Automation in Surgery: The Surgeons' Perspective on Human Factors Issues of Image-Guided Navigation

J.E. Bahner-Heyne[a], S. Roettger[a], D. Schulze-Kissing[b] and D. Manzey[a]
[a]Berlin Institute of Technology, Berlin, Germany
[b]German Aerospace Center, Hamburg, Germany

Image-guided navigation (IGN) systems support the surgeon in navigating through the patients' anatomy. Previous research on IGN has focused on technical feasibility and clinical applications. Yet, as the introduction of IGN corresponds to a partial automation of the surgeon's task, well known issues of human-automation interaction might play a crucial role for the success of IGN as well. The present study represents a first attempt to assess the impact of IGN on four key issues of human automation-interaction, i.e., workload, situation awareness, trust, and skill degradation, from the surgeons' perspective. A nation-wide survey among 213 German surgeons from 94 different hospitals was conducted. Results revealed (1) a workload-shift due to IGN rather than a reduction of workload, (2) benefits of IGN with respect to situation awareness, (3) comparatively high levels of perceived reliability, trust and reliance, and (4) skill degradation as a possible risk, albeit only for inexperienced surgeons.

INTRODUCTION

Technological advances continue to revolutionize medical care by enabling earlier diagnoses, safer treatments and more and more sophisticated surgical interventions. Enhancements of minimally invasive surgery allow for access to difficult-to-reach anatomy, shorter hospital stays, and less pain. Yet, the trade-off for these minimal invasive approaches involves decreased visibility and impaired spatial orientation for the surgeon. This is especially the case if two-dimensional endoscopic images without any depth cues serve as the main source of information for navigating through the patient's anatomy. As a consequence, the surgeon has to rely on his general anatomical knowledge and also a little bit of guesswork to determine the location of vital structures. Yet, the development of image-guided navigation (IGN) systems bears the potential to facilitate this demanding task essentially. IGN enables the surgeon during ongoing surgery to pinpoint the position of surgical instruments in relation to the patient's anatomy. Core elements of IGN systems include a registration device to align the patient's current location with preoperative images (e.g., based on computer tomography) and a tracking device (based on an electromagnetic or optical camera) to allow for an intra-operative tracking of surgical instruments as well as the patient's position. The actual position of the tip of the instrument is then displayed on a screen in relation to the patient's anatomy. Whenever there emerges any uncertainty about the current localization, the surgeon just needs to inspect this screen. Hence, from a human factors point of view IGN can be referred to as automation of one of the main tasks of the surgeon: The spatial localization of the surgical instrument within the three-dimensional space of the patient's anatomy is delegated partially to a computer-based system (Manzey, Strauss, Trantakis, Lueth, Roettger, Bahner-Heyne et al., in press). According to the framework model of Parasuraman, Sheridan, and Wickens (2000), IGN systems can be classified as a low to medium level of information automation: The systems support to some degree the information acquisition and analysis of visual information by providing additional information to localize the surgical instrument. Yet, it is still completely up to the surgeon to draw any decisions from this information and to take any appropriate surgical actions (Strauss et al., 2006). One of the most prominent areas where IGN has been applied includes functional endoscopic sinus surgery and other skull base surgeries (e.g., Caversaccio, Nolte, & Haeusler, 2002). Although IGN is already in clinical use today, navigation technology is continuously being enhanced. Hence, it is not surprising that previous research on IGN has focused on technological issues and basic questions of the clinical use of IGN, as for example general feasibility questions (Brown, Sadoughi, Cuellar, von Jako, & Fried, 2007), the accuracy of the registration (Labadie, Davis, & Fitzpatrick, 2005) the suitability of IGN for different surgical interventions (Reijnders et al., 2007) or the benefits of these systems with respect to surgical time, precision or postoperative complications (Gong, Mohr, & Vézina, 2007; Strauss et al., 2006). However, human factors issues which are associated with the underlying process of automation have not been considered so far. Many years of experience with automation in other domains (e.g., aviation and process control) as well as more than three decades of research on human-automation interaction have demonstrated that automation does not only provide performance benefits but also may involve cost effects. More specifically, four aspects have been identified as key issues of human-automation interaction (Parasuraman et al., 2000), and, it can be expected that these also play a crucial role for the successful introduction of IGN in surgery:

(1) Mental workload. A beneficial effect which is usually intended by the introduction of automation is the reduction of workload for the human operator. Yet, the unloading of the human by providing support may be counterbalanced by effects related to the operation of a new tool (Roettger, Bali, & Manzey, in press). In this case, the use of an automated system would only lead to a workload-shift instead of a workload reduction. Examples from aviation show that sometimes the operation of automation is so complex that it even leads to an overall increase of workload ("clumsy automation", Wiener,

1989). Furthermore, there exist hints that using IGN represents an additional and time-consuming task (Metson, Cosenza, Gliklich, & Montgomery, 1999). Thus, effects on workload need to be carefully considered when implementing IGN.

(2) Situation awareness. Endsley (1995) defined this concept 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" (p. 36). Transferred to the surgeon's task, "surgical situation awareness" might comprise how well s/he perceives and understands all relevant cues needed to assess the current status of the surgical procedure, and how well appropriate predictions can be derived about the further dynamic development of the operation. IGN systems may support this awareness by providing additional spatial information, albeit this has not been investigated empirically, so far.

(3) Trust and reliance. A high reliability of an automated system is usually mirrored by a high level of trust on the part of the user (Muir & Moray, 1996). However, as automation is always fallible, it has to be taken into account that people might overly rely on such devices before rushing to automate medical procedures (Vicente, 2003). Due to a very high level of trust, the human might neglect an appropriate monitoring of the automation and therefore fail to detect occurring automation failures. This effect has been termed "automation induced complacency" (Parasuraman, Molloy, & Singh, 1993). Another effect of over-trust which has been referred to as "automation bias" (Mosier & Skitka, 1996) is reflected in following the information provided by the system even if contradicting information is available from other sources (e.g., the endoscopic image). Accordingly, besides the perceived reliability of IGN and trust in these systems, critical questions are to what extent surgeons still perform cross-checks when supported by IGN, whether possible malfunctions (e.g., data base from the wrong patient) are detected in due time, and whether there is a risk of automation bias.

(4) Degradation of skills. Delegating tasks to an automated system puts the user from an active into a more passive role and on-the-job training is usually considerably reduced. As a possible consequence, the human operator might lose the skills to perform the supported task manually (Wickens & Hollands, 2000). Hence, with respect to the use of IGN systems one important question regards whether or not the users of IGN still possess enough knowledge of the relevant anatomical structures and have enough navigation skills to be able to detect possible failures of IGN systems. A second question involves whether it will still be possible to smoothly convert to conventional (i.e., manual) surgical techniques in case that the IGN system is not available.

The present study represents a first attempt to assess these four issues of human-automation interaction with respect to the use of IGN systems. For this purpose, a nation-wide survey study was performed in Germany. In this survey surgeons familiar with computer-assisted surgeries were asked to evaluate the IGN system they use regarding the issues outlined above.

METHOD

Questionnaire

A specific questionnaire (Human Factors Evaluation Questionnaire for Computer-Assisted-Surgery-Systems; HFEQ-CASS) was developed for the survey. The HFEQ-CASS includes a total of 48 items and can be answered within 10 to 15 minutes. Besides the four key issues of human-automation interaction, the questionnaire also addresses the usability of the IGN system, several other perceived performance consequences, and background information on clinical experience, which are not subject of this contribution. Possible costs and benefits of IGN directly associated with human-automation-interaction are covered in the following way by the questionnaire:

(1) Workload is assessed with respect to the five different dimensions of mental demands, physical demands, time pressure, effort, and frustration and stress. The surgeons are requested to indicate for each of these dimensions on a 5-point rating scale whether they perceive workload during a surgery as lower (scale levels 1-2), equal (3), or higher (4-5) when using an IGN system as compared to the standard (manual, i.e., unsupported) approach which still represents the "golden standard" of surgery. The selected dimensions directly correspond to those assessed by the NASA Task-Load-Index (NASA TLX) which represents the internationally most used method for subjective workload assessment (Hart, 2006).

(2) Situation awareness is assessed by three items addressing all aspects of the concept as defined by Endsley (1995). Again, the surgeons are requested to indicate whether each aspect of situation awareness is lower, equal, or higher when supported by IGN compared to the approach without navigation support. In contrast, all following items simply require ratings on 5-point rating scales with 1 expressing total agreement and 5 expressing total disagreement with a given statement.

(3) Trust and reliance related issues are addressed with nine Items. Two items assess the surgeon's perception of the reliability of the IGN system in terms of accuracy and precision. The overall level of trust in IGN systems is inquired by two items assessing to what extent surgeons generally trust in the correct functioning of IGN systems, and, to what extent they themselves would opt for the application of the system if they had to undergo a minimally invasive surgery. Three items address questions regarding the monitoring of the system: Surgeons have to indicate whether they often think about consequences of possible malfunctions of the IGN system, whether they usually check the system's proper functioning repeatedly during longer lasting surgery, and whether they usually check the proper function before conducting risky surgical interventions in particular. One item explicitly addresses the question to what extent surgeons think that they would detect any malfunction of the IGN system. A last item deals with the possible occurrence of automation bias by asking how the surgeon would decide if the automatically generated navigation information conflicts with their own assessment of the situation.

(4) Possible effects of skill degradation are addressed by two items which ask for the impact of repeated use of IGN on the surgeon's own skills, and on the skill development of junior surgeons, respectively.

Sample

Five questionnaires each were sent to the clinical heads of 292 German hospitals specialised in ENT surgery, neurosurgery, orthopaedic surgery, or trauma surgery. The heads were asked to distribute the questionnaire among surgeons experienced with the clinical use of IGN systems. These surgeons were asked to think of the last five surgeries where they have used an IGN system and then to answer the different questions with respect to this system. Questionnaires from 112 hospitals and 213 surgeons were returned, which corresponds to a response rate of 38.4%. However, 18 hospitals just stated that they would not apply IGN. Therefore, the statistical analyses are finally based on the responses of 213 surgeons from 94 different hospitals.

Statistical Analysis

Ratings were tested for significant deviations from the midpoint of scale (3) by means of one-sample t-tests. For the surgeons' ratings of workload and situation awareness significant deviations from the neutral midpoint indicated a clearly perceived benefit (> 3) or cost (< 3) effect of computer-assisted surgery compared to the non-supported ("golden") standard approach. For all other items, significant deviations from the midpoint were interpreted as a clear tendency to agree (> 3) or disagree (< 3) with the given statement. To prevent from an inflation of the probability of alpha errors due to the high number of t-tests, alpha-levels were Bonferroni-adjusted whenever several tests focused on the same human factors aspect (family-wise approach). In accordance with this procedure effects will be reported as significant of the given alpha-level only if their probabilities are equal or smaller than the adjusted alpha.

RESULTS

Workload

Results related to the comparison of perceived workload during surgeries with and without IGN-support are shown in Figure 1. As becomes evident from the first column, the data do not reveal a difference in overall workload between surgeries with and without IGN support, $t(212) = -0.9$, $p = .35$. However, a more detailed inspection of the data revealed a trade-off of workload effects among the different workload dimensions: Beneficial effects in favour of IGN were reported with respect to effort, $t(212) = -8.6$, $p < .001$, as well as frustration and stress, $t(212) = -5.1$, $p < .001$. Yet, these effects were balanced by a perceived increase of mental demands, $t(212) = 2.94$, $p = .004$, and time pressure during the surgery, $t(212) = 5.9$, $p < .001$.
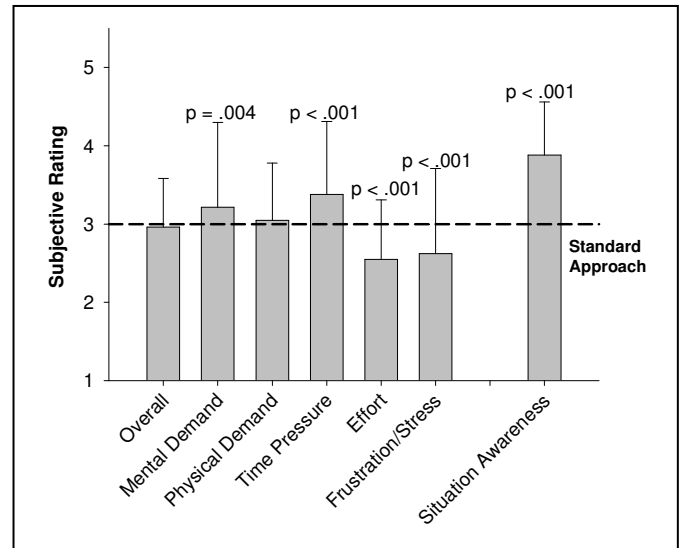

Fig. 1: Means and standard deviations of the surgeons' workload and situation awareness with support of an IGN system as compared to the manual standard procedure (lower = 1, higher = 5). P-values indicate the significance of the deviation from the standard approach.

Situation Awareness

The surgeons' average ratings of situation awareness are also depicted in Figure 1. As becomes evident, situation awareness is perceived significantly higher during surgery supported by IGN technology as compared to the standard approach $t(212) = 19.0$, $p < .001$. This was the case for each of the three aspects of situation awareness.

Trust and Reliance

Results on trust in IGN systems and reliance on the proper functioning of these systems are shown in Figure 2. First of all, surgeons perceive IGN systems as rather reliable, $t(211) = 12.1$, $p < .001$, and affirm that they, in general, trust in the proper function of the system, $t(211) = 14.8$, $p < .001$. Yet, despite their rather high levels of trust, the surgeons reported that they usually still test the proper function of the IGN system (i.e., its correct calibration) repeatedly during longer lasting surgeries, and before performing any critical intervention, $t(212) = 14.7$, $p < .001$. Because these cross-checks seem to be of particular importance for patient safety, a more detailed analysis of the distribution of ratings was performed. This analysis revealed that in fact 81.4% of the surgeons affirmed that they usually perform routine checks of the system's function in case of longer lasting surgeries. 96.8% of the surgeons reported to conduct such tests at least before particularly critical interventions. However, after all, a total of 42 surgeons (19.6%) reported that they would not perform any further checks of the system after pre-operative calibration, and seven surgeons (3.2%) even seem to refrain from such tests before a critical surgical step.
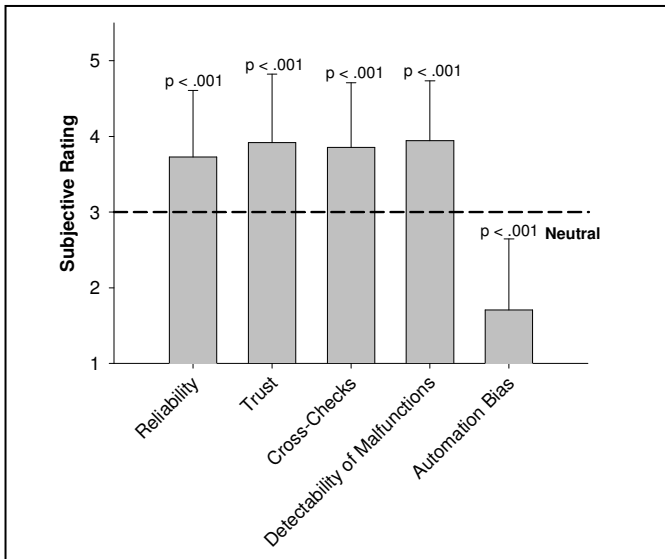
Fig. 2: Means and standard deviations of perceived reliability, trust and different aspects of reliance. P-values indicate the significance of the deviation from the neutral mid point of the scale (total disagreement = 1; total agreement = 5).
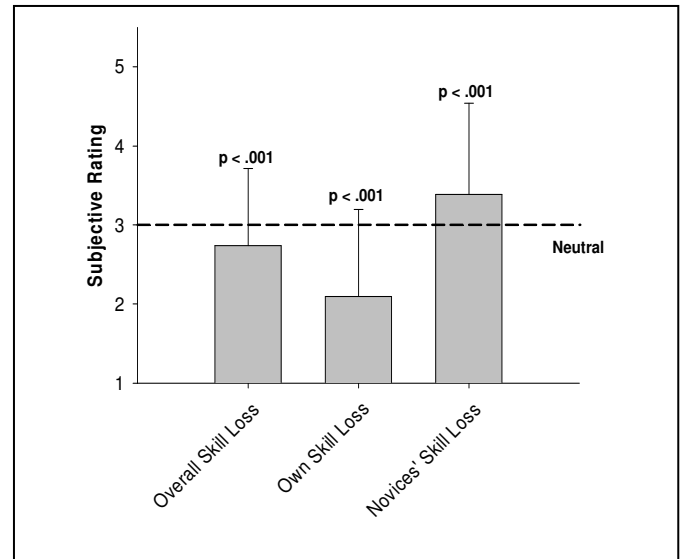

Fig. 3: Means and standard deviations of the rated loss of surgical skills due to the use of IGN. P-values indicate the significance of the deviation from the neutral mid point of the scale (total disagreement = 1; total agreement = 5).

Another aspect inquired addressed how fast surgeons think they would detect a "malfunction" of their navigation system while using it. On average, the respondents turned out to be very confident that they would quickly realize such malfunctions, $t(210) = 17.3$, $p < .001$. Nevertheless, ten surgeons (4.7%) expressed doubts in this respect.

With respect to automation bias, the surgeons reported that they would rely more on their own assessment than on the navigation aid, $t(212) = -20.1$, $p < .001$. Even though, there were 15 surgeons (7%) reporting that they (rather or definitely) would follow the information provided by the system even if it would be in conflict with their own assessment.

Skill Degradation

Figure 3 shows the surgeons' ratings with regard to skill-decrements that may be associated with an ongoing use of IGN systems.

Overall, the surgeons do not see any risk for skill maintenance associated with use of IGN, $t(212) = -3.9$, $p < .001$. However, looking at the individual items of this scale, it becomes apparent that this only holds for the evaluation of the impact of repeated IGN-use on their own navigational skills. With respect to novices, it is assumed that the use of IGN systems might entail a negative effect on the development of proper navigational skills, $t(212) = 12.0$, $p < .001$.

DISCUSSION

The present study provides insights into pivotal aspects of human-automation interaction associated with the use of computer assisted surgery as they are perceived by experienced surgeons. Four different issues were considered: mental workload, situation awareness, trust and reliance, and degradation of skills.

With respect to effects on workload and stress during surgeries, a rather inconsistent picture emerged. Taken as a whole, the use of IGN technology does not seem to reduce the overall workload of the surgeon during surgery. On first sight this result is surprising because a reduction of workload usually is regarded as one important driver for the introduction of automation. And in fact, a closer inspection of the data revealed that using IGN leads to the expected beneficial effects on perceived effort and stress during surgery. Yet, these positive effects are counterbalanced by the perception of increased mental demand and time pressure during computer-assisted surgeries. This seems to be related to the fact that using IGN means interaction with another technical system which does not only provide extra demands on processing and integrating additional information (i.e., displayed navigation information) but also involves additional steps in the work flow which need time to complete. This is also suggested by earlier results from Metson et al. (1999) which revealed that the use of IGN considerably prolongs the time needed for a sinus surgery. As long as these elevated time demands are taken into account in the planning and scheduling of surgeries, this won't represent a problem. Otherwise, additional time demands will directly increase the time-pressure of the surgeons.

Perceived benefits of IGN emerged with respect to situation awareness during surgery. As expected, results revealed a clear advantage of IGN including an improvement of situation awareness on all three levels, i.e., the correct perception of relevant information from the surgical site, the proper interpretation of this information, and the correct anticipation of the future development of the surgical process. This positive evaluation confirms earlier reports from studies pointing to beneficial effects of IGN on the clinical outcome of surgeries (e.g., Gong et al., 2007; Strauss et al., 2006).

Furthermore, the surgeons rated the perceived reliability of IGN-systems as rather high. In line with that, the surgeons also reported that they have a high level of trust in the proper functioning of the systems. One issue which might arise from the generally high trust in the systems includes the possible development of over-trust and overreliance. The present study provides some indications that this might be a problem with the current systems. While 20% of the surgeons indicated that they refrain from checking the proper function repeatedly during a surgery, 3% refrain from doing so even before critical surgical interventions. In addition, 5% of the surgeons indicated that they might have problems in the detection of malfunction of the system und 7% committed that they would trust the automation more than their own assessment in case that the displayed position of the surgical instrument conflicts with their own appraisal. The overall count of surgeons indicating some kind of overreliance is certainly not high, albeit not negligible. This suggests not to leave the decision about cross-checks in the responsibility of the surgeon but to define them as an unconditional requirement during computer-assisted surgeries.

Finally, surgeons do not perceive skill degradation as a problem of IGN for themselves. However, skill loss or, better, lack of skill development is perceived as a possible issue if novices work with the systems already from the beginning of their clinical training. So far, no studies are available which have addressed this possible effect, and it must be left to future research to investigate the impact of IGN systems on the development and acquisition of navigational skills of surgeons.

In summary, the results of the present survey acknowledge IGN as a very beneficial tool. Yet, there are two issues that have to be considered when implementing these systems: Firstly, the risk of a workload-shift rather then a reduction of workload, and secondly, the risk of overreliance on the system's proper functioning. Although this seems to include only a minority of surgeons, the consequences that a single non-detected malfunction of an IGN system can bring are enough to require the development of effective countermeasures such as appropriate trainings and prescribed cross-checks of calibration.

In evaluating these results, two constraints need to be taken into account. The first one concerns possible limits arising from the analysis of subjective questionnaire data. It is obvious that these data can only be regarded as a first screening of possible issues related to this kind of technology. The second constraint relates to the fact that the IGN systems evaluated only represent relatively low levels of automation. It remains to be seen whether the results of the present study also apply to more advanced systems which are currently under development, e.g., navigated control systems, which may take over decision making and control functions as well. These further developments in the area of image-guided surgery need to be carefully evaluated not only with respect to its technical capabilities but also with respect to human factors issues involved.

REFERENCES

Brown, S.M., Sadoughi, B., Cuellar, H., von Jako, R., & Fried, M.P. (2007). Feasibility of near real-time image-guided sinus surgery using intraoperative fluoroscopic computed axial tomography. *Otolaryngology – Head and Neck Surgery, 136*, 268-273.

Caversaccio, M., Nolte, L.-P., & Haeusler, R. (2002). Present state and future perspectives of computer aided surgery in the field of ENT and skull base. *Acta Oto-Rhino-Laryngologica Belgica, 56*, 51-60.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Gong, J., Mohr, G., & Vézina, J.L. (2007). Endoscopic pituitary surgery with and without image guidance: an experimental comparison. *Surgical Neurology, 67*, 572-578.

Hart, S.G. *(*2006*)*. NASA-Task Load Index (NASA-TLX): 20 years later. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 904-908)*.* Santa Monica, CA*:* Human Factors & Ergonomics Society.

Labadie, R.F., Davis, B.M., & Fitzpatrick, J.M. (2005). Image-guided surgery: what is the accuracy? *Current Opinion in Otolaryngology & Head and Neck Surgery, 13*, 27-31.

Manzey, D., Strauss, G., Trantakis, C., Lueth, T., Roettger, S., Bahner-Heyne, J.E., Dietz, A., & Meixensberger, J. (in press). Automation in surgery: A systematic approach. *Surgical Technology International XVIII*.

Metson, R., Cosenza, M., Gliklich, R.E., & Montgomery, W.W. (1999). The Role of Image-Guidance Systems for Head and Neck Surgery. *Archives of Otolaryngology – Head & Neck Surgery, 125*, 1100-1104.

Mosier, K.L., & Skitka, L.J. (1996). Human decision-makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum Associates.

Muir, B., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*, 429-460.

Parasuraman, R., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*, 1-23.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30*, 286-297.

Reijnders, K., Coppes, M.H., van Hulzen, A.L.J., Gravendeel, J.P., van Ginkel, R.J., & Hoekstra, H.J. (2007). Image guided surgery: New technology for surgery of soft tissue and bone sarcomas. *European Journal of Surgical Oncology, 33*, 390-398.

Roettger, S., Bali, K., & Manzey, D. (in press). Impact of automated decision aids on performance, operator behaviour, and workload in a simulated supervisory control task. *Ergonomics*.

Strauss, G., Koulechov, K., Roettger, S., Bahner, J., Trantakis, C., Hofer, M. et al. (2006). Evaluation of a navigation system for ENT with surgical efficiency criteria. *The Laryngoscope, 116*, 564-572.

Vicente, K. (2003). Less is (sometimes) more in cognitive engineering: The potential role of technology in improving patient safety. *Quality and Safety in Health Care, 12*, 291-294.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance*. Englewood Cliffs, NJ: Prentice-Hall.

Wiener, E.L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. NASA Contractor Report 177528*. Moffet Field, CA: NASA Ames Research Center Report.