Anaesthetists and patients faced with the risk of human error: development of a methodology for evaluating computerised systems of monitoring and control

Prof. V. DE KEYSER[1] (coordinator)
Dr. A-S NYSSEN[1]
Prof. M. LAMY[2]
Prof. A. D'HOLLANDER[3]

[1] Department of Wok Psychology
[2] Departement of Anesthesie - Reanimation
University of Liège – ULg

[3] Department of Anesthesie
Free University of Brussels - ULB

Introduction

In many domains, human error is evoked as the major contributing factor or cause of accidents. For example, surveys of anaesthesia incidents in the operating room have attributed 70% to the human element[1]. Similar surveys in highly technical systems such as aviation or nuclear power plants show the same percentage[2]. Accident investigations are traditionally based on statistical data and epidemiological methods rather than on detailed analyses of individual cases. These studies often classify accidents in mutually exclusive categories: human error, equipment failure, or complication. The result is widespread perception of the human error problem. The typical belief is that the human element is separate from the system in question. In practice, things prove not to be so simple. Detailed investigations of recent catastrophes, such as airline accidents, Three Mile Island, or the Challenger explosion have highlighted that these catastrophes are almost always caused by poor coordination between human and machine. Even when human error is directly involved, it is always possible to point to problems in the design, manufacture, installation, or maintenance of some part of the system[3]. Our fascination for the benefits of technology has often obscured the fact that technology also creates new demands on individual practitioners or groups of practitioners responsible for operating and managing the system. The new demands can include new or modified tasks (setup, initialisation, operating sequences, etc.) as well as new cognitive demands. There are new knowledge requirements (e.g., how the device functions), new communication tasks (instructing the automated device), new management tasks (finding the relevant data on the device), and new demands on attention (monitoring the state and performance of the automated device). The presence of these demands creates opportunities for new forms of human error and failure that can be classified as design-induced. The term "clumsy automation"[4] is used to illustrate this kind of poor coordination between human and machine.

The goal of this study is to develop an evaluation methodology for assessing the impact of technology changes on practitioner cognition and behaviour, in order to better orientate the design and integration process and reduce the potential for human-machine interaction deficiencies.

---

1 CHOPRA, V., BOVILL, J.G., SPIERDIYK, J. & KOORNEEF, F. (1992). Reported significant observations during anesthesia: a prospective analysis over an 18-month period. *British Journal of Anaesthesia*, 68, 13-17.

2 AMALBERTI, R. (1993). Safety in flight operations. In B. Wilpert & T. Qvale (Eds), *Reliability and safety in hazardous work systems: approaches to analysis and design.* Hillsdale: Erlbaum.

3 REASON J. (1997). *Organizational Accidents: The management of Human and Organizational Factors in Hazardous Technologies.* Cambridge, England: Cambridge University Press.

4 WIENER, E.L. & CURRY, R.E. (1980). Flight-deck automation: promises and problems. *Ergonomics*, 23, 995-1011.

Many evaluation studies have focused on the ergonomic aspects of the display: letter size, digital versus analogue information. Yet the above-mentioned problems occur not because of the characteristics of the device per se but because of how it is used in a given context. Context-free evaluations are unlikely to uncover the important problems, determine why they are important, or identify criteria that more successful systems should meet[5]. To ensure that our methodology goes beyond the superficial structure of the interface, we developed it through a series of studies in the field of anaesthesia, looking at anaesthetist interactions with new information technology in the operating room.

Materials and methods

In what follows, we will first present these studies and then explain the evaluation methodology.

The investigation started with analysis of the human-machine interaction system. Naturalistic observations (more than 400 hours) were conducted in an operating room selected for the use of two new devices: a new monitoring device and a new automatic infusion device. The aim was to identify:

- characteristics of the context of anaesthetists interacting with the new device

- characteristics of the device making it difficult to operate and error prone.

A prior evaluation methodology was developed. Given the observation data, the analyst decomposes the device into its functions and then applies evaluation criteria to each identified function. Twenty evaluation criteria were constructed. They assess the functional structure of the computer information system: where are the desired data located in the display space? Which menu provides access? How does one navigate to that location? Are the navigating rules coherent? Is there any feedback in case of error? Is it possible to recover from an error? Is it easy to predict the next behaviour of the device, etc. The evaluation of each function is mainly based on practitioner appreciation and on the use of experimental tasks in order to better assess problems in interacting with the device. For both devices, it was reported that different functions didn't meet the critical value of 75% positive appreciation. The documentation function and the visibility and feedback in case of error had to be changed to better meet user preferences. Such a methodology can help collect data about problems in the system's functional design. Yet the analytical approach is demanding in terms of resources. For the monitoring device, for example, more than 400 functions were identified by the analyst and evaluated. Other limitations were identified, mainly the fact that impacts of the device on cognitive activities were not clearly analysed. This needs to be further examined from an empirical perspective.

We conducted three studies from this perspective. The first is devoted to collecting problem situations in the field and to analysing them in order to identify contributing factors, including critical cognitive activities. The second is an experimental study analysing in detail how anaesthetists handle crisis situations in a full-scale anaesthesia simulator. The third study assesses the immediate and delayed impacts of a new automatic infusion device in order to better understand how practitioners adaptively respond to implementation of the device. Results of these studies are important for the analyst in order to identify which problems are important in the context and define criteria for evaluation and design.

Collection and analysis of problem situations

Although data collection is still going on, preliminary analyses are available, based on a sample of 30 cases reported over a period of 16 months. The results reveal the importance of different time characteristics of the situation. Order and duration of events and actions, time pressure, delays in the feed-back of actions, drug latencies and duration of efficacy, dynamicity, conflicting temporal reference systems affect the subjects' performance, especially that of novices. Our cognitive analysis shows that diagnosis is far from always being the critical phase in the decision-making process. Most often were mentioned failure to anticipate and failure to perceive information during surgery. The analysis of our

---

5 SARTER, N.B. & WOODS, D.D. (1995). Strong, silent, and out-of-the-loop: properties of advanced (cockpit) automation and their impact on human-automation interaction. *CSEL Report-95-TR-01*. Columbus, Ohio State University.

data also reveals different cognitive difficulties for different degrees of expertise. We observed that mainly anaesthetists in their 3rd year of training experienced diagnostic difficulties. This analytical perspective is interesting if we want to predict cognitive failures connected with particular work conditions (prototypical risk situations) and limit the risk through means such as training or technical or organisational improvements.

Analysis of crisis management in a full-scale simulator

We compared and analysed the responses of more or less experienced trainee anaesthetists to 5 simulated problem situations of different complexity. In all scenarios, the time required for diagnosis was greater for novices than for experts. The accuracy of diagnosis varied according to the complexity of the problem situation (speed of evaluation, number of variables to supervise, frequency of occurrence) and the level of expertise. In the diagnostic process, novice subjects proposed more hypotheses than the more experienced subjects. We also noted in the more experienced group a high frequency of planning behaviours and of observations of the efficiency of the treatment set-up. Analysis of behavioural sequences reveals differences in how the two groups react when the problem arises during execution of a highly proceduralised sequence. The experienced subjects interrupt the sequence to treat the problem on the basis of anticipation of an unfavourable evolution. These results may help shed light on the special abilities that medical experts possess that enable them to respond to problem situations as well as they do. This should provide a basis for defining criteria for evaluating the human-machine system.

Repeated observations of problem interactions with the new automatic infusion device

Impacts of technology changes on work can be immediate or delayed. Assessment implies a comprehensive evaluation of effects as the technology is being implemented and during its application. In this study, we observe the work situation before implementation of the device, during the implementation phase, and one month later. Understanding how practitioners respond adaptively to implementation of the device and the limits of their adaptation is critical for understanding how automation creates the potential for new forms of error and system breakdown.

Given these results, we identify the important problems in actual work situations and define evaluation criteria to cover the cognitive and operational costs of more and more complex devices. There are four critical points in our evaluation methodology:

1. evaluating device in context
2. collecting and integrating data on important issues in the field (e.g. performance issues)
3. assessing effects on the human-machine system, including on cognitive activities
4. evaluating both direct and delayed effects to gain insight into adaptation abilities
5. involving users and designers in the evaluation process

The following paragraphs describe the technology evaluation methodology that we developed. The objective of this methodology is to assess the impact of technology changes on all the components of a work situation: technical, clinical, cognitive, organisational, economic, and other dimensions. Which dimensions and criteria are developed in more detail depends on the type of technology and on the objective of the assessment. The list proposed below is not exhaustive. For each dimension, we describe:

- the evaluator (user, expert, designer, etc)
- the dimension and the criteria measures
- the source of data used
- the phase of the evaluation

1. Technical dimension
"Evaluating the technical objectives and checking to what extent they are met in the context."

Multicriteria evaluation is used for this dimension. The expert (analyst) can evaluate the compatibility of the use of the system in the context with the technical prerequisites of the device on the basis of the documentation and observation. He can also analyse the reliability of the device in the context. Two

kinds of studies can be carried out: an experimental study often conducted by the designer in collaboration with the users and a field study based on observations of problem interactions. Although the results of our studies show the interest of such systematic observations, they are rarely organised by the designer. Most of the time, feedback returns to the designer whenever a critical problem has occurred.

2. Clinical dimension
"Evaluating the compatibility of the device with the clinical objectives pursued"

Many information technologies do not have clinical effects even in the health domain (monitoring devices, communicative tools, etc.). In this case, this dimension does not have to be considered. If the device does have a clinical effect, the analyst must assess to what extent the clinical objectives are met in the context. This dimension requires the collaboration of users and designer.

3. Cognitive dimension`
"Evaluating the impact of the device on practitioner strategies analysed in a cognitive framework".

To evaluate this dimension, the analyst assesses for each practitioner the impact of the device on information processes. He measures the impact of the device on activities related to information observation, decision-making, and execution. Depending on the type of technology, he will develop a detailed analytical evaluation of the functionality of the information display using the criteria developed in the prior methodology. In some cases, observations of problem interactions before, during, and after the phase of implementation are more important. These require an analyst familiar with the work situation and with the device.

4. Organisational dimension
"Evaluating the impact of the device on the team and on how the work is organised"

The development and complexity of techniques and the intensification of specialisations have contributed to making teamwork indispensable. The implementation of new technology changes the team and how the work is organised. Multicriteria evaluation is used for this dimension. It is possible to measure the impact on the roles of the actors, on communication and cooperation strategies.

5. Economic dimension.
"Evaluating the impact of the device in terms of economic effects"

Each use of technology has an economic effect. This effect can influence the usability of the device. Effects measured in this dimension are numerous. Often cost-benefit analyses are done by economists and focus on direct running costs. Benefits derived from cognitive or organisational or clinical effects are more difficult to quantify.

Conclusion

Clearly, the problems we have identified by applying this methodology to three new devices can only be identified by taking into account all the dimensions of the human-machine system: human, machine, and context. It implies using different methods of analysis to identify the important problems in the context and to define relevant criteria for evaluation and design.