

Maria Luz, Gero Strauss, Dietrich Manzey

Impact of image-guided surgery on surgeons' performance: a literature review

Journal article | Accepted manuscript (Postprint)

This version is available at <https://doi.org/10.14279/depositonce-10984>



Luz, M., Strauss, G., & Manzey, D. (2016). Impact of image-guided surgery on surgeons' performance: a literature review. *International Journal of Human Factors and Ergonomics*, 4(3/4), 229. <https://doi.org/10.1504/ijhfe.2016.083516>

Terms of Use

Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

Impact of image-guided surgery on surgeons' performance: a literature review

Maria Luz

Research Scientist
Otto-von-Guericke University Magdeburg
Faculty of Computer Science
Computer-Assisted Surgery Group
P.O. 4120
39016 Magdeburg, Germany
maria.luz@ovgu.de

Gero Strauss

Managing Director
IRDC GmbH International Reference and Development Centre for Surgical Technology
Käthe-Kollwitz-Str. 64
04109 Leipzig, Germany
gero.strauss@irdc-leipzig.de

Dietrich Manzey

Professor
Technische Universität Berlin
Institut für Psychologie und Arbeitswissenschaft
Fachgebiet Arbeits-, Ingenieur- und Organisationspsychologie
Sekretariat F 7
Marchstr. 12
10587 Berlin, Germany
dietrich.manzey@tu-berlin.de

Maria Luz is a research scientist at the Computer-Assisted Surgery group in Otto-von-Guericke University Magdeburg. Her research addresses human-automation interaction in the medical field. She obtained a diploma in psychology from the Berlin Institute of Technology in 2008. In 2011 she has won the Human Factors Prize for Excellence for Human Factors/Ergonomics Research for her work about human factors issues of image-guided navigation systems. Currently she is a PhD candidate. The topic of her PhD thesis is “The impact of image-guided navigation with different degrees of automation on performance, workload and situation awareness of surgeons”.

Dietrich Manzey is professor of Work, Engineering and Organisational Psychology at Technische Universität Berlin, Germany. He completed his Diploma and PhD in Psychology at University of Kiel and got his habilitation from the University of Marburg, Germany. His research addresses issues of human-automation interaction in medicine, aviation, and process

control; multitasking, and cockpit display design. In 2011 he has won the Human Factors Prize for Excellence for Human Factors/Ergonomics Research for a work on human factors issues of image-guided navigation systems. He is also the responsible coordinator of the Human Factors master program at Technische Universitaet Berlin.

Gero Strauss is a director of the International Reference and Development Centre for Surgical Technology in Leipzig. He received his MD from the University of Leipzig, Germany, in 1999 and his habilitation in ENT from the same university in 2006. His areas of expertise are otology, ENT and plastic surgery.

Impact of image-guided surgery on surgeons' performance: a literature review

Abstract

The goal of current work is to provide a literature review of the performance consequences of image-guided surgery (IGS) use. Regarding IGS system functionalities the pattern of results suggests to distinguish between IGS systems which only provide information support (e.g. pointer-based systems) and those which directly intervene in surgeons' decision-making and actions (e.g. instrument disablement). The former offer benefits for patient safety and surgical outcome, as well as improved intraoperative orientation for surgeons, helping them to identify anatomical structures. Furthermore, IGS systems providing information support seem to shorten the time needed for surgery and reduce the subjective workload. IGS systems which intervene in a surgeon's decision-making and actions also have positive impact on patient safety and surgical outcome. In addition, these systems seem to reduce the physiological effort of surgeons. However, they can also prolong surgery and increase the subjective workload.

Key words: image-guided surgery, image-guided navigation, computer-assisted surgery, automation, types and levels of automation, patient safety, workload, situation awareness, skills, human factors

Introduction

Image-guided surgery (IGS) systems represent an advanced technology to support a surgeon when performing demanding operations. However, IGS research has focused either on technical issues (Cleary and Peters, 2010; Hassfeld and Mühling, 2001; Kingdom and Orlandi, 2004; Markelj et al., 2012; Ruppin et al., 2008; Widmann and Bale, 2006) or on the clinical outcomes (Dalgorf et al., 2013; Kosmopoulos and Schizas, 2007; Ramakrishnan et al., 2013; Shin et al., 2012; Smith et al., 2007; Sunkaraneni et al., 2013; Tian et al., 2011; Vreugdenburg et al., 2016). Because of this, most of the reviews have failed to provide any evidence that IGS actually has an impact on how the surgeon performs surgery or that IGS can contribute to make surgery cognitively easier and less stressful. As is known from other areas, e.g. aviation (Ferris et al., 2010), the introduction of computer-based assistance might not only provide advantages and benefits but could also entail new risks and challenges for the user which need to be considered in the overall evaluation of the systems. For example, apart from beneficial effects in terms of navigational performance, IGS might also introduce new aspects of workload, issues of inappropriate trust in its function, and/or may present new cognitive challenges for the surgeon. The consequences of these factors might not be directly reflected in variables of clinical outcome but could nevertheless represent important aspects to be taken into account when evaluating the overall impact of the system. This suggests that an evaluation of IGS should not be based on a consideration of clinical outcomes alone. To fully understand the impact of IGS other aspects also need to be taken into account, including

human factors issues arising from the complex surgeon-system interaction. Figure 1 illustrates the impact of interaction between IGS system, the patient, and the surgeon on clinical outcomes.

Insert Figure 1

Furthermore, the consequences might also be moderated by the specific functionalities of a given IGS system. This is suggested by work from basic automation research which suggests that the impact of computer-based assistance systems is highly dependent on the extent to which an assistance system provides only support for information acquisition and analysis or even higher cognitive functions like decision making (Onnasch et al., 2014). Based on these results it might be assumed that the performance consequences of IGS systems which just provide navigational information, like pointer-based IGS, might be different from systems which directly intervene in the intra-operative decision-making of surgeons, like sophisticated systems which automatically disable surgical instrument when the instrument approaches sensitive anatomical structures (Labadie and Fitzpatrick, 2011; Strauss et al., 2005; Strauss et al., 2007).

A first attempt to assess the impact of IGS systems on surgeons' performance compared to unassisted surgery taking surgeon-system interaction into account was made by Manzey et al. (2009a), drawing on a German survey of surgeons who had experience with pointer-based IGS. The results revealed several positive consequences of IGS, including subjective reports of improved patient safety and quality of surgical outcome, less time needed for the surgery, lower levels of perceived effort and stress during surgery, and increased intraoperative situation awareness, i.e. a better awareness and judgement of the current surgical situation. However, possible negative side effects were also identified, including increased perceived time pressure and mental demands, issues of overreliance on the computer assistance and possible detrimental effects on the acquisition of surgical skills. But because these results were based on subjective reports, the conclusiveness of this research appears to be limited.

The present review systematically analyzes available empirical research that has addressed performance consequences of IGS. We consider clinical studies with patients as well as studies conducted with cadavers or phantoms in the laboratory. Going beyond previous reviews and meta-analyses, we have included studies which address not only effects of IGS on clinical outcomes such as patient safety, surgical outcome, and surgery duration but also effects on the surgeon's cognitive performance using an IGS system, e.g. effects on workload and situation awareness while performing an intervention, as well as consequences for the maintenance of surgical skills. Furthermore, we analyzed the extent to which specific findings depend on the functionality of IGS systems, i.e. whether a system provided only navigational information, or more sophisticated direct intervention in the decision-making and actions of surgeons.

Methods

Search strategy and selection criteria

Articles for the literature review were identified by the first author through a PubMed search of the English-language literature from 1950 till August 2016 and additionally from the author's files and reviews of reference lists. Selection criterion for clinical studies was the consideration of data addressing aspects which go beyond describing consequences for patient safety and surgical outcome. Selection criterion for cadaver and phantom studies was the presence of data apart from technical or workflow data. Key inclusion criterion for publication was that the research addressed comparison of IGS with a control group performing the same task without an IGS system or a comparison of different IGS system functionalities. With respect to the types of interventions considered, we excluded IGS applications for surgery of soft tissue, joint replacement, dental treatment, and maxillofacial surgery because of the characteristics of these sorts of surgery and resulting specific issues when using IGS (e.g. tissue shift in soft tissue surgery or the need for high accuracy in dental treatment).

Publications were included independent of the type of data information (objective or subjective data) and data quality (whether the research involved purely descriptive variables or variables analyzed by statistical tests). Moreover, we included publications independent of whether the data was collected and statistically treated run-related, surgeon-related or patient-related. This differentiation is dependent on the purpose and conditions of data collection, which are associated with specific characteristics in terms of analyses by means of statistical tests, and should be taken into account in study design. Data is called *run-related* if one specific dataset is linked to a certain run (i.e. a procedure like registration) and the result is not dependent on the human participants. This is the case when a new technology is evaluated in terms of technical features across a number of applications. For example, the registration accuracy of an IGS system can be tested by performing registration several times (i.e. runs) according to a standardized procedure which makes the application independent of both the user (e.g. nurse) and the receiver of this procedure (patient). Data is called *surgeon-related* if a specific dataset is originated from a specific surgeon and the surgeon is considered as the relevant study participant. This kind of data collection is used when any kind of new surgical technique or technology is evaluated in terms of performance consequences, for example the evaluation of IGS impact on the intraoperative orientation of surgeons. In this case the surgical outcome is a direct result of the surgeon's actions (Fig. 1) and the patients, with their different anatomical characteristics, are considered as part of the surgical task. Finally, we call data *patient-related* if one specific dataset is originated from a certain patient (patient-based outcome, e.g. severity of the symptoms) and each patient is considered individually as a study participant. Patient-related data collection is appropriate for evaluating medication effects on patients. In this case, the physician providing the medication plays a negligible role. The distinction regarding data collection and treatment seems to be important because an inappropriate statistical data treatment might bias the results of a study. We report the type of data information and quality features in results section and discuss these later.

Types of performance consequences

In analyzing the literature, the following five aspects of performance consequences of IGS were considered:

Patient safety and surgical outcome. Two aspects reflect the efficacy of a surgery: patient safety and surgical outcome. According to the World Health Organization “Patient safety is the absence of preventable harm to a patient during the process of health care” [World Health Organization]. We consider as variables of patient safety any objective or subjective assessment of intraoperative and postoperative complications, expert assessments of possible complications, and unwanted injuries and damages of risk structures. Other variables of efficacy include all aspects determining the quality of the surgical outcome such as maintenance of safety margins, precision of tissue removal, missed paranasal sinuses, number of retained ethmoid cells and partitions, number of unopened frontal recess cells, or expert assessments of the overall quality of surgical outcome.

Surgery duration. This performance aspect includes any variables used to assess how long it takes a surgeon to accomplish a given intervention or a particular surgical step with and without IGS system or with different sorts of IGS systems. It directly reflects an aspect of efficiency of the surgeon-system interaction.

Situation awareness. Situation awareness has been defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” [Endsley, (1995), p.36]. The most frequent argument that justifies the IGS is that it improves intraoperative orientation of surgeons and identification of anatomical structures (e.g. Caversaccio et al., 2002; Dalgorf et al., 2013; Dubin et al., 2008; Eliashar et al., 2003; Hassfeld and Mühling, 2001; Metson, 2003; Ramakrishnan et al., 2013; Tabaei et al., 2003). We consider intraoperative orientation and identification of anatomical structures as aspects of situation awareness. However, situation awareness also includes other aspects, e.g. knowledge of characteristics of the current surgical situation.

Workload and stress. One of the goals of IGS is the reduction of surgeons’ cognitive workload and stress. However, if the new technology is too complex or requires the operation of an additional tool, it can have the opposite effect. Two groups of workload variables are distinguished for this review: subjective workload and objective physiological indicators of effort and stress (e.g. ECG or blood pressure).

Acquisition and maintenance of surgical skills. IGS use during training of young surgeons may result in the development of inappropriate surgical skills. Moreover, IGS may lead to skill degradation of experienced surgeons. This can have negative consequences when the IGS system is for any reason not available. All variables used to assess the impact of IGS on the acquisition or maintenance of surgical skills were included in this category.

IGS system functionalities

With respect to the diversity of IGS systems in terms of functionalities, two broad classes of IGS systems were distinguished: (1) IGS systems which mainly support information acquisition and information analysis functions, e.g. by providing some sort of navigation

information, (2) More sophisticated IGS systems which support surgeons' decision-making and intervene in the surgeon's actions (see Table 1).

The most common type of the first class of IGS systems are *pointer-based systems* (PB-IGS) which provide the surgeon with basic information about instrument position by integrating it within triplanar image data. In order to obtain this information, the surgeon uses a special instrument, the "pointer", which must be actively directed to the current position of an instrument. More recent developments, referred to as *instrument navigation* (IN-IGS) (Caversaccio et al., 2002; Sindwani, 2008), make it possible to directly and continuously track surgical instruments (e.g. shaver, aspirator, forceps, endoscope, drill). Using one of these sorts of IGS systems, the surgeon no longer needs to interrupt the surgery for instrument change to obtain navigational information. PB-IGS has been used in the operating room for over 25 years. Some IN-IGS systems are already available for clinical use. However, instrument navigation is impractical for some types of surgeries (e.g. sinus surgery) and therefore less widespread.

A number of more advanced functionalities provide the surgeon with information which goes beyond a basic presentation of the instrument localization within image data. Accordingly there are various other specific approaches which we have assigned to the group of information support IGS, including systems providing intraoperative visualization of cumulative instrument positions corresponding to removed tissue providing the surgeon with information about the surgical process over time (*process visualization*, PV-IGS) (Hong et al., 2009; Voormolen et al., 2012; Woerdeman et al., 2009a), visualization of uncertainty arising from registration, instrument calibration and tracking (*uncertainty visualization*, UV-IGS) (Simpson et al., 2014), visualization of distances between the surgical instrument and sensitive anatomical structures that need to be protected during the surgery (*distance visualization*, DV-IGS) (Cho et al., 2013; Voormolen et al., 2012), auditory or visual alerts if the instrument approaches such a structure (*proximity warnings*, PW-IGS) (Cho et al., 2013; Dixon et al., 2014a; Voormolen et al., 2012; Willems et al., 2005; Woerdeman et al., 2009b), augmentation of risk structures included directly into the whole endoscopic video image or just on its background (*augmented risk structures*, ARS-IGS) (Dixon et al., 2012; Li et al., 2016), augmentation of target tissue included directly in the microscopic view (*augmented target volume*, ATV-IGS) (Woerdeman et al., 2009b), augmented cues for small targets (e.g. crosshair) included in the endoscopic video image (*augmented target*, AT-IGS) (DeLisi et al., 2014; DeLisi et al., 2015), augmented pathways which represent the best way to bring the instrument to a certain target area or tumor directly included in the endoscopic video image (*augmented pathway*, AP-IGS) (Caversaccio and Frysinger, 2003; Freysinger et al., 1997), or three-dimensional endoscopic virtual visualizations of anatomical structures (*three-dimensional virtual image guidance*, 3DV-IGS) (Dixon et al., 2016; Dixon et al., 2014a).

The second class of more advanced IGS systems includes systems which go beyond just providing information support. In addition to some navigational information presented on a navigation monitor, they also support intraoperative decision-making of the surgeon and even intervene directly in the surgeon's action. Thus, they constitute a considerably higher level of automation support than the information support systems (Manzey et al., 2009b). Examples are IGS systems which reduce the speed of the surgical instrument as soon as sensitive anatomical structures are approached too closely or disable it entirely (*instrument*

disablement, ID-IGS) (Labadie and Fitzpatrick, 2011; Strauss et al., 2005; Strauss et al., 2007), or restrict the movement of the instrument attached to a robotic arm (*movement restriction*, MR-IGS) (Lim et al., 2016). Other systems automatically control the cutting depth based on navigation information while the surgeon freely moves the instrument on the skull surface (*semiautomatic trepanation system*, STS-IGS) (Follmann et al., 2010).

In contrast to standard information support systems, advanced IGS systems usually require some additional pre-operative steps (see Table 1). Most of these systems only exist as research prototypes and their use is limited to the context of some clinical pilot studies or laboratory studies on cadavers or phantoms.

Insert Table 1

Results

Database

A total of 36 studies (34 publications) were included in this review. Two general groups of studies were identified: 22 studies (21 publications) which compare IGS with conventional, unsupported surgery and 11 studies (10 publications) which compare different IGS system functionalities (Table 2). Three studies include data addressing both sorts of comparison (Luz et al., 2015; Marcus et al., 2015; Wise et al., 2008).

All relevant areas of application are represented in this set of studies: paranasal sinuses and anterior skull base (17 studies), temporal bone and lateral skull base (9), orbital surgery (4), neurosurgery (5), and spinal surgery (1). The studies were performed on patients (8), cadavers (12), phantoms (15), and animals (1). The number of participating surgeons included in the studies varies from one (Cho et al., 2013; DeLisi et al., 2015; Lim et al., 2016) to 50 surgeons (Dixon et al., 2014b; Marcus et al., 2015). Three publications gave no information about participating surgeons (Gong et al. 2007; Hofer et al., 2008; Woerdeman et al. 2009b). The experience of participants included in the studies as “surgeons” varied from non-physicians to trained surgeons with various level of clinical experience.

We report the results of comparisons of IGS with unsupported surgery for each type of performance consequences. The results of comparison of different IGS system functionalities are presented afterwards in a single section.

Insert Table 2

Patient safety and surgical outcome

Twelve of 24 studies were identified which included 27 different variables to evaluate the impact of IGS on patient safety and surgical outcome as compared to non-IGS performance. Table 3 provides an overview of the studies, the variables considered, and a qualitative evaluation of effects of IGS on patient safety and surgical outcome according to IGS system functionality. Additionally, we separated IGS functionalities which provide support for information acquisition and analysis from IGS functionalities which intervene in surgeons’ decision-making and actions by horizontal dashed line. We indicated whether the impact of

IGS system on specific variables was positive, neutral, or negative according to literature reports.

Generally, the overall pattern of effects points to an advantage of IGS in comparison to unsupported surgeries independent of IGS system functionality. Fifteen variables (including three variables that were analyzed by means of a statistical test) suggest higher efficacy when surgeons were supported by an IGS system compared to unsupported performance. Ten out of 27 variables (including six variables that were analyzed statistically) indicated no differences between the IGS and the control condition. Two variables which were reported descriptively revealed a negative effect of IGS use on surgeons' performance. However, one result was considered by the authors as a coincidence (Luz et al., 2014) and another result was not considered by the authors as clinically relevant (Follmann et al., 2010).

Insert Table 3

Surgery duration

A total of 13 studies were identified which included 18 different variables to evaluate the impact of IGS on surgery duration or surgery speed as compared to unsupported performance. Table 4 provides an overview of the studies, the variables considered and a qualitative evaluation of effects of surgery duration.

Overall, the results provide a mixed picture, with five studies (six variables) pointing to an advantage of IGS in terms of faster surgery or speed, six studies (six variables) pointing to a disadvantage, and five studies (six variables) reporting no difference between IGS and unsupported surgeries in this respect. However, closer examination reveals that all studies in which disadvantages of IGS were found either concentrated on preparation times or involved IGS systems which directly intervene in the surgeon's action (e.g. ID-IGS). In contrast, benefits of IGS were specifically found for systems that only provide information support to the surgeon. This benefit seems to emerge specifically in non-routine situations, as is suggested by the results of the study of Gong et al. (2007), who found that the use of PB-IGS does not affect duration of particular surgical steps when the anatomy is normal, but shortens the duration, when the anatomy exhibits anomalies.

Insert Table 4

Situation awareness

A total of 12 studies (33 variables) were identified which addressed the impact of IGS on situation awareness. Table 5 provides an overview of the studies and effects. Overall, the results do not show a clear picture: 13 variables point to a positive impact of IGS on situation awareness, four variables point to a negative impact, and 16 variables report no difference between IGS and unsupported surgeries in this respect. Closer examination of results reveals that all studies showing advantages of IGS had concentrated on IGS systems that provide information support using mostly simple localization and identification task to explore the research question. Only two studies involved complete surgery (Casiano and Numa; 2000; Stelter et al., 2011) investigating the impact of IGS on performance of sinus surgery. These studies support the common argument that IGS improves the intraoperative orientation of

surgeons. That is, surgeons are better able to identify and to localize anatomical structures when using IGS systems as compared to the non-IGS control condition.

The general positive pattern of results is clouded by results pointing to a possible new risk of attention tunneling (Dixon et al., 2013; Marcus et al., 2015). Dixon et al. (2013) and Marcus et al. (2015) compared the impact of IGS with conventional endoscopic display on the performance of surgeons performing a landmark localization task while also assessing the extent to which surgeons working with or without an IGS system were able to detect an anomaly, e.g. a foreign body embedded in the surgical site. The results show that surgeons supported by an IGS system localized a relevant landmark significantly better and more quickly. However, they also missed a foreign body or an injury of optic nerve significantly more often compared to the surgeons working with the conventional endoscopic display. Obviously, the augmented risk structures captured their attention to such extent that they failed to detect the anomalies in the surgical site. One variable showed a negative impact of IGS on target identification time (Ingram et al., 2011). However, the latter result was only reported descriptively and might not be significant.

Three studies (15 variables) evaluated the impact of IGS on other aspects of situation awareness. As becomes evident from this list, the vast majority of variables was evaluated as part of laboratory research addressing the impact of advanced IGS systems (ID-IGS). No effects were found – neither positive nor negative – on the situation awareness of surgeons as a consequence of computer-based assistance compared to conventional surgery. The only exception points to impaired assessment of anatomical characteristics using IGS systems (Manzey et al., 2011).

Insert Table 5

Workload

Forty-three variables were used in eleven studies to evaluate possible impacts of IGS on subjective workload. A detailed overview of these variables is presented in Table 6. At first sight, the results provide a mixed picture. More than half of the reported findings (23 of 43 variables) show no differences between IGS and non-IGS, 11 variables point to advantages and nine variables show disadvantages of IGS in comparison to non-IGS. A closer inspection of the data reveals that disadvantageous effects of IGS were mainly reported from studies involving IGS systems that directly intervene in the workflow of surgeons, and were mainly reflected in increased subjective assessments of frustration (Follmann et al., 2010; Luz et al., 2014; Manzey et al., 2011). Contradictory results were only reported in the study conducted by Hofer et al. (2008), who used a one-dimensional rating of cognitive strain for workload assessment. However, the authors do not describe how many surgeons participated in the study and they only present their results descriptively. Positive results (10 of 21) or neutral results (11 of 21) were reported for all other IGS system functionalities (PB-IGS, ARS-IGS, and PW-IGS).

Insert Table 6

A total of six studies addressed the impact of IGS on objective physiological effort. The 22 different variables used in these studies and the effects are summarized in Table 7. The studies can be divided into two subgroups. The first subgroup involves three clinical studies

(9 variables) comparing effects of PB-IGS with a control group (Alobid et al., 2011; Stelter et al., 2015; Theodoraki et al., 2015). All of these studies failed to find significant differences between IGS and unsupported sinus surgery in terms of cardiovascular and endocrine indicators of workload and stress.

The second subgroup includes three laboratory studies (13 variables) using phantoms to simulate a mastoidectomy and addressed effects of advanced ID-IGS. In two of these studies a significant reduction of physiological effort was found while operating with support of ID-IGS, reflected in heart rate and respiration rate (Manzey et al., 2011), systolic blood pressure, and heart rate variability (Luz et al., 2014).

Insert Table 7

Skill acquisition and maintenance

A total of three studies were identified which included five different variables to evaluate the impact of IGS on skill acquisition as compared to unsupported training: these are summarized in Table 8. Two studies (four variables) were performed on patients using basic IGS systems (PB-IGS). Both studies only included subjective assessments by surgeons with respect to possible effects of IGS on this performance aspect. The results in terms of skill acquisition are contradictory: one study (two variables) reports positive expectations with respect to the impact of IGS on skill acquisition of young surgeons (Casiano and Numa, 2000). In contrast, another study raised doubts as to whether young surgeons would develop appropriate surgical skills using IGS systems during training (Stelter et al., 2011). The third study is the only available which has experimentally investigated the impact of an advanced ID-IGS on skill acquisition using phantoms (Manzey et al., 2011). Analysis by means of a statistical test of objective data revealed neither positive nor negative effects of IGS on skill acquisition.

Only one study has addressed the issue of possible skill degradation in relatively experienced surgeons due to frequent use of IGS (Stelter et al., 2011). Based on subjective assessments, the degradation hypothesis was not supported. The overall pattern of results does not provide support for concerns about possible negative effects of IGS on skill acquisition and maintenance. However, the overall number of studies addressing this issue is small, which makes any decisive conclusion difficult.

Insert Table 8

Comparison of different IGS system functionalities

In this section, we present the results of comparison of different IGS system functionalities.

A total of 14 studies (13 publications) were identified which included 42 different variables to evaluate the impact of different IGS system functionalities on surgeons' performance. The results are presented in Table 9 where the impact is related to more advanced IGS functionality. The number of variables for each performance aspect is relatively small, making any distinct consideration of these aspects impossible and any distinct conclusions difficult. Overall, the results suggest that providing information which goes

beyond the basic navigation information provided by PB-IGS and IN-IGS systems can significantly improve the surgeons' performance. However, it might also introduce some new risks. For example, Woerdeman et al. (2009b) report worse subjective resection quality and objective resection precision if the preoperatively segmented target volume was augmented into the microscope. Dixon et al. (2014b) report worse recognition of unexpected anomalies if augmented risk structures are presented on the same endoscopic video screen. Thus the positive effects of providing more distinct information might be off-set by negative side effects, which should be carefully considered in future research. Moreover, one study points to differences between the two classes of IGS systems. Participants needed more time to complete surgery and reported higher subjective workload when using decision and action supporting IGS system compared to systems only providing information support (Luz et al., 2015).

Insert Table 9

Discussion

Our goal was to review the literature in terms of IGS impact on surgeons' performance. Besides effects of patient safety, surgical outcome, and surgery duration, which have been addressed in previous reviews, we also considered situation awareness, workload, and skill maintenance. In the following, we summarize and discuss the results for the various performance aspects, followed by a discussion of the impact of different IGS system functionalities on surgeons' performance. We then consider the quality of the research data and the typical study environments used to investigate the impact of IGS in the studies included in this review. We conclude with a consideration of review limitations, future challenges, and a summary.

Patient safety and surgical outcome

The reported data provide consistent evidence for a positive impact of IGS on patient safety and surgical outcome, independent of specific IGS system functionality. This finding is consistent with results of four previous reviews and meta-analyses which have considered effects of IGS on surgical complications, mostly based on non-randomized clinical trials (Dalgorf et al., 2013; Shin et al., 2012; Tian et al., 2011; Vreugdenburg et al., 2016). Our review broadens the empirical basis in this respect.

Surgery duration

Surgery duration is an important economic factor. Our review suggests that the possible advantages of IGS on surgery duration are moderated by the functionality of IGS systems. IGS systems which only provide information support for the surgeon were generally found to speed up surgery in most studies. This result is in agreement with survey results of Manzey et al. (2009a), who found similar effects based on subjective reports from experienced surgeons. However, when looking at the more advanced ID-IGS or STS-IGS, surgery times were generally found to be prolonged as compared to unassisted performance. This seems to be related to the fact that these systems actively interrupt the surgeon by disabling the instrument

in proximity of risk structures or in case of line-of-sight – it then takes time to check and to re-activate the instrument. Thus, the possible benefits in terms of patient safety and surgical outcome seem to be traded off against a prolongation of the surgery when using these more advanced IGS systems.

Situation awareness

Our results suggest that IGS systems which provide information support improve intraoperative orientation and the identification of anatomical structures and, thus, improve the situation awareness of surgeons. This observation supports findings of the previous survey study of Manzey et al. (2009a). In contrast, IGS systems which intervene in surgeons' decision-making and actions do not seem to have a comparable effect. One finding even suggests that using these more advanced IGS systems may reduce situation awareness, at least in novice surgeons (Manzey et al., 2011). This suggests that increasing automation support of IGS systems beyond information support can reduce the positive effects seen on situation awareness. However, looking at this conclusion it must be taken into account that our comparison of the two classes of IGS systems was confounded with a number of other factors such as study environment, application, and specific aspects of situation awareness (simple localization and identification tasks vs. questions about the dynamic situation of the surgery). Thus, it is not certain whether the differences found in this review really reflect differences related to the functionality of the two classes of IGS systems, or are just related to other differences between the studies.

The most common reasons advanced for the application of IGS are that it improves intraoperative orientation of surgeons and helps them to identify anatomical structures that lead to improved patient safety and surgical outcome (e.g. Caversaccio et al., 2002; Dalgorf et al., 2013; Dubin et al., 2008; Eliashar et al., 2003; Hassfeld and Mühling, 2001; Metson et al., 2003; Ramakrishnan et al., 2013; Tabaee et al., 2003). Our review suggests that both arguments are correct. However, we could not find any work that explicitly investigated the relationship between intraoperative orientation and clinical outcomes.

Workload

The impact of IGS on surgeons' subjectively perceived workload seems to be highly dependent on the sort of support the IGS provides. IGS systems which only provide information support were consistently found to reduce the subjectively perceived workload during surgery. However, for the more advanced ID-IGS and STS-IGS systems an increase in the subjective workload was found compared to unassisted conditions. These were mostly reflected in elevated frustration scores. One obvious reason for this effect is the interruptions of surgeons' workflow caused by "false alarms" due to technically related issues such as lack of line-of-sight and system inaccuracy.

With respect to physiological effort, IGS systems which provide information support do not seem to have any effects on surgeons' workload and stress. However, clear benefits were found for the more advanced IGS systems which automatically monitor the surgeons' actions and intervene when detecting a possible risk of injuring sensitive anatomical structures. This suggests that although the ID-IGS systems might lead to elevated levels of frustration, they nevertheless reduce the overall stress level during the surgery. This most likely reflects a

direct effect of the automated protection function implemented in these systems, which directly reduces the risk of unwanted injuries of risk structures and which marks a clear contrast to systems that only provide information support.

Skill acquisition and maintenance

Possible effects of computer-assisted surgery on the acquisition and maintenance of surgical skills is a controversial issue. While some authors argue that IGS can help to develop and acquire proper surgical skills (Baudoin et al., 2013; Dalgorf et al., 2013; Eliashar et al., 2003) others have expressed concerns that the routine use of IGS systems might increase the risk of skill loss and dependence on the system (e.g. Mueller and Caversaccio, 2010). Our review revealed only three studies that have addressed the possible impact of IGS on skill acquisition in surgical training, and only one of these studies was experimental in nature and reported objective data. The results do not show any effects of IGS on skill acquisition in either direction. But the number of studies is still small and does not provide a proper data basis for any decisive conclusion. Remarkably, we did not find any studies which addressed the impact of routine use of IGS systems on the retention of navigation skills of surgeons. Given the issue of automation induced skill degradation in other areas (e.g. aviation, Ferris et al., 2010), such studies seem to be urgently called for.

Comparison of different IGS system functionalities

With respect to a comparison of IGS systems providing different sorts of information support, the results show that any kind of additional information support beyond that provided by the most basic PB-IGS and IN-IGS systems can improve surgical performance and may contribute to making surgery quicker and more accurate. However, it seems that the implementation of the functionalities, e.g. how the information is presented to the surgeon, plays an important role in this respect. For example, the use of augmented reality which directly presents augmented information in the endoscopic or microscopic view leads to attentional tunneling and biases which actually increase risks for patient safety as compared to other IGS functionalities (Dixon et al., 2013; Marcus et al., 2015; Woerdeman et al., 2009b). A presentation of the information on a separate screen or in the form of acoustic signals (e.g. proximity warnings) might represent a better solution in this respect (Dixon et al., 2014b; Woerdeman et al., 2009b). A general conclusion which can be drawn from this finding is that less augmentation is sometimes better than more augmentation and that implementing new and more sophisticated formats of information presentation in IGS systems functionalities should always go along with a careful consideration of possible negative consequences for surgeons' attention.

With respect to a comparison of IGS systems providing information support with more advanced systems also providing decision and actions support, the available research suggest that one obvious benefit of the latter systems can be seen in a reduction of the physiological stress of surgeons during surgery. This effect seems to be mainly due to the automated protection function for injuries of risk structures implemented in these systems. However, at least in the current generation of prototypes this benefit seems to be achieved only at expense of an increased subjective frustration level and prolongation of surgery, induced by a considerable number of "false alarms". Further developments of these systems need to

address the issue of “false alarm” in order to improve the overall beneficial effects of using such systems.

Data quality

A review of the data quality of the available studies on human performance consequences of IGS systems reveals a surprisingly low quality, often reflected in a lack of statistical analysis and small sample sizes resulting in low statistical power. Another point is that the run-related and patient-related data treatment can lead to an underestimation of the variance in the data and an artificial increase in the sample size, compared to the more appropriate surgeon-related data collection. Both of these effects directly impact the outcome of statistical testing of effects by making the testing considerably more liberal, i.e. effects observed in such studies might become statistically significant although the variance in the underlying data is in fact only random variation.

Study environment

Two kinds of study environments have been used to assess the IGS impact on surgeons’ performance – field studies with patients and laboratory studies with cadavers or phantoms. From the review, it becomes evident that phantom studies usually produce data of higher quality than cadaver and clinical studies, i.e. objective, surgeon-related data which was analyzed by means of statistical tests. This increases the internal validity of the results of phantom studies. Phantom studies provide a high level of experimental control and standardization of anatomy that reduces the influence of confounding factors and decreases data variance, and may therefore make it easier to detect weak effects. In contrast, clinical studies are characterized by high variation of patients’ anatomy that can only be controlled for by patient randomization. However, it is considered unethical to randomize patients if experts consider particular surgical treatment advantageous and another risky (Marple and Setzen, 2006; Smith et al., 2007). This seems to be the main reason why meta-analyses about IGS impact on clinical outcomes have been performed only on the basis of predominantly non-randomized studies (Dalgorf et al., 2013; Shin et al., 2012; Tian et al., 2011; Vreugdenburg et al., 2016). This only corresponds to evidence level 2a according to Centre for Evidence-Based Medicine (CEBM) Levels of Evidence Working Group. In retrospective studies, the use of IGS was dependent on its availability. In other studies the use of IGS was dependent on the severity of disease: based on recommendations of American Academy of Otolaryngology Head and Neck more severe cases were treated with an IGS system (Stelter et al., 2015; Strauss et al., 2006). The position statement of the American Academy of Otolaryngology Head and Neck (AAO-HN) regarding intra-operative use of computer-assisted surgery is based on expert opinions and has not been verified experimentally, also due to ethical issues.

However, phantom studies also have disadvantages. Because they do not exactly reflect the conditions and risks of “real” surgery, their external validity can be challenged. Ethical reasons prohibit the investigation of some potential risks related to IGS on human patients (e.g. skill loss, impairments of situation awareness). In order to minimize these risks, comprehensive phantom studies should be conducted prior to any clinical studies on human patients.

The role of surgeons in clinical research

Although IGS has been used for more than 20 years, the impact on surgeons has rarely been investigated. Only 34 publications could be identified as relevant for this review. Taking into consideration that all studies were included which address human factors issues as their primary or secondary purpose and the review considers numerous aspects of these issues, this is relatively small number. However, the interest in this topic seems to be growing: 25 of 34 articles have been published within the past six years.

The small number of studies investigating IGS impact on surgeons' performance is possibly due to a general neglect of the importance of the individual surgeons in clinical research. It seems that clinical research implies that IGS would directly affect the surgical outcome and that surgical outcome is related only to the individual characteristics of the patient. The surgeons with their specific skills and experience seem to be absent from this mental model despite their active role in using IGS and integrating it in the workflow during the surgery. This is reflected in statistical and methodological neglect of surgeons in study design (e.g. as an independent variable if patient-related data collection is required), the disregard of a highly-controlled phantom or cadaver studies, and lack of appreciation of the surgeon as a buffer for possible differences between surgical techniques.

The hierarchy of evidence-based medicine is also limited to a consideration of patients' data. With respect to evaluation of surgical treatments it does not take the surgeon into account as an important agent (OCEBM Levels of Evidence Working Group). However, there is no doubt that objective data on surgeons' performance collected in highly controlled experimental phantom studies provide stronger evidence than expert opinions (evidence level 5) and case-series (evidence level 4). It is also debatable whether they provide stronger evidence than the non-randomized clinical trials (evidence level 3b, 2b, and 2c). Particularly for surgical treatment, the hierarchy of evidence-based medicine should be expanded by methodological and statistical consideration of surgeons and consideration of laboratory studies. This would make it possible to include phantom studies in evidence-based reviews and meta-analyses and allow more insight and definitive statements in terms of IGS impact.

Surgeons are a crucial factor which may significantly influence the clinical outcome. Because they are aware of their responsibility for the patients' life and health, they possibly compensate for weaknesses of certain surgical technique (including the use of specific technologies) by investing more cognitive and physical effort. Thus, performance consequences of different surgical techniques might not always be seen in the overall outcomes of a surgery but in more subtle aspects such as effort, cognitive spare capacity or aftereffects like fatigue (Hockey, 1997). Consequently, it makes sense to determine the advantages of one surgical technique compared with another by measuring multiple aspects of surgeons' performance. Moreover, this possible phenomenon should be considered by statutory health insurers in their decision to pay for specific surgical intervention.

Limitations of the review

In our review, we included all available studies addressing performance consequences of IGS, independent of kinds of surgery, IGS system functionalities, study environment, and data quality. This was primarily done to work with a most comprehensive data base. However, the

heterogeneity of studies has led to some confounding factors which may have biased some of our conclusions. For example, practically relevant PB-IGS systems were evaluated mostly for anterior skull base and sinus surgery on patients or cadavers. Few of these studies reported objective, surgeon-related data that were analyzed by means of statistical tests. In contrast, almost all studies which evaluated ID-IGS were performed in the laboratory and included phantom-based simulations of lateral skull base and temporal bone surgery. These were usually carefully designed experimental studies with objective, surgeon-related data, which were analyzed by means of statistical tests. Thus, the differences which have emerged between these classes of systems in this review could have been caused by the differences in study protocols.

Unfortunately, no meta-analysis could be performed due to the limited number of studies providing high quality data and due to the above mentioned heterogeneity of included studies. No distinction between results obtained in different study environments was made neither.

We did not consider whether the IGS systems were optical or magnetic. The issue of line-of-sight in optical IGS systems can have a large impact on surgeons. The role of IGS system accuracy and surgical experience were also not considered in this review. Furthermore, this review was limited to the IGS application for otorhinolaryngology and neurosurgery. Other IGS applications such as soft tissue surgery, joint replacement, dental treatment, and maxillofacial surgery were not included. These applications have specific characteristics and may raise various issues.

As mentioned in the introduction, the implementation of computer-based assistance systems not only provides advantages and benefits but also entails new risks and challenges for the user. Most studies we considered direct their work predominantly to positive IGS impact. A much smaller number of studies investigated possible risks and problems. Thus, this review gives only limited insight into possible risks arising from IGS use.

Future challenges

What conclusions can be drawn from the present review concerning future challenges in investigating and evaluating the benefits and costs of IGS in terms of human performance consequences? First, it has become evident that some aspects and possible risks which might arise from long term IGS use have not yet been thoroughly investigated. These aspects include possible risks of (manual) skill degradation. Other risks such as development of overtrust and overreliance which possibly may result in complacency and automation bias effects (Parasuraman & Manzey, 2010) have not been studied at all and for this reason could not been considered in this review. All of these effects can off-set the benefits of IGS at least to some extent and may even compromise patient safety. Future studies should address these issues in order to better assess the risks related to them.

Second, the current review has disclosed several weaknesses of the current research with respect to research designs, data treatment and collected data. Studies have often used research designs without proper control conditions, data treatments leading to liberal statistical testing, or subjective variables as primary dependent variables (e.g. surgeons' subjective assessments of systems). Future challenges will include it to conduct more controlled studies based on cadavers or simulations beside clinical trials. Moreover, data-treatment in these studies should be surgeon-related, i.e. based on a sufficient number of

surgeons as participants, and dependent variables should not only be based on subjective data but also include objective data like performance data or physiological data. A number of effective variables which can be considered in this respect are provided by respective studies considered in this review.

Summary

The review reveals apparent benefits of IGS in terms of patient safety and surgical outcome. Moreover, IGS systems which provide information support seem to shorten surgery duration. The reason for this effect is less clear. It could be, as several studies included in this review show, that IGS improves surgeons' situation awareness and reduces their workload and stress. Moreover, the review reveals new risks that might arise from IGS. Some efforts have already been made to investigate risks such as attention tunneling and insufficient skill acquisition.

References

Alobid, I., De Pablo, J., Mullol, J., Centellas, S., Parramon, G., Carrasco, J., Armario, A. and Bernal-Sprekelsen, M. (2011) 'Increased cardiovascular and anxiety outcomes but not endocrine biomarkers of stress during performance of endoscopic sinus surgery: a pilot study among novice surgeons', *Archives of Otolaryngology – Head & Neck Surgery*, Vol. 137, No. 5, pp.487–492.

American Academy of Otolaryngology – Head and Neck Surgery. [online] <http://www.entnet.org/content/intra-operative-use-computer-aided-surgery> (Accessed 4 July 2016).

Baudoin, T., Grgić, M.V., Zadavec, D., Geber, G., Tomljenović, D. and Kalogjera, L. (2013) 'Algorithm for navigated ESS', *Rhinology*, Vol. 51, No. 4, pp.335–342.

Casiano, R.R. and Numa, W.A. (2000) 'Efficacy of computed tomographic image-guided endoscopic sinus surgery in residency training programs', *The Laryngoscope*, Vol. 110, No. 8, pp.1277–1282.

Caversaccio, M. and Freysinger, W. (2003) 'Computer assistance for intraoperative navigation in ENT surgery', *Minimally Invasive Therapy & Allied Technologies*, Vol. 12, No. 1-2, pp.36–51.

Caversaccio, M., Nolte, L.P. and Häusler, R. (2002) 'Present state and future perspectives of computer aided surgery in the field of ENT and skull base', *Acta Oto-Rhino-Laryngologica Belgica*, Vol. 56, No. 1, pp.51–59.

CEBM Centre for evidence-based medicine. Oxford Centre for Evidence-based Medicine – Levels of Evidence (March 2009). [online] <http://www.cebm.net/index.aspx?o=1025> (Accessed 4 July 2016).

Cho, B., Oka, M., Matsumoto, N., Ouchida, R., Hong, J. and Hashizume, M. (2013) 'Warning navigation system using real-time safe region monitoring for otologic surgery', *International Journal of Computer Assisted Radiology and Surgery*, Vol. 8, No. 3, pp.395–405.

Cleary, K. and Peters, T.M. (2010) 'Image-guided interventions: technology review and clinical applications', *Annual Review of Biomedical Engineering*, Vol. 12, pp.119–142.

Dalgorf, D.M., Sacks, R., Wormald, P.J., Naidoo, Y., Panizza, B., Uren, B., Brown, C., Curotta, J., Snidvongs, K. and Harvey, R.J. (2013) 'Image-guided surgery influences perioperative morbidity from endoscopic sinus surgery: a systematic review and meta-analysis', *Otolaryngology – Head and Neck Surgery*, Vol. 149, No. 1, pp.17–29.

Davis, K.S., Vosler, P.S., Yu, J. and Wang, E.W. (2016) 'Intraoperative image guidance improves outcomes in complex orbital reconstruction by novice surgeons', *Journal of Oral and Maxillofacial Surgery*, Vol. 74, No. 7, pp.1410–1415.

DeLisi, M.P., Mawn, L.A. and Galloway Jr., R.L. (2014) 'Image-guided transorbital procedures with endoscopic video augmentation', *Medical Physics*, Vol. 41, No. 9, pp.091901-1–091901-9.

DeLisi, M.P., Mawn, L.A. and Galloway Jr., R.L. (2015) 'Transorbital target localization with augmented ophthalmologic surgical endoscopy', *International Journal of Computer Assisted Radiology and Surgery*, Vol. 10, No. 7, pp.1141–1148.

Dixon, B.J., Chan, H., Daly, M.J., Qiu, J., Vescan, A., Witterick, I.J. and Irish, J.C. (2016) 'Three-dimensional virtual navigation versus conventional image guidance: A randomized controlled trial', *The Laryngoscope*, Vol. 126, No. 7, pp.1510–1515.

Dixon, B.J., Chan, H., Daly, M.J., Vescan, A.D., Witterick, I.J. and Irish, J.C. (2012) 'The effect of augmented real-time image guidance on task workload during endoscopic sinus surgery', *International Forum of Allergy & Rhinology*, Vol. 2, No. 5, pp.405–410.

Dixon, B.J., Daly, M.J., Chan, H., Vescan, A., Witterick, I.J. and Irish, J.C. (2011) 'Augmented image guidance improves skull base navigation and reduces task workload in trainees: a preclinical trial', *The Laryngoscope*, Vol. 121, No. 10, pp.2060–2064.

Dixon, B.J., Daly, M.J., Chan, H., Vescan, A.D., Witterick, I.J. and Irish, J.C. (2013) 'Surgeons blinded by enhanced navigation: the effect of augmented reality on attention', *Surgical Endoscopy*, Vol. 27, No. 2, pp.454–461.

Dixon, B.J., Daly, M.J., Chan, H., Vescan, A., Witterick, I.J. and Irish, J.C. (2014a) 'Augmented real-time navigation with critical structure proximity alerts for endoscopic skull base surgery', *The Laryngoscope*, Vol. 124, No. 4, pp.853–859.

Dixon, B.J., Daly, M.J., Chan, H.H., Vescan, A., Witterick, I.J. and Irish, J.C. (2014b) 'Inattention blindness increased with augmented reality surgical navigation', *American Journal of Rhinology & Allergy*, Vol. 28, No. 5, pp.433–437.

- Dubin, M.R., Tabae, A., Scruggs, J.T., Kazim, M. and Close, L.G. (2008) 'Image-guided endoscopic orbital decompression for Graves' orbitopathy', *The Annals of Otolaryngology, Rhinology, and Laryngology*, Vol. 117, No. 3, pp.177–185.
- Eliashar, R., Sichel, J.Y., Gross, M., Hocwald, E., Dano, I., Biron, A., Ben-Yaacov, A., Goldfarb, A. and Elidan, J. (2003) 'Image guided navigation system—a new technology for complex endoscopic endonasal surgery', *Postgraduate Medical Journal*, Vol. 79, No. 938, pp.686–690.
- Endsley, M.R. (1995) 'Toward a theory of situation awareness in dynamic systems', *Human Factors*, Vol. 37, No. 1, pp.32–64.
- Ferris, T., Sarter, N. and Wickens, C.D. (2010) 'Cockpit automation: still struggling to catch up...', in Salas, E. and Maurino, D. (Eds), *Human Factors in Aviation*, Academic Press, Amsterdam, pp.479–504.
- Follmann, A., Korff, A., Fürtjes, T., Lauer, W., Kunze, S.C., Schmieder, K. and Radermacher, K. (2010) 'Evaluation of a synergistically controlled semiautomatic trepanation system for neurosurgery' in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Buenos Aires, Argentina, pp.2304–2307.
- Freysinger, W., Gunkel, A.R. and Thumfart, W.F. (1997) 'Image-guided endoscopic ENT surgery', *European Archives of Oto-Rhino-Laryngology*, Vol. 254, No. 7, pp.343–346.
- Gong, J., Mohr, G. and Vézina, J.L. (2007) 'Endoscopic pituitary surgery with and without image guidance: an experimental comparison', *Surgical Neurology*, Vol. 67, No. 6, pp.572–578.
- Haerle, S.K., Daly, M.J., Chan, H., Vescan, A., Witterick, I., Gentili, F., Zadeh, G., Kucharczyk, W. and Irish, J.C. (2015) 'Localized Intraoperative Virtual Endoscopy (LIVE) for surgical guidance in 16 skull base patients', *Otolaryngology – Head and Neck Surgery*, Vol. 152, No. 1, pp.165–171.
- Hassfeld, S. and Mühling, J. (2001) 'Computer assisted oral and maxillofacial surgery—a review and an assessment of technology', *International Journal of Oral and Maxillofacial Surgery*, Vol. 30, No. 1, pp.2–13.
- Hockey, G.R.J. (1997) 'Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework', *Biological Psychology*, Vol. 45, No. 1, pp.73–93.
- Hofer, M., Dittrich, E., Scholl, C., Neumuth, T., Strauss, M., Dietz, A., Lüth, T. and Strauss, G. (2008) 'First clinical evaluation of the navigated controlled drill at the lateral skull base', *Studies in Health Technology and Informatics*, Vol. 132, pp.171–173.
- Hong, J., Matsumoto, N., Ouchida, R., Komune, S. and Hashizume, M. (2009) 'Medical navigation system for otologic surgery based on hybrid registration and virtual intraoperative

computed tomography', *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 2, pp.426–432.

Ingram, M.C., Atuegwu, N., Mawn, L. and Galloway, R.L. (2011) 'Transorbital therapy delivery: phantom testing', in *Proceedings of SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling*, International Society for Optics and Photonics, Lake Buena Vista, USA, pp.79642A-1–79642A-7.

Kingdom, T.T. and Orlandi, R.R. (2004) 'Image-guided surgery of the sinuses: current technology and applications', *Otolaryngologic Clinics of North America*, Vol. 37, No. 2, pp.381–400.

Kosmopoulos, V. and Schizas, C. (2007) 'Pedicule screw placement accuracy: a meta-analysis', *Spine*, Vol. 32, No. 3, pp.E111–E120.

Labadie, R.F. and Fitzpatrick, J.M. (2011) 'System and method for surgical instrument disablement via image-guided position feedback', U.S. Patent No. 7,899,512.

Li, L., Yang, J., Chu, Y., Wu, W., Xue, J., Liang, P. and Chen, L. (2016) 'A novel augmented reality navigation system for endoscopic sinus and skull base surgery: A feasibility study', *PloS one*, Vol. 11, No. 1, e0146996, doi:10.1371/journal.pone.0146996.

Lim, H., Matsumoto, N., Cho, B., Hong, J., Yamashita, M., Hashizume, M. and Yi, B.J. (2016) 'Semi-manual mastoidectomy assisted by human–robot collaborative control–A temporal bone replica study', *Auris Nasus Larynx*, Vol. 43, No. 2, pp.161–165.

Luz, M., Manzey, D., Modemann, S. and Strauss, G. (2015) 'Less is sometimes more: a comparison of distance-control and navigated-control concepts of image-guided navigation support for surgeons', *Ergonomics*, Vol. 58, No. 3, pp.383–393.

Luz, M., Manzey, D., Mueller, S., Dietz, A., Meixensberger, J. and Strauss, G. (2014) 'Impact of navigated-control assistance on performance, workload and situation awareness of experienced surgeons performing a simulated mastoidectomy', *The International Journal of Medical Robotics and Computer Assisted Surgery*, Vol. 10, No. 2, pp.187–195.

Manzey, D., Luz, M., Mueller, S., Dietz, A., Meixensberger, J. and Strauss, G. (2011) 'Automation in surgery: The impact of navigated-control assistance on performance, workload, situation awareness, and acquisition of surgical skills'. *Human Factors*, Vol. 53, No. 6, pp.584–599.

Manzey, D., Röttger, S., Bahner-Heyne, J.E., Schulze-Kissing, D., Dietz, A., Meixensberger, J. and Strauss, G. (2009a) 'Image-guided navigation: the surgeon's perspective on performance consequences and human factors issues', *The International Journal of Medical Robotics and Computer Assisted Surgery*, Vol. 5, No. 3, pp.297–308.

- Manzey, D., Strauss, G., Trantakis, C., Lueth, T., Röttger, S., Bahner-Heyne, J.E., Dietz, A. and Meixensberger, J. (2009b) 'Automation in surgery: A systematic approach', *Surgical Technology International*, Vol. 18, pp.37–45.
- Marcus, H.J., Pratt, P., Hughes-Hallett, A., Cundy, T.P., Marcus, A.P., Yang, G.Z., Darzi, A. and Nandi, D. (2015) 'Comparative effectiveness and safety of image guidance systems in neurosurgery: a preclinical randomized study', *Journal of Neurosurgery*, Vol. 123, No. 2, pp.307–313.
- Markelj, P., Tomaževič, D., Likar, B. and Pernuš, F. (2012) 'A review of 3D/2D registration methods for image-guided interventions', *Medical Image Analysis*, Vol. 16, No. 3, pp.642–661.
- Marple, B.F. and Setzen, M. (2006) 'Image-guidance technology: What type of information best guides its appropriate use?' *Otolaryngology – Head and Neck Surgery*, Vol. 135, No. 1, pp.73–75.
- Metson, R. (2003) 'Image-guided sinus surgery: lessons learned from the first 1000 cases', *Otolaryngology – Head and Neck Surgery*, Vol. 128, No. 1, pp.8–13.
- Mueller, S.A. and Caversaccio, M. (2010) 'Outcome of computer-assisted surgery in patients with chronic rhinosinusitis' *The Journal of Laryngology & Otology*, Vol. 124, No. 5, pp.500–504.
- Onnasch, L., Wickens, C.D., Li, H. and Manzey, D. (2014) 'Human performance consequences of stages and levels of automation an integrated meta-analysis', *Human Factors*, Vol. 56, No. 3, pp.476–488.
- Parasuraman, R. and Manzey, D.H. (2010) 'Complacency and bias in human use of automation: An attentional integration', *Human Factors*, Vol. 52, No. 3, pp.381–410.
- Ramakrishnan, V.R., Orlandi, R.R., Citardi, M.J., Smith, T.L., Fried, M.P. and Kingdom, T.T. (2013) 'The use of image-guided surgery in endoscopic sinus surgery: an evidence-based review with recommendations', *International Forum of Allergy & Rhinology*, Vol. 3, No. 3, pp.236–241.
- Ruppin, J., Popovic, A., Strauss, M., Spüntrup, E., Steiner, A. and Stoll, C. (2008) 'Evaluation of the accuracy of three different computer-aided surgery systems in dental implantology: optical tracking vs. stereolithographic splint systems', *Clinical Oral Implants Research*, Vol. 19, No. 7, pp.709–716.
- Shin, B.J., James, A.R., Njoku, I.U. and Härtl, R. (2012) 'Pedicule screw navigation: a systematic review and meta-analysis of perforation risk for computer-navigated versus freehand insertion: A review', *Journal of Neurosurgery: Spine*, Vol. 17, No. 2, pp.113–122.
- Simpson, A.L., Ma, B., Vasarhelyi, E.M., Borschneck, D.P., Ellis, R.E. and Stewart, A.J. (2014) 'Computation and visualization of uncertainty in surgical navigation' *The*

International Journal of Medical Robotics and Computer Assisted Surgery, Vol. 10, No. 3, pp.332–343.

Sindwani, R. (2008) ‘Image-guided surgery of the paranasal sinuses and skull base’, *Missouri Medicine*, Vol. 105, No. 3, pp.257–261.

Smith, T.L., Stewart, M.G., Orlandi, R.R., Setzen, M. and Lanza, D.C. (2007) ‘Indications for image-guided sinus surgery: the current evidence’, *American Journal of Rhinology*, Vol. 21, No. 1, pp.80–83.

Stelter, K., Ertl-Wagner, B., Luz, M., Muller, S., Ledderose, G., Siedek, V., Berghaus, A., Arpe, S. and Leunig, A. (2011) ‘Evaluation of an image-guided navigation system in the training of functional endoscopic sinus surgeons. A prospective, randomised clinical study’, *Rhinology*, Vol. 49, No. 4, pp.429–437.

Stelter, K., Theodoraki, M.N., Becker, S., Tsekmistrenko, V., Olzowy, B. and Ledderose, G. (2015) ‘Specific stressors in endonasal skull base surgery with and without navigation’, *European Archives of Oto-Rhino-Laryngology*, Vol. 272, No. 3, pp.631–638.

Strauss, G., Koulechov, K., Hofer, M., Dittrich, E., Grunert, R., Moeckel, H., Müller, E., Korb, W., Trantakis, C., Schulz, T., Meixensberger, J., Dietz, A. and Lueth, T. (2007) ‘The navigation-controlled drill in temporal bone surgery: A feasibility study’, *The Laryngoscope*, Vol. 117, No. 3, pp.434–441.

Strauss, G., Koulechov, K., Richter, R., Dietz, A., Trantakis, C. and Lüth, T. (2005) ‘Navigated control in functional endoscopic sinus surgery’, *The International Journal of Medical Robotics and Computer Assisted Surgery*, Vol. 1, No. 3, pp.31–41.

Strauss, G., Koulechov, K., Röttger, S., Bahner, J., Trantakis, C., Hofer, M., Korb, W., Burgert, O., Meixensberger, J., Manzey, D., Dietz, A. and Lüth, T. (2006) ‘Evaluation of a navigation system for ENT with surgical efficiency criteria’, *The Laryngoscope*, Vol. 116, No. 4, pp.564–572.

Sunkaraneni, V.S., Yeh, D., Qian, H. and Javer, A.R. (2013) ‘Computer or not? Use of image guidance during endoscopic sinus surgery for chronic rhinosinusitis at St Paul's Hospital, Vancouver, and meta-analysis’, *The Journal of Laryngology & Otology*, Vol. 127, No. 4, pp.368–377.

Tabaee, A., Kacker, A., Kassenoff, T.L. and Anand, V. (2003) ‘Outcome of computer-assisted sinus surgery: a 5-year study’, *American Journal of Rhinology*, Vol. 17, No. 5, pp.291–297.

Theodoraki, M.N., Ledderose, G.J., Becker, S., Leunig, A., Arpe, S., Luz, M. and Stelter, K. (2015) ‘Mental distress and effort to engage an image-guided navigation system in the surgical training of endoscopic sinus surgery: a prospective, randomised clinical trial’, *European Archives of Oto-Rhino-Laryngology*, Vol. 272, No. 4, pp.905–913.

- Tian, N.F., Huang, Q.S., Zhou, P., Zhou, Y., Wu, R.K., Lou, Y. and Xu, H.Z. (2011) 'Pedicle screw insertion accuracy with different assisted methods: a systematic review and meta-analysis of comparative studies', *European Spine Journal*, Vol. 20, No. 6, pp.846–859.
- Voormolen, E.H.J., Woerdeman, P.A., van Stralen, M., Noordmans, H.J., Viergever, M.A., Regli, L. and van der Sprenkel, J.W.B. (2012) 'Validation of exposure visualization and audible distance emission for navigated temporal bone drilling in phantoms', *PloS one*, Vol. 7, No. 7, p.e41262, doi:10.1371/journal.pone.0041262.
- Vreugdenburg, T.D., Lambert, R.S., Atukorale, Y.N. and Cameron, A.L. (2016) 'Stereotactic anatomical localization in complex sinus surgery: A systematic review and meta-analysis', *The Laryngoscope*, Vol. 126, No. 1, pp.51–59.
- Widmann, G. and Bale, R.J. (2006) 'Accuracy in computer-aided implant surgery – a review', *International Journal of Oral & Maxillofacial Implants*, Vol. 21, No. 2, pp.305–313.
- Willems, P.W.A., Noordmans, H.J., van Overbeeke, J.J., Viergever, M.A., Tulleken, C.A.F. and van der Sprenkel, J.B. (2005) 'The impact of auditory feedback on neuronavigation', *Acta Neurochirurgica*, Vol. 147, No. 2, pp.167–173.
- Wise, S.K., Harvey, R.J., Goddard, J.C., Sheahan, P.O. and Schlosser, R.J. (2008) 'Combined image guidance and intraoperative computed tomography in facilitating endoscopic orientation within and around the paranasal sinuses', *American Journal of Rhinology*, Vol. 22, No. 6, pp.635–641.
- Woerdeman, P.A., Willems, P.W., Noordmans, H.J., Tulleken, C.A. and Van der Sprenkel, J.W. (2009a) 'The impact of workflow and volumetric feedback on frameless image-guided neurosurgery', *Neurosurgery*, Vol. 64, No. 3, pp.ons170–ons176.
- Woerdeman, P.A., Willems, P.W., Noordmans, H.J. and van der Sprenkel, J.W.B. (2009b) 'Auditory feedback during frameless image-guided surgery in a phantom model and initial clinical experience: Technical note', *Journal of Neurosurgery*, Vol. 110, No. 2, pp.257–262.
- World Health Organization*. [online] <http://www.who.int/patientsafety/about/en/> (Accessed 4 July 2016).
- Zuckerman, J.D., Wise, S.K., Rogers, G.A., Senior, B.A., Schlosser, R.J. and DelGaudio, J.M. (2009) 'The utility of cadaver dissection in endoscopic sinus surgery training courses', *American Journal of Rhinology & Allergy*, Vol. 23, No. 2, pp.218–224.

Figure 1 Relationship between IGS system, patient, surgeon, and clinical outcomes. A surgeon uses IGS dependent on their attitude, experience, expectations, and skills. In turn, IGS used by the surgeon influences their situation awareness, workload, and skill acquisition and loss (gray upper box). This surgeon-system-interaction, as well as anatomical characteristics of a patient, have impact on clinical outcomes (gray lower box)

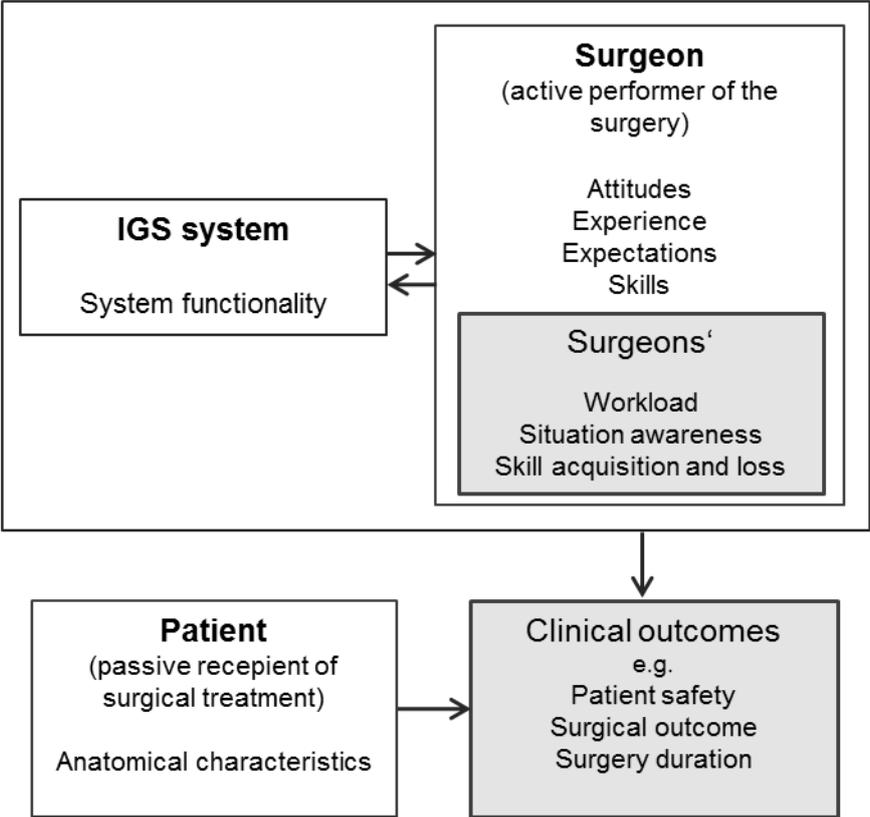


Table 1 Classification of IGS system functionalities

	IGS system which provides information support		IGS system which provides decision and action support	
<i>Basic functionalities</i>				
Routinely used in operating room	PB-IGS IN-IGS	pointer-based systems instrument navigation		
<i>Advanced functionalities</i>				
	PV-IGS UV-IGS	process visualization uncertainty visualization	STS-IGS	semiautomatic trepanation system
Preoperative segmentation of risk structures or target required	DV-IGS PW-IGS ARS-IGS ATV-IGS 3DV-IGS	distance visualization proximity warnings augmented risk structures augmented target volume three-dimensional virtual image guidance	ID-IGS MR-IGS	instrument disablement movement restriction
Preoperative target indication required	AT-IGS	augmented target		
Preoperative planning required	AP-IGS	augmented pathway		

Table 2 Characteristics of included studies

Study	Comparison	App	SE	n	Experience	Comments
Alobid et al. (2011)	PB vs. non	PS/ASB	p	15	R	5 non-surgery ^a , 5 IGS, 5 non-IGS
Casiano and Numa (2000)	PB vs. non	PS/ASB	c	4	R	
Cho et al. (2013)	DV/PW vs. non	TB/LSB	s	1	U	
Davis et al. (2016)	PB vs. non	OS	c	10	R, F	
DeLisi et al. (2014)	IN vs. AT	OS	s	16	Different	
DeLisi et al. (2015)	IN vs. AT	OS	a	1	E	
Dixon et al. (2011)	ARS vs. non	PS/ASB	c	12	R, F	
Dixon et al. (2012)	PW/ARS vs. non	PS/ASB	c	8	R, F	
Dixon et al. (2013)	ARS vs. non	PS/ASB	c	32	E, U	17 non-IGS, 15 IGS
Dixon et al. (2014a)	IN vs. PW/3DV	PS/ASB	c	7	F	
Dixon et al. (2014b)	ARS vs. ARS+	PS/ASB	c	50	E, U	
Dixon et al. (2016)	PB vs. 3DV	PS/ASB	c	37	E, U	
Follmann et al. (2010)	STS vs. non	NS	s	6	E	
Gong et al. (2007)	PB vs. non	PS/ASB	c	n.i.	n.i.	
Haerle et al. (2015)	PW/ARS vs. non	PS/ASB	p	4	E	
Hofer et al. (2008)	ID vs. non	TB/LSB	p	n.i.	n.i.	
Ingram et al. (2011)	IN vs. non	OS	s	28	Different	
Li et al. (2016)	PB vs. ARS	PS/ASB	c	15	E, U	
Lim et al. (2016)	MR vs. non	TB/LSB	s	1	U	Non-physician
Luz et al. (2014)	ID vs. non	TB/LSB	s	7	F	
Luz et al. (2015)*	DV/PW, DV/PW/ID vs. non DV/PW vs. DV/PW/ID	TB/LSB	s	18	R	
Manzey et al. (2011)	ID vs. non	TB/LSB	s	14	MS	
	ID vs. non	TB/LSB	s	21	MS	10 non-IGS, 11 IGS
Marcus et al. (2015)*	PB, ARS vs. non	NS	s	50	MS, R	10 in each study group
	PB vs. ARS			40	MS, R	10 in each study group
Simpson et al. (2014)	IN vs. UV	SS	s	13	Different	
Stelter et al. (2011)	PB vs. non	PS/ASB	p	8	R	
Stelter et al. (2015)	PB vs. non	PS/ASB	p	4	E	
Strauss et al. (2006)	PB vs. non	PS/ASB	p	7	E, U	n only for IGS group
Strauss et al. (2007)	ID vs. non	TB/LSB	s	5	E	
Theodoraki et al. (2015)	PB vs. non	PS/ASB	p	8	R	Same participants as Stelter 2011
Voormolen et al. (2012)	IN vs. PV/DV/PW	TB/LSB	s	5	E, U	
Willems et al. (2005)	IN vs. PW	NS	s	3	n.i.	
Wise et al. (2008)*	IN, PV ^b vs. non IN vs. PV ^b	PS/ASB	c	11	R	
Woerdeman et al. (2009b)	IN vs. PW vs. ATV	NS	s	4	n.i.	
	IN vs. PW	NS	p	n.i.	n.i.	
Zuckerman et al. (2009)	PB vs. non	PS/ASB	c	29/22	Different	2 day course

Abbreviations. **Comparison:** non, non-IGS; PB, pointer-based; IN, instrument navigation; PV, process visualization; UV, uncertainty visualization; DV, distance visualization; PW, proximity warning; ARS, augmented risk structures; ARS+, augmented risk structures and conventional endoscopic video display; ATV, augmented target volume; AT, augmented target; 3DV, three-dimensional virtual image guidance; ID, instrument disablement; MR, movement restriction; STS, semiautomatic trepanation system. **App (Application):** PS/ASB, paranasal sinuses and anterior skull base; TB/LSB, temporal bone and lateral skull base; OS, orbital surgery; NS, neurosurgery; SS, spinal surgery. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation); a, animal. **n (sample size/number of participating surgeons):** n.i., not indicated. **Experience:** R, residents; F, fellows; E, experienced; U, unexperienced; MS, medical students; n.i., not indicated. **Comments:** ^a non-surgery – participants were evaluated on non-surgical day (control day); ^b process visualization through intraoperative CT; * studies which compare both IGS with conventional, i.e. unsupported surgery, and different IGS system functionalities.

Table 3 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on patient safety and surgical outcome

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Casiano and Numa (2000)	PB	c	Number of complications	o	p	non > IGS	+
			Incidence of serious complications	s	s	non = IGS	0
			Precision of tissue removal	s	s	non < IGS	+
Davis et al. (2016)	PB	c	Orbital volume	o	s	non = IGS*	0
Stelter et al. (2011)	PB	p	Intraoperative complications	o	p	non = IGS	0
			Patient safety	s	p	non < IGS	+
			Missed paranasal sinuses	o	p	non > IGS	+
Zuckerman et al. (2009)	PB	c	Number of unopened ethmoid cells	o	s	non = IGS*	0
			Number of retained ethmoid partitions	o	s	non = IGS*	0
			Number of unopened frontal recess cells	o	s	non = IGS*	0
Cho et al. (2013)	DV/PW	s	Facial nerve injury	o	p	non > IGS	+
			Maintenance of safe margin	o	p	non < IGS*	+
Luz et al. (2015)	DV/PW, DV/PW/ID	s	Possible complications	o	s	non > IGS	+
			Injuries of risk structures	o	s	non = IGS	0
			Quality of surgery	o	s	non = IGS*	0
Hofer et al. (2008)	ID	p	Risk structure damage	o	p	non = IGS	0
Luz et al. (2014)	ID	s	Possible complications	o	s	non > IGS	+
			Injuries of risk structures	o	s	non < IGS	-
			Quality of surgery	o	s	non < IGS*	+
Manzey et al. (2011)	ID	s	Possible complications	o	s	non > IGS*	+
			Injuries of risk structures	o	s	non > IGS	+
			Quality of surgery	o	s	non = IGS*	0
Strauss et al. (2007)	ID	s	Injuries of risk structures	o	s	non > IGS	+
			Precision in execution regarding planned volume	o	s	non < IGS	+
Lim et al. (2016)	MR	s	Facial nerve injuries	o	p	non > IGS	+
Follmann et al. (2010)	STS		Resection accuracy	o	s	non > IGS	-
			Cutting gap ^a	o		non > IGS	+

Abbreviations. **IGS:** PB, pointer-based; DV, distance visualization; PW, proximity warning; ID, instrument disablement; MR, movement restriction; STS, semiautomatic trepanation system. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related. **Result:** *, analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS. **Comment:** ^a reported as example.

Table 4 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on surgery duration

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Casiano and Numa (2000)	PB	c	Surgery duration	o	p	non = IGS*	0
Davis et al. (2016)	PB	c	Surgery duration	o	s	non = IGS*	0
Gong et al. (2007)	PB	c	Duration of ostium sphenoidale exposure	o	p	non = IGS*	0 ^a
			Duration of ostium sphenoidale exposure	o	p	non > IGS*	+ ^b
			Duration of sellar floor bone window creation	o	p	non = IGS*	0 ^a
			Duration of sellar floor bone window creation	o	p	non > IGS*	+ ^b
Stelter et al. (2011)	PB	p	Surgery duration	o	p	non = IGS	0
Strauss et al. (2006)	PB	p	Preparation time	o	p	non < IGS	-
			Surgery duration	o	p	non > IGS	+
Dixon et al. (2012)	PW/ARS	c	Surgery speed	s	s	non ≤ IGS	+
Haerle et al. (2015)	PW/ARS	p	Surgery speed	s	p	non < IGS	+
Luz et al. (2015)	DV/PW, DV/PW/ID	s	Surgery duration	o	s	non < IGS*	-
Hofer et al. (2008)	ID	p	Surgery duration	o	p	non = IGS	0
Luz et al. (2014)	ID	s	Surgery duration	o	s	non < IGS*	-
Manzey et al. (2011)	ID	s	Surgery duration	o	s	non < IGS*	-
Strauss et al. (2007)	ID	s	Surgery duration	o	s	non < IGS	-
			Drilling speed	o	s	non < IGS	+
Follmann et al. (2010)	STS	s	Surgery duration	o	s	non < IGS	-

Abbreviations. **IGS:** PB, pointer-based; DV, distance visualization; PW, proximity warning; ARS, augmented risk structures; ID, instrument disablement; STS, semiautomatic trepanation system. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related. **Result:** *, analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS. **Comments:** ^a normal anatomy, ^b anatomical variation.

Table 5 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on situation awareness

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Casiano and Numa (2000)	PB	c	Correctly identified landmarks	o	r	non < IGS*	+
			Ability to identify landmarks	s	s	non < IGS	+
Ingram et al. (2011)	IN	s	Accuracy of finding the target	o	r	non < IGS	+
			Time-to-target	o	r	non > IGS	-
Stelter et al. (2011)	PB	p	Global subjective assessment of intraoperative situation awareness	s	p	non < IGS	+
Wise et al. (2008)	IN, PV	c	Identification of anatomic site	o	s	non < IGS*	+ ^a
Marcus et al. (2015)	PB, ARS	s	Duration of landmark localization	o	s	non > IGS*	+
			Tool path length	o	s	non > IGS*	+
			Recognition of unexpected findings	o	s	non > IGS*	-
Dixon et al. (2011)	ARS	c	Localization precision	o	s	non < IGS	+
			Aid in localization	s	r	non < IGS	+
			Confidence in localization	s	r	non < IGS	+
Dixon et al. (2013)	ARS	c	Accuracy of landmark localization	o	s	non < IGS*	+
			Duration of landmark localization	o	s	non = IGS*	0
			Recognition of unexpected findings	o	s	non > IGS*	-
Dixon et al. (2012)	PW/ARS	c	Assessment of proximity to critical structures	s	s	non < IGS	+
Haerle et al. (2015)	PW/ARS	p	Assessment of proximity to critical structures	s	p	non < IGS	+
			Assessment of degree of resection	s	p	non = IGS	0
Luz et al. (2015)	DV/PW, DV/PW/ID	s	Assessment of achievement of surgical steps	o	s	non = IGS*	0
			Assessment of anatomical characteristics	o	s	non = IGS*	0
			Assessment of distances to structures	o	s	non = IGS*	0
			Assessment of complications	o	s	non = IGS*	0
			Assessment of remaining time	o	s	non = IGS*	0
Luz et al. (2014)	ID	s	Assessment of achievement of surgical steps	o	s	non = IGS*	0
			Assessment of anatomical characteristics	o	s	non = IGS*	0
			Assessment of distances to structures	o	s	non = IGS*	0
			Assessment of complications	o	s	non = IGS*	0
			Assessment of remaining time	o	s	non = IGS*	0
Manzey et al. (2011)	ID	s	Assessment of achievement of surgical steps	o	s	non = IGS*	0
			Assessment of anatomical characteristics	o	s	non > IGS*	-
			Assessment of distances to structures	o	s	non = IGS*	0
			Assessment of complications	o	s	non = IGS*	0
			Assessment of remaining time	o	s	non = IGS*	0

Abbreviations. **IGS:** PB, pointer-based; IN, instrument navigation; PV, process visualization; DV, distance visualization; PW, proximity warning; ARS, augmented risk structures; ID, instrument disablement. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related; r, run-related. **Result:*** analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS. **Comment:** ^a based on global assessment of four conditions.

Table 6 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on subjective workload

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Davis et al. (2016)	PB	c	Workload (overall)	s	s	non = IGS*	0
Stelter et al. (2011)	PB	p	Workload	s	p	non = IGS	0
Strauss et al. (2006)	PB	p	Workload	s	p	non < IGS	+
Dixon et al. (2011)	ARS	c	Mental demand	s	s	non > IGS*	+
			Physical demand	s	s	non = IGS*	0
			Temporal demand	s	s	non = IGS*	0
			Performance	s	s	non > IGS*	+
			Effort	s	s	non > IGS*	+
			Frustration	s	s	non > IGS*	+
Dixon et al. (2012)	PW/ARS	c	Mental demand	s	s	non > IGS*	+
			Physical demand	s	s	non = IGS*	0
			Temporal demand	s	s	non > IGS*	+
			Performance	s	s	non > IGS*	+
			Effort	s	s	non > IGS*	+
			Frustration	s	s	non > IGS*	+
Haerle et al. (2015)	PW/ARS	p	Mental demand	s	p	non = IGS*	0
			Physical demand	s	p	non = IGS*	0
			Temporal demand	s	p	non = IGS*	0
			Performance	s	p	non = IGS*	0
			Effort	s	p	non = IGS*	0
			Frustration	s	p	non = IGS*	0
Luz et al. (2015)	DV/PW, DV/PW/ID	s	Workload (overall)	s	s	non = IGS*	0
Hofer et al. (2008)	ID	p	Cognitive strain	s	p	non > IGS	+
Luz et al. (2014)	ID	s	Workload (overall)	s	s	non < IGS*	-
			Mental demand	s	s	non = IGS*	0
			Physical demand	s	s	non = IGS*	0
			Temporal demand	s	s	non = IGS*	0
			Performance	s	s	non = IGS*	0
			Effort	s	s	non = IGS*	0
			Frustration	s	s	non < IGS*	-
Manzey et al. (2011)	ID	s	Workload (overall)	s	s	non < IGS*	-
			Mental demand	s	s	non = IGS*	0
			Physical demand	s	s	non = IGS*	0
			Temporal demand	s	s	non = IGS*	0
			Performance	s	s	non = IGS*	0
			Effort	s	s	non = IGS*	0
			Frustration	s	s	non < IGS*	-
Follmann et al. (2010)	STS	s	Mental demand	s	s	non < IGS	-
			Physical demand	s	s	non < IGS	-
			Temporal demand	s	s	non = IGS	0
			Performance	s	s	non < IGS	-
			Effort	s	s	non < IGS	-
			Frustration	s	s	non < IGS	-

Abbreviations. **IGS:** PB, pointer-based; DV, distance visualization; PW, proximity warning; ARS, augmented risk structures; ID, instrument disablement; STS, semiautomatic trepanation system. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related; **Result:** *, analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS.

Table 7 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on physiological effort

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Alobid et al. (2011)	PB	p	Systolic blood pressure	o	s	non = IGS*	0
			Diastolic blood pressure	o	s	non = IGS*	0
			Heart rate	o	s	non = IGS*	0
			Plasma cortisol	o	s	non = IGS*	0
			Plasma prolactin	o	s	non = IGS*	0
Stelter et al. (2015)	PB	p	Heart rate	o	p	non = IGS*	0
			Heart rate variability	o	p	non = IGS*	0
Theodoraki et al. (2015)	PB	p	Heart rate	o	s	non = IGS*	0
			Heart rate variability	o	s	non = IGS*	0
Luz et al. (2015)	DV/PW,	s	Heart rate	o	s	non = IGS*	0
	DV/PW/ID		Respiration rate	o	s	non = IGS*	0
			MF/HF ratio	o	s	non = IGS*	0
			Systolic blood pressure	o	s	non = IGS*	0
Luz et al. (2014)	ID	s	Heart rate	o	s	non = IGS*	0
			Respiration rate	o	s	non = IGS*	0
			HRV, high-frequency	o	s	non < IGS*	+
			HRV, low-frequency	o	s	non < IGS*	+
Manzey et al. (2011)	ID	s	Systolic blood pressure	o	s	non > IGS*	+
			Heart rate	o	s	non > IGS*	+
			Respiration rate	o	s	non > IGS*	+
			MF/HF ratio	o	s	non = IGS*	0
			Systolic blood pressure	o	s	non = IGS*	0

Abbreviations. **IGS:** PB, pointer-based; DV, distance visualization; PW, proximity warning; ID, instrument disablement. **SE (study environment):** p, patients; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related. **Result:***, analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS.

Table 8 Overview of the studies, variables considered and a qualitative evaluation of effects of IGS on skill acquisition and maintenance

Study	IGS	SE	Variables	S/O	DT	Result	Impact
Casiano and Numa (2000)	PB	p	Endoscopic surgical skills	s	s	non < IGS	+
			Educational benefit	s	s	non < IGS	+
Stelter et al. (2011)	PB	p	Loss of own skills	s	p	non = IGS	0
			Novices' skill acquisition	s	p	non > IGS	- ^a
Manzey et al. (2011)	ID	s	Novices' skill acquisition	o	s	non = IGS*	0 ^b

Abbreviations. **IGS:** PB, pointer-based; ID, instrument disablement. **SE (study environment):** p, patients; s, phantoms (simulation). **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-related. **Result:** *, analyzed by means of statistical test. **Impact:** +/-/0, positive/negative/neutral impact of IGS. **Comments:** ^a opinion of more experienced surgeons, ^b several variables.

Table 9 Overview of the studies, variables considered and a qualitative evaluation of effects of different IGS system functionalities on surgeons' performance. Impact is related to more advanced IGS functionality

Study	SE	Variables	S/O	DT	Result	Impact		
Wise et al. (2008)	c	Identification of anatomic site	o	s	IN < PV*	+ ^a		
Simpson et al. (2014)	s	Insertion precision	o	r	IN = UV	0		
		Degree of breach of pedicle	o	r	IN = UV	0		
		Time to insert screw	o	r	IN < UV*	+ ^b		
Willems et al. (2005)	s	Resection precision	o	s/r	IN < PW	+		
Voormolen et al. (2012)	s	Damages of risk structures	o	s	IN > PV/DV/PW	+		
		Surgery duration	o	s	IN = PV/DV/PW	0		
		Surgical orientation	s	s	IN < PV/DV/PW	+		
Dixon et al. (2014a)	c	Surgery speed	s	s	IN < PW/3DV	+		
		Mental demand	s	s	IN > PW/3DV *	+		
		Physical demand	s	s	IN = PW/3DV *	0		
		Temporal demand	s	s	IN = PW/3DV *	0		
		Performance	s	s	IN = PW/3DV *	0		
		Effort	s	s	IN > PW/3DV *	+		
		Frustration	s	s	IN > PW/3DV *	+		
		Assessment of proximity to critical structures	s	s	IN < ARS/PW	+		
		Woerdeman et al. (2009b)	s	Surgery duration	o	s	IN = PW = ATV*	0
				Resection quality	s	s	PW > IN > ATV	-
Resection precision	o			s	IN = PW > ATV*	-		
DeLisi et al. (2014)	s	Target identification accuracy	o	s	IN < AT	+		
		Procedure time	o	s	IN > AT*	+		
DeLisi et al. (2015)	a	Target identification accuracy	o	p	IN = AT	0		
		Procedure time	o	p	IN > AT*	+		
Li et al. (2016)	c	Surgery duration	o	s	PB > ARS*	+		
		Subjective workload	s	s	PB > ARS*	+		
Marcus et al. (2015)	s	Duration of landmark localization	o	s	PB = ARS*	0		
		Tool path length	o	s	PB > ARS*	+		
		Recognition of unexpected findings	o	s	PB = ARS*	0		
Dixon et al. (2014b)	c	Accuracy of landmark localization	o	s	ARS+ = ARS*	0		
		Duration of landmark localization	o	s	ARS+ = ARS*	0		
		Recognition of unexpected findings	o	s	ARS+ > ARS*	-		
Dixon et al. (2016)	c	Accuracy of landmark localization	o	s	PB = 3DV	0		
		Duration of landmark localization	o	s	PB = 3DV	0		
		Workload	s	s	PB = 3DV	0		
Luz et al. (2015)	s	Possible complications	o	s	DV/PW = DV/PW/ID*	0		
		Injuries of risk structures	o	s	DV/PW = DV/PW/ID	0		
		Possible complications	o	s	DV/PW = DV/PW/ID	0		
		Surgery duration	o	s	DV/PW < DV/PW/ID*	-		
		Subjective workload	s	s	DV/PW < DV/PW/ID*	-		
		Spare capacity	o	s	DV/PW = DV/PW/ID*	0		
		Physiological effort	o	s	DV/PW = DV/PW/ID*	0		
Situation awareness	o	s	DV/PW = DV/PW/ID*	0				

Abbreviations. **IGS:** PB, pointer-based; IN, instrument navigation; PV, process visualization; UV, uncertainty visualization; DV, distance visualization; PW, proximity warning; ARS, augmented risk structures; ARS+, augmented risk structures and conventional endoscopic video display; ATV, augmented target volume; AT, augmented target; 3DV, three-dimensional virtual image guidance; ID, instrument disablement. **SE (study environment):** p, patients; c, cadavers; s, phantoms (simulation); a, animals. **S/O:** s, subjective data; o, objective data. **DT (data treatment):** p, patient-related; s, surgeon-

related; r, run-related. **Result:** *, analyzed by means of statistical test. **Impact:** +/-0, positive/negative/neutral impact of IGS related to more advanced IGS functionality. **Comments:** ^a global assessment of four conditions by means of statistical test, ^b only specific visualization.