68] Perry SJ. An overlooked alliance: using human factors engineering to reduce patient harm. Jt Comm J Qual Saf. 2004;30(8):455–9.

[69] Rule AM, Drincic A, Galt KA. New technology, new errors: how to prime an upgrade of an insulin infusion pump. Jt Comm J Qual Patient Saf 2007;33(3):155–62.

[70] Syed S, Paul JE, Hueftlein M, Kampf M, McLean RF. Morphine overdose from error propagation on an acute pain service. Can J Anesth 2006;53(6):586–90.

[71] Vicente KJ. What does it take? A case study of radical change toward patient safety. Jt Comm J Qual Saf 2003;29(11):598–609.

[72] Baron J, Hershey J. Outcome bias in decision evaluation. J Pers Soc Psych 1988;54(4):569–79.

[73] Caplan R, Poser K, Cheney F. Effect of outcome on physician judgments of appropriateness of care. JAMA 1991;265:1957–60.

[74] Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. JEP: HPP 1975;1(3):288–99.

[75] Reason J. Human error. Cambridge, England: Cambridge University Press; 1990.

[76] Dekker SWA. On the epistemology and ethics of communicating a Cartesian consciousness. Safety Sci, in press. http://dx.doi.org/10.1016/j.ssci.2012.05.028.