How a Clinical Decision Support System Changed the Diagnosis Process: Insights from an Experimental Mixed-Method Study in a Full-Scale Anesthesiology Simulation

Sara Wolf*
Chair of Psychological Ergonomics
Julius-Maximilians-Universität Würzburg
Würzburg, Germany
sara.wolf@uni-wuerzburg.de

Tobias Grundgeiger*
Chair of Psychological Ergonomics
Julius-Maximilians-Universität Würzburg
Würzburg, Germany
tobias.grundgeiger@uni-wuerzburg.de

Raphael Zähringer
Julius-Maximilians-Universität Würzburg
Würzburg, Germany
raphael.zaehringer@stud-mail.uni-wuerzburg.de

Lora Shishkova
Julius-Maximilians-Universität Würzburg
Würzburg, Germany
lora.shishkova24@gmail.com

Franzisca Maas
Chair of Psychological Ergonomics
Julius-Maximilians-Universität Würzburg
Würzburg, Germany
franzisca.maas@uni-wuerzburg.de

Christina Dilling
Universitätsklinikum Würzburg
Würzburg, Germany
dilling_c1@ukw.de

Oliver Happel
Universitätsklinikum Würzburg
Würzburg, Germany
happel_O@ukw.de

Abstract

Recent advancements in artificial intelligence have sparked discussions on how clinical decision-making can be supported. New clinical decision support systems (CDSSs) have been developed and evaluated through workshops and interviews. However, limited research exists on how CDSSs affect decision-making as it unfolds, particularly in settings such as acute care, where decisions are made collaboratively under time pressure and uncertainty. Using a mixed-method study, we explored the impact of a CDSS on decision-making in anesthetic teams during simulated operating room crises. Fourteen anesthetic teams participated in high-fidelity simulations, half using a CDSS prototype for comparative analysis. Qualitative findings from conversation analysis and quantitative results on decision-making efficiency and workload revealed that the CDSS changed team structure, communication, and diagnostic processes. It homogenized decision-making, empowered nursing staff, and introduced friction between analytical and intuitive thinking. We discuss whether these changes are beneficial or detrimental and offer insights to guide future CDSS design.

*These authors contributed equally to this work.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/2025/04
https://doi.org/10.1145/3706598.3713372

CCS Concepts

• **Information systems** → Information systems applications; Decision support systems; • **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI; • **Applied computing** → Life and medical sciences; Healthcare information systems.

Keywords

Clinical decision support system, decision-making process, artificial intelligence, anesthesiology, conversation analysis, high-fidelity simulation, acute care, teamwork

ACM Reference Format:
Sara Wolf, Tobias Grundgeiger, Raphael Zähringer, Lora Shishkova, Franzisca Maas, Christina Dilling, and Oliver Happel. 2025. How a Clinical Decision Support System Changed the Diagnosis Process: Insights from an Experimental Mixed-Method Study in a Full-Scale Anesthesiology Simulation. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3706598.3713372

1 Introduction

In high-stakes fields like anesthesiology and emergency care, decision-making is frequently a collaborative effort [16] that includes collecting information, integrating information, considering different options, and deciding on treatment (e.g., [100]). Split-second decisions made under time pressure and uncertainty can mean the difference between life and death. It is therefore not surprising that diagnostic errors are one of the main factors leading to unfavorable results [84, 97].

Quantitative studies have shown that clinical decision support systems (CDSSs) can make decision-making more effective and efficient (e.g., [45]), and qualitative studies have investigated users' experience of CDSSs (e.g., [67]). Recently, Zhang et al. [114] highlighted that CDSS research focused too much on the later steps of the decision-making process (i.e., deciding on treatment) but neglected earlier steps, including the collection and integration of information and the consideration of different options. Zhang et al.'s [114] framework considers (1) generating hypotheses, (2) gathering data and the integration of information, and (3) the consideration of different hypotheses as earlier steps. They make the point that current CDSS too often provide a "final" and competing decision in (4) the step of final decision-making. In high-uncertainty, high-stakes, time-sensitive medical decision-making, CDSS designs are needed that support the early steps of the decision-making process [114]. For example, presenting uncertainty while generating hypotheses, visualizing important features during data gathering, and reducing uncertainty by presenting needed information [114]. Although some existing CDSSs support these early decision-making steps (e.g., [67]), the steps were seldom the focus of CDSS research. Current CDSS research lacks a detailed understanding of how CDSSs change the actual decision-making process, whether these changes are beneficial or detrimental, and also lacks a profound understanding of design implications—especially for the earlier steps of the decision-making process.

The earlier steps of the decision-making process require team members to have a common understanding of the situation. Such understanding is achieved by conversational grounding: the collective effort of team members to achieve mutual understanding [18]. Conversational analysis provides a method for understanding how humans achieve conversational grounding and to describe the decision-making process [47, 72]. It can therefore help generate a detailed understanding of how CDSSs affect the decision-making process. To the best of our knowledge, however, there is only limited research applying conversation analysis in CDSS research (e.g., [82]).

In the present study, we addressed a critical gap in the existing literature through an experimental mixed-method study to understand the effects of a CDSS on the decision-making processes in acute care, especially looking at the earlier decision-making steps that require conversational grounding. To the best of our knowledge, this is the first investigation specifically aimed at understanding how a CDSS affects and changes the way anesthetic teams make sense of a situation and achieve a common understanding during an operating room crisis. To this end, we ran a full-scale anesthesiology simulation, including a life-threatening event during general anesthesia (a malignant hyperthermia crisis). Anesthetic teams consisting of one anesthesiologist and one anesthetic nurse either used an existing CDSS prototype [67] to support decision-making during an operating room crisis or worked without the CDSS. Unlike previous work that reported the design of the CDSS and investigated its user experience via interviews [67], we aimed to contribute empirical insights into how the CDSS influenced the diagnosis decision-making processes in detail by conducting an ethnomethodologically informed and video-based conversation analysis [47]. Hence, our primary methodological approach was qualitative. In addition, we analyzed the time until a final diagnosis decision and collected subjective workload measures from the anesthetic teams. Since CDSSs are expected to improve clinical performance in acute care (e.g., [46]), we wanted to test the hypotheses that the CDSS under study can make the diagnosis decision-making processes more efficient (i.e., faster) and that CDSS use reduces workload compared to no CDSS use. Our findings highlight three major changes in relation to team structure, communication, and diagnostic process and suggest that CDSS use homogenized decision-making, empowered nursing staff, and introduced friction between analytical and intuitive thinking. To assess whether these changes due to the CDSS can be considered beneficial or detrimental to the quality of the decision-making process, we discuss our findings through the lens of Crew Resource Management (CRM) [30, 57]. Overall, the novel contributions are (1) detailed descriptions and quantitative data of how the decision-making process of anesthetic teams was affected by the CDSS, (2) an analysis and discussion of whether the observed changes were to the better or worse, and (3) implications for future CDSS design and evaluation.

2 Related Work

CDSSs are computer-based information systems intended to enhance clinicians' decision-making by supporting clinicians during critical decision points [10]. CDSSs encompass all systems that help with information retrieval from databases, automatically match patient-specific information to the underlying knowledge base, or provide AI-based analysis of a patient's status and output a recommendation to the user [8, 9, 102, 113]. In the following, we consider what is known about the effects of CDSSs on the decision-making process and what methods have been used to generate this knowledge in the human–computer interaction (HCI) and healthcare literature. Next, we consider clinical decision-making in acute care (i.e., domains such as anesthesiology, trauma resuscitation bays, and intensive care where patients cannot survive without medical expertise and medical technology) and finally the paper's focal topic of CDSSs in acute care.

2.1 Effects of Clinical Decision Support Systems and Evaluation Methods in Clinical Research and HCI

Clinical research has shown the effectiveness of CDSSs [45]. For example, researchers have reported that CDSS use reduced medication errors [41, 60, 93], increased adherence to clinical guidelines [36, 63], and reduced cost and environmental impact in relation to inhalation anesthetics [83]. Although CDSSs can improve clinical decision-making, when not carefully designed or introduced, CDSS uptake might be low [68] or CDSSs might even have negative consequences [102]. Low uptake has frequently been attributed to a mismatch between the potential of the CDSSs to support the actual clinical decision-making process [107], a poor fit of the CDSSs with the actual clinical context [23, 59, 64, 92], or a lack of support for the motivational orientation of the users [38].

Most clinical research has investigated the effectiveness of CDSSs by conducting quantitative research using randomized control trials in the actual clinical contexts and has not considered the changes in the decision-making process (e.g., [27, 45, 59, 63, 83]). However, some clinical research also includes qualitative research on clinicians' experience of CDSSs using interviews [49] and conversation analysis [25, 81]. An outstanding example of qualitative clinical research is Murdoch et al.'s [82] study, which investigated the consequences of computer-mediated decision-support software (the triage nurse in a general practitioner practice performed telephone triage either with or without a CDSS) on the interaction between nurses and the callers using conversational analysis. The results showed that nurses had to listen and comprehend the diverse symptoms presented by the patients and navigate and consider the CDSS, which should support safe triage outcomes. Murdoch et al. [82] used Goffman's [32, 33] four production formats and the resulting roles of an agent in the conversation (animator, author, principal, figure) to explain role changes due to the CDSS. The constraints imposed on the nurses by the CDSS led the authors to consider the CDSS as a living questionnaire animated by the nurses that affected the nurse–caller interaction and authored the telephone conversations between the human agents. Although telephone triage is different from this paper's context of acute care in regard to the physical proximity of human agents and the number of agents, the study demonstrated the impact of CDSS on conversation and agency and the value of comparisons to the status quo (no CDSS) and in-depth analysis of interactions with CDSSs as they unfold.

HCI research on CDSS has different focuses from clinical research and can roughly be structured according to three main approaches. First, HCI researchers have shown the principal benefits of CDSSs, such as improved decision-making performance when working with a CDSS compared to not working with a CDSS or performance differences when working with a CDSS that provides different interaction forms (e.g., [15, 61, 70, 71, 90]). For example, Rajashekar et al. [90] compared a standard graphical user interface and a large language model interface and conducted a comprehensive usability survey. Most of the studies are quantitative evaluations of CDSSs in simulated healthcare settings.

Second, HCI researchers have generated insights about clinical decision-making processes without CDSSs in real contexts (e.g., [16, 31, 95, 111]). These studies provide valuable insights into the integration of possible CDSSs, such as general barriers to computer use in the decision trajectory [111]. Frequently, researchers have followed up such insights from the field with the design of a prototype (e.g., [62]) or the design and evaluation of a prototype (e.g., [4, 62, 67, 90, 98, 99, 109, 110]). Most of the time, these evaluations have been qualitative and included simulated or hypothetical scenarios for prototype testing. These studies frequently used interviews in which potential users contemplated how CDSSs would change the work and enabled first insights into possible changes that CDSSs introduce. Such evaluations are valuable, but detailed observations of interactions with the prototypes and comparisons between CDSS use and the status quo (no CDSS use) are missing.

Third, HCI researchers have studied already-implemented CDSSs (e.g., [7, 106]). These studies have focused on evaluating the CDSSs' effects on workflow, possible technical limitations, usability issues, mismatches between work-as-planned and work-as-done, and issues related to the transparency and trustworthiness of the CDSSs. For example, Beede et al. [7] conducted pre-post CDSS-introduction observations but focused more on macro changes in workflow and post-introduction problems (e.g., lighting conditions, image quality not meeting CDSS requirements, internet speed) rather than micro changes concerning the actual interaction with the CDSS and the effects on decision-making processes as they unfold. Similarly to the studies above investigating the decision-making process without a CDSS, studies on implemented CDSSs used interviews to assess the possible changes the CDSS introduced but lack analysis of CDSS use in context and a comparison with the status quo (no CDSS use).

Overall, HCI research has mainly focused on analyzing the potential of CDSSs, designing CDSS prototypes, and evaluating prototypes in user studies using interviews. The earlier steps of the decision-making process including collecting and integrating information and considering different options have received less attention [111, 114]. To the best of our knowledge, no HCI study has used conversation analysis and an experimental design to examine the change in conduct and sense-making when working with a CDSS compared to the status quo (no CDSS use). However, conversation analysis and, in particular, ethnomethodologically informed and video-based conversation analysis [46, 47] are well-known methods in HCI (e.g., [48, 79]). Heath and colleagues' approach [46, 47] follows ethnomethodology and conversation analytic orientations by considering the actions and activities as inseparable from the immediate context. The immediate context shapes action and is used for action, and, in turn, the context is affected and created by action (i.e., action is both context-shaped and context-renewing [51]). Researchers have also used conversation analysis in studies in the operating room (e.g., [54, 55, 77–79]). In the context of surgery, for example, Mentis [77] demonstrated how imaging systems for supporting better surgical work practices created additional work to progress the case (so-called articulation work) due to the need to consider the images and integrating the gained knowledge back into the ongoing surgical process. In the context of anesthesiology, Hindmarsh and Pilnick [55] used video-based conversation analysis to analyze the bodily conduct of the anesthesia team during tracheal intubations. They showed that colleagues' actions are sensitive to the body of the other team members and argue that coordination is an embodied conduct in this context. Technical artifacts such as interactive CDSSs are a substantial manipulation of the context. Because anesthetic teams work in proximity and in a visually and auditive-rich environment, we consider video-based conversation analysis [46, 47] as a suitable approach to analyze how CDSSs affect actions and are affected by team members' actions in the clinical decision-making process.

2.2 Decision-Making Processes in Acute Care

In contrast to many clinical settings, however, decisions in acute care settings must be made within minutes or even seconds [101]. Acute care shares several characteristics with aviation, such as highly qualified staff, a complex socio-technical system, and a fast-paced and uncertain environment [22]. To improve safety, anesthesiology adapted the concept of CRM from aviation [30, 57]. CRM addresses non-technical skills by considering the management of all technological and human resources at hand to respond to problems that arise and perform necessary patient care tasks. Non-technical skills are at the core of CRM and include situation awareness (gathering information, understanding the situation, anticipating future states), decision-making (gathering information, considering options, re-evaluating), task management (planning, prioritizing, maintaining standards, using resources), and teamwork (exchanging information, coordinating, using authority, assessing capabilities, supporting others) [28].

Considering the literature (e.g., [69, 100, 114]), a textbook-based, analytic diagnostic process should start with (1) information gathering: that is, perceiving, gathering, and sharing all available information and identifying missing information (which may be followed by task management). Next, (2) the information must be integrated and interpreted. In this step, single data points are put into relation to each other to understand what, for example, the combination of a fast heartbeat, low blood pressure, and a bleeding wound indicate. Furthermore, (3) hypotheses about possible diagnoses are generated and the different options are evaluated, and eventually (4) a decision is made on a diagnosis and an appropriate treatment. As indicated in CRM [28], this "final" diagnosis is re-evaluated in the future. Zhang et al.'s framework [114] describes similar steps in slightly different order and considers (1) generating hypothesis, (2) information gathering and interpretation, and (3) evaluating hypothesis as early steps in the decision-making process.

As one can easily imagine, in situations such as an ongoing operating room crisis, trauma resuscitation, or a quickly deteriorating intensive care patient, stress and time pressure affect decision-making in diagnostic processes. A recent review indicated that healthcare staff relies more frequently on so-called intuitive processes in situations with time pressure, uncertainty, and dynamic event development [91]. Using the dual-process theory [24, 42], clinical researchers (e.g., [21, 88]) have contrasted the analytic process described above with an intuitive process. The intuitive process is driven by heuristics and experience and is frequently described as matching the present situation to patterns that have been learned or experienced before [66]. The intuitive process is fast, less effortful, and frequently helpful, because "common diseases are common" [5]; but it is also reliant on the quality of the cues in the situation, the emotional state of the individual, and the experience of the individual [88]. As a result, the intuitive process may be more prone to cognitive biases than the analytic process [101]. Cognitive biases such as premature closure (i.e., deciding on one diagnosis rather than considering alternatives or re-evaluating), omission bias (i.e., tendency toward inaction rather than action), and confirmation bias (i.e., tendency to seek supporting evidence rather than attempting to falsify a diagnosis) are the most frequent contributors to diagnostic errors [35]. Finally, Pelaccia et al. [87] showed that the analytic and intuitive processes work simultaneously when emergency physicians treat patients and that a constant re-evaluation while gathering information is critical in the diagnosis decision-making process. In light of these cognitively challenging environments [101], diagnostic error rates [84, 97], and the high-stakes nature of the decisions, it is unsurprising that healthcare and HCI researchers have considered technology to support staff in acute care.

2.3 Clinical Decision Support Systems in Acute Care

CDSSs may support the full process of gathering and integrating data, considering hypotheses, and selecting diagnoses in the context of acute care. However, many CDSSs do not support all steps but focus heavily on the final decision or treatment [111, 114]. Decision-making processes in acute care may be conducted by single users for preliminary evaluations (e.g., [98, 99, 109, 114]), but frequently, the process is a team effort. In the following, we focus on findings and open questions from CDSS research in team-based acute care settings structured along three categories of CDSSs in acute care.

The first category of technologies that support teams' decision-making in acute care is information displays that provide information about, for example, patient scheduling, occupancy, or other data clinical information systems. Information displays have been suggested to support decision-making in the emergency department (e.g., [29, 53, 112]). In the form of electronic whiteboards, however, they are used for coordination and management decisions rather than diagnosis decision-making processes. Similarly, head-worn information displays have been tested in the operating room to support the supervision of junior staff [40] or monitoring of vital signs [65, 73] rather than diagnosis decision-making.

A second category is cognitive aids and checklists. Cognitive aids and checklists are digital or paper-based artifacts that support a user while performing a medical task such as cardiopulmonary resuscitation [12, 74]. Most of the time, these systems are intended to help clinicians to consider specific tasks or actions by providing reminders or guiding attention to specific tasks at specific times. Research has shown that these systems can improve, for example, guideline adherence during resuscitation [26, 36, 86], and recent research has investigated the requirements for transitioning cognitive aids to decision support platforms [75]. In a randomized control trial, Fitzgerald et al. [27] showed that an aid in form of a large display, including patient demographics, action prompts, vital signs, stated diagnoses, and interventions, reduced errors during trauma resuscitation. The researchers speculated that the display supported shared awareness among team members, facilitated information exchange, and helped to clarify diagnostic and therapeutic decisions. Such speculation is further supported by a recent HCI review on collaboration supported by large interactive displays [76]. In mostly laboratory-based studies with student participants [76], large interactive displays had a positive effect on collaboration processes such as the understanding of another person's interaction with the workspace (workspace awareness), coordination of the team (coordination flow), and the involvement of single members of a group in the reasoning process (level of reasoning), whereas such displays showed no general advantage in relation to verbal and gestural communication of the groups and inconclusive evidence in relation to the level of involvement of individuals.

A third category is systems that provide decisions based on data. Rajashekar et al. [90] developed a CDSS for risk assessment of upper gastrointestinal bleeding and evaluated the CDSS in a randomized control trial by comparing the CDSS alone with the CDSS plus a large language model chat interface in a full-scale medical simulation. In general, both CDSS versions were well perceived by clinicians but also affected team dynamics. In interviews, some team members stated that they were assessing the patient and did not get to interact with the CDSS and discussed whether the CDSS was better for single or team use. Rajashekar et al. concluded that it is important to consider team composition and team dynamics in future research [90]. Similarly, in the context of volume therapy in the intensive care unit, Kaltenhauser et al. [62] highlight the requirement of CDSSs to facilitate collaboration. In the context of anesthesiology, Klüber et al. [67] followed a user-centered design process with a specific focus on user experience and contributed a prototype to support anesthesiologists' diagnosis decision-making process during medical crises. The researchers evaluated the CDSS with six anesthetic teams in a full-scale medical simulation and conducted interviews with the users to assess their user experience. Klüber et al. [67] reported that the CDSS affected user experience by increasing psychological need satisfaction such as competence (attaining or exceeding a standard in one's performance) and relatedness (experiencing a sense of belonging, attachment, and closeness). Klüber et al. [67] argued that the effect on relatedness was induced by improved communication. However, due to the chosen method (usability evaluation with interviews in a single group design) and their research question (user experience in CDSS design), Klüber et al. did not aim at understanding what improved communication meant (e.g., more/less communication, less articulation work related to communication, changes in authoring roles within the team). Furthermore, due to the single group design and the focus on user experience, Klüber et al. did not consider how the CDSS changed the decision-making process and, in particular, how it affected specific steps of the decision-making process (i.e., generating hypotheses, information gathering and interpretation, considering options, making a diagnosis; [114]). Klüber et al. [67] suggested further experimental work to investigate whether CDSS use results in more communication, how CDSSs affect communication, and the trajectory of the diagnosis processes.

2.4 Gaps in knowledge

Summarizing the above literature, research in healthcare and HCI indicates that CDSSs can change decision-making processes by possibly affecting communication, coordination, and team dynamics (e.g., [27, 62, 67, 90]), and research approaches using conversation analysis (e.g., [79, 82]) have demonstrated the general effect of technology on communication and coordination in clinical work. However, no previous study endeavored to produce empirical insights into how CDSSs actually change decision-making processes (e.g., by applying conversation analysis). Addressing this gap is essential to inform the design of useful and successful CDSSs for acute care. Therefore, in the present study, we aim to understand how a CDSS changes the decision-making process in acute care compared to decision-making processes without CDSS (the status quo), whether these changes are beneficial or detrimental, and contribute design guidelines, especially for earlier steps of the decision-making process including collecting information, integrating information, and considering options. Beyond Klüber et al. [67], who contributed a novel artifact [108], we make novel empirical contributions by uncovering how a CDSS affected team structure, communication, and diagnostic process in acute care.

3 Method

3.1 The intraoperative anesthesia setting

The research was conducted in the simulation center of a large teaching university hospital in Germany. In the German healthcare system, a patient is typically taken care of by an anesthetic team consisting of one anesthetic nurse and one anesthesiologist in training (i.e., resident level) or specialized anesthesiologists (i.e., consultant level) and supervising senior anesthesiologists (experienced and specialized anesthesiologists). The anesthetic nurse prepares the setup, connects the patient to the monitoring systems, assists during the induction of anesthesia, and, if needed, assists during the operation or is available on call. The anesthesiologist has the medical responsibility for the patient. As per German law [6], the team is supported by additional senior physicians (who attend during specific procedures, help out, or are in call distance in case of complications). In the hospital under study, the ratio is one senior physician for two or three operating rooms. With local variations, such a team composition is common in many countries. For example, in the US, certified anesthesiologist assistants and certified registered nurse anesthetists are supervised by physician anesthesiologists [3, 14], and in the UK, anesthesia associates or anesthesiologists in training are supervised by anesthesiologists [1, 2].

During elective operations with a stable patient, only one qualified and supervised staff member may attend the patient. The staff member needs to keep the patient in a state of amnesia and akinesia (absence of awareness), analgesia (pain-free), and autonomic and sensory block (muscle relaxation) while maintaining physiologic homeostasis (i.e., hemodynamic stability, oxygenation, ventilation, temperature) [80]. However, if there are concerns about patient management, an anesthetic team takes care of the patient. Working in teams is international practice (e.g., [1, 3]), for example, the American Society of Anesthesiologists "Statement on the Anesthesia Care Team" states that "the physician anesthesiologist must ensure that quality of care and patient safety are not compromised, participate in critical parts of the anesthetic, and remain immediately available for management of emergencies." [3]. In these situations, the Standford Emergency Manual first step is always "inform team" [34], and good CRM, including situation assessment, sharing of information, and teamwork, is essential [28].

We chose the case of malignant hyperthermia (MH) for the simulated operating room crisis because it is a rare incident whose early symptoms are non-specific and could also fit many differential diagnoses [96]. MH is an incident that can occur in patients with a genetic defect of the ryanodine receptor after being exposed to specific trigger substances such as drugs that are used for general anesthesia (e.g., sevoflurane, desflurane, isoflurane, halothane). A common early symptom of MH is the sudden increase of tidal CO_2, with additional symptoms such as tachycardia, muscular rigidity, or desaturation. To verify some symptoms (e.g., acidosis, hyperkalemia, hypoxemia, and hypercapnia), it is essential to perform an arterial blood gas analysis (ABG). Since the symptoms of MH can appear decelerated and may be elicited from other causes, the clinical decision to treat MH is likely to be made under remaining uncertainties. The course of a fulminant MH becomes life-threatening within minutes. As soon as there is a strong suspicion of MH, it is essential that the trigger substance is discontinued, an alternative drug is used to maintain anesthesia, ventilation is intensified with higher oxygen levels, and, finally, the drug dantrolene sodium is administered. Throughout the MH scenario, it is essential for the anesthetic team to (1) recognize that the patient is deteriorating, (2) act to stabilize the patient, (3) collect all available information (e.g., patient data, medication, vital signs, blood values), (4) consider various potential differential diagnoses, and (5) finally decide on a diagnosis and treat it accordingly.

Figure 1: Staged photograph to illustrate the setup of the full-scale simulation room and the Cassandra clinical decision support system (CDSS, [67]). On the left, the surgical team (two confederates; in the photograph, only one is visible) perform a laparoscopic appendectomy. On the right, the anesthetic team (one anesthesiologist and one anesthesia nurse) take care of the simulated patient in general anesthesia. The CDSS screen is integrated into the drape that separates the sterile surgical area (left) and the non-sterile side where the anesthetic team is (right). The CDSS display shows the currently selected diagnosis (malignant hyperthermia), the symptoms of the diagnosis, and the six most likely alternative diagnoses. Due to the surgical procedure, the operating room lighting is reduced.

3.2 The Clinical Decision Support System Cassandra

Most clinical support tools designed for operating room crises in anesthesia focus on the treatment rather than the whole diagnostic process (e.g., [50, 74]). Klüber et al. [67] designed a CDSS that focused on the whole diagnostic process, including early (collecting and integrating information, considering options) and late decision-making steps (decision on diagnosis and associated treatment). The CDSS, called Cassandra (**C**linical **ass**essment **and r**easoning in **a**nesthesia), is a research prototype that can be operated in a Wizard-of-Oz manner and that has demonstrated the importance of considering user experience for medical device design [67]. We obtained Cassandra for the present study and extend the work of Klüber et al. [67] by studying the CDSS's effects on decision-making processes and clinical performance in an experimental study. Cassandra combines possible diagnoses, respective symptoms, and live information about the intraoperative status of a patient, such as their vital signs or anamnestic data. It can be accessed upon demand through a screen positioned above the patient and in front of the drape, separating the surgical and anesthetic team (Figure 1).

To activate Cassandra, the anesthetic team touches the display and enters a tentative diagnosis via a keyboard (on the CDSS screen or a physical keyboard next to the ventilator) or uses speech recognition with the keyword "Hey Cassandra, show me [diagnosis]." The entered diagnosis is then displayed at the center of the screen alongside six differential diagnoses in order of probability in the lower part of the screen (Figure 1). Even if the entered diagnosis is the most probable, the display of additional differential diagnoses invites the anesthetic team to consider other diagnoses. All diagnoses are visualized as container bars with boxes representing the respective symptoms. The symptom boxes are arranged from left to right in order of importance for a diagnosis and can be green and in the container (symptom present), gray and out of the container (symptom absent), or blue and on the edge of the container (undetectable by the system and must be entered by the team; e.g., transpiration). The anesthetic team can change the state of each symptom by swiping them in or out of the container. When a symptom is assessed by Cassandra, a single checkmark is displayed for the symptom. Two checkmarks signify that the anesthetic team has checked and confirmed a symptom. Since Cassandra receives live information about available patient data, it can automatically update the symptoms and differential diagnoses. This real-time update of patient data is realized in a Wizard-of-Oz manner. If a team decides on a diagnosis, treatment information based on existing checklists [34] can be accessed. For more information on the design process of Cassandra, see Klüber et al. [67].

Table 1: Demographic data of participants. Values indicate *M* (*SD*) or frequencies.

	No-CDSS teams (n = 7)		CDSS teams (n = 7)	
	Anesthesiologists	**Anesthetic nurses**	**Anesthesiologists**	**Anesthetic nurses**
Age in years	31 (5)	35 (13)	34 (7)	31 (7)
Work experience in months	60 (40)	116 (74)	72 (76)	115 (83)
Gender (male/female)	6/1	1/6	3/4	1/6
Work experience with team partner[a]	Yes = 5, No = 2		Yes = 6, No = 1	

[a] Partners worked together on at least two occasions.

3.3 Participants

The study was conducted in a large teaching hospital in Germany. In total, 16 anesthetic teams (i.e., 16 anesthetic nurses and 16 anesthesiologists) participated. Two teams were excluded (one team piloted the scenario and the procedure, and during one scenario, the control of the patient manikin failed). The demographics of the final 14 teams are provided in Table 1. There were no statistically significant differences in any demographic variable between the anesthesiologists or the anesthetic nurses in the No-CDSS and the CDSS teams. All participants participated according to their work schedule, and we had no control over team composition. All but five participants had previously taken part in simulation-based training. Three participants had experienced a case of MH. The local ethics committee approved the study protocol, and all participants provided written and informed consent.

3.4 Design

The teams were randomly assigned to the CDSS or No-CDSS condition. All teams were confronted with the same scenario. For the qualitative conversation analysis, we recorded the view of three cameras and the audio in the simulation room, captured with the recording setup of the simulation center (SIMStation GmbH, Wien, Austria), and the CDSS screen. For the quantitative analysis, we recorded the time until a final diagnosis decision and the NASA-TLX [43, 44].

3.5 Material and Procedure

First, participants provided informed consent. Second, the CDSS teams received a short, standardized introduction to familiarize themselves with the simulation room and the CDSS. The CDSS was presented on an Iiyama T2735 touchscreen (13.2” × 23.5”), which was positioned in front of the green drape separating the sterile surgical area from the non-sterile side where the anesthetic team was. Participants could enter a diagnosis using the on-screen keyboard, a Microsoft Wireless Desktop 900 keyboard, which was placed on the side board of the ventilation machine, or via speech recognition (simulated in a Wizard-of-Oz manner). For the No-CDSS teams, the touchscreen was covered with a drape (Figure 2).

Third, the MH scenario started. The scenario included two confederates acting as the surgical team who simulated a laparoscopic appendectomy and one confederate anesthesiologist who provided a handover of the already-anesthetized simulated patient (patient-simulator manikin HPS Human Patient Simulator; CAE Healthcare, Sarasota, FL, USA). The Dräger Perseus A500 served as the anesthetic machine during the simulation. However, due to safety and environmental impact, the gas tank was empty, so the participants did not have a valid display for administering the anesthetic gas quantity. The vital signs of the simulated patient were displayed using a Philipps IntelliVue MX750. Both devices were known to the participants through their daily practice. The participants were introduced to the scenario via a handover from the confederate anesthesiologist, who then left the simulation room and did not re-enter the scenario. During this handover, the participants received paper-based patient information, including information from the pre-anesthesia assessment and all drugs administered so far.

Two minutes into the scenario, the previously stable patient deteriorated. The deterioration was indicated by a rise in heart rate and a drop in oxygen saturation. Depending on the actions of the teams, the patient entered a life-threatening state indicated by a severe rise in expired CO_2, extreme arrhythmic tachycardia, or further desaturation. The patient remained in a life-threatening state until the team decided to administer dantrolene sodium, which indicates that the team settled on the MH diagnosis. The scenario was finished after the team definitively ordered dantrolene sodium.

During the scenario, one simulation instructor (an anesthesiologist) controlled the patient manikin, and an HCI researcher operated the CDSS from an adjunct control room. The HCI researcher imitated the automatic detection of changing symptoms according to the state of the patient manikin. The confederate surgical team simulating the laparoscopic appendectomy was equipped with in-ear headphones, through which the simulation instructor could give instructions or provide answers to medically relevant questions from the anesthetic team during the scenario. Furthermore, the teams received a stationary telephone to simulate communication with people outside the operating room, for example, with a senior physician, as is common during operating room crises. Information that cannot be simulated, such as the patient's temperature, cyanosis of the patient's skin, or the gas quantity, was provided by the simulation instructor via loudspeakers in the simulation room upon request.

Fourth, after the scenario, we conducted a medical debriefing with the teams. Furthermore, each anesthesiologist and each nurse completed a demographic questionnaire and the NASA-TLX questionnaire. A single session lasted about 60 minutes.

3.6 Qualitative Data Analysis

For the qualitative analysis, we adapted the video-based analytical practice outlined by Heath et al. [47]. First, we preliminarily

Figure 2: Study setup for CDSS teams (left image) and No-CDSS teams (right image). In the background, the upper screen (black frame) shows the vital signs, and the screen below (white frame) belongs to the ventilation machine. The trolley on the right side contains medications and other equipment.

reviewed the captured videos in the annotation software ELAN (https://archive.mpi.nl/tla/eland). For the CDSS teams, we added simple descriptions of when and how the CDSS was used. The videos of the No-CDSS teams were annotated in the same way for comparable events. All annotations were performed by LS. After the preliminary review of all videos, authors LS and TG outlined five common events of the decision-making process in which most of the CDSS teams used the CDSS: (1) suspecting an emergency, (2) monitoring and collecting available information, (3) weighing up diagnosis options and consultation, (4) interpreting the results of the ABG, and (5) making a final diagnosis decision and initiating treatment. Next, LS conducted a detailed conversation analysis transcription of the events, including utterances, pauses, inflection, and overlap, as well as a notation of visible conduct and gaze as proposed by Heath et al. [47].

Second, we engaged in an iterative substantive review process. In weekly meetings, authors LS and TG reviewed and discussed the events in both CDSS and No-CDSS teams using used an online collaboration board (https://miro.com). Events lasted 20 to 80 seconds and were fully transcribed and illustrated with anonymized screenshots.

Third, we conducted five so-called data sessions [47] with at least one anesthesiologist (authors CD or OH) and three HCI researchers per session (author LS, TG, SW, or RZ). As suggested by Heath et al. [47], our data sessions included several individuals viewing, commenting on, and analyzing videos and transcripts together to gain new perspectives, collect immediate feedback from colleagues about the data, and clarify any technical questions (e.g., about the medical procedure). LS facilitated and documented all data sessions on the online collaboration board. In each session, we focused on one of the abovementioned five events of the decision-making process. For each event, we chose one excerpt from the CDSS and No-CDSS conditions. All data sessions started by watching the video footage of a No-CDSS team several times, with all attendees noting observations and impressions on the online collaboration board. This step was then followed by an in-depth discussion of the detailed transcription to understand the anesthetic team's communication and behavior concerning the decision-making process. In these discussions, we focused on describing the approaches, interactions, and particularities evident in the excerpt. Next, we performed the same steps on a CDSS excerpt of the same event. The data session members were instructed to stay close to what they could see in the video sequence and be aware and reflective about interpretations that are not evident from the actual video [47]. After we gained an in-depth understanding of each excerpt (CDSS and No-CDSS), we started going back and forth between excerpts, considering similarities and differences between conditions. This was initially performed at a very detailed level (i.e., interactions, individual phrases, pauses in speech, etc.) and then increasingly extended to a more abstract level (i.e., the overall configuration of the situation, more structural changes through the CDSS, etc.). As the number of data sessions increased, references were also made to findings from previous data sessions on other events, uncovering persistent changes. In addition, technical questions about the diagnostic process and local procedures were repeatedly discussed with the anesthesiologists to avoid misinterpretations.

Fourth, in an analytic review, author LS summarized the observations of the data sessions and progressively refined and compared them with a focus on the teams' work structure, how they worked together as a team, if and how decisions were made, and how the CDSS affected the decision-making process of CDSS teams. Here, all data (video and transcripts) on all events for both CDSS and No-CDSS teams was reviewed again and used to refine the insights generated in the data sessions. The process was similar to the data sessions and began by reviewing all excerpts from No-CDSS teams on a particular event, followed by all excerpts from CDSS teams on the same event. The initial focus was on gaining an in-depth understanding of each excerpt, followed by an in-depth comparison between conditions, and concluded by comparisons between conditions at a more abstract level. After all data on all events had been reviewed, comparisons were drawn between insights from different events. As a result, this analytic review revealed the most striking similarities and differences between CDSS and No-CDSS teams, which informed the next step. All observations and insights of the analytic review were documented on the online collaboration board and summarized in written paragraphs.

Fifth, authors SW, LS, and TG reviewed the insights from the previous steps documented on the online collaboration board and in written paragraphs individually and then met to discuss their perspectives to identify a focused set of how the CDSS affected the

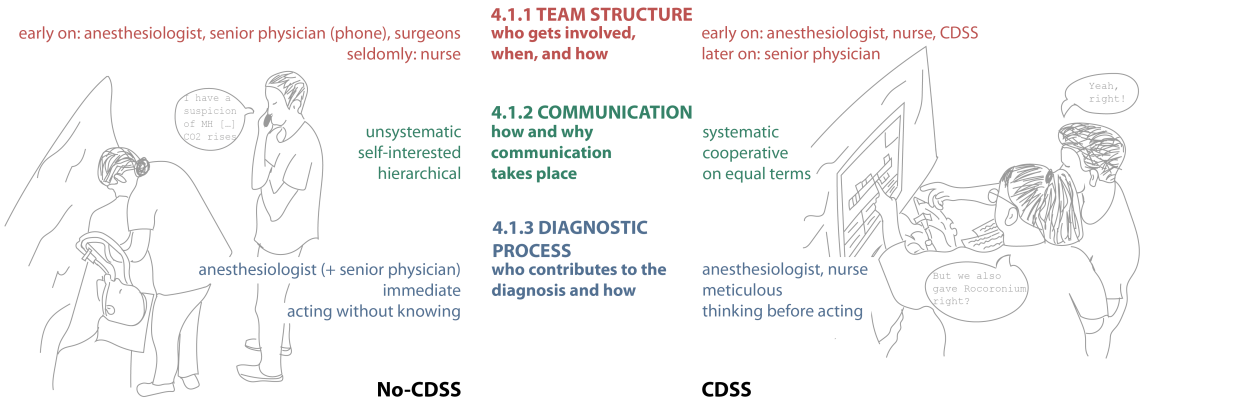

Figure 3: Summary of the main qualitative findings: Three novel ways in which the CDSS affected the decision-making process.

decision-making processes of teams. These discussion meetings were repeated several times until they agreed on a set of changes that could best be justified from the data and presented in the final write-up. We follow the approach of Mentis et al. [79] and present not only descriptions of how CDSS use affected the decision-making process but also excerpts from our transcripts as evidence. Each excerpt was chosen carefully to illustrate a recurring phenomenon in the data as an example. For this paper, the German excerpts were translated into English.

3.7 Quantitative Data Analysis

Based on the videos, we extracted the specific times of (1) the start of the patient's deterioration, (2) the patient entering the life-threatening condition, and (3) the verbal statement to order dantrolene sodium. The period from (1) to (3) was defined as the time until a final diagnosis decision. For the NASA-TLX, we separately analyzed the anesthesiologists' and the anesthetic nurses' data. For the inferential analysis, we calculated the Raw TLX score by averaging the six single scales [43]. For descriptive purposes, we also provide the results of the single scales. All statistical tests were two-sided with α =.05. If not stated otherwise, all parametric tests met the required statistical assumptions.

4 Results

4.1 Qualitative Results

Our analysis uncovered three novel ways in which CDSS use affected the decision-making process, which were particularly present in the data and relate to essential aspects of managing operating room crises. The three changes relate to the anesthetic teams' structure, communication, and the diagnostic process which we have summarized in Figure 3. To describe the changes, we first present examples from a No-CDSS team for each change, along with supporting excerpts from the transcripts, and then corresponding examples from a CDSS team. The transcript excerpts include the anesthesiologists' and nurses' verbal conversation (e.g., utterances, pauses, inflection, overlap) and behavior (e.g., visible conduct and gaze) and provide snapshots from the videos to support understanding of the context. Each section ends with a small summary of the described changes.

4.1.1 Changes in team structure: who gets involved, when, and how. In the **No-CDSS teams**, we observed that the anesthetic team (anesthesiologist and nurse) was often extended once it was realized that the patient's condition was deteriorating. For example, the No-CDSS teams often included the surgical team as extended team members to gather information about the surgical procedure. In addition, the senior physician was called in as an extended team member for telephone counseling. Most frequently, the anesthesiologist called the senior physician and used this call as an opportunity to summarize evidence for a particular diagnosis. During this process, the nurse was left out of the conversation but could hear what the anesthesiologist said on the telephone.

Before the moment depicted in the excerpt in Figure 4, the No-CDSS team noticed that the patient was deteriorating, and the team started collecting information about the patient and their status. During the excerpt, the team was about to weigh up potential diagnoses. This process primarily occurred on a telephone call between the anesthesiologist and the senior physician, not within the anesthetic team (Figure 4, from second 2). The anesthesiologist briefly introduced themselves and stated the patient's name (here: Stenosis, Adam) to the senior physician. Next, the anesthesiologist immediately presented the suspected diagnosis (Figure 4, second 7). Following that, they presented a few (but not all) relevant symptoms to verify the diagnosis and repeated their suspicion (Figure 4, from second 9). The telephone call remained one-sided for long stretches, and the senior physician only got a chance to speak when the anesthesiologist had nothing more to say (Figure 4, second 16). As a result, the anesthesiologist seemed to have already decided on the diagnosis without consulting with the nurse and before hearing the senior physician's opinion. The diagnosis of MH was primarily based on the thought processes of the anesthesiologist, who used the telephone call with the extended team member, the senior physician, to communicate and justify their diagnosis. During this process, the nurse did not participate in the conversation and the decision-making process but could hear what the anesthesiologist said on the telephone.

In the **CDSS teams**, some anesthesiologists also considered extending the team by calling the senior physician. Calling the senior physician seems to be a deeply rooted procedure in the

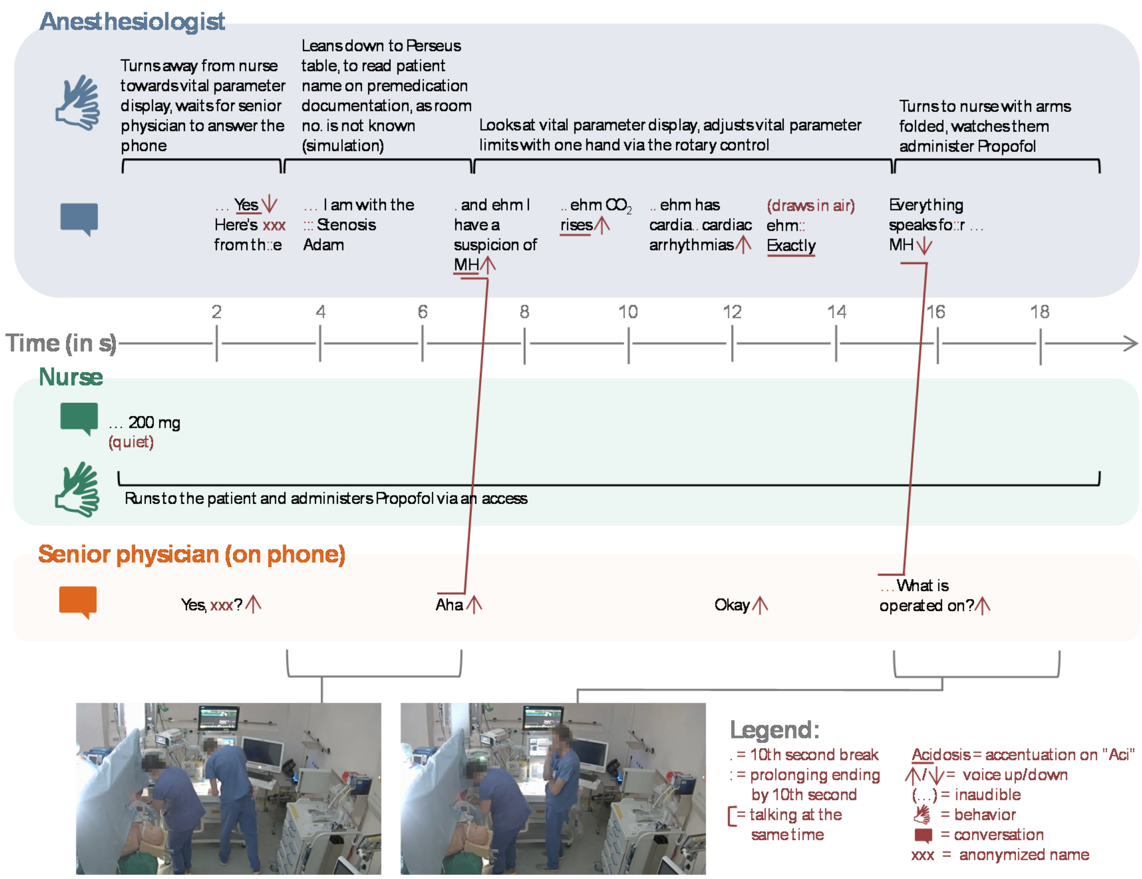

Figure 4: Transcript and snapshots of a No-CDSS team excerpt. In the snapshot, the nurse is in the foreground, and the anesthesiologist is in the background. The excerpt reflects an event in the middle of the malignant hyperthermia (MH) scenario, which is about weighing up diagnosis options and consultation. Here, the anesthesiologist justifies their diagnosis to the senior physician on the telephone.

event of an operating room crisis. In one CDSS team, for example, we observed that an anesthesiologist held the telephone in their hand while weighing up diagnoses, indicating that they considered extending the team by calling the senior physician. If the senior physician was called at this point in time, the call was similar to that of No-CDSS teams. However, at this point, most CDSS teams decided to consult the CDSS as a kind of extended team member instead of the senior physician. When the CDSS was consulted, this was often done by both members of the anesthetic team with shared tasks or at least in such a way that both team members remained informed about the current status. If the senior physician was called later, anesthesiologists stated their diagnosis and requested further staff for medication preparation.

In the excerpt shown in Figure 5, the anesthesiologist of the CDSS team was occupied with preparing the materials for the continuous narcotic infusion. Therefore, after verbal consultation with the anesthesiologist, the nurse took over the task of interacting with the CDSS, typed in "Malignant Hyperthermia," and started reading aloud the first two symptoms from the overview (Figure 5, from second 8). The nurse then confirmed the tachycardia symptom by tapping the symptom box on the screen and continued explaining aloud that this symptom could already be verified from their point of view. At the same time, the anesthesiologist informed the nurse about their actions verbally (Figure 5, second 12) and then joined the interaction with the CDSS. The anesthesiologist agreed with the nurse on the verified symptoms and began to read the following symptoms from the overview out loud to process and discuss them.

As summarized in Figure 3, like the No-CDSS team, the CDSS team's anesthesiologist expressed a suspicion of MH early on. However, when using the CDSS, both anesthetic team members were more involved in the thought and review process of the individual symptoms and differential diagnoses. In some teams, the CDSS with its additional "opinion" was humanized by calling it "her" opinion or even calling the CDSS "the aunt." In other CDSS teams, the CDSS was acknowledged as an application or tool to use during work but not as a team member. Nevertheless, this additional "opinion" of the CDSS led to the change that in CDSS teams, senior physicians were not involved until later, and the surgical team was also less

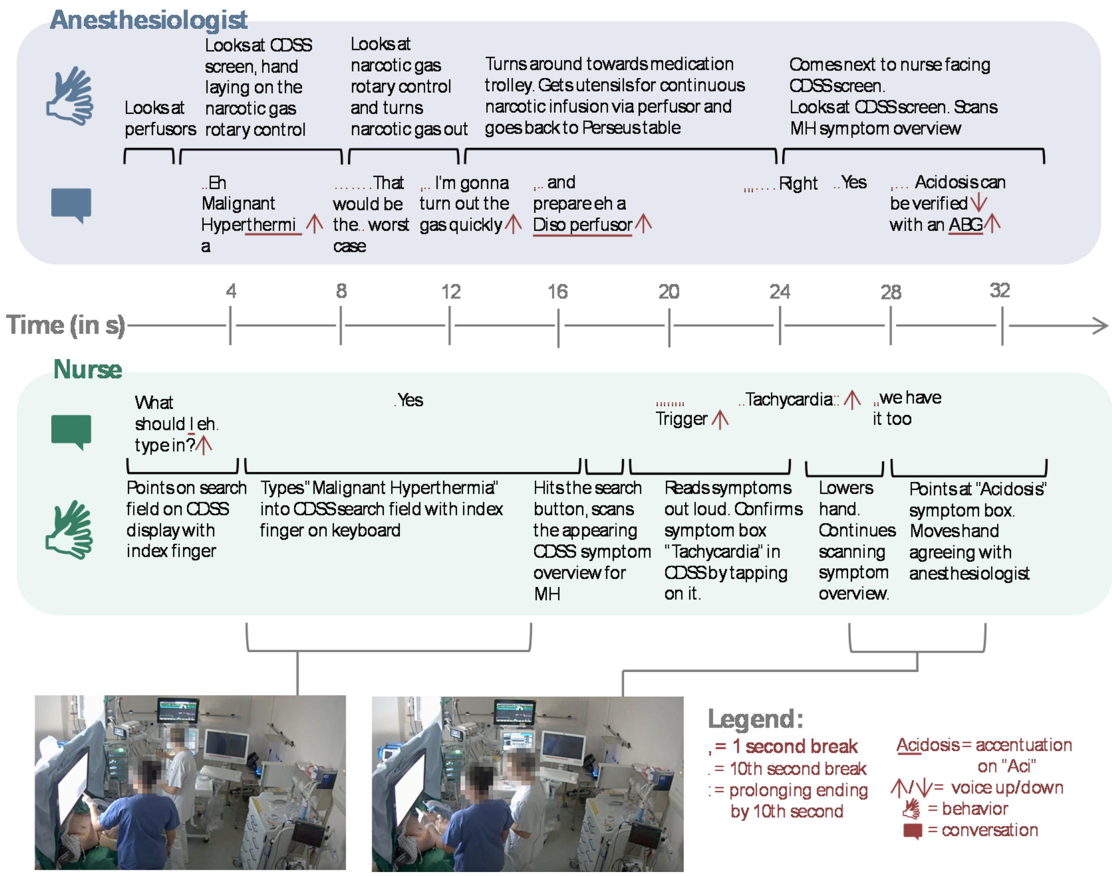

Figure 5: Transcript and snapshots of a CDSS team excerpt. In the snapshot, the nurse is in the foreground, and the anesthesiologist is in the background. Similar to the excerpt in Figure 4, this excerpt reflects an event in the middle of the malignant hyperthermia (MH) scenario, which is about weighing up diagnosis options and consultation. Here, the team checks the likelihood of their potential diagnosis using the CDSS.

frequently engaged as part of the extended team (e.g., in asking questions about the surgical procedure to rule out any surgery-related incident). In CDSS teams, communication, work, and the diagnostic process took place more within the anesthetic team (including the CDSS) and less in the extended team, changing the team structure. Compared to No-CDSS teams, there was a change in who was involved during operating room crises, when, and how.

4.1.2 Changes in communication: how and why communication takes place. In **No-CDSS teams**, we observed a large discrepancy between anesthetic teams regarding verbal communication about the diagnosis and decision-making. While some anesthesiologists and nurses communicated extensively, others rarely exchanged thoughts relevant to the diagnosis or decision-making process, with the nurses being more passive and less involved and more communication taking place between the anesthesiologist and the extended team members (e.g., senior physician on the telephone, the surgical team). Verbal communication often appeared unsystematic and focused on confirming an initially assumed diagnosis, not on jointly developing a diagnosis. In addition, communication within the anesthetic team was often organized hierarchically, with the nurses only expressing opinions if they were actively asked to do so by the anesthesiologists. If anesthetic team members talked about the possible diagnoses, they often started expressing thoughts out loud but not finishing their sentences, not answering questions, or reacting to questions with delays. This type of communication appears to reflect individualized thought processes.

In Figure 6, the No-CDSS team just received results from the ABG and tried making sense of them. The anesthesiologist took the initiative and invited the nurse to share their opinion and thus fostered team communication by asking two questions (Figure 6, second 3, second 5). The nurse only responded to the second question, and their response was delayed. In turn, the anesthesiologist expressed further ABG results as statements rather than questions, supporting their own thought process (self-interest) rather than further inviting the nurse to join the conversation (Figure 6, from second 9). The anesthesiologist's strategy for analyzing the ABG results (e.g., most important first, from top to bottom) remained unclear from the outside. The nurse was not invited to join the anesthesiologist's thought process. At the end of this excerpt, the anesthesiologist verbalized a diagnosis they reached through their

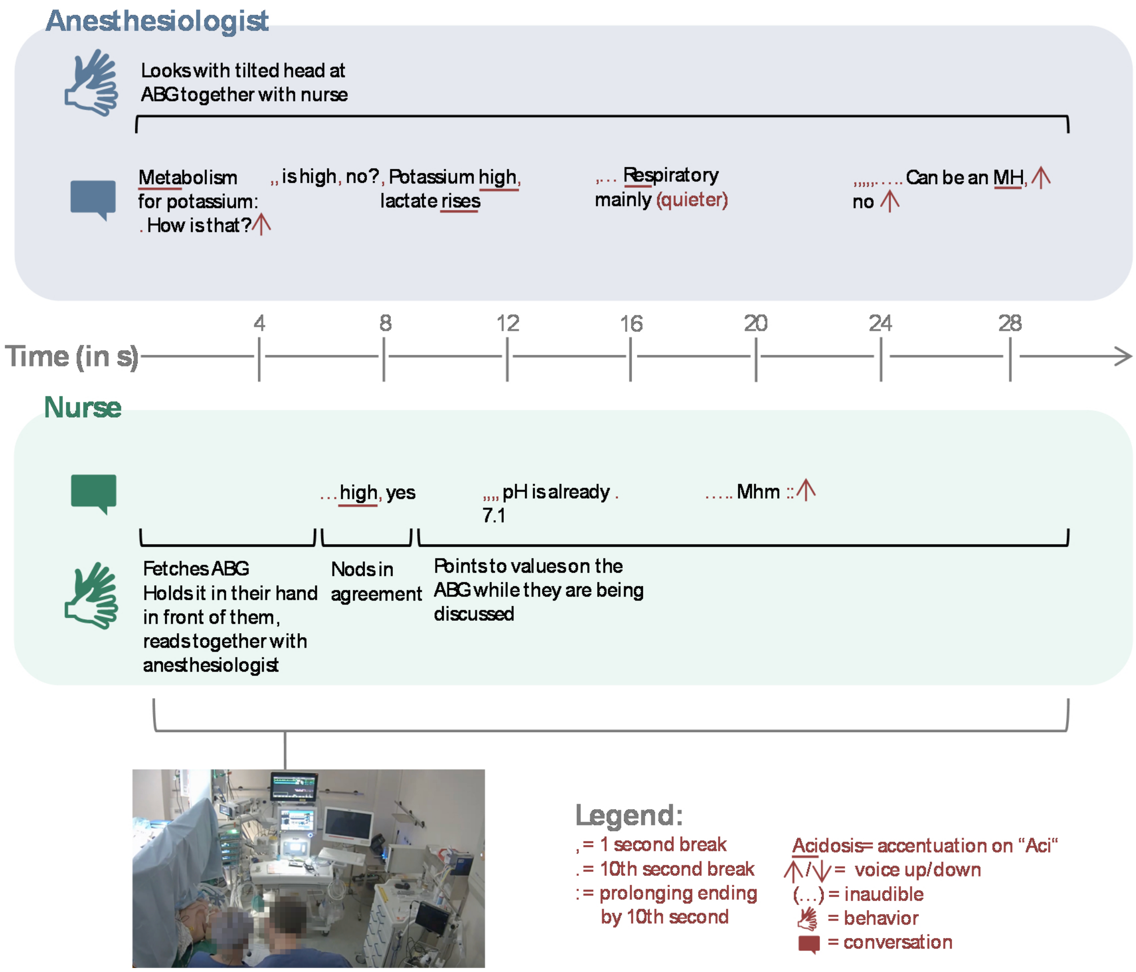

Figure 6: Transcript and snapshots of a No-CDSS team excerpt. In the snapshot, the anesthesiologist (right) and the nurse (left) look at the arterial blood gas (ABG) paper printout (not visible in the snapshot). The excerpt reflects an event later in the malignant hyperthermia (MH) scenario, where teams interpreted the results of the ABG.

own thought process and, finally, invited the nurse to respond to this suggestion for validation (Figure 6, second 24); however, the nurse never responded.

In **CDSS teams**, however, communication about the diagnosis decision-making between team members was more structured and intended to inform the other team member about each other's thought processes. To do so, both team members used verbal and non-verbal communication and followed the CDSS suggestions for structuring (e.g., symptoms are ordered in descending order of importance for a diagnosis).

As in the No-CDSS excerpt, the CDSS team just received results from the ABG and is positioned so both can look at the ABG (Figure 7). Here, the anesthesiologist also started analyzing the ABG, verbally expressing their thoughts. What was different in the CDSS team compared to the No-CDSS team, however, was the team members' positioning toward the CDSS and the extensive use of non-verbal communication, such as pointing or circling with a finger to share their own thought processes with the other team members. In doing so, the CDSS team members could adjust to each other's thinking or interpretation pace and style by looking at which symptom was currently the focus of the other team member.

The team's communication was cooperative. In addition, the CDSS team followed the CDSS's symptom overview order, which gave them a predefined, shared structure for analyzing the ABG that both team members could follow. In this case, the nurse was not only able to follow the thought process of the anesthesiologist by seeing which symptom they were currently checking on the CDSS, but also expressed their opinion and ideas to which the anesthesiologist then reacted. As a result, communication in CDSS teams took place on equal terms.

As summarized in Figure 3, unlike No-CDSS teams, communication in CDSS teams was more structured inspired by the CDSS and the teams used it as a checklist to keep track of the ABG results (still to be) checked. The team members could return to results they were not sure about and memorize which results were checked already, which helped them to structure their work, reduce repetition, and

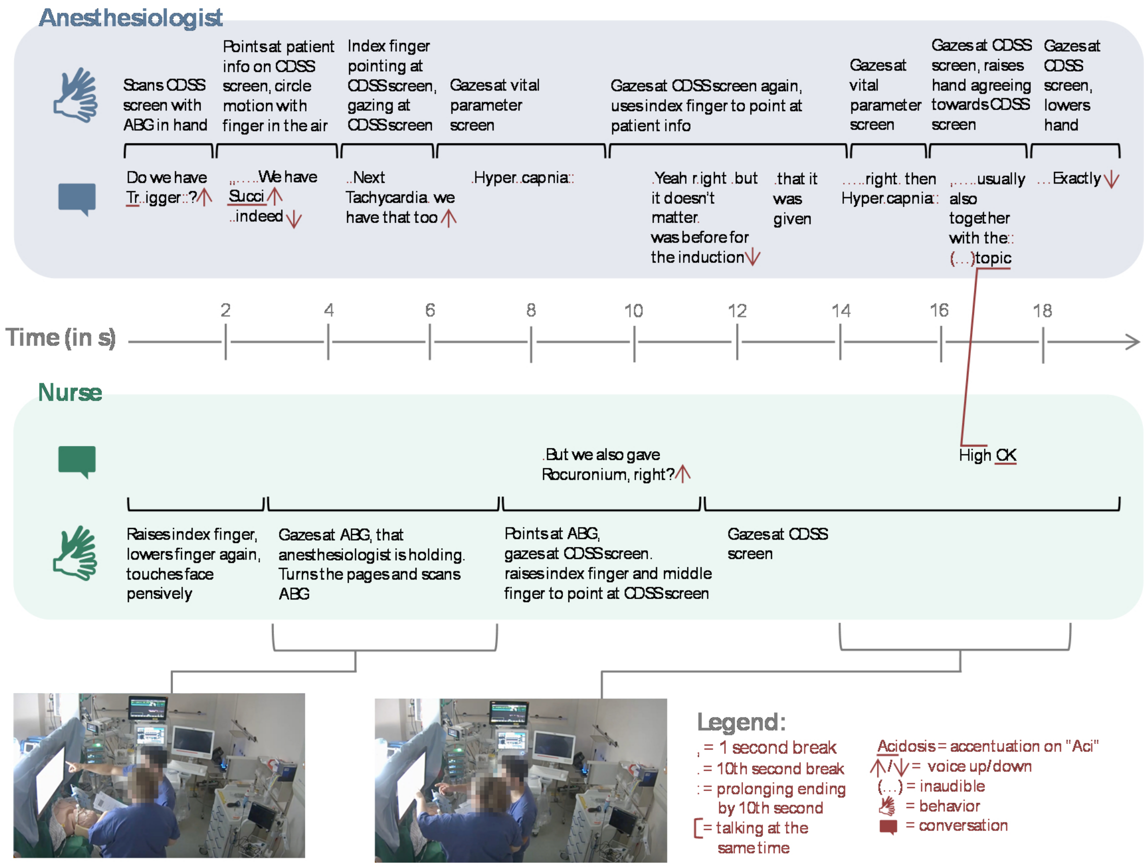

Figure 7: Transcript and snapshots of a CDSS team excerpt. In the snapshot, the nurse is in the foreground, and the anesthesiologist is in the background. Similar to the excerpt in Figure 6, this excerpt reflects an event later in the malignant hyperthermia (MH) scenario, where teams interpreted the results of the arterial blood gas (ABG). The team discusses the injected medication, e.g., succinylcholine ("Succi"), and the patient's symptoms retrieved from the ABG, e.g., creatine kinase ("CK").

support their fellow team members. While the anesthesiologists in both teams shared their thoughts and opinions on the ABG results, the nurses in the CDSS teams did so more frequently, leading to communication on equal terms and a cooperative thought process within the team.

4.1.3 Changes in the diagnostic process: who contributes to the diagnosis and how. Among the **No-CDSS teams**, we observed substantial differences in when and how the teams decided on a diagnosis. Many teams made the decision early, after realizing that the expired CO_2 level of the patient had risen alarmingly and without ordering and analyzing an ABG, which is necessary to make a more reliable diagnosis. In turn, the diagnosis was based on monitoring the patient's vital parameters and gathering information about the administered drugs. As seen in the previous examples (e.g., Figure 6), there was little verbal consideration of differential diagnoses in the No-CDSS teams. We suspect that the No-CDSS teams considered other differential diagnoses, but only in their individual, unspoken thought processes, which were not accessible to us as researchers and also not to the other team members in the scenario. A recurring pattern in the No-CDSS teams was the telephone call with the senior physician, which was used to gather all the "evidence" for diagnosing MH (e.g., Figure 4). However, we observed that in five out of six No-CDSS teams, the decision was not influenced by the telephone call with the senior physician. The anesthesiologists, the primary decision-makers in all cases, had already decided on the diagnosis before the call and just communicated to the senior physician why they came to a decision, summarizing the occurring symptoms and what they needed to counteract the MH.

In **CDSS teams**, however, the diagnostic process was characterized by a more traceable and transparent consideration of differential diagnoses and a more structured, homogenous diagnostic process across teams. Prompted by the CDSS, most teams meticulously checked all symptoms displayed by the CDSS, often in the order suggested by the CDSS. This systematic check of all essential symptoms is also reflected by the fact that all CDSS teams ordered an ABG during their diagnostic process, presumably triggered by symptoms displayed by the CDSS that can only be verified through an ABG. The CDSS was also used as mental support for symptoms already checked and for arguing in favor of a specific diagnosis

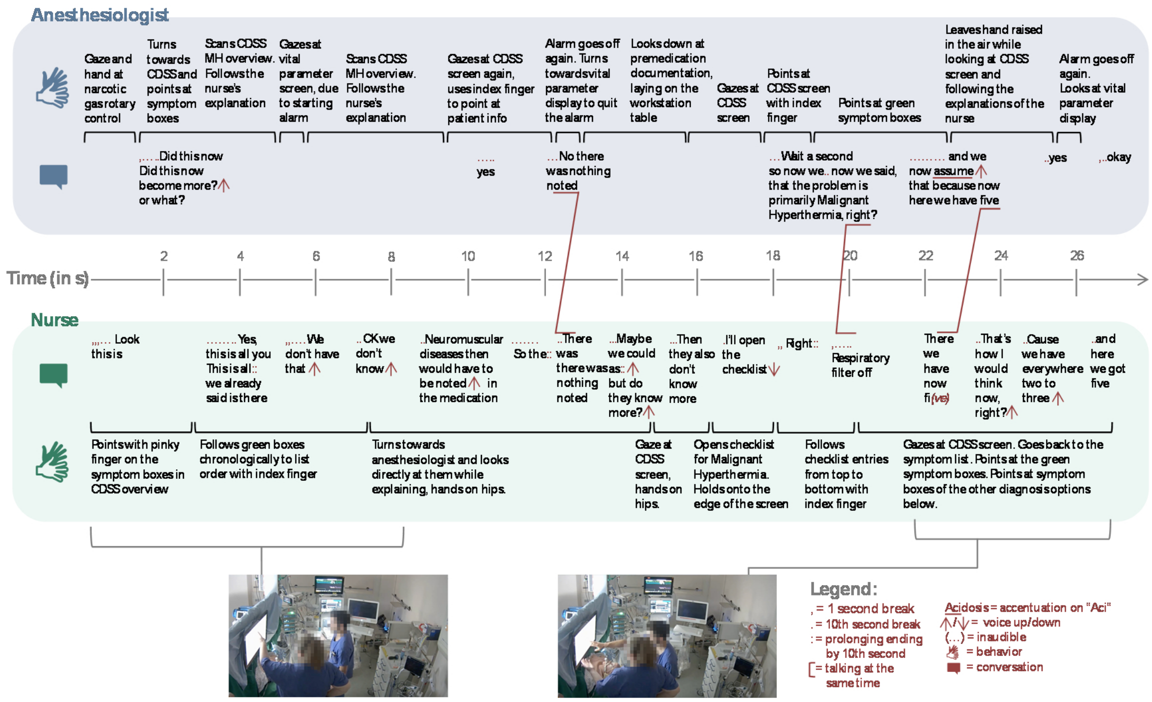

Figure 8: Transcript and snapshots of a CDSS team excerpt. In the snapshot, the nurse is in the foreground, and the anesthesiologist is in the background. The excerpt reflects an event close to the end of the malignant hyperthermia (MH) scenario, where teams made final diagnosis decisions and initiated treatment. Here, the team discusses the probability of the diagnosis of MH based on the number of validated symptom boxes.

based on the estimated probability (i.e., the number of green boxes displayed for each of the alternative diagnoses).

In the excerpt shown in Figure 8, the CDSS team is about to decide on the diagnosis and subsequent treatment. At the start of the excerpt, the anesthesiologist realized that many of the symptoms from the MH symptom overview were valid (i.e., green boxes). However, the anesthesiologist was trying to decide if they should explore other diagnoses, in this case sepsis, or if they should continue proceeding with the MH overview, which they had already been working on for a longer time (Figure 8). In contrast, the nurse seemed affected by the time pressure and wanted to initiate preventive actions instead of continuing to weigh further diagnosis options. The nurse hinted to the anesthesiologist about making a definite decision by pointing out the number of green symptom boxes in the MH overview, expressing a high probability of the diagnosis (Figure 8, from second 4). The nurse explained the progress to the anesthesiologist by pointing at each symptom box in the order indicated by the CDSS and explaining which symptom they "had," meaning which symptoms they could already verify. The anesthesiologist was able to follow this explanation; however, they still did not seem sufficiently convinced to commit to the diagnosis. Similarly, we observed in other CDSS teams that it was not the probability visualization alone that convinced the teams to make the final decision for the MH diagnosis. The visualization made the decision more visually apparent for the teams, but all teams had the assumption for a longer time and were already quite sure of the diagnosis, based on the closer monitoring of the patient, the documented anesthetic drugs, and the results of the ABG.

In the excerpt in Figure 8, the nurse uses the symptom visualization of the CDSS again to support the argument for the MH diagnosis decision by pointing at the green symptom boxes, counting them out loud, and comparing the number of boxes to those of other differential diagnoses (Figure 8, from second 22). This is when the anesthesiologist is finally convinced that they should commit to the decision and start a treatment.

As summarized in Figure 3, unlike No-CDSS teams, CDSS teams approached the diagnostic process meticulously and tried to be particularly confident about the diagnosis, weigh it up, and consider all the options before acting. They often did not decide on a final diagnosis lightly. As a result, the decision-making process in CDSS teams was also characterized by uncertainty and not committing to a diagnosis for some time ("thinking before acting"). However, this process also led to more reliable and sound diagnoses than in No-CDSS teams, because more symptoms were checked and differential diagnoses were considered. Unlike in No-CDSS teams, the senior physician was less involved in the diagnostic process via telephone in CDSS teams but was consulted afterward. In CDSS teams, the diagnostic process took place within the anesthetic team.

4.2 Quantitative Results

Contrary to our expectation, an independent t-test of the time until a final diagnosis decision showed no difference between the CDSS

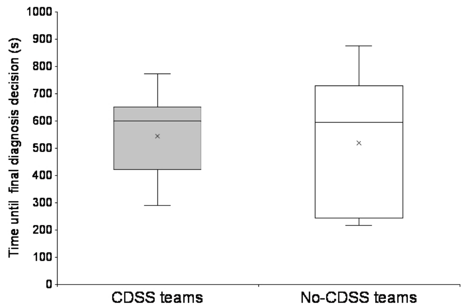

Figure 9: Box plots of the time until a final diagnosis decision separated into CDSS teams and No-CDSS teams. Lines indicate the median and × indicates the mean of the teams.

(M=545 s, SD=173 s) and the No-CDSS teams (M=519 s, SD=284 s), $t(12)$=0.205, p=.412, d=0.11. Despite a large descriptive difference in the variance between the team conditions (Figure 9), a Levene's test did not indicate significant differences in the variance for the time until a final diagnosis decision, $F(1,12)$=3.996, p=.069.

Because the start of the patient entering the life-threatening condition depended on the teams' actions, some teams entered the state earlier than others. To test whether entering the state early might speed up the time until a final diagnosis decision, we correlated the period from (1) the start of patient deterioration to (2) the patient entering the life-threatening condition with the period from (2) to (3) the verbal statement to order dantrolene sodium. The resulting Spearman correlation indicated no association ($r = -0.045$, p=.946). Hence, the transition time did not affect the time until a final diagnosis decision.

The results of the NASA TLX subjective workload assessment are illustrated in Figure 10. For the anesthesiologists, an independent t-test of the Raw TLX scores showed no significant difference between the CDSS (M=9.86, SD=2.45) and the No-CDSS teams (M=11.45, SD=1.54), $t(12) = -1.459$, p=.170, d=0.78. For the anesthetic nurses, an independent t-test of the Raw TLX scores showed a significant difference between the CDSS (M=7.71, SD=3.07) and the No-CDSS teams (M =10.74, SD=1.00), $t(12) = -2.483$, p=.029, d=1.33, with nurses in the No-CDSS teams experiencing significantly higher levels of subjective workload.

5 Discussion

In this study, we addressed a critical gap in the existing literature by exploring how CDSS use affects and changes the decision-making processes in acute care teams and what we might learn from these insights for the design and evaluation of future CDSSs. We employed a mixed-method approach to capture both qualitative insights on conversational grounding and quantitative insights on clinical performance and workload of anesthetic teams facing a medical crisis in a full-scale medical simulation with and without a CDSS. In the discussion, we (1) highlight three overarching patterns evident in our quantitative and qualitative data (homogenization, nurse empowerment, and friction in decision thinking) that change our understanding of CDSS effects and assess whether observed changes are detrimental or beneficial by considering CRM, (2) reflect on the strengths and limitations of the present work, and (3) summarize novel implications for future CDSS design and evaluation.

5.1 Homogenization

We found a recurring pattern of homogenization across both quantitative and qualitative data. In contrast to our expectations, the quantitative results of the time until a final diagnosis decision showed no significant differences between the CDSS and the No-CDSS teams. However, the box plots in Figure 9 indicate that the time until a final diagnosis decision had a wide distribution for No-CDSS teams, whereas the times were more homogenous for CDSS teams. Similarly, homogenization was also evident in the qualitative data, especially in relation to communication and symptom review.

Following up Klüber et al.'s [67] report of improved communication when using the CDSS, the present two-group design enabled a detailed comparison of communication patterns within the CDSS and No-CDSS teams. While the verbal communication differed greatly between different teams in the No-CDSS teams, communication was more frequent and more homogenous among teams in the CDSS teams. A possible explanation might be that in No-CDSS teams, communication was more based on individuals' preferences, characters, habits, and prevalent work culture, whereas in CDSS

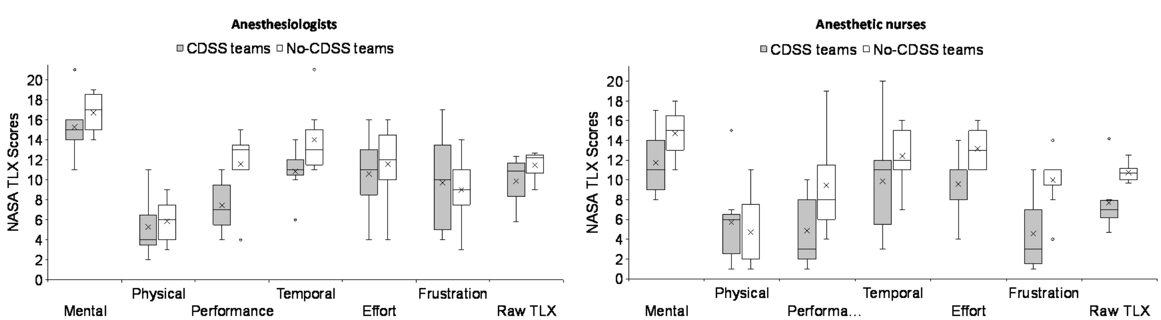

Figure 10: Box plots of the NASA TLX scores for anesthesiologists (left) and anesthetic nurses (right) separated into CDSS teams and No-CDSS teams. Note that lower performance ratings indicate higher subjective success in accomplishing the goals of the task. Line in the box indicate median, × indicates mean.

teams, the CDSS might have disrupted these structures and, through its specific design, invited a certain kind of team communication that then led to more homogeneity between teams. Activities, especially those connected to the CDSS (e.g., reviewing a symptom), were more verbally coordinated, and both team members were more involved in the thought and review process of individual symptoms and differential diagnoses. From a CRM perspective, this communication style can be considered good practice, as exchanging information, coordinating activities, and supporting each other are essential aspects of good teamwork [28] and support the generation of collaborative mental models [16].

Another qualitative homogenization effect in the CDSS teams was related to the symptom review processes. While the number and order of symptoms reviewed in the No-CDSS teams differed greatly between teams (e.g., some ordered an ABG, some did not), the symptom review process was more homogenous in CDSS teams (e.g., all teams ordered an ABG). Prompted by the CDSS's visualization of symptoms to review, CDSS teams followed a more homogenous approach to symptom review. For example, most teams followed the order of symptoms presented by the CDSS. In this way, the CDSS also served as a visualization of the teams' situation awareness during early decision-making steps (e.g., information gathering and reasoning) and a place to offload and externalize memory. This frees cognitive capacity for, for example, combining the symptoms and generating a bigger picture [88]. Furthermore, given that the CDSS prompted users to gather particular information and served as a documentation aid (i.e., green, gray, or blue boxes indicating the status of a symptom and checkmarks indicating who confirmed the status), the CDSS might have also reduced the amount of articulation work needed during the decision-making process. In this way, the CDSS use resembled how whiteboards or information displays are used as tools to gather and understand information within teams [27]. Communication around symptom reviews was structured and focused on gathering and understanding information about symptoms, which can be considered good practice from a clinical perspective [28], especially considering that decisions must be made within minutes or seconds [101]. The more homogenous symptom review process that followed the CDSS's visualization of symptoms also led to a more thorough review of all relevant symptoms and can be considered good practice from a CRM perspective, since it supports maintaining standards [28].

The observations of the symptoms review process echo previous findings in relation to wall mounted, handheld, and paper-based cognitive aids in the potential to increase adherence to clinical guidelines [26, 36, 63, 86]. However, there is a fundamental difference between the use of cognitive aids described in the literature [12, 37] and the use of the CDSS in our study. Burian et el. [12] describe a "sampling fashion" use of cognitive aids in which staff members first act based on individual knowledge and sample the cognitive aid to double-check, generate new ideas, or seek specific information. One explanation for this difference might be that, as described above, the symptom review process in the present study was much more homogenized across teams and aligned to the structure of the CDSS, and we suggested that the well-aligned symptom structure and the visualization reduced the amount of articulation work that is usually needed to discuss the symptoms, gather information, and reach a common understanding in the teams. Another explanation might be the phase of the decision-making process. Cognitive aids such as the Stanford Emergency Manual [34] are designed for late stages of diagnostic processes such as treatment decisions (e.g., what steps must one take if the diagnosis is MH?), whereas the present CDSS also addresses early stages of diagnostic processes such as integrating and collecting information (i.e., what is wrong with the patient?). The cognitive processes in treatment decisions might be more related to memory retrieval of the correct actions and their execution, whereas the diagnosis decisions require more complex and effortful cognitive processes when gathering information, discussing the symptoms, and reaching a common understanding.

Taken together, the CDSS homogenized team communication and work steps in symptom review, which might explain the greater homogeneity in the times to a final diagnosis in CDSS teams. Guided by CRM, we interpret the homogenization as something positive but want to highlight that the homogenization has been the result of discretionary CDSS use, and homogenization (or even standardization) was not the aim of the initial design by Klüber et al. [67]. Indeed, standardization might not be ideal in the fast-paced and complex decision-making process in such emergency situations or healthcare settings in general [11, 19], and tailoring a CDSS to one specific way of reaching decisions has been problematic [106].

5.2 Nurse Empowerment

Across both conditions, anesthesiologists took the lead in communication (e.g., initiating discussions and asking questions), which can be considered good practice from a CRM perspective, since it confirms roles and responsibilities in teamwork [28]. However, the way communication and teamwork unfolded over time differed between conditions. Some of these changes seem to have affected subjective workload ratings, especially among nurses. Nurses in the No-CDSS teams experienced significantly higher workload levels than nurses in the CDSS teams (Figure 10). The subscale ratings indicate that the differences are mainly produced by less frustration (e.g., how insecure, discouraged, irritated, stressed, and annoyed the users felt during the task) and less effort (e.g., how mentally and physically hard the users had to work) in CDSS teams than in No-CDSS teams. This result aligns with findings from studies on electronic information displays in emergency departments, which demonstrated that these displays are used for collaborative tasks such as timeouts and handovers, and nurses—but not physicians—experienced reduced subjective workload [53].

From our qualitative data, several observations might explain these results. Regarding the team structure, anesthesiologists in the No-CDSS teams often expanded the anesthetic team by talking to the surgical team or calling senior physicians early on. During these telephone calls, there was no direct verbal communication between the anesthesiologists and the nurses (Figure 4), so nurses had to take a passive role but listen actively to stay updated on the anesthesiologists' current thoughts and intentions and potential tasks that could result from the conversation. In addition, nurses could only speculate about what the senior physicians on the telephone were saying to the anesthesiologists. This ongoing parallel task of actively listening might have led to higher effort for nurses and increased frustration because of being left out of the conversation and decision-making process. In contrast, anesthesiologists in the CDSS teams relied less on people outside the anesthetic team. In light of CRM, not considering all potential information sources and not actively integrating the surgical team can be viewed critically [28, 30]. However, essential information about the surgery, such as the type, could also be gathered without verbally involving the surgical team, and not integrating the nurse as an anesthetic team member, as observed in No-CDSS teams, is also not favorable practice.

CDSS use changed the nurses' role within the anesthetic team drastically, with nurses taking a much more active role in decision-making. For example, nurses started using the CDSS after verbal consultations with the anesthesiologists (Figure 5). Activities were more verbally coordinated, especially activities around the CDSS, and both anesthetic team members were more involved in the thought and review process of individual symptoms and differential diagnoses. This exchanging of information, coordination of activities, and support of each other can be considered beneficial from a CRM perspective, because it may enhance situation awareness and support teamwork [28].

Furthermore, it was not only the "what" but also the "how" of communication in the CDSS teams that promoted the empowerment of nursing staff. Communication in CDSS teams was more focused on informing each other on equal terms and more cooperative than in No-CDSS teams. In CDSS teams, both members expressed and reacted to each other's opinions. Interestingly, nurses in CDSS teams used the CDSS as an additional argument for their position in decision-making, highlighting, for example, the number of symptoms in favor of a specific diagnosis (Figure 7). In turn, the CDSS corroborated the nurse's position and lent it more credibility. Generating options for decisions and getting different opinions on options are considered good CRM practices. The present results are also in line with the finding that large interactive displays enhance common ground, help to establish a joint conceptual problem space, and also enable better workspace awareness (e.g., understanding another person's actions in a shared workspace; [76]). Furthermore, the results add to the missing evidence that interactive displays enhance verbal and gestural communication of groups and the level of involvement of individuals [76].

Nurse empowerment is especially interesting in light of Goffman's [32, 33] four production formats (animator, author, principal, figure). The introduced CDSS can be considered an author in the sense that the content of the CDSS affected, for example, the symptom review process but did not dominate the conversation as the CDSS did in Murdoch et al.'s [82] study on nurse-led telephone triage. More interestingly, the CDSS seemed to provide nurses with the opportunity to animate the content of the CDSS and become an active author of the situation. Previous research has described that nurse interaction style with doctors becomes more direct in case of emergency events [89]; however, in our study, the CDSS seemed to enable or encourage nurses to change to an authoring role. In line with Prowse and Allen [89], we also observed that the principal role (i.e., the person responsible for the conversation) was not challenged. The nurse played the role of author in the conversation by putting forward arguments ("we have a trigger"), summarizing results ("there are more green boxes"), and eventually hinting at the diagnosis without explicitly stating it. In light of CRM, acceptance of a leading and non-leading role is important for good teamwork. Despite empowering the role of nurses, the CDSS did not result in challenging the roles within the team. Our findings, therefore, provide empirical evidence for recent CSCW research on empowering nurses in the clinical context by using technology [103].

A final aspect relates to the role of the CDSS itself and the perspective of the anesthetic team on the role of the CDSS. In the CDSS teams, the senior physician was consulted later compared to the No-CDSS teams. One may speculate that the CDSS was (1) treated and considered as an "AI senior physician" and (2) consulted instead of a "human senior physician". Regarding the first point, some CDSS teams humanized the CDSS, a tendency that has been reported before in the context of acute care technology [56]. Recent work on designing human-AI teaming also investigated the personality traits of a possible AI team agent in acute care, and participating anesthesiologists envisioned an experienced physician who is not neurotic and very conscientious [58]. Consulting the CDSS in our study resulted in a calm, ordered, and very diligent consideration of symptoms. The present CDSS may have had the envisioned effect (i.e., analytic, comprehensive, and calm consideration of the situation) when considering the personality traits of future acute care AI agent teammates. Regarding the second point, we can compare the content and purpose of the No-CDSS anesthesiologist phone interaction with the senior physician's and the CDSS anesthesiologist's interaction with the CDSS. The No-CDSS anesthesiologists stated their hypothesis and summarized their observed symptoms, whereas the CDSS anesthesiologists deliberated on symptoms in a nurse-CDSS-anesthesiologist triad. Finally, when the CDSS anesthesiologists contacted the senior physician, their call was more about informing the senior physician about the event and request staff support. The role of the current CDSS may be described as a calm, rational, and diligent thinker, and the anesthetic team assigned the CDSS an equally qualified role rather than a senior physician role. As a result, the CDSS was not consulted instead but in addition to the senior physician.

Overall, the CDSS did not affect leadership and role responsibilities, but the qualitative and quantitative results indicated a general empowerment of the nursing role. We explain this effect by the visualization of the decision-making processes (i.e., collecting information, deciding on the presence or absence of symptoms) in the CDSS teams. Unlike in No-CDSS teams, where the decision-making process was more dependent on the individual cognition and style of the anesthesiologist in charge, visualization in CDSS teams made this process accessible to nurses. A final aspect is the positioning of the introduced CDSS within the team. The CDSS was considered as an agent on the same hierarchical level with profound knowledge. In light of CRM [28, 30], empowering nurses, getting further opinions from other staff, or incorporating external agents in the form of cognitive aids [50] or CDSS to take part in the decision-making process can be considered a positive, cooperation-fostering element with the aim of creating a mutual working environment.

5.3 Friction Between Analytical and Intuitive Thinking

In the CDSS condition, all teams verbally considered various differential diagnoses as prompted by the CDSS, which can be considered good practice from a CRM perspective, because it fosters team work and re-evaluation [28, 30]. What and how the CDSS teams communicated verbally indicated that they engaged in analytical thinking (i.e., a more rational, deliberate process of collecting information and conscious application of rules; [88]). In contrast, the more individualized approaches of No-CDSS teams to dealing with the operating room crisis and how anesthesiologists communicated indicate that No-CDSS teams engaged in more intuitive thinking (i.e., recognition of single signs or patterns that can be matched with previous encounters or general knowledge; [88]). This is an interesting finding, because it underlines the CDSS's capacity to change the clinicians' style of thinking who would, according to the literature [91], rely more frequently on intuitive thinking in situations with time pressure, uncertainty, and dynamic event development.

The CDSS not only triggered analytical thinking but also helped mitigate common biases, which is beneficial from a CRM perspective. The CDSS teams showed no signs of premature closure, and the teams questioned considered alternative diagnoses, suggesting no indication of confirmation bias. While previous work on strategies to mitigate biases through CDSS designs with ophthalmologists [4] suggested an efficiency–thoroughness trade-off (i.e., decisions might be less biased but require more time), the CDSS teams in our study (within the limits of null-hypothesis testing) did not need more time until the final diagnosis than the No-CDSS teams.

These results sound positive, and previous work described that intuitive and analytical thinking "became seamlessly interwoven" through the interaction with the CDSS ([67], p.1528). However, our conversation analysis and the two-group design make us question this all-round positive picture. The fast, intuitive thinking of No-CDSS teams often resulted in a tentative diagnosis early on, and the No-CDSS teams initiated preventive action such as turning off the anesthetic gas very quickly. Such an approach fits in with what is typically described in the literature, such as a tendency for unsupported, natural diagnosis decision-making in the emergency department, in which 77% of tentative diagnosis were made during (or even before) the first patient encounter [87]. While such an approach may not be considered textbook good practice from a CRM perspective and comes with problematic side effects, such as committing to a diagnosis early on and confirming (rather than falsifying) the diagnosis, the quick preventive actions were effective in the present scenario (i.e., removing the trigger substance of the MH). It is these preventive actions that CDSS teams initiated later rather than sooner because they focused so much on making the right decision first (i.e., considering various differential diagnoses and questioning considered diagnoses). As mentioned above, the average time until the final diagnosis did not differ between conditions, but the CDSS teams heavily relied on analytical thinking, as triggered by the CDSS's design that supports consideration of differential diagnoses and questioning of initial diagnoses, which seemed to have been superimposed on the more intuitive thinking.

The analytical thinking might have resulted in a form of omission bias (i.e., the tendency toward inaction because harm as a result of commission is considered more negative than harm as a result of omission) triggered by the CDSS. Alternatively, CDSS teams may have focused their attention too much on making the right diagnosis decision and had no first tentative diagnosis decision. In favor of the latter explanation, recent research [39] showed that a cognitive aid for in-hospital resuscitation support had positive effects by shifting attention to organizing high-quality advanced cardiopulmonary life support. However, the cognitive aid also kept the team leader's attention on the life support during a later phase of the event in which diagnosis of the arrest and planning of how to proceed with the patient would have been more appropriate. In a similar way, the CDSS in our study might have shifted attention to the diagnosis and away from treatment. In light of CRM, the CDSS resulted in a focus on recognizing and understanding the situation and a neglect of anticipating and asking "what if" questions. Asking the question of "what if this is MH?" may have resulted in acting on such a tentative diagnosis with actions that would help and have no drawbacks, such as removing the trigger substance of the MH and switching to a different drug to maintain anesthesia. Overall, the CDSS has triggered a more analytical diagnosis decision-making process with positive (mitigating common biases, better-informed decision-making) and negative (delayed preliminary treatment actions) effects. Our and other results on cognitive aids [39] indicate that technology may capture and keep attention on a specific task or a specific decision mode. As we highlight below, a future challenge will be accommodating and integrating analytic and intuitive thinking.

5.4 Strengths and Limitations

Our study has several strengths and limitations. First, we were not able to access clinical performance in relation to patient outcomes. Studying patient outcomes is inherently difficult in a simulation environment, and we could only use principles of CRM and the associated non-technical skills (i.e., situation awareness, decision-making, task management, team work; [28]) as assessment methods for best practice. Our primary methodological approach was qualitative. Our secondary quantitative methodological approach provided us with insights but did not meet the standards of a full medical randomized control trial [17]. Second, the anesthetic teams only used the CDSS once and for a single crisis event, and it has been shown that the use of technology—for example, head-mounted information displays in acute care [40]—changes over time and tasks. Third, conversation analysis assumes that only communicative acts can be studied (although Heritage's concept of institutional talk widens this strict assumption [52]). Although we followed this line of thinking in the ethnomethodologically informed and video-based conversation analysis [47], we took a more pragmatic approach in the discussion and enriched the qualitative results with the quantitative results for a more comprehensive picture. We think that the combination of methods resulted in interesting insights that no single method could have generated on its own. Fourth, we provided a detailed description of three changes that occurred when comparing the sense-making process of the anesthetic teams working with and without the CDSS. While we are confident that the observed behavior is the result of acute care staff's education, mindsets, and culture in action (i.e., other anesthetic teams would have shown a similar reaction to the circumstances in the scenario), different healthcare disciplines have different education, mindsets, and culture [20]. Similarly, even so homogenization and friction between analytic and intuitive thinking in the context of CDSS use may not be specific to team settings, our insights are based on a team setting. Consequently, our results and the following design insights are likely to generalize to anesthesiology but may only be used as insights to stimulate thinking about design solutions, guide future research, and generally highlight the role of technology in the decision-making process rather than straightforward and ready-to-implement design suggestions for other healthcare domains or single user settings.

5.5 Implications for CDSS Design and Evaluation

Overall, our study has several novel implications for future CDSS design and evaluation. We principally agree with Zhang et al.'s [114] idea of supporting all steps of the decision-making process, encompassing hypothesis generation, data gathering, hypothesis testing, and decision-making. However, our in-depth analysis, especially of No-CDSS teams (representing the status quo), showed that the process is much less linear and analytic than suggested by Zhang et al. [114]. This is also echoed by observations in other medical domains [87, 91]. It was the CDSS under study that introduced a more textbook-like and linear decision-making process. Therefore, our first design implication is that CDSS designers should start with the premise that staff have a tendency toward intuitive rather than analytical thinking and that decision-making processes in acute care might be much messier and more diverse than reflected by theoretical models of decision-making.

Like other CDSSs [4], the CDSS used in this study was designed to de-bias decision-making processes and encourage clinicians to think analytically [67]. According to our results, this seems to have worked well and a recent opinion paper in the Communications of the ACM argues that the role of AI should be to challenge and not obey [94]. However, we also observed that clinicians using the CDSS were concerned with and focused on making the right decision and therefore took preventive actions later rather than sooner. This observation suggests that the strong focus on analytical thinking may have overridden intuitive thinking. Errors happen with both analytic and intuitive thinking [85], but CDSS designs seem to focus on supporting and fostering only analytic thinking. Our second design implication is therefore to consider and support both processes in CDSS design rather than solely focusing on analytical thinking. In our case, treatment options that have no harm but possible benefits (i.e., removing the trigger substance) could be displayed alongside the symptoms. A more general approach would be to follow the recognition-primed decision model [66] and aim at designing the patterns in the environment to support the recognition process. For example, principles of ecological interface design [13, 105] may enable the design of big data CDSS outputs in such a way that clinicians can recognize known patterns rather than needing to spend time reading and understanding decision suggestions.

Notably, clinicians using the CDSS and following more analytical thinking did not take longer to reach a final diagnosis, even though they also interacted with the CDSS. This shows the effectiveness and efficiency of the CDSS under study [67]. Restating previous insights (for a review, see [59]), many CDSSs, including attempts with large language model interfaces [90], require too much data gathering, data entry, or data comprehension and generate additional workload. Our third design implication is to consider how not to increase workload further when designing novel CDSSs. This may be achieved by different means, and designers can take inspiration from our example. First, the present CDSS mainly relied on direct manipulation via a touch interface, which resulted in effective and efficient interaction. Second, the present CDSS reduced data input requirements. Rather than requiring inputting symptoms, the CDSS presented symptoms and required the user to manipulate their status (present, unknown, absent). Third, the present CDSS visualized the symptoms of the selected diagnosis and the six most likely alternative diagnoses, reducing the need to keep all symptoms and alternative diagnoses in memory. The latter point is crucial when considering the work of Pelaccia et al. [87], who highlighted working memory capacity as the main factor for limited alternative hypothesis generation in the emergency department. In addition, the visualization reduces the ancillary workload for staff who do not lead the team (e.g., the nurses in our study), as they do not have the problem of following the thought processes of the team leader.

Our study showed that the CDSS has structured team communication. Therefore, designing the visual CDSS layout also means designing conversation and possibly affecting Goffman's production formats [32, 33]. In this way, the CDSS can be conceptualized as setting the context and affecting decision-making. The importance of purposefully setting the context and considering this in the design of CDSSs is also highlighted by Croskerry [21], who argued that context might be one of the main constraints on reasoning in medical decision-making. CDSS designers should consider not only the design of human-technology interaction and the improvement of clinical tasks, but also how the design of the technology affects the people who interact with it. Because basically, „designing technology is designing human beings" (p. 29, [104]). Overall, this perspective highlights that it is essential for CDSS design to study "what actually happens," because how a CDSS sets a (different) context and how this change affects conversations and, more globally, decision-making processes can only truly be understood in context. Our fourth design implication, therefore, is to evaluate novel CDSSs in fully simulated contexts early on and observe resulting conversations and actions as they unfold.

6 Conclusion

In this paper, we aimed to gain an in-depth understanding of how CDSSs affect the decision-making processes in acute care. To this end, we conducted a mixed-method study in a high-fidelity, full-scale anesthesiology simulation investigating teams' decision-making processes with and without CDSS support. The combined consideration of the quantitative and qualitative data revealed three patterns that open up novel perspectives on CDSSs. CDSS use led to (1) homogenization (e.g., homogenization of times until a final diagnosis, communication, and symptom review), (2) nurse empowerment, and (3) friction between analytical and intuitive thinking. Introducing a novel CDSS into existing workflows and processes had profound effects on team structure, communication, and diagnostic processes that went well beyond commonly observed effects on effectiveness and efficiency. The CDSS fundamentally changed how diagnoses were made and how teams worked together. Our insights significantly advance previous work and urge researchers to reconsider existing standards in CDSS design and evaluation, such as the intense focus on (only) supporting analytical thinking through CDSS design or on interviews and workshops as means of CDSS evaluation. Future work should consider supporting both analytical and intuitive thinking and evaluating novel designs in the field early on to understand how novel CDSSs change the setting of the context and, thus, decision-making or teamwork more generally.

Acknowledgments

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – (grant number 425868361) and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments. We thank Cordula Baur for helping to improve the figures in the result section.

References

[1] Royal College Of Anaesthetists. (2023). Chapter 1: Guidelines for the Provision of Anaesthesia Services: The Good Department 2024. In *Guidelines for the Provision of Anaesthetic Services*: Royal College of Anaesthetists (RCoA). Retrieved from https://www.rcoa.ac.uk/gpas/chapter-1

[2] Royal College Of Anaesthetists. (2024). Guidance on supervision arrangements for anaesthetists. In *Guidelines for the Provision of Anaesthetic Services*: Royal College of Anaesthetists (RCoA). Retrieved from https://www.rcoa.ac.uk/gpas/chapter-1

[3] American Society Of Anesthesiologists. (2023). Statement on the Anesthesia Care Team. Retrieved from https://www.asahq.org/standards-and-practice-parameters/statement-on-the-anesthesia-care-team

[4] Anne Kathrine Petersen Bach, Trine Munch Nørgaard, Jens Christian Brok, and Niels Van Berkel. 2023. *"If I Had All the Time in the World": Ophthalmologists' Perceptions of Anchoring Bias Mitigation in Clinical AI Support.* In the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Hamburg, Germany. https://doi.org/10.1145/3544548.3581513

[5] Edouard Battegay and Stefano Bassetti. 2017. Grundlagen der Differenzialdiagnose. In *Differenzialdiagnose Innere Krankheiten.* New York, NY, USA: Georg Thieme Verlag.

[6] Geschäftsführung Bda/Dgai. (2015). Behandlungsqualität und Patientensicherheit: Eckpunkte zur ärztlich-personellen Ausstattung anästhesiologischer Arbeitsplätze in Krankenhäusern [Treatment quality and patient safety: Guidelines for adequate staffing at anesthesiologic workplaces in hospitals]. *Anästh Intensivmed.* Retrieved from https://www.dgai.de/alle-docman-dokumente/entschliessungen-vereinbarungen/1762-mindestanforderungen-an-den-anaesthesiologischen-arbeitsplatz.html

[7] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. 2020. *A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy.* In the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376718

[8] Patrick E. Beeler, David W. Bates, and Balthasar L. Hug. 2014. Clinical decision support systems. *Swiss Medical Weekly, 144*, w14073. https://doi.org/10.4414/smw.2014.14073

[9] Eta S. Berner. 2007. *Clinical decision support systems* (Vol. 233): Springer. https://doi.org/10.1007/978-0-387-38319-4

[10] Eta S. Berner and Tonya J. La Lande. 2016. Overview of Clinical Decision Support Systems. In *Clinical Decision Support Systems: Theory and Practice.* Cham: Springer International Publishing.

[11] Joshua Biro, David M. Neyens, Candace Jaruzel, Catherine D. Tobin, Myrtede Alfred, Sarah Coppola, . . . Ken R. Catchpole. 2022. "One size" doesn't "fit all": understanding variability in anesthesia work practices. *Human Factors in Healthcare, 2*, 100026. https://doi.org/10.1016/j.hfh.2022.100026

[12] Barbara K. Burian, Anna Clebone, Key Dismukes, and Keith J. Ruskin. 2018. More than a tick box: medical checklist development, design, and use. *Anesthesia and Analgesia, 126*(1), 223-232. https://doi.org/10.1213/ANE.0000000000002286

[13] Catherine M. Burns and John Hajdukiewicz. 2004. *Ecological interface design*: CRC Press.

[14] Michael L. Burns, Leif Saager, Ruth B. Cassidy, Graciela Mentz, George A. Mashour, and Sachin Kheterpal. 2022. Association of Anesthesiologist Staffing Ratio With Surgical Patient Morbidity and Mortality. *JAMA surgery, 157*(9), 807-815. https://doi.org/10.1001/jamasurg.2022.2804

[15] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, . . . Michael Terry. 2019. *Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making.* In the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, Scotland Uk. https://doi.org/10.1145/3290605.3300234

[16] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact., 3*(CSCW), Article 104. https://doi.org/10.1145/3359206

[17] Adam Cheng, David Kessler, Ralph Mackinnon, Todd P. Chang, Vinay M. Nadkarni, Elizabeth A. Hunt, . . . Investigators Education Reporting Guidelines. 2016. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Advances in Simulation, 1*(1), 25. https://doi.org/10.1186/s41077-016-0025-y

[18] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In L. B. Resnick, R. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA Press.

[19] Richard I. Cook, Marta Render, and Daniel D. Woods. 2000. Gaps in the continuity of care and progress on patient safety. *British Medical Journal, 320*(7237), 791-794. https://doi.org/10.1136/bmj.320.7237.791

[20] Andy Crabtree, Peter Tolmie, and Mark Rouncefield. 2013. *"How Many Bloody Examples Do You Want?" Fieldwork and Generalisation.* In the Proceedings of the 13th European Conference on Computer Supported Cooperative Work (ECSCW), Paphos, Cyprus. https://doi.org/10.1007/978-1-4471-5346-7_1

[21] Pat Croskerry. 2009. Context is everything or how could I have been that stupid. *Healthcare Quarterly, 12*(Spec No Patient), e171-176. https://doi.org/10.12927/hcq.2009.20945

[22] Francis T. Durso and Frank A. Drews. 2010. Health care, aviation, and ecosystems: A socio-natural systems perspective. *Current Directions in Psychological Science, 19*(2), 71-75. https://doi.org/10.1177/0963721410364728

[23] Glyn Elwyn, Isabelle Scholl, Caroline Tietbohl, Mala Mann, Adrian G. K. Edwards, Catharine Clay, . . . Dominick L. Frosch. 2013. "Many miles to go . . .": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC Medical Informatics and Decision Making, 13*(2), S14. https://doi.org/10.1186/1472-6947-13-S2-S14

[24] Seymour Epstein. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709.

[25] Daphne C. Erkelens, Tessa C. Van Charldorp, Vera V. Vinck, Loes T. Wouters, Roger A. Damoiseaux, Frans H. Rutten, . . . Esther De Groot. 2021. Interactional implications of either/or-questions during telephone triage of callers with chest discomfort in out-of-hours primary care: A conversation analysis. *Patient Education and Counseling, 104*(2), 308-314. https://doi.org/https://doi.org/10.1016/j.pec.2020.07.011

[26] Larry C. Field, Matthew D. Mcevoy, Jeremy C. Smalley, Carlee A. Clark, Michael B. Mcevoy, Horst Rieke, . . . Cory M. Furse. 2014. Use of an electronic decision support tool improves management of simulated in-hospital cardiac arrest. *Resuscitation, 85*(1), 138-142. https://doi.org/10.1016/j.resuscitation.2013.09.013

[27] Mark Fitzgerald, Peter Cameron, Colin Mackenzie, Nathan Farrow, Pamela Scicluna, Robert Gocentas, . . . Linas Dziukas. 2011. Trauma resuscitation errors and computer-assisted decision support. *Archives of Surgery, 146*(2), 218-225. https://doi.org/10.1001/archsurg.2010.333

[28] Rhona Flin, Rona Patey, Ronnie Glavin, and Nicola Maran. 2010. Anaesthetists' non-technical skills. *BJA: British Journal of Anaesthesia, 105*(1), 38-44. https://doi.org/10.1093/bja/aeq134

[29] Amy Franklin, Swaroop Gantela, Salsawit Shifarraw, Todd R. Johnson, David J. Robinson, Brent R. King, . . . Nnaemeka G. Okafor. 2017. Dashboard visualizations: Supporting real-time throughput decision-making. *Journal of Biomedical Informatics, 71*, 211-221. https://doi.org/https://doi.org/10.1016/j.jbi.2017.05.024

[30] David M. Gaba. 2010. Crisis resource management and teamwork training in anaesthesia. *British Journal of Anaesthesia, 105*(1), 3-6. https://doi.org/10.1093/bja/aeq124

[31] Pratik Ghosh, Karen L. Posner, Stephanie L. Hyland, Wil Van Cleve, Melissa Bristow, Dustin R. Long, . . . Kenton O'hara. 2023. Framing Machine Learning Opportunities for Hypotension Prediction in Perioperative Care: A Socio-technical Perspective. *ACM Trans. Comput.-Hum. Interact., 30*(5), Article 79. https://doi.org/10.1145/3589953

[32] Erving Goffman. 1974. *Frame analysis: An essay on the organization of experience*: Harvard University Press.

[33] Erving Goffman. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.
[34] Sara N. Goldhaber-Fiebert, Naola Austin, E. Sultan, Barbara K. Burian, A. Burden, Steven K. Howard, . . . Kyle K. Harrison. 2021. Emergency Manual: Cognitive aids for perioperative crises. *Stanford Anesthesia Cognitive Aid Program*.Retrieved from http://emergencymanual.stanford.edu
[35] Mark L. Graber, Nancy Franklin, and Ruthanna Gordon. 2005. Diagnostic Error in Internal Medicine. *Archives of Internal Medicine, 165*(13), 1493-1499. https://doi.org/10.1001/archinte.165.13.1493
[36] Tobias Grundgeiger, Felix Hahn, Thomas Wurmb, Patrick Meybohm, and Oliver Happel. 2021. The use of a cognitive aid app supports guideline-conforming cardiopulmonary resuscitations: A randomized study in a high-fidelity simulation. *Resuscitation Plus, 7*, 100152. https://doi.org/10.1016/j.resplu.2021.100152
[37] Tobias Grundgeiger, Stephan Huber, Daniel Reinhardt, Andreas Steinisch, Oliver Happel, and Thomas Wurmb. 2019. *Cognitive aids in acute care: investigating how cognitive aids affect and support in-hospital emergency teams.* In the Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI '19), Glasgow, UK. https://doi.org/10.1145/3290605.3300884
[38] Tobias Grundgeiger, Jörn Hurtienne, and Oliver Happel. 2021. Why and how to approach user experience in safety-critical domains: The example of healthcare. *Human Factors, 63*(5), 821–832. https://doi.org/10.1177/0018720819887575
[39] Tobias Grundgeiger, Annabell Michalek, Felix Hahn, Thomas Wurmb, Patrick Meybohm, and Oliver Happel. 2023. Guiding attention via a cognitive aid during a simulated in-hospital cardiac arrest scenario: a SEEV model analysis. *Human Factors, 65*(8), 1689-1701. https://doi.org/10.1177/00187208211060586
[40] Tobias Grundgeiger, Alea Münz, Paul Schlosser, and Oliver Happel. 2023. Supervising Multiple Operating Rooms Using a Head-Worn display: A Longitudinal Evaluation of the Experience of Supervising Anesthesiologists and Their Co-Workers. In the Proceedings of the 2023 Conference on Human Factors in Computing Systems (CHI '23), Hamburg, Germany. https://doi.org/10.1145/3544548.3581180
[41] Sadrieh Hajesmaeel Gohari, Kambiz Bahaadinbeigy, Shahrad Tajoddini, and Sharareh R. Niakan Kalhori. 2021. Effect of computerized physician order entry and clinical decision support system on adverse drug events prevention in the emergency department: a systematic review. *Journal of Pharmacy Technology, 37*(1), 53-61. https://doi.org/10.1177/87551225209581
[42] Kenneth R. Hammond. 1996. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*: Oxford University Press.
[43] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. https://doi.org/10.1177/154193120605000909
[44] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9
[45] Robert Harutyunyan, Sean D. Jeffries, Jose L. Ramírez-Garcíaluna, and Thomas M. Hemmerling. 2023. Clinical performance of decision support systems in anesthesia, intensive care, and emergency medicine: a systematic review and meta-analysis. *Anesthesia and Analgesia, 136*(6), 1084-1095. https://doi.org/10.1213/ANE.0000000000006500
[46] Christian Heath and Jon Hindmarsh. 2002. Analysing interaction: Video, ethnography and situated conduct. In T. May (Ed.), *Qualitative research in action* (pp. 99-121). London: Sage.
[47] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in qualitative research*: Sage Publications.
[48] Christian Heath and Paul Luff. 1992. Collaboration and control Crisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW), 1*, 69-94. https://doi.org/doi.org/10.1007/BF00752451
[49] Katharine E. Henry, Rachel Kornfield, Anirudh Sridharan, Robert C. Linton, Catherine Groh, Tony Wang, . . . Suchi Saria. 2022. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *npj Digital Medicine, 5*(1), 97. https://doi.org/10.1038/s41746-022-00597-7
[50] David L. Hepner, Alexander F. Arriaga, Jeffrey B. Cooper, Sara N. Goldhaber-Fiebert, David M. Gaba, William R. Berry, . . . Angela M. Bader. 2017. Operating Room Crisis Checklists and Emergency Manuals. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 127*(2), 384-392. https://doi.org/10.1097/aln.0000000000001731
[51] John Heritage. 2013. *Garfinkel and ethnomethodology*: John Wiley & Sons.
[52] John Heritage. 2017. *Conversation analysis and institutional talk: Analyzing distinctive turn-taking*. In the Proceedings of the International Association for Dialog Analysis, Tübingen, Germany.
[53] Morten Hertzum and Jesper Simonsen. 2016. Effects of electronic emergency-department whiteboards on clinicians' time distribution and mental workload. *Health Informatics Journal, 22*(1), 3-20. https://doi.org/10.1177/1460458214529678
[54] Jon Hindmarsh and Alison Pilnick. 2002. The Tacit Order of Teamwork: Collaboration and Embodied Conduct in Anesthesia. *The Sociological Quarterly, 43*(2), 139-164. https://doi.org/10.1111/j.1533-8525.2002.tb00044.x
[55] Jon Hindmarsh and Alison Pilnick. 2007. Knowing Bodies at Work: Embodiment and Ephemeral Teamwork in Anaesthesia. *Organization Studies, 28*(9), 1395-1416. https://doi.org/10.1177/0170840607068258
[56] Anna Hohm, Oliver Happel, Jörn Hurtienne, and Tobias Grundgeiger. 2023. *"When the beeping stops, you completely freak out" - How acute care teams experience and use technology.* In the Proceedings of the ACM on Human-Computer Interaction 6.CSCW2 (2022): 1-32, Minneapolis, MN USA. https://doi.org/10.1145/3579590
[57] Steven K. Howard, David M. Gaba, Kevin J. Fish, George Yang, and Frank H. Sarnquist. 1992. Anesthesia crisis resource management training: Teaching anesthesiologists to handle critical incidents. *Aviation Space and Environmental Medicine, 63*(9), 763-770.
[58] Stephan Huber, Nathalie Papenfuss, Valentina Wohlfart, Lea Weppert, Johannes Basch, Oliver Happel, and Tobias Grundgeiger. 2024. *Hiring an AI: Incorporating Personnel Selection Methods in User-Centered Design to Design AI Agents for Safety Critical Domains.* In the Proceedings of the Nordic Conference on Human-Computer Interaction, Uppsala, Sweden. https://doi.org/10.1145/3677045.3685418
[59] Monique W. M. Jaspers, Marian Smeulers, Hester Vermeulen, and Linda W Peute. 2011. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association, 18*(3), 327-334. https://doi.org/10.1136/amiajnl-2011-000094
[60] Pengli Jia, Longhao Zhang, Jingjing Chen, Pujing Zhao, and Mingming Zhang. 2016. The effects of clinical decision support systems on medication safety: an overview. *PLoS ONE, 11*(12), e0167683. https://doi.org/10.1371/journal.pone.0167683
[61] Zhuochen Jin, Shuyuan Cui, Shunan Guo, David Gotz, Jimeng Sun, and Nan Cao. 2020. CarePre: An Intelligent Clinical Decision Assistance System. *ACM Trans. Comput. Healthcare, 1*(1), Article 6. https://doi.org/10.1145/3344258
[62] Annika Kaltenhauser, Verena Rheinstädter, Andreas Butz, and Dieter P. Wallach. 2020. *"You Have to Piece the Puzzle Together": Implications for Designing Decision Support in Intensive Care.* Paper presented at the Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS), Eindhoven, Netherlands. https://doi.org/10.1145/3357236.3395436
[63] Teus H. Kappen, Karel G.M. Moons, Leo Van Wolfswinkel, Cornelis J. Kalkman, Yvonne Vergouwe, and Wilton A. Van Klei. 2014. Impact of risk assessments on prophylactic antiemetic prescription and the incidence of postoperative nausea and vomiting: a cluster-randomized trial. *Anesthesiology, 120*(2), 343-354. https://doi.org/10.1097/ALN.0000000000000009
[64] Kensaku Kawamoto, Caitlin A. Houlihan, E Andrew Balas, and David F. Lobach. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj, 330*(7494), 765. https://doi.org/10.1136/bmj.38398.500764.8F
[65] Simon Kimmel, Vanessa Cobus, and Wilko Heuten. 2021. *opticARe - Augmented Reality Mobile Patient Monitoring in Intensive Care Units.* In the Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, Osaka, Japan. https://doi.org/10.1145/3489849.3489852
[66] Gary A Klein. 2008. Naturalistic decision making. *Human Factors, 50*(3), 456-460. https://doi.org/10.1518/001872008X288385
[67] Sara Klüber, Franzisca Maas, David Schraudt, Gina Hermann, Oliver Happel, and Tobias Grundgeiger. 2020. *Experience matters: Design and evaluation of an anesthesia support tool guided by user experience theory.* In the Proceedings of the ACM Conference on Designing Interactive Systems (DIS), Eindhoven, Netherlands. https://doi.org/10.1145/3357236.3395552
[68] Andrew Kouri, Janet Yamada, Jeffrey Lam Shin Cheung, Stijn Van De Velde, and Samir Gupta. 2022. Do providers use computerized clinical decision support systems? A systematic review and meta-regression of clinical decision support uptake. *Implementation Science, 17*(1), 21. https://doi.org/10.1186/s13012-022-01199-3
[69] John D Lee, Christopher D Wickens, Yili Liu, and Linda Ng Boyle. 2017. *Designing for people: An introduction to human factors engineering*: CreateSpace.
[70] Min Hun Lee and Chong Jun Chew. 2023. Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. *Proc. ACM Hum.-Comput. Interact., 7*(CSCW2), Article 369. https://doi.org/10.1145/3610218
[71] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez Bermúdez I Badia. 2021. *A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment.* In the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), Yokohama, Japan. https://doi.org/10.1145/3411764.3445472
[72] Anthony J. Liddicoat. 2021. *An Introduction to Conversation Analysis.* London, UK: Bloomsbury Publishing Plc.
[73] Dave Liu, Simon A. Jenkins, Penelope M. Sanderson, Perry Fabian, and John W. Russell. 2010. Monitoring with head-mounted displays in general anesthesia: A clinical evaluation in the operating room. *Anesthesia and Analgesia, 110*(4), 1032-1038. https://doi.org/10.1213/ANE.0b013e3181d3e647

[74] Stuart D. Marshall. 2013. The use of cognitive aids during emergencies in anesthesia: a review of the literature. *Anesthesia and Analgesia, 117*(5), 1162-1171. https://doi.org/10.1213/ANE.0b013e31829c397b

[75] Angela Mastrianni, Aleksandra Sarcevic, Allison Hu, Lynn Almengor, Peyton Tempel, Sarah Gao, and Randall S. Burd. 2023. Transitioning Cognitive Aids into Decision Support Platforms: Requirements and Design Guidelines. *ACM Trans. Comput.-Hum. Interact., 30*(3), Article 41. https://doi.org/10.1145/3582431

[76] Magdalena Mateescu, Christoph Pimmer, Carmen Zahn, Daniel Klinkhammer, and Harald Reiterer. 2021. Collaboration on large interactive displays: a systematic review. *Human–Computer Interaction, 36*(3), 243-277. https://doi.org/10.1080/07370024.2019.1697697

[77] Helena M. Mentis. 2017. Collocated Use of Imaging Systems in Coordinated Surgical Practice. *Proc. ACM Hum.-Comput. Interact., 1*(CSCW), Article 78. https://doi.org/10.1145/3134713

[78] Helena M. Mentis, Yuanyuan Feng, Azin Semsar, and Todd A. Ponsky. 2020. *Remotely Shaping the View in Surgical Telementoring*. In the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376622

[79] Helena M. Mentis, Ahmed Rahim, and Pierre Theodore. 2016. *Crafting the image in surgical telemedicine.* In the Proceedings of the Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. https://doi.org/10.1145/2818048.2819978

[80] Chris Mowatt. 2016. Conduct of anaesthesia. In T. Lin, T. Smith, & C. Pinnock (Eds.), *Fundamentals of Anaesthesia* (4 ed., pp. 29-45). Cambridge: Cambridge University Press.

[81] Jamie Murdoch, Rebecca Barnes, Jillian Pooler, Val Lattimer, Emily Fletcher, and John L. Campbell. 2014. Question design in nurse-led and GP-led telephone triage for same-day appointment requests: a comparative investigation. *BMJ open, 4*(3), e004515. https://doi.org/10.1136/bmjopen-2013-004515

[82] Jamie Murdoch, Rebecca Barnes, Jillian Pooler, Valerie Lattimer, Emily Fletcher, and John L. Campbell. 2015. The impact of using computer decision-support software in primary care nurse-led telephone triage: Interactional dilemmas and conversational consequences. *Social Science and Medicine, 126*, 36-47. https://doi.org/https://doi.org/10.1016/j.socscimed.2014.12.013

[83] Bala G. Nair, Gene N. Peterson, Moni B. Neradilek, Shu Fang Newman, Elaine Y. Huang, and Howard A. Schwid. 2013. Reducing wastage of inhalation anesthetics using real-time decision support to notify of excessive fresh gas flow. *Anesthesiology, 118*(4), 874-884. https://doi.org/10.1097/ALN.0b013e3182829de0

[84] David E Newman-Toker, Susan M. Peterson, Shervin Badihian, Ahmed Hassoon, Najlla Nassery, Donna Parizadeh, . . . Saraniya Tharmarajah. 2022. Diagnostic errors in the emergency department: a systematic review. https://doi.org/10.23970/AHRQEPCCER258

[85] Geoffrey R. Norman and Kevin W. Eva. 2010. Diagnostic error and clinical reasoning. *Medical Education, 44*(1), 94-100. https://doi.org/https://doi.org/10.1111/j.1365-2923.2009.03507.x

[86] Samantha E. Parsons, Elizabeth A. Carter, Lauren J. Waterhouse, Jennifer Fritzeen, Deirdre C. Kelleher, Karen J. O'connell, . . . Randall S. Burd. 2014. Improving ATLS performance in simulated pediatric trauma resuscitation using a checklist. *Annals of Surgery, 259*(4), 807-813. https://doi.org/10.1097/sla.0000000000000259

[87] Thierry Pelaccia, Jacques Tardif, Emmanuel Triby, Christine Ammirati, Catherine Bertrand, Valérie Dory, and Bernard Charlin. 2014. How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Annals of Emergency Medicine, 64*(6), 575-585. https://doi.org/10.1016/j.annemergmed.2014.05.003

[88] Thierry Pelaccia, Jacques Tardif, Emmanuel Triby, and Bernard Charlin. 2011. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Medical Education Online, 16*(1), 5890. https://doi.org/10.3402/meo.v16i0.5890

[89] Morag Prowse and Davina Allen. 2002. 'Routine' and 'emergency' in the PACU: the shifting contexts of nurse—doctor interaction. In D. Allen & D. Hughes (Eds.), *Nursing and the Division of Labour in Healthcare* (pp. 75-97). London: Macmillan Education UK.

[90] Niroop Channa Rajashekar, Yeo Eun Shin, Yuan Pu, Sunny Chung, Kisung You, Mauro Giuffre, . . . Dennis Shung. 2024. *Human-Algorithmic Interaction Using a Large Language Model-Augmented Artificial Intelligence Clinical Decision Support System.* In the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642024

[91] Carrie Reale, Megan E. Salwei, Laura G. Militello, Matthew B. Weinger, Amanda Burden, Christen Sushereba, . . . William R. Mcivor. 2023. Decision-making during high-risk events: a systematic literature review. *Journal of Cognitive Engineering and Decision Making, 17*(2), 188-212. https://doi.org/10.1177/15553434221147415

[92] Nicholas Riches, Maria Panagioti, Rahul Alam, Sudeh Cheraghi-Sohi, Stephen Campbell, Aneez Esmail, and Peter Bower. 2016. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PLoS ONE, 11*(3), e0148991. https://doi.org/10.1371/journal.pone.0148991

[93] Nadia Roumeliotis, Jonathan Sniderman, Thomasin Adams-Webber, Newton Addo, Vijay Anand, Paula Rochon, . . . Christopher Parshuram. 2019. Effect of electronic prescribing strategies on medication error and harm in hospital: a systematic review and meta-analysis. *Journal of General Internal Medicine, 34*, 2210-2223. https://doi.org/10.1007/s11606-019-05236-8

[94] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM, 67*(10), 18–21. https://doi.org/10.1145/3649404

[95] Sheree May Saßmannshausen, Nazmun Nisat Ontika, Aparecido Fabiano Pinatti De Carvalho, Mark Rouncefield, and Volkmar Pipek. 2024. *Amplifying Human Capabilities in Prostate Cancer Diagnosis: An Empirical Study of Current Practices and AI Potentials in Radiology.* In the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642362

[96] Daniel Schneiderbanger, Stephan Johannsen, Norbert Roewer, and Frank Schuster. 2014. Management of malignant hyperthermia: diagnosis and treatment. *Therapeutics and Clinical Risk Management*, 355-362. https://doi.org/10.2147/TCRM.S47632

[97] Christian M. Schulz, Amanda Burden, Karen L. Posner, Shawn L. Mincer, Randolph Steadman, Klaus J. Wagner, and Karen B. Domino. 2017. Frequency and Type of Situational Awareness Errors Contributing to Death and Brain Damage: A Closed Claims Analysis. *Anesthesiology, 127*(2), 326-337. https://doi.org/10.1097/ALN.0000000000001661

[98] Noa Segall, David B. Kaber, Jeffrey M. Taekman, and Melanie C. Wright. 2013. A cognitive modeling approach to decision support tool design for anesthesia provider crisis management. *International Journal of Human-Computer Interaction, 29*(2), 55-66. https://doi.org/10.1080/10447318.2012.681220

[99] Venkatesh Sivaraman, Leigh A Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. 2023. *Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care.* In the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '24), Hamburg, Germany. https://doi.org/10.1145/3544548.3581075

[100] Harold C. Sox, Michael C. Higgins, Douglas K. Owens, and Gillian Sanders Schmidler. 2024. *Medical decision making*: John Wiley & Sons.

[101] Marjorie Podraza Stiegler, Jacques P. Neelankavil, Cecilia Canales, and Anahat K. Dhillon. 2012. Cognitive errors detected in anaesthesiology: a literature review and pilot study. *British Journal of Anaesthesia, 108*(2), 229-235. https://doi.org/10.1093/bja/aer387

[102] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. 2020. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine, 3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

[103] Angelique Taylor, Hee Rin Lee, Alyssa Kubota, and Laurel D. Riek. 2019. Coordinating clinical teams: Using robots to empower nurses to stop the line. *Proceedings ACM Human-Computer Interaction, CSCW, 3*(CSCW), 1-30. https://doi.org/10.1145/3359323

[104] Peter-Paul Verbeek. 2015. Beyond interaction: a short introduction to mediation theory. *interactions, 22*(3), 26–31. https://doi.org/10.1145/2751314

[105] Kim J. Vicente and Jens Rasmussen. 1992. Ecological interface design: Theoretical foundations. *IEEE Transactions on systems, man, and cybernetics, 22*(4), 589-606. https://doi.org/10.1109/21.156574

[106] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, . . . Feng Tian. 2021. *"Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment.* In the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), Yokohama, Japan. https://doi.org/10.1145/3411764.3445432

[107] Robert L. Wears and Marc Berg. 2005. Computer Technology and Clinical Work-Still Waiting for Godot. *JAMA, 293*(10), 1261-1263. https://doi.org/10.1001/jama.293.10.1261

[108] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions, 23*(3), 38-44.

[109] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. *Harnessing Biomedical Literature to Calibrate Clinicians' Trust in AI Decision Support Systems.* In the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Hamburg, Germany. https://doi.org/10.1145/3544548.3581393

[110] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. *Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes.* In the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, Scotland Uk. https://doi.org/10.1145/3290605.3300468

[111] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F. Antaki. 2016. *Investigating the Heart Pump Implant Decision Process: Opportunities for Decision Support Tools to Help.* In the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), San Jose, California, USA. https://doi.org/10.1145/2858036.2858373

[112] Junsang Yoo, Kwang Yul Jung, Taerim Kim, Taerim Lee, Sung Yeon Hwang, Hee Yoon, . . . Won Chul Cha. 2018. A Real-Time Autonomous Dashboard for the Emergency Department: 5-Year Case Study. *JMIR Mhealth Uhealth, 6*(11), e10666. https://doi.org/10.2196/10666

[113] Minfan Zhang, Daniel Ehrmann, Mjaye Mazwi, Danny Eytan, Marzyeh Ghassemi, and Fanny Chevalier. 2022. *Get To The Point! Problem-Based Curated Data Views To Augment Care For Critically Ill Patients*. In the Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), New Orleans, LA, USA. https://doi.org/10.1145/3491102.3501887

[114] Shao Zhang, Jianing Yu, Xuhai Xu, Changchang Yin, Yuxuan Lu, Bingsheng Yao, . . . Dakuo Wang. 2024. *Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis*. In the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642343