Cite as:

Assessing distractors and teamwork during surgery: Developing an event-based method for direct observation

Julia C. Seelandt (1) *
Franziska Tschan (1)*
Sandra Keller (1)
Guido Beldi (3)
Nadja Jenni (1)
Anita Kurmann (3)
Daniel Candinas (3)
Norbert K. Semmer (2)

(1) University of Neuchâtel, Institute for Work and Organizational Psychology, Neuchâtel, Switzerland
(2) University of Bern, Institute for Psychology, Bern, Switzerland
(3) University Hospital of Bern, Department of Visceral Surgery and Medicine, Bern, Switzerland

*These authors contributed equally to the paper.

Corresponding author: Franziska.tschan@unine.ch; University of Neuchâtel, Rue Emile Argand 11, 2000 Neuchâtel, Switzerland;
PHONE +41 32 718 13 96, FAX +41 32 718 13 92

Keywords: human factors in surgery; behavior observation; distractors; communication; teamwork
ABSTRACT

Objective

To develop a behavioural observation method to simultaneously assess distractors and communication/teamwork during surgical procedures through direct, on-site observations; to establish the reliability of the method for long (> 3 h) procedures.

Methods

Observational categories for an event-based coding system were developed based on expert interviews, observations and a literature review. Using Cohen’s kappa and the Intraclass Correlatoin Coefficient (ICC), inter-observer agreement was assessed for 29 procedures. Agreement was calculated for the entire surgery, and for the first hour. In addition, inter-observer agreement was assessed between two tired observers and between a tired and a non-tired observer after three hours of surgery.

Results

The observational system has five codes for distractors (door openings, noise distractors, technical distractors, side conversations and interruptions), eight codes for communication/teamwork (case-relevant communication, teaching, leadership, problem solving, case-irrelevant communication, laughter, tension, and communication with external visitors) and five contextual codes (incision, last stich, personnel changes in the sterile team, location changes around the table and incidents). Based on five-minute intervals, Cohen’s kappa was good to excellent for distractors (0.74 to 0.98) and for communication/teamwork (0.70 to 1). Based on frequency counts, ICC was excellent for distractors (0.86 to 0.99) and good to excellent for communication/teamwork (0.45 to 0.99). After three hours of surgery, Cohen’s kappa was 0.78 to 0.93 for distractors, and 0.79 to 1 for communication/teamwork.
Discussion

The observational method developed allows a single observer to simultaneously assess distractors and communication/teamwork. Even for long procedures, high inter-observer agreement can be achieved. Data collected with this method allow for investigating separate or combined effects of distractions and communication/teamwork on surgical performance and patient outcomes.
INTRODUCTION

It is increasingly accepted that human factors play an important role in surgical performance, and more research is needed to assess their influence [1-4]. The most often discussed human factors in surgery are distractions in the operating room and intra-surgical teamwork [5-9]. Although distractions and teamwork are both recognized as important influences, they have rarely been assessed simultaneously. Observational methods exist to study either teamwork or distractions, but to our knowledge, there is currently no established method that allows a single observer to simultaneously assess both aspects. Furthermore, most observational studies have assessed relatively short surgeries. Because long surgeries bear higher risks for patient complications [10, 11] it is important to include long procedures in human factor research, and this inclusion may require the development of new methods. To address these gaps, we developed Simultaneous Observation of Distractors and Communication in the Operating Room (SO-DIC-OR), an event-based behavioural observation method that can be used in the operating room (OR). This method simultaneously captures distractors and teamwork and can be used to observe short and long procedures.

We first provide a short introduction into the characteristics of different approaches to behavioural observation in OR settings, as well as their advantages and disadvantages. We then present the development of the observational method and provide information about inter-observer reliability, including reliability after three hours of continuous observation.

Distractors and teamwork in the OR

There are many potential sources of distractions in the OR (e.g., noise from machines and manipulations, alarms, incoming phone calls, or conversations outside the sterile team). Distractions are very common: even for short procedures, a distraction occurs every one to three minutes [12-15]. Distractions have been found to negatively affect surgical performance [16-18], as they
threaten the concentration of the surgical team members, particularly the concentration of less experienced surgeons [18-20].

Teamwork and communication in the OR are another important influence on surgical quality [21-25]. Surgeons, nurses, and anaesthetists have to cooperate closely and effectively; this requires complex collaboration [26]. Good teamwork and optimal communication in the OR increase the quality of surgeries, whereas poor or ineffective communication jeopardizes patient outcomes [26-29].

**Methods for observing behaviour in the OR**

From a research perspective, the gold standard for investigating the relationship among distractions, teamwork, and surgical outcomes is behavioural observation. Behavioural observation does not rely on self-reports or on retrospective analyses. Retrospective analyses are based on memory processes, which may contain errors [30] and can be biased [1], particularly if the outcome is known [31]. Behaviours during surgical procedures can be observed based on videos [32-34] or by direct observation in the OR. Although videotaping has many advantages [35-37], legal and ethical issues and technological constraints often limit filming in the OR. Therefore, much research in this field still relies on observers present in the OR.

Direct observation presents several challenges. First, observers have to record behaviour and events as they happen [35], which requires a high degree of concentration and constant attention. Attention is limited, therefore, a single observer can only assess a limited number of different behaviours. Second, observers have to make fast and immutable decisions during the on-going process. There is a risk that observers miss or misinterpret behaviours. Third, if the observation time is long, fatigue can influence the quality of the observation. It is thus crucial to develop methods that are well suited for direct observations and to show that high inter-observer agreement can be achieved [38].

Before developing a new observational system, it is useful to assess existing methods, as it is an advantage to use an established methodology. Our literature search revealed that observational
methods exist to assess teamwork [6, 39-42], communication [27, 43-47], and distractors [12, 15, 48] in the OR (for an overview, see web appendix). We found only two papers that combine observations of teamwork and distractions. One of these studies used two very different methodological approaches [14], and the other one limited observed distractors to a few categories [49]; none of the studies reported inter-observer agreements.

One of the reasons that observations of teamwork and distractors have not yet been combined may lie in the different methodological approaches that have traditionally been used in this field. Research groups observing communication in the OR have most often used field notes [27, 36, 43, 50]; research groups assessing the quality of teamwork and non-technical skills have often relied on behavioural markers [7, 9, 51-53]; and research on distractions in the OR has most often used event-coding methodology [12, 13]. The following section compares these three approaches.

Most common in ethnographic research [54], field notes have been used in studies assessing communication [27, 43] and leadership [47] in the OR. Observers take extensive notes in a free text form [55-57]. In addition to a general thematic focus, observers normally have few restrictions in terms of what they observe and take notes on. The advantage of field notes is that they can be flexibly used in almost every situation. When using this methodology, researchers should well understand the situation they observe; otherwise, they may overlook or misinterpret important events. The use of field notes is most appropriate if it is difficult to define behavioural categories in advance, which generally occurs when little is known about a situation [47] and when the situations observed are very diverse. Because they allow a wide angle on a situation, field notes are well adapted for observing non-routine situations and are particularly useful for explorative studies[47]. Field notes are often the basis for qualitative analyses. It is, however, possible to code and categorize field note contents after the observation, which allows the derivation of quantitative data [27], although to a limited degree. The disadvantage of field notes is that they cannot easily be used
for quantitative research, and it is difficult to assess inter-observer reliability for the initial taking of field notes.

Most systems that assess the quality of teamwork in the OR use *behavioural marker* methodology [6, 39, 40]. When using *behavioural marker* methodology, the observers are instructed to assess “behaviour classes”. This classes are defined in advance, based on a thorough analysis of the non-technical skills required for a specific situation or professional role[58]. Within behavioural classes, exemplar behaviours that represent good or poor behaviours are defined [7, 59]. An example is the “exchanging information” behavioural class within the Non-Technical Skills for Surgeons (NOTSS) observational system [60]. Optimal information exchange is described as “talk about the progress of the operation”, whereas poor information exchange is described as “fail to communicate concerns with others” [61, p.17]. In behavioural marker-based observations, observers do not report or note single behaviours; they instead provide an integrative quality score for each behaviour class for the whole procedure [62] or for a predefined observational period.

Behavioural marker-based systems have to be specific to the role or the situation. Methods have been developed for non-technical skills of surgeons [60], anaesthetists [63], scrub nurses [52] and the entire surgical team [64].

The advantage of behavioural marker systems is that they focus on desired and undesired behaviour in a specific situation, and that observers provide a summary score. It is thus possible to assess the quality of teamwork and to provide immediate feedback after the observation. The disadvantage is that such integrative judgments are vulnerable to hindsight effects and observer biases [62].

Observers need to (a) continuously assess the quality of behaviours, (b) relate these behaviours to the predefined classes, and (c) mentally integrate their observations into an overall qualitative judgment for each behaviour class. This complex and cognitively demanding process requires extensive training and domain-specific knowledge [65]. It is thus rather challenging to achieve high inter-observer agreement [62, 64, 66, 67]. If teamwork quality is only assessed once using behavioural markers during the entire intraoperative phase, this approach may have limited
usefulness in long surgeries; as different phases of the surgery have different coordination requirements, and as the quality of teamwork may not be consistent for the whole procedure [68, 69].

**Event coding** is the continuous real-time observation and registration of specific, predefined events or behaviours. This methodology has been used to observe communication in the OR [39, 46], but it is most common for assessing distractors in the OR [12, 15, 49]. Some examples of observed events are “door to the OR opens” or “an alarm sounds”. Observers note events as they happen. Event coding can be as simple as keeping a tally; more complex methods use time codes (i.e., noting the event as well as the time it occurs). To develop an event-coding system, researchers define specific behaviours or events to be observed based on conceptual considerations. Each event is defined and described in a coding manual [70]. For example, Healey and colleagues [12] coded “case irrelevant communication” as a distractor each time the team engaged in communication that was not related to the patient or the procedure; they coded “Monitor-F” each time someone moved in front of the video display monitor in the OR ([12], page 596). Event coding requires extensive observer training [70, 71].

The advantage of event coding is that observed events and behaviours are specific. The clear definition of events requires little integrative judgment from observers; thus, observers can simultaneously chart several categories [35]. If events are time-stamped, event coding allows to assess frequency, timing, and sequences of events; it is thus well suited for detecting behavioural patterns [70]. In addition, event coding allows for analyses and comparisons of different phases of a surgery [72]. The disadvantage of event coding is that only predefined behaviours are captured; thus, some methods combine event coding with the possibility of providing open comments [49]. Event coding is of limited use for immediate feedback, as it does not provide an integrative quality score.
Table 1: Comparison of field notes, behavioural markers and event-based observations for observation in the OR.

<table>
<thead>
<tr>
<th>Is the system suitable for direct observations in the OR?</th>
<th>Field notes</th>
<th>Behavioural markers</th>
<th>Event coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the OR setting traditionally used for...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication, method development</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Teamwork, non-technical skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distractors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is an elaborate theoretical basis needed to develop the observational system?</td>
<td>Yes</td>
<td>Possible</td>
<td></td>
</tr>
<tr>
<td>Not necessary, but useful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Based on extended previous analysis of optimal behaviour in a specific situation and role</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Necessary; can be based on general assumptions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are subject matter experts needed for observations?</td>
<td>Useful</td>
<td>Yes, to a very high degree</td>
<td>Yes, but to a limited degree</td>
</tr>
<tr>
<td>Can the system capture unusual events?</td>
<td>Yes</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Is the quality of behaviour assessed?</td>
<td>Possible</td>
<td>Yes</td>
<td>Possible</td>
</tr>
<tr>
<td>Can the system be used for immediate feedback to the observed party after the observation?</td>
<td>Not systematically</td>
<td>Yes, quality is assessed</td>
<td>Limited to frequencies of events and behaviours</td>
</tr>
<tr>
<td>Is timing of events possible?</td>
<td>Limited</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Can sequences of events be assessed?</td>
<td>Limited</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Where is the attentional focus of the observer?</td>
<td>On the general process</td>
<td>On behaviour classes</td>
<td>On single behaviours and events</td>
</tr>
<tr>
<td>What type of observer training is needed?</td>
<td>Training with regard to the general method</td>
<td>Training with regard to behavioural markers, exemplar events, quality ratings, integration</td>
<td>Training with regard to observational categories</td>
</tr>
</tbody>
</table>
There is no a priori advantage for one particular observational method; method choice depends on the specific research goal. Nevertheless, in the OR setting, researchers have traditionally chosen different methods to observe communication, teamwork or distraction. Each method requires the observer to focus his or her attention on different aspects: Field notes require attention to the whole process and to its narrative structure; behavioural markers require the observer to make quality assessments by mentally integrating specific observations into overall judgments; and event coding requires attention to the occurrence of single events. Because different observation methods require different attentional foci, it is difficult to combine two existing methodological approaches.

The web appendix provides an overview of the observational methods used for direct observation in the OR. We included papers focusing on the presentation of an observational method and papers focusing on specific content that also provide information about the observational system in the methods section. We excluded methods that focus solely on adverse events (e.g., [73]) and papers based on subsamples from earlier published research. We also excluded systems focusing on anaesthesia (i.e., ANTS -Behavioural Marker System for Anaesthetists’ Non-Technical Skills [63]) or on the preoperative or postoperative phases (i.e., [74]). The overview contains information on method type, observed behaviour or events, observers, observation targets, and procedure type and duration. In addition, if provided, information on inter-observer reliability is presented.

The review of existing systems revealed two gaps that we aim to address with this paper. First, there is no observational method that combines the assessment of communication/teamwork and distractors as potential influences on the surgical process and has been shown to be reliable. Existing behavioural marker systems and event coding systems require each a different attentional focus from observers, thus, combining two existing methods would overburden observers and most likely result in low inter-observer agreement. Second, most current knowledge with regard to communication/teamwork and distractors during surgeries is based short procedures. An
observational system that is suitable for observing procedures that last several hours allows to extending research to procedures that bear a particularly high risk of patient complications. We thus developed an event-based coding system that allows to simultaneously assessing communication and distractors in the OR and can be used for short, but also for long procedures.

Research goals

We address the following research questions:

Q1: Is it possible to reliably assess distractions and aspects of teamwork simultaneously during surgery using an event-sampling methodology?

Q2: Is the observational method suitable for the observation of long procedures (3 hours or longer) by maintaining acceptable inter-observer reliability over time?

METHODS

Sample

The sample consisted of 29 elective open abdominal procedures that were entirely or partially observed by two observers. These surgeries were a sub-sample of 103 procedures observed over a period of 12 months at a university hospital in a western European country. General inclusion criteria for observed surgeries were elective open abdominal surgery and the observers’ availability. Throughout the study period, about every fourth procedure was observed by two researchers to assess inter-observer reliability; these 29 observations are included in this study. The 29 procedures related to the digestive tract, intestines, rectum, liver, pancreas and oesophagus. There were major liver resections and minor liver resections (i.e., resections of less than three liver segments); surgeries of the duodenum/pancreas, including duodenopancreatectomies and segmental duodenectomies; procedures related to the upper gastrointestinal tract, including gastrectomies (total or partial), oesophagectomies (including transhiatal) and hiatalplasties; endocrine procedures, including adrenalectomies; procedures related to the lower gastrointestinal tract, including hemicolecetomies (right or left) and resections of enterocutaneous fistula; and spleen surgeries,
including splenectomies. These surgeries are representative of the surgeries performed in the
department where the study took place; they were carried out in two identically designed and
equipped ORs.
Ten trained observers with at least a bachelor’s degree in industrial psychology participated in the
study. The local institutional review board approved the study.

Procedure

Development of the observational system

Our main goal was to develop and test an observational system to assess distractors and aspects of
teamwork during surgery (SO-DIC-OR; Simultaneous Observation of Distractions and
Communication in the OR). Each observational method has to satisfy the validity criteria (i.e., the
method measures what it is supposed to measure; thus, the observational categories have to be
meaningful and adapted to the situation) and reliability criteria (i.e., the observations must be
consistent across observers and over time; thus, inter-observer agreement has to be established).

To satisfy the criterion of construct validity, we developed a list of events to be observed based on
expert interviews, observations of five surgical procedures, and a literature review (Figure 1). We
performed seven in-depth expert interviews with senior and junior surgeons, anaesthetists, scrub
nurses, and circulating nurses about their perceptions of potential sources of distractions during the
intraoperative phase and their assessment of helpful and problematic communication and teamwork
in the OR. Using a guided field-note method (i.e., instructing observers to concentrate on teamwork,
communication, and distractors), we observed five open abdominal procedures. The field notes
were reviewed to extract observational categories. We also conducted an extensive literature search
on observational systems already in use in the OR (web appendix). Unsurprisingly, the behaviours
that were mentioned in expert interviews, extracted from field notes and described in the literature
largely overlapped. Two observers tested a first version of the observational system during eight
surgical procedures; they were advised to write comments on the coded events. After each surgery,
the observers compared their observations event by event, and differences were discussed. Code definitions and descriptions were revised, and the final system was developed (Table 2).

We chose a timed-event sampling methodology (i.e., recording the event and the specific time at which the event occurs) for several reasons. First, clearly defining events and behaviours to observe does not require observers to make integrative judgments over time. Therefore, the system is cognitively less demanding than behavioural marker methodology, allowing the inclusion of more observational categories without overburdening observers. Second, for long procedures, an overall integrative assessment, as is customary in behavioural marker-based observations, is very difficult to make. Furthermore, and event-based system allows to assess teamwork quality separately for the different phases of a procedure, and allows for analyzing sequential patterns; therefore, it is particularly suitable for the observation of long procedures.

The observational system contains five distraction-related codes: door openings, noise distractors, technical distractors, side conversations, and interruptions; these are largely based on the system developed by Healey and colleagues [12] and were adapted for open procedures. The system contains eight teamwork-related codes that focus on communication within the sterile team and between sterile team members and anaesthetists. The observational codes include case-relevant communication (i.e., short-term planning), teaching, leadership, and problem solving. These codes are related to the patient and procedure (i.e., task-related communication). We included task-related communication because it helps a team build and maintain a shared understanding of the task and may thus facilitate coordination [3, 75, 76]. We also included case-irrelevant communication (e.g., laughter and tension) because they represent social aspects of teamwork and may influence team building and team climate in the OR [41]. For this reason, case-irrelevant communication within the sterile team is considered a teamwork code, not a distractor as in other systems [12]. In addition, note that talking among anaesthetists or among circulating nurses/visitors is coded as side conversation and categorized as a potential distractor for the sterile team, despite the fact that they
could these conversations could be case-relevant. The observational system also contains several *contextual codes* (e.g., time of incision, time of the last stitch, personnel changes within the sterile team, and personnel location changes around the operating table). Unusual incidents (e.g., an X-ray after an inconclusive sponge count) are described using an open-text option and “other” code. The open text option allows observers to describe any observation that is not covered by the predefined event codes but that they regard as important or interesting. Table 2 presents the codes and a short description of each code (a full codebook is available on request). Codes are entered into a laptop using a spreadsheet (Microsoft Excel); a macro is used to automatically time-stamp each event the moment it is entered. Observations started at incision and ended with the last stitch.

Observers were seated behind a small moveable tray close to the wall. They were about two metres away from the sterile field at the left side of the patient, thus facing the primary surgeon for most procedures. This position allowed a good view of the room, the sterile team and the anaesthetic team, including the patient monitor; all doors were in sight of the observers. The observers were sufficiently close to the sterile team to overhear communication; however, they were sufficiently far away to not to be an obstacle for the OR personnel.

**Observer training**

Observers underwent a four-step training procedure that lasted between 25 and 35 hours. The training started with an informal visit to the OR that included instructions about dress codes, hygiene procedures, and behavioural guidelines in the OR, as well as an unstructured observation of one procedure. The second step was a 4-hour off-site training session during which trainees received general information about the setting (e.g., roles and functions of OR team members, formal working procedures, and spatial arrangements in the OR), followed by a structured introduction into the observational system (e.g., explanations for each code and short video clips as behavioural examples). Trainees were then handed an information packet and asked to familiarize
themselves with the coding system. The third training step consisted of observing two procedures under the direct guidance of an expert observer. In the fourth step, trainees observed two to four surgical procedures independently, but at the same time as an expert coder. After each of the surgeries, disagreements between expert and trainee were discussed. Training was considered complete if agreement between trainees and expert coders (Cohen’s kappa) was $\geq 0.75$ for all codes, which was typically the case after three or four independent observations.

**Inter-observer reliability**

Many studies based on observational data refer to relatively short procedures (cf. web appendix). SO-DIC-OR was developed to observe long procedures with a scheduled duration of three to seven hours. Long continued observation bears a high risk of potential quality loss due to observer fatigue. We therefore tested inter-observer reliability for different time periods, and we assessed fatigue effects. Reliabilities were calculated (a) for the whole procedure, (b) for the early (i.e., the first hour) and late phases (i.e., three hours after the incision until the end of the procedure). To test for fatigue effects, we assessed inter-observer reliability for the late phase using an observer present from the beginning of the procedure ("tired") or an observer who joined three hours into the surgery ("non-tired").

**Statistical Analyses**

Cohen’s kappa and intra-class correlation coefficients (ICCs) were estimated to assess inter-observer reliability. Cohen’s kappa is well suited for nominal scales and expresses the proportion of agreement in terms of a given category being coded or not, controlling for chance agreement. It ranges from -1.00 to +1.00, with zero indicating no agreement [77]. Values between 0.41 and 0.59 are defined as fair, values between 0.60 and 0.80 are defined as substantial, and values above 0.81 indicate very good agreement [78]. We calculated Cohen’s kappa for the occurrence versus non-occurrence of each observational code for every five-minute segment of the observational period.
To assess inter-observer reliability for frequency counts, we calculated one-way random ICCs for each code between two observers for the different observational periods [79, 80]. ICC normally ranges from 0 to 1 but can also be negative. Values higher than 0.75 indicate very good inter-observer reliability [81, 82].
RESULTS

The mean duration of the 29 surgeries was 302 minutes (Median: 290, Standard Deviation: 121, Range: 119-643 minutes). All but five surgeries lasted more than 3 hours. Table 2 displays the results for inter-observer agreement for different time periods. Cohen’s kappa values indicate good to excellent inter-observer agreement for the whole procedure, for the first hour, and after three hours of coding (for surgeries lasting 4 hours or more), as well as for a tired and a non-tired observer who joined three hours into the procedure (all kappas >0.74). Similarly, most ICCs are above 0.75. Exceptions are frequency ratings of tension in the first hour of coding (ICC = 0.703) and after three hours of coding between two “tired” observers (ICC = 0.667). The frequency agreement of two tired observers was also below 0.75 for teaching activities (ICC = 0.555) and for communication with external visitors (ICC = 0.446).
Table 2: SO-DIC-OR event codes, short descriptions, and inter-observer reliabilities

<table>
<thead>
<tr>
<th>Event-code</th>
<th>Description (example)</th>
<th>Entire procedure</th>
<th>Early (first h)</th>
<th>Late &gt; 3h (tired-tired)</th>
<th>Late &gt; 3h (tired-not tired)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=29</td>
<td>M p.h. (SD)</td>
<td>κ</td>
<td>ICC</td>
<td>κ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distractors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doors</td>
<td>A door to the OR (operating room) opens and closes</td>
<td>33.8(7.1)</td>
<td>0.887</td>
<td>0.971</td>
<td>0.926</td>
</tr>
<tr>
<td>Noise distractors</td>
<td>Events (excluding communication) that are loud enough to be potential distractors and were produced by non-sterile team members (noise from putting away instruments or from an instrument falling on the floor)</td>
<td>10.9(4.8)</td>
<td>0.789</td>
<td>0.877</td>
<td>0.740</td>
</tr>
<tr>
<td>Technical distractors</td>
<td>A technical device requires attention (incoming phone call, beeper, alarms from technical devices)</td>
<td>6.7(3.4)</td>
<td>0.892</td>
<td>0.976</td>
<td>0.843</td>
</tr>
<tr>
<td>Side conversations</td>
<td>Conversations between members outside of the sterile team that can be well overheard but do not imply a member of the sterile team (an anaesthesiologist talks with an external surgeon about the next case; two circulating nurses are talking and laughing)</td>
<td>10.8(6.3)</td>
<td>0.783</td>
<td>0.983</td>
<td>0.746</td>
</tr>
<tr>
<td>Interruptions</td>
<td>The surgery is interrupted, and the surgeons are not operating (a visitor to the OR asks question; the surgical team waits for pathology results)</td>
<td>1.9(1.0)</td>
<td>0.855</td>
<td>0.946</td>
<td>0.965</td>
</tr>
<tr>
<td>Teamwork/Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-relevant communication</td>
<td>Communication involving a member of the sterile team about the patient or the procedure: planning next steps, explanations of own actions, loud thinking, or talking to the room (e.g. the primary surgeon announces which part she is dissecting next)</td>
<td>16.8(6.1)</td>
<td>0.863</td>
<td>0.958</td>
<td>0.924</td>
</tr>
<tr>
<td>Teaching</td>
<td>A member of the sterile team engages in a teaching-conversation (explaining, asking questions) with a trainee concerning aspects of the procedure, including anatomy, disease, and surgical techniques (a resident explains how to do sutures; a surgeon explains which percentage of patients has a similar anatomical structure as the current patient)</td>
<td>1.2(2.0)</td>
<td>0.885</td>
<td>0.954</td>
<td>1</td>
</tr>
<tr>
<td>Leadership</td>
<td>A surgeon’s explicit instructions about what to do or not to do, except for demands to hand an instrument already on the table (a surgeon asks scrub nurse to dial a phone number; orders the</td>
<td>2.6(1.9)</td>
<td>0.916</td>
<td>0.936</td>
<td>0.905</td>
</tr>
</tbody>
</table>
### Problem-solving

Focused discussion about a problem of the case within the sterile team or with external experts. This is only coded if the surgery is interrupted during the discussion and the discussion focuses on clearly problematic aspects of the case (the sterile team gathers around the CT on the screen and discusses the next steps).

|            | 0.4(0.14) | 1 | 0.923 | 1 | 0.832 | NO | NO | NO | NO |

### Case-irrelevant communication

Communication not related to the actual patient or procedure within the sterile team (the resident talks about his children).

|            | 2.2(2.1) | 0.847 | 0.954 | 0.905 | 0.984 | 0.828 | 0.956 | 0.830 | 0.893 |

### Laughter

Joking or laughter within the sterile team (the surgeon makes a joke about an overweight dog).

|            | 3.1(2.9) | 0.834 | 0.979 | 0.815 | 0.974 | 0.843 | 0.937 | 0.939 | 0.982 |

### Tension

Open conflict or tense conversations involving a member of the sterile team (the scrub nurse yells at the resident to not start suturing before the sponge count is completed; two surgeons angrily disagree about the next step of the procedure).

|            | 0.22(0.36) | 0.696 | 0.703 | 0.815 | 0.830 | 1 | 0.667 | 1 | - |

### Communication with visitors

A member of the sterile team talks with a person temporarily in the OR and not part of the surgical team (a surgeon from another OR asks for a consult).

|            | 0.84(0.61) | 0.899 | 0.741 | 0.964 | 0.949 | 0.844 | 0.446 | 0.925 | 0.896 |

### Contextual codes

#### Incision

Time of incision

|            | NA | NA | NA | NA | NA | NA | NA | NA | NA |

#### Personnel changes in sterile team

A member of the sterile team leaves the table or arrives at the table (surgeons or scrub nurses).

|            | 1.3(0.55) | 0.938 | 0.968 | 0.946 | 0.979 | 0.930 | 0.992 | 1 | NA |

#### Location changes around the table

Position changes around the table within the sterile team (the senior surgeon and resident surgeon change places).

|            | 0.95(0.75) | 0.904 | 0.930 | 0.978 | 0.973 | 0.890 | 0.890 | 0.951 | 0.968 |

#### Last stitch

Time of last stitch

|            | NA | NA | NA | NA | NA | NA | NA | NA | NA |

#### Incident

A special, unforeseen incident happens (the sponge count is inconclusive and an X-ray is performed).

|            | 2/29d | 0.650 | 0.695 | NO | NO | NO | NO | 1 | NO |

#### Other

Any observation or thought of coders that is not captured by a code but judged to be worth noting

|            | NA | NA | NA | NA | NA | NA | NA | NA | NA |

**NOTE:** \( \kappa \) = Cohen’s kappa, reflecting whether a given category is coded within a predefined 5-minute interval and based on the number of units included. ICC = intra-class correlation coefficient, reflecting agreement of the number of codes within a specific category.

NO = not observed; NA = reliability measures do not apply.

\(^a\) Sixteen of the 18 procedures are the also included in the estimation of inter-observer reliability for the whole procedure. \(^b\) The first number refers to the number of different surgeries included; the second number represents the number of five-minute intervals assessed. \(^d\) An incident occurred in two of the 29 procedures; we thus do not report descriptive statistics other than overall frequency.
DISCUSSION

We developed and tested SO-DIC-OR, an observational system that allows observers to simultaneously observe distractors and aspects of teamwork and communication in the OR. The importance of human factors in surgeries is uncontested, and research on communication, teamwork and distractors in the OR has become increasingly important. Thus far, their co-occurrence and potential mutual influence have not yet been evaluated. For example, it could well be that noise distractors influence task-relevant communication in the sense that distractors may lower the rate of task-relevant communication, which, in turn, may influence performance. Such research questions can now be addressed, because SO-DIC-OR provides timed observational data for distractors as well as for communication throughout the whole process.

Observers using SO-DIC-OR achieved high inter-observer agreement, a crucial indicator of the quality of the system. Of the 22 studies summarized in the web appendix, only 12 reported results about observer agreement. Compared with the values reported there, inter-observer agreement of SO-DIC-OR is similar or higher. This is a good result, given that 17 different event types had to be observed and given that behavioural observation is a difficult task, requiring constant attention and often quick decision-making.

We developed SO-DIC-OR to be suitable for the observation of long surgical procedures. Inter-observer agreement was acceptable to excellent for all time phases tested, with the exception of two ICC values (teaching and communication with visitors) between two “tired” observers (i.e. after three hours of observation). Note that both events occurred with low frequency, implying that any discrepancy had a rather strong influence on ICC. Apart from these two codes, there were no substantial
signs of fatigue effects after three hours of continuous observation, making the system well suited for direct observation of short as well as long procedures.

The high inter-observer agreement of SO-DIC-OR may be due to several reasons. First, we chose well-defined categories and described them as unambiguously as possible. We defined specific, rather than combined, categories because they are easier to code. For example, we distinguished between teaching and case-relevant communication, although both are examples of a broader “task-relevant communication” category. More specific categories require less cognitive effort from observers because they do not have to relate different behaviours to the same category. For later analyses, categories can be used separately but can also be combined into larger categories. Second, we chose event coding, which does not require observers to judge the quality of the behaviour observed or to integrate behaviours over time. This choice reduces cognitive load and interpretational biases; we can therefore expect higher inter-observer agreement and fewer differences between novices and experts than in behavioural marker-based systems [46, 64, 66]. Third, observers underwent intense training which included theoretical aspects, coding at least five procedures with an expert present, and post-observation discussions. This training is a considerable investment, but it is not unusual for observing group interactions [83, 84].

This study has limitations. First, SO-DIC-OR has only been tested in elective surgeries; emergency procedures have not yet been included, nor have laparoscopic procedures. Second, our data do not allow us to assess observer-specific biases. To assess such biases, multiple observers would have to observe the same procedure. Due
to space limitations, it was not possible to install more than two observational stations in the OR. Third, aiming at demonstrating the reliability of our system, the current study does not allow us to establish predictive validity; doing so would have required us to comparing the observations with external performance standards. Fourth, to limit the number of different categories to observe, the level of code differentiation is limited. For example, the communication categories are relatively general—future research will have to show whether these categories are sensitive enough to detect meaningful differences between high and low performing teams. In addition, some categories may not be unambiguous with regard to their categorization. For example, side conversations (e.g., among anaesthetists) may not always have a distracting effect. They could contain important information that—when overheard by the sterile team—could have positive effects on coordination and the procedure. Unlike observational systems based on behavioural markers, SO-DIC-OR does not allow an immediate assessment of teamwork quality. To be used for training purposes, it would need to be adapted. However, it is easy to produce frequency counts for the whole procedures or for specific time periods. These can serve as bases for training-related discussions within surgical teams.

CONCLUSIONS AND OUTLOOK

Our study showed that it is possible to reliably observe both teamwork and distractors simultaneously in the OR, even for long procedures. Data collection is relatively straightforward and based on an easily adaptable spreadsheet; no specialized observational software is needed. SO-DIC-OR is conceived primarily for research purposes. Data collected with SO-DIC-OR allows assessing combined influences of distractors and communication on surgical performance and outcomes.
### Webster Appendix – Overview Observational Systems for Direct Observations in the OR

#### Methods for observing Communication and Teamwork in the OR

<table>
<thead>
<tr>
<th>Observers</th>
<th>Number and type of procedures; duration (minutes)</th>
<th>Who is observed</th>
<th>Observational method</th>
<th>Behaviours observed</th>
<th>Inter-observer reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undre et al, 2006 [6]</td>
<td>Surgeons Psychologists 50 mixed general surgical procedures from hernia repairs to ileostomy reversals, open and laparoscopic, elective and emergency. Duration: M=136, SD=not reported; Range: 61-240</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers</td>
<td>OTAS - Communication, Cooperation, Coordination, Leadership, Monitoring</td>
<td>None reported</td>
</tr>
<tr>
<td>Sevdalis et al, 2009 [64]</td>
<td>Psychologist Human factor experts 12 elective urological procedures Duration: Not reported</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers</td>
<td>OTAS - Same categories as Undre et al. (2006)</td>
<td>Pearson r correlation between 0.72 and 0.76</td>
</tr>
<tr>
<td>Hull et al., 2011 [58]</td>
<td>Psychologists 30 general surgical procedures Duration: M/SD not reported; Range: 30-240</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers</td>
<td>OTAS - Same general categories as Undre et al. (2006); identified 114 specific behaviours</td>
<td>Intra-class correlation for OTAS &gt;.63</td>
</tr>
<tr>
<td>Russ et al., 2012 [67]</td>
<td>Surgeons Psychologists 14 general surgical procedures open/laparoscopic Duration: M=94, SD=36; Range: 30-150</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers</td>
<td>OTAS - Same categories as Undre et al. (2006)</td>
<td>ICC: 0.68 and &gt;0.70 for trained observers</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Procedures</td>
<td>Duration</td>
<td>Behavioural Markers</td>
<td>Scoring System</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>----------</td>
<td>---------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Frankel et al., 2007 [39]</td>
<td>Not specified</td>
<td>17 mixed surgical procedures (gastric bypass, interdisciplinary surgeries, caesarean sections)</td>
<td>Not specified</td>
<td>Behaviour markers; event coding</td>
<td>CATS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yule et al. 2008, 2009 [40] [85]</td>
<td>Surgeons</td>
<td>43 cases [85]; 6 videotaped simulated scenarios (general and orthopaedic surgery) [40]</td>
<td>M/SD not reported; Range: 2.3-5.4</td>
<td>Behavioural markers</td>
<td>NOTSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crosseley, et al., 2011 [62]</td>
<td>Participants observation by Anaesthetists Scrub nurses Surgical care practitioners Independent assessors</td>
<td>404 cases of 15 index procedures</td>
<td>Surgeons in training</td>
<td>Behavioural markers</td>
<td>NOTSS, see Yule et al., 2008 [40]</td>
</tr>
<tr>
<td>Mishra et al., 2009 [41]</td>
<td>Surgeons Human factor experts</td>
<td>36 laparoscopic cholecystectomies and carotid endarterectomies</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers</td>
<td>Oxford NOTECHS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Study</th>
<th>Participants/Expertise</th>
<th>Procedures/Duration</th>
<th>Sub-teams/Behavioural markers</th>
<th>Instrument/Team Skills</th>
<th>Agreement/Inter-rater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robertson et al., 2014 [42]</td>
<td>Surgical trainees or practitioners Human factor experts</td>
<td>297 surgeries, Duration: M=117, SD and Range not reported</td>
<td>Sub-teams: Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers: Oxford NOTECHS II</td>
<td>Per cent agreement: 45-78%; ICC: 343-881</td>
</tr>
<tr>
<td>Lingard et al., 2002 [43]</td>
<td>Researcher (not specified)</td>
<td>35 mixed surgical procedures (general, urology, otolaryngology, cardiac surgery), Duration: M=219.42; SD, Range not reported</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Field notes Post-observational coding of field notes Assessing level of tension</td>
<td>Communication: Time, Safety and sterility, Resources, Roles, Situation control Consensus through discussion</td>
</tr>
<tr>
<td>Lingard et al., 2004 [86]</td>
<td>Field Researcher (not specified)</td>
<td>32 general surgical procedures (not specified), Duration: Not reported</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Field notes Post-observational coding of field notes Assessing level of tension</td>
<td>Communication: Content, Non-verbal cues, Tone of voice, Use of repetition and emphasis, Outcome of exchange Consensus through discussion</td>
</tr>
<tr>
<td>Mazzocco et al., 2009 [44]</td>
<td>Registered nurses</td>
<td>300 mixed general surgical procedures (from biopsy to aortic aneurysm repair, open/laparoscopic), Duration: Not reported</td>
<td>Surgeons Anaesthetists Nurses Others</td>
<td>Behavioural markers: Communication /team skills</td>
<td>Briefing, Information sharing, Inquiry, Assertion, Vigilance and awareness, Contingency management RWG for observer calibration between 0.85 and 0.90</td>
</tr>
<tr>
<td>Nurok et al., 2010 [87]</td>
<td>Anaesthetists Nurses Safety specialists</td>
<td>Type of surgical procedures not reported, Duration: Not reported</td>
<td>Surgeons Anaesthetists Nurses</td>
<td>Behavioural markers: Communication / team skills</td>
<td>Briefing/re-briefing, Verbal knowledge sharing, Structured problem solving, Closed-loop communication, Conflict resolution and assertion, Debriefing Kappa between 0.41 and 0.99</td>
</tr>
<tr>
<td>Study Reference</td>
<td>Data Collection Methodology</td>
<td>Human Factor Experts</td>
<td>Procedures</td>
<td>Surgeons</td>
<td>Anaesthetists</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>Schraagen et al., 2010 [46]</td>
<td>Human factor experts</td>
<td>19 paediatric cardiac surgeries</td>
<td>19</td>
<td>Surgeons</td>
<td>Anaesthetists</td>
</tr>
<tr>
<td>Schraagen et al., 2011 [68]</td>
<td>Human factor experts</td>
<td>40 procedures performed by the same team</td>
<td>40</td>
<td>Surgeons</td>
<td>Anaesthetists</td>
</tr>
<tr>
<td>Parker et al., 2011 [47]</td>
<td>Psychologists</td>
<td>29 general, orthopaedic and vascular surgical procedures</td>
<td>29</td>
<td>Surgeons</td>
<td>Anaesthetists</td>
</tr>
</tbody>
</table>

- ** Duration: Not reported **
- ** M=136, SD=92; Range: 20-305 **
- ** Teamwork classification **
- ** Teamwork and cooperation **
- ** Leadership **
- ** Situation awareness **
- ** Decision-making **
## Methods for observing DISTRACTIONS

<table>
<thead>
<tr>
<th>Observers</th>
<th>Number and type of procedures; duration (minutes)</th>
<th>Who is observed</th>
<th>Observational method</th>
<th>Behaviours observed</th>
<th>Inter-observer reliability</th>
</tr>
</thead>
</table>
| Healey et al., 2006 [12]  | Psychologists 50 mixed general surgical procedures from anterior resection to cholecystectomy (open and laparoscopic) Duration: M=55.62, SD=5.44; Range: 13-217 | Surgeons Anaesthetists Nurses | Event coding Intensity rating | • Telephone calls  
• Bleeper  
• Radio  
• External staff  
• Equipment  
• Work environment  
• Procedural  
• Movement in front/behind monitors  
• Case irrelevant communication  
• Communication difficulties | ICC: 0.85; 0.65 for specific events |
| Healey et al., 2007 [13]  | Medical student 30 urology day-case procedures Duration: M=52.35, SD=10.70; Range: 7.4-312.7 | Surgeons Anaesthetists Nurses | Event coding Intensity rating | Same categories as Healey et al. (2006); not observed: radio, external staff, communication difficulties | None reported |
| Parikh et al., 2010 [48]  | Medical Student 26 paediatric orthopaedic surgical procedures Duration: Not reported | Surgeons Anaesthetists Nurses | Event coding | • Number of door swings  
• Changes of OR personnel during procedure | None reported |
| Persoon et al., 2011 [15] | Intern, researcher (not specified) 78 endourological procedures, not conventional or laparoscopic Duration: M=35, SD not reported; Range: 8-107 | Surgeons | Event coding Intensity rating | • Telephone calls  
• Bleeper  
• Radio  
• Door movement  
• Equipment  
• Procedure  
• Patient-irrelevant communication  
• Medically irrelevant communication | Percentage of agreement a 80% for first 17 procedures |
**Methods for observing Distractions as well as teamwork in the OR**

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Participants</th>
<th>Procedures Description</th>
<th>Personnel</th>
<th>Event Coding</th>
<th>Distractors</th>
<th>Teamwork</th>
<th>Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gillespie et al., 2012 [49]</td>
<td>160</td>
<td>Registered nurses 160 mixed surgical procedures, 10 specialties, elective and emergencies</td>
<td>Surgeons, Anaesthetists, Nurses</td>
<td>Event coding, combined with open comments</td>
<td>Procedural interruptions (distractor), Conversational interruptions (distractor), Miscommunications</td>
<td>None reported</td>
<td></td>
</tr>
<tr>
<td>Healey et al., 2008 [14]</td>
<td>22</td>
<td>Not specified laparoscopic cholecystectomies</td>
<td>Surgeons, Anaesthetists, Nurses</td>
<td>Event coding of distractors, behavioural markers for teamwork</td>
<td>Distractors: same codes as Healey et al. (2006), Teamwork: OTAS</td>
<td>None reported</td>
<td></td>
</tr>
</tbody>
</table>

M = Mean SD = Standard deviation

NOTE: OTAS=Observational Teamwork Assessment for Surgery; NOTECHS=Non-technical skills, NOTSS=Non-Technical Skills for Surgeons; CATS = Communication and Teamwork Skills; Kappa = Cohen's kappa; ICC = intra-class correlation coefficient; RWG = within group inter-rater agreement; videotaped simulated scenarios used to test reliability, tool developed for direct observations; introduction of a behaviour-based tool that can be used for direct observations and videotaped simulations.
REFERENCES


Acknowledgements

We thank Brigitte Dubach (head nurse), Uwe Klopfch and Melanie Bolliger for their support and Guillaume Crot, Christa Gfeller, Simon Huber, Moana Monnier, Irene Mühlemann, and Anna Püschel for help in data collection.

Conflicts of interest

None of the authors declares any conflicts of interest.

Authors’ contributions

Study concept and design: Tschan, Beldi, Semmer, Candinas
Method development: Tschan, Seelandt, Semmer, Keller
Data acquisition: Seelandt, Keller, Jenni, Tschan,
Drafting of the manuscript: Tschan, Seelandt, Semmer
Critical revision of the manuscript: Beldi, Jenni, Keller, Kurmann, Candinas
Statistical analyses: Seelandt, Tschan, Jenni
Administrative, technical and material support: Beldi
Study supervision: Tschan, Beldi, Semmer,

Financial Disclosure:

This study was supported by a grant from the Swiss National Science Foundation (grant 138273).
Figure 1: Development process of the observational system.