



Published in final edited form as:

Ann Biomed Eng. 2018 October ; 46(10): 1548–1557. doi:10.1007/s10439-018-2099-2.

Clinical Translation of the *LevelCheck* Decision Support Algorithm for Target Localization in Spine Surgery

Amir Manbachi^{1,2}, Tharindu De Silva¹, Ali Uneri¹, Matthew Jacobson¹, Joseph Goerres¹, Michael Ketcha¹, Runze Han¹, Nafi Aygun³, David Thompson⁴, Xiaobu Ye², Sebastian Vogt⁵, Gerhard Kleinszig⁵, Camilo Molina², Rajiv Iyer², Tomas Garzon-Muvdi², Michael R. Raber², Mari Groves², Jean-Paul Wolinsky², Jeffrey H. Siewerdsen^{1,2,3,4,6}

¹Department of Biomedical Engineering, Johns Hopkins University, 3400 N. Charles Street, Wyman Park Building, Suite 400 West, Baltimore, MD 21218, USA

²Department of Neurosurgery, Johns Hopkins University School of Medicine, The Johns Hopkins Hospital, 1800 Orleans Street, Baltimore, MD 21287, USA

³Russell H. Morgan Department of Radiology, Johns Hopkins University, The Johns Hopkins Hospital, 1800 Orleans Street, Baltimore, MD 21287, USA

⁴Armstrong Institute for Patient Safety and Quality, Johns Hopkins University School of Medicine, 750 E Pratt St, 15th Floor, Baltimore, MD 21202, USA

⁵Siemens Healthineers, Henkestraße 127, 91052 Erlangen, Germany

⁶Department of Biomedical Engineering, Johns Hopkins University, Traylor Building, Rm 622, 720 Rutland Avenue, Baltimore, MD 21205, USA

Abstract

Recent work has yielded a method for automatic labeling of vertebrae in intraoperative radiographs as an assistant to manual level counting. The method, called *LevelCheck*, previously demonstrated promise in phantom studies and retrospective studies. This study aims to: (#1) Analyze the effect of *LevelCheck* on accuracy and confidence of localization in two modes: (a) *Independent Check* (labels displayed after the surgeon's decision) and (b) *Active Assistant* (labels presented before the surgeon's decision). (#2) Assess the feasibility and utility of *LevelCheck* in the operating room. Two studies were conducted: a laboratory study investigating these two workflow implementations in a simulated operating environment with 5 surgeons, reviewing 62 cases selected from a dataset of radiographs exhibiting a challenge to vertebral localization; and a clinical study involving 20 patients undergoing spine surgery. In Study #1, the median localization error without assistance was 30.4% (IQR = 5.2%) due to the challenging nature of

Address correspondence to Jeffrey H. Siewerdsen, Department of Biomedical Engineering, Johns Hopkins University, Traylor Building, Rm 622, 720 Rutland Avenue, Baltimore, MD 21205, USA. jeff.siewerdsen@jhu.edu.

This research was conducted in accordance with the Institutional Review Board protocol NA_00078717, approved by Johns Hopkins Medical Institution: (Principal Investigator: Jean-Paul Wolinsky M.D.).

CONFLICT OF INTEREST

The work was supported by the National Institutes of Health (NIH R01-EB-017226) and academic-industry partnership with Siemens Healthineers (XP Division, Erlangen, Germany). The authors have publicly available intellectual property associated with this work, yet apart from paid employment by Siemens employees (Sebastian Vogt and Gerhard Kleinszig), the authors have no personal financial interest in the work reported in this paper.

the cases. *LevelCheck* reduced the median error to 2.4% for both the *Independent Check* and *Active Assistant* modes ($p < 0.01$). Surgeons found *LevelCheck* to increase confidence in 91% of cases. Study #2 demonstrated accuracy in all cases. The algorithm runtime varied from 17 to 72 s in its current implementation. The algorithm was shown to be feasible, accurate, and to improve confidence during surgery.

Keywords

Spine surgery; Clinical translation; Surgical workflow; Intraoperative imaging; *LevelCheck*; Image-guided surgery

INTRODUCTION

Approximately 10–12% of surgical patients are reported to experience adverse events during hospitalization, with close to half considered preventable.^{1,5,8} In the US alone, operation on the wrong body part may occur as often as 40 times a week.¹⁰ Wrong-level (alternatively, unintended-level) spine surgery is one such occurrence, with studies suggesting an occurrence of 1 in 3110 procedures.¹¹

Vertebral localization is often accomplished using intraoperative x-ray projection images (*via* mobile radiography or C-arm fluoroscopy) in which the surgeon counts vertebral levels relative to a known anatomical landmark (e.g., counting “up” from the sacrum or “down” from C1).⁶ The process is subject to a variety of sources of potential error, including anatomical variations (e.g., lumbarized sacrum or thirteenth thoracic vertebra) and poor radiographic image quality. Some institutions conduct a separate preoperative procedure under CT guidance to “tag” the target vertebrae with a radiographically conspicuous marker.

Prior work yielded a method for automatic labeling of vertebrae in intraoperative radiographs as an assistant to manual level counting. The method, called *LevelCheck*, enables the vertebral labels identified in preoperative CT or MR images to be overlaid onto corresponding locations in intraoperative radiographs.^{2-4,7,9,12-14} *LevelCheck* has been previously assessed in phantom and retrospective studies, motivating evaluation in clinical studies.

As an important precursor to larger scale, multi-center studies, we investigated the implementation of *LevelCheck* in two possible modes: (a) as an *Independent Check*, in which labels are displayed only after the surgeon commits a decision on localization; and (b) as an *Active Assistant*, in which labels are immediately applied to the radiograph, and the surgeon’s decision is made concurrently with the displayed labels. The former is more conservative and analogous to the implementation of computer-aided detection systems in mammography.¹⁵ The latter is potentially more streamlined but carries potential bias.

The work reported below includes two main studies: (#1) to analyze the effect of *LevelCheck* on the accuracy and confidence of surgical localization when implemented as an *Independent Check* or as an *Active Assistant*; and (#2) for the first time, to assess the utility of the algorithm in patient studies in the operating room. Larger scale, multi-center

clinical studies are the subject of future work and are important to fully evaluate the role of the *LevelCheck* in routine workflow.

MATERIALS AND METHODS

Details of the 3D-2D registration algorithm underlying *LevelCheck* was described in previous work.^{4,12-14} In principle, the algorithm solves for the 3D orientation (“pose”) that relates a preoperative 3D image to an intraoperative 2D image—viz., a preoperative CT or MRI and an intraoperative x-ray projection image, respectively. The solution comprises 6 degrees of freedom (3 translational and 3 rotational) in orientation of the 3D preoperative image such that its projection (a digitally reconstructed radiograph, “DRR”) matches the 2D intraoperative radiograph. The match is quantified in terms of a similarity metric that forms the objective function in an iterative optimization. In the current work, the similarity metric was taken as the Gradient Orientation (GO) between the DRR and radiograph, which was shown to provide robustness against image mismatch (e.g., instrumentation in the radiograph but not in the CT) as well as anatomical deformation.⁴ The optimization method was the covariance matrix adaptation evolution strategy (CMA-ES), a stochastic optimizer that is robust against local minima, parallelizable (for faster runtime), and has demonstrated a fairly large capture range.² Given a 3D-2D registration solution, labels defined in the 3D image (e.g., labels corresponding to individual vertebrae) can be projected *via* the resulting transformation onto the coordinate system of the 2D radiograph and overlaid as image augmentation.

In the studies reported below, initialization of the 3D-2D registration was performed manually simply to provide overlap between the preoperative CT and the intraoperative radiograph. The region of interest was limited to a rectangular mask about the region of the spine to avoid strong gradients presented by the skin line. The algorithm was implemented on a Dell Precision T7910 workstation with Intel Xeon processor (3.5 GHz), 64 GB RAM, and GeForce GTX TITAN X graphical processing unit (GPU, Nvidia, Santa Clara CA).

Laboratory Study

Modes of LevelCheck Implementation—Two distinct modes of *LevelCheck* implementation were evaluated. As illustrated in Fig. 1a, implementation as an *Independent Check* involves displaying spine labels after the surgeon makes a decision on localization. In such a scenario, the labels are typically confirmatory of the surgeon’s decision, and a discrepancy requires resolution by double-checking and/or additional consultation. As illustrated in Fig. 1b, implementation as an *Active Assistant* involves immediate display of the labels concurrent with the surgeon’s decision-making process.

Case Selection and Registration—A previous retrospective study⁴ evaluated *LevelCheck* performance in 398 intraoperative radiographs and identified five main conditions for which vertebral localization was most challenging: anatomical complexities; poor radiographic image quality; lack of anatomical landmarks; vertebrae obscured by other anatomy; and long spine segments. From that dataset, a subset of 62 images was selected for the laboratory study. These images were identified in the prior multireader study as the most challenging cases to localize. Example images with *LevelCheck* labels are depicted in Fig. 2.

The true location of each vertebra was defined in preoperative CT for each patient by a board-certified neuroradiologist. Standard level designations (C7, T2, etc.) were used. Each CT image and radiograph pair was registered *via* 3D-2D transformation that overlaid vertebral labels on their corresponding locations in the radiograph.

The study involved evaluation by 5 surgeons on 62 cases (7 cervical, 32 thoracic, 23 lumbar). Reading order was randomized for each surgeon. Five cases were repeated at the end of each reading to assess intra-observer repeatability and possible effects due to learning or fatigue.

To ensure rigorous evaluation by surgeons, 16 (/62) cases were purposely presented with a labeling error—e.g., shift by one vertebra. For the *Active Assistance* mode, this also ensured that surgeons did not become complacent during the study and simply “trust” the *LevelCheck* labels without exercising their own expertise.

Experimental Setup—The laboratory study was conducted in a semi-realistic mock operating room environment (Fig. 3). The images displayed were drawn from the 62 clinical cases described above, and the surgeon was free to interact with a mobile fluoroscopic C-arm and body phantom to illustrate challenges in setup and with the image display to modify displayed contrast etc. The surgeons were unrestricted in reading time.

For each case, the surgeon localized a particular vertebral level and evaluated confidence and challenging factors *via* questionnaire (below). For the *Independent Check* mode (Fig. 1a), the surgeon’s localization was recorded before and after display of the labels. In separate trials using the *Active Assistance* mode (Fig. 1b), the surgeon’s localization was recorded just once.

Questionnaire and Data Analysis—The accuracy of surgeons’ localization was analyzed as the fraction of cases for which their decision agreed with the true vertebral level (as defined by the neuroradiologist). Accuracy was measured for localizations performed without labels (conventional approach) and with *LevelCheck* implemented as an *Independent Check* (Fig. 1a) or *Active Assistant* (Fig. 1b).

A questionnaire was used to assess the utility of *LevelCheck* in each mode of implementation. Table 1 summarizes the questionnaires. Question 1 (Q1) assessed image quality on a 5-point ordinal scale. Question 2 (Q2) assessed potential utility by rating the degree of challenge in localization and whether additional decision support would be helpful, and Question 3 (Q3) provided free response regarding the nature of the challenge. Question 4 (Q4) assessed the effect of decision support on the surgeon’s confidence in localization, and Question 5 (Q5) provided free response regarding what specific challenges such decision support would help to overcome.

The median, interquartile range (IQR), and full range in responses were computed for Q1, Q2, and Q4 and displayed as violin plots. Correlation of confidence (Q4) with image quality (Q1) and/or challenge levels (Q2) helped identify the scenarios under which *LevelCheck* decision support was most useful.

For each surgeon, the error rate in localization was calculated for different modes of implementation. The statistical significance of the difference in error rates among the three modes was analyzed by a paired student *t* test.

Clinical Study

Case Selection—The utility of *LevelCheck* was assessed in the operating room under an IRB-approved protocol. For each procedure, a mobile C-arm identical to that in the laboratory setup was used to acquire intraoperative radiographs. Informed consent was obtained in writing from each patient. The study involved 20 patients (8 cervical, 4 thoracic, 8 lumbar spine) with evaluation by 2 surgeons according to the questionnaire described below.

Study Procedures—The *Independent Check* mode was selected for the clinical studies as the more conservative of the two modes investigated in the laboratory study. To preserve the standard of care, the surgeons localized target vertebrae according to conventional level counting procedures (without *LevelCheck*), and the case proceeded accordingly—i.e., the display of *LevelCheck* labels did not alter the surgical decision exercised in the case. Following target localization by conventional means, the *LevelCheck* results were displayed to the surgeon, and the utility was evaluated by questionnaire (below). Any potential discrepancy between the surgeon's original localization and that indicated by *LevelCheck* required resolution by independent means—e.g., consultation with a neuroradiologist.

Questionnaire and Data Analysis—The questionnaire summarized in Table 2 was provided to each surgeon. Questions 1–4 were posed prior to considering the *LevelCheck* labeled radiograph. Question 1 (Q1) gauged the purpose of the image acquisition, and Question 2 (Q2) assessed image quality *via* the same 5-point scale as in the laboratory study. Question 3 (Q3) gauged potential utility with respect to degree of challenge in localization and whether additional decision support would be useful, with Question 4 (Q4) providing free response regarding the nature of the challenge.

Questions 5–7 were posed following consideration of the *LevelCheck* labeled radiograph. Question 5 (Q5) assessed the accuracy of the labels—i.e., whether the *Independent Check* agreed with the surgeon's original decision and whether they were, in fact, anatomically accurate. Question 6 (Q6) assessed the effect of the labels on confidence in localization, and Question 7 (Q7) assessed the amount of time that the surgeon would be willing to expend (in algorithm runtime) for this particular case.

Analysis of performance was similar to that described above for the laboratory study, with pooled ordinal values evaluated in terms of median and interquartile range (IQR), and correlation between confidence (Q3) and purpose (Q1) and image quality (Q2) was analyzed to determine under what conditions *LevelCheck* could provide greatest utility.

RESULTS

Laboratory Study

Results from the questionnaire of Table 1 are summarized in Fig. 4, presenting the responses regarding (Q1) image quality, (Q2) degree of challenge, and (Q4) confidence. The median image quality was rated as [2] “poor,” with some rated as [3] “fair,” indicative of the challenging nature of images selected for the laboratory study. Accordingly, the median degree of challenge in localization was [3] “challenging,” and all cases were judged [2] or [3] such that decision support was considered helpful. For both modes of implementation, the median confidence was [3] (improved confidence by reassurance of original decision) with some cases rated [4] (improved confidence by providing additional information).

Figure 5 shows the error rate in vertebral localization in the laboratory study. The challenging nature of cases selected is evident from high median error rate of 30.4% (IQR = 5.2%) for manual level counting. Such a high rate is not indicative of clinical practice and is a product of cases selected to exhibit poor image quality, lack of clear anatomical landmarks, etc. Median error rate was reduced to 2.4% (IQR = 10.6%) for localization with *LevelCheck* implemented as an *Independent Check* and to 2.4% (IQR = 3%) for implementation as an *Active Assistant*. The improvement was statistically significant ($p < 0.01$) for each case. The *Independent Check* appeared to exhibit a larger range in error (possibly suggesting that surgeons are more likely to “stick to” an erroneous decision), but the difference was not statistically significantly different from the *Active Assistance* mode.

Clinical Study

Example labeled radiographs from the clinical study are shown in Fig. 6. In Fig. 6a, *LevelCheck* is seen to provide accurate labeling of the cervical spine despite strong attenuation in the shoulders and a fairly strong difference in spinal curvature (neck flexion) between the preoperative CT and intraoperative radiograph. Figure 6b shows a thoracolumbar case exhibiting a 13th rib (attachment at L1 attachment) that could present a confounding factor in conventional level counting. Figure 6c illustrates labeling in an image of the sacrolumbar region showing a lumbarized sacrum—another potentially confounding factor that could lead to a counting error in conventional methods.

Figure 7 summarizes responses to the clinical study questionnaire (Table 2). As shown in Fig. 7a, most images were acquired for purposes (Q1) of localization at [1]–[2] early or [3] middle stages in the procedure for localization of anatomy and/or hardware placement. Image quality (Q2) exhibited a median value of [5] “excellent visibility” with a broad distribution that included some instances of [2] “poor” or [1] “very poor,” which is representative of image quality as encountered in clinical practice (c.f., the laboratory study, which purposely included a preponderance of cases with poor image quality).

The degree of challenge (Q3) shown in Fig. 7c exhibited a median value of [2] “unambiguous/decision support helpful,” with a roughly equal range of [1] “unambiguous/decision support not helpful” and [3] “challenging/decision support helpful.” These results are somewhat different from the laboratory study (median value [3]) due to case selection and are likely more representative of cases encountered in clinical practice.

Geometric accuracy (Q5) was uniformly [2] confirmatory of the surgeon's original decision. In the current study, there were no instances of discrepancy ([1] or [3]) from the surgeon's original decision—i.e., there were no wrong-level localizations on the part of the surgeon, and there were no registration errors on the part of *LevelCheck*. Responses with respect to utility of the algorithm (Q6) indicated a median value of [3] “improve confidence by providing reassurance of original decision.”

Figure 7f summarizes the responses as to the amount of time that would be tolerable for adding such decision support to the radiographic image (median value [3], 20–60 s) compared to the actual runtime of the *LevelCheck* algorithm (17–72 s, consistent with [3]).

DISCUSSION

Development of the *LevelCheck* algorithm for translation to multi-center trials and routine clinical use benefits from rigorous assessment of its performance and manner of implementation. The studies reported above considered two potential modes of implementation that differ in workflow and potential for error/bias and—for the first time—translated *LevelCheck* to the operating theatre to assess its feasibility and utility in a small ($N=20$) pilot study.

The laboratory study elucidated possible differences in implementation of the algorithm as an Independent Check or as an Active Assistant: the former being more conservative, whereas the latter carries the potential for bias. In these studies, however, there was no difference measured in either the error rate or surgeons' confidence between the two modes, suggesting that surgeons rightly relied upon their experience and observation rather than trusting the labels. This behavior was reinforced by informing surgeons in advance of the study that the registered labels may not be accurate every time and including ($N=16$) purposely misregistered results.

The challenging nature of cases in the laboratory study was evident from the high median value of 30.4% in the error rate in localization. *LevelCheck* reduced the median error rate to 2.4% in both the Independent Check and Active Assistant modes ($p < 0.01$). No statistical significance was found between Independent Check versus Active Assistance ($p \sim 0.15$), although the latter appeared to exhibit a reduced IQR in error rate. In the challenging cases selected for the laboratory study, *LevelCheck* was found to increase confidence in 91% of all readings (373/410), while confidence was unchanged in 5.8% of the cases (24/410). The remaining 3% of cases (13/410) judged to “degrade confidence” may indicate frustration with challenging cases—i.e., cases of strong ambiguity in which assessment based on the surgeon's expertise alone was very challenging, and for which decision support did not truly add to certainty. Intraobserver repeatability showed no obvious effects of learning or fatigue, with fair agreement (intra-observer κ -value = 0.54) in the confidence level rankings between the first and second readings.

The results of the laboratory study supported translation of *LevelCheck* to clinical studies, adopting Independent Check as the more conservative mode of implementation. The study demonstrated 100% accuracy in labeling and was uniformly judged to improve confidence

while demonstrating compatibility with surgical workflow. The clinical scenarios for which the method was most beneficial included lack of anatomical landmarks or the presence of abnormal anatomy, as illustrated in Fig. 6.

In 65% of the clinical cases (12/20), *LevelCheck* increased confidence *via* reassurance of the surgeon's original decision. In 20% of the cases (4/20), confidence was improved by providing additional information (e.g., pointing out anatomical landmarks that were otherwise not easily visible in a poor-quality image), and in 15% of the cases (3/20) it was judged to have no effect on confidence (e.g., cases of conspicuous, unambiguous landmarks with good image quality).

Runtime in clinical data ranged from 17 to 72 s (initial implementation on a desktop workstation with a single GPU as described above) depending primarily on image quality (which affects the rate of convergence in the underlying optimization algorithm). This was reasonably aligned with the amount of time that surgeons indicated they were willing to wait (median 20–60 s). Correlation of responses to (Q2) image quality and (Q7) showed that surgeons were more willing to wait for labels when the image quality was poor. Accordingly, the utility of *LevelCheck* was judged to be higher in cases with lower image quality and higher degree of localization challenge.

The current studies are an important step toward larger scale, multi-center clinical trials. The laboratory studies help to identify the most appropriate mode of implementation, suggesting equivalent performance for either and relative freedom from bias among the observers (i.e., the surgeons did not blindly trust the algorithm). The clinical pilot study demonstrated both the safety and feasibility of implementation in real clinical workflow—an essential step toward a larger scale prospective study.

CONCLUSIONS

LevelCheck was found to improve surgical decision-making in vertebral level localization when implemented as either an Independent Check or Active Assistant. Clinical studies demonstrated for the first time that the algorithm is feasible in the operating room, accurate in its labeling, judged to be useful in all cases (even when confirmatory of the surgeon's original localization), and consistent with workflow.

ACKNOWLEDGMENTS

This work would have not been possible without the cooperation of Kelly Menon, Tangie Gaither-Bacon and Samantha Ernest as the Neurosurgical nursing team who supported the study in the operating rooms. Similarly, the support received from the radiology team is acknowledged—in particular: Rebecca Engberg, Jessica Enwright Wood, Charles Arterson, Joshua Shannon, Aris Thompson, and Lauryn Hancock. Finally, the authors would like to thank Ian Suk for his help and support.

REFERENCES

1. Adverse Events, Near Misses, and Errors. Patient Safety Primer. Patient Safety Network, Agency for Healthcare Research and Quality, 2016. <https://psnet.ahrq.gov/primers/primer/34/adverse-events-near-misses-and-errors> Accessed April 22, 2017.

2. De Silva T, Lo SL, Aygun N, Aghion DM, Boah A, Petteys R, Uneri A, Ketcha MD, Yi T, Vogt S, Kleinszig G, Wei W, Weiten M, Ye X, Bydon A, Sciubba DM, Witham TF, Wolinsky JP, and Siewerdsen JH. Utility of the LevelCheck algorithm for decision support in vertebral localization. *Spine (Phila Pa 1976)* 41(20):E1249–E1256, 2016. 10.1097/brs.0000000000001589. [PubMed: 27035579]
3. De Silva T, Uneri A, Ketcha MD, Reaungamornrat S, Goerres J, Jacobson M, Vogt S, Kleinszig G, Khanna AJ, Wolinsky JP, and Siewerdsen JH. Registration of MRI to intraoperative radiographs for target localization in spinal interventions. *Phys. Med. Biol* 62(2):684–701, 2017. 10.1088/1361-6560/62/2/684. [PubMed: 28050972]
4. De Silva T, Uneri A, Ketcha MD, Reaungamornrat S, Kleinszig G, Vogt S, Aygun N, Lo SF, Wolinsky JP, and Siewerdsen JH. 3D–2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. *Phys. Med. Biol* 61(8):3009–3025, 2016. 10.1088/0031-9155/61/8/3009. [PubMed: 26992245]
5. Hogan LJ, Rutherford BK, Governor L, Mitchell VT, and Nay PT. Maryland Hospital Patient Safety Program Annual Report Fiscal Year 2015 Department of Health and Mental Hygiene; Office of Health Care Quality, 2015. <https://health.maryland.gov/ohcq/hos/Documents/Maryland%20Hospital%20Patient%20Safety%20Program%20Report,%20FY%2015.pdf>, FY 15.pdf. Accessed April 16, 2017.
6. Hsiang J Wrong-level surgery: a unique problem in spine surgery. *Surg. Neurol. Int* 2:47, 2011. 10.4103/2152-7806.79769. [PubMed: 21660270]
7. Ketcha MD, De Silva T, Uneri A, Jacobson MW, Goerres J, Kleinszig G, Vogt S, Wolinsky JP, and Siewerdsen JH. Multi-stage 3D–2D registration for correction of anatomical deformation in image-guided spine surgery. *Phys. Med. Biol* 2017. 10.1088/1361-6560/aa6b3e.
8. Levinson DR Adverse events in hospitals: national incidence among medicare beneficiaries. Washington, DC: US Department of Health and Human Services, Office of the Inspector General; November 2010. Report No. OEI-06-09-00090. <https://psnet.ahrq.gov/resources/resource/19811> Accessed April 22, 2017.
9. Lo SF, Otake Y, Puvanesarajah V, Wang AS, Uneri A, De Silva T, Vogt S, Kleinszig G, Elder BD, Goodwin CR, Kosztowski TA, Liauw JA, Groves M, Bydon A, Sciubba DM, Witham TF, Wolinsky JP, Aygun N, Gokaslan ZL, and Siewerdsen JH. Automatic localization of target vertebrae in spine surgery. *Spine (Phila Pa 1976)* 40(8):E476–E483, 2015. 10.1097/brs.0000000000000814. [PubMed: 25646750]
10. Makary MA How to Stop Hospitals from Killing Us—WSJ. *The Wall Street Journal*. <https://www.wsj.com/articles/SB10000872396390444620104578008263334441352>. Published 2012. Accessed April 13, 2017.
11. Mody MG, Nourbakhsh A, Stahl DL, Gibbs M, Alfawareh M, and Garges KJ. The prevalence of wrong level surgery among spine surgeons. *Spine (Phila Pa 1976)* 33(2):194–198, 2008. 10.1097/brs.0b013e31816043d1. [PubMed: 18197106]
12. Otake Y, Schafer S, Stayman JW, Zbijewski W, Kleinszig G, Graumann R, Khanna AJ, and Siewerdsen JH. Automatic localization of vertebral levels in x-ray fluoroscopy using 3D–2D registration: a tool to reduce wrong-site surgery. *Phys. Med. Biol* 57(17):5485–5508, 2012. 10.1088/0031-9155/57/17/5485. [PubMed: 22864366]
13. Otake Y, Wang AS, Stayman JW, Uneri A, Kleinszig G, Vogt S, Khanna AJ, Gokaslan ZL, and Siewerdsen JH. Robust 3D–2D image registration: application to spine interventions and vertebral labeling in the presence of anatomical deformation. *Phys. Med. Biol* 58(23):8535–8553, 2013. 10.1088/0031-9155/58/23/8535. [PubMed: 24246386]
14. Otake Y, Wang AS, Uneri A, Kleinszig G, Vogt S, Aygun N, Lo SF, Wolinsky JP, Gokaslan ZL, and Siewerdsen JH. 3D–2D registration in mobile radiographs: algorithm development and preliminary clinical evaluation. *Phys. Med. Biol* 60(5):2075–2090, 2015. 10.1088/0031-9155/60/5/2075. [PubMed: 25674851]
15. Samulski M, Hupse R, Boetes C, Mus RDM, den Heeten GJ, and Karssemeijer N. Using computer-aided detection in mammography as a decision support. *Eur. Radiol* 20(10):2323–2330, 2010. 10.1007/s00330-010-1821-8. [PubMed: 20532890]

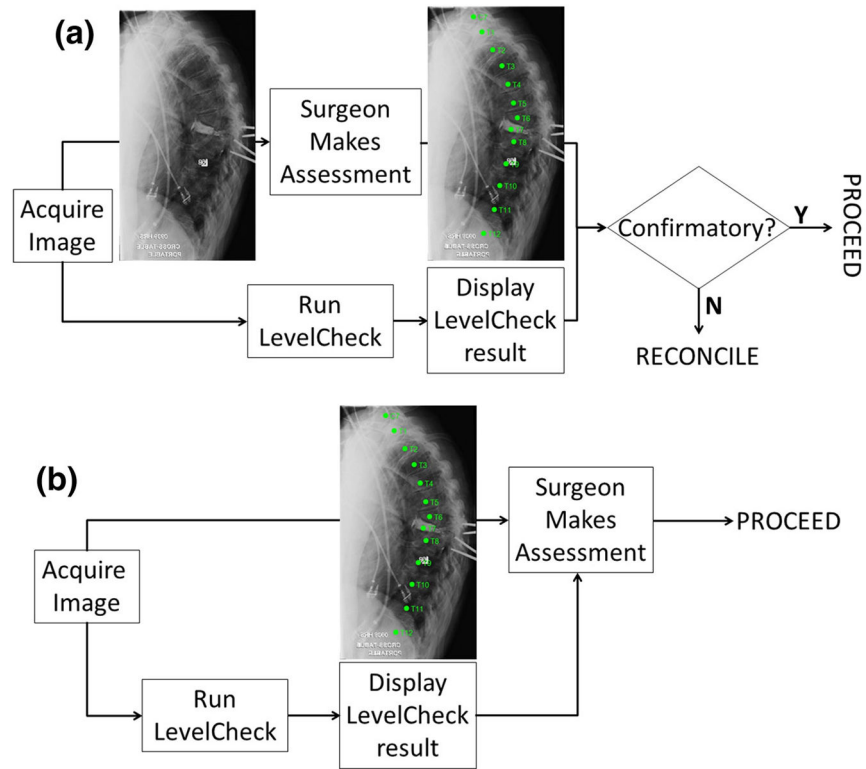


FIGURE 1. Two modes of *LevelCheck* implementation. (a) *Independent Check*, in which labels are displayed subsequent to the surgeon's decision. (b) *Active Assistance*, in which labels are displayed concurrent with the surgeon's decision.

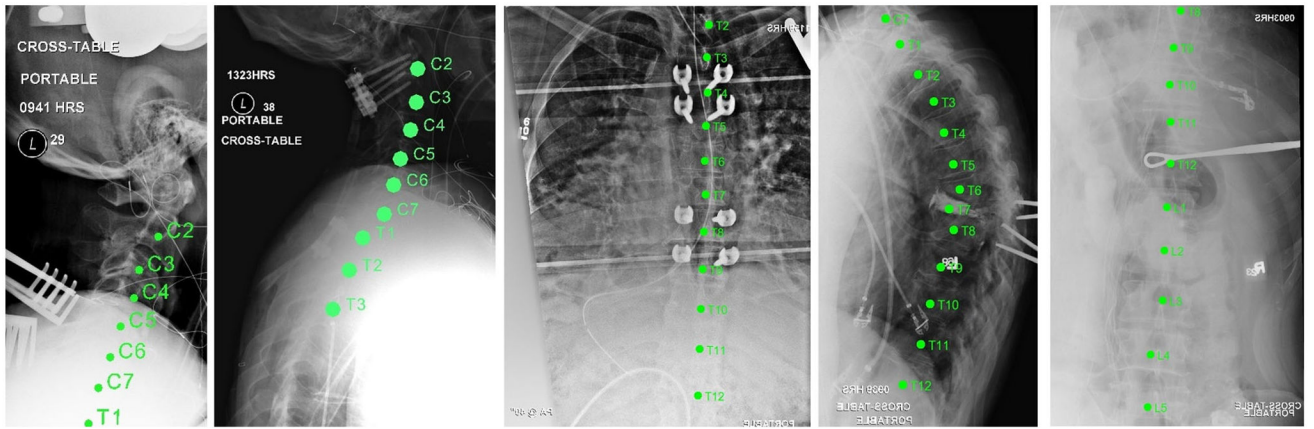


FIGURE 2. Example images from the laboratory study. Each exhibits one or more challenges to accurate level location—e.g., poor image quality, lack of anatomical landmarks, etc.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

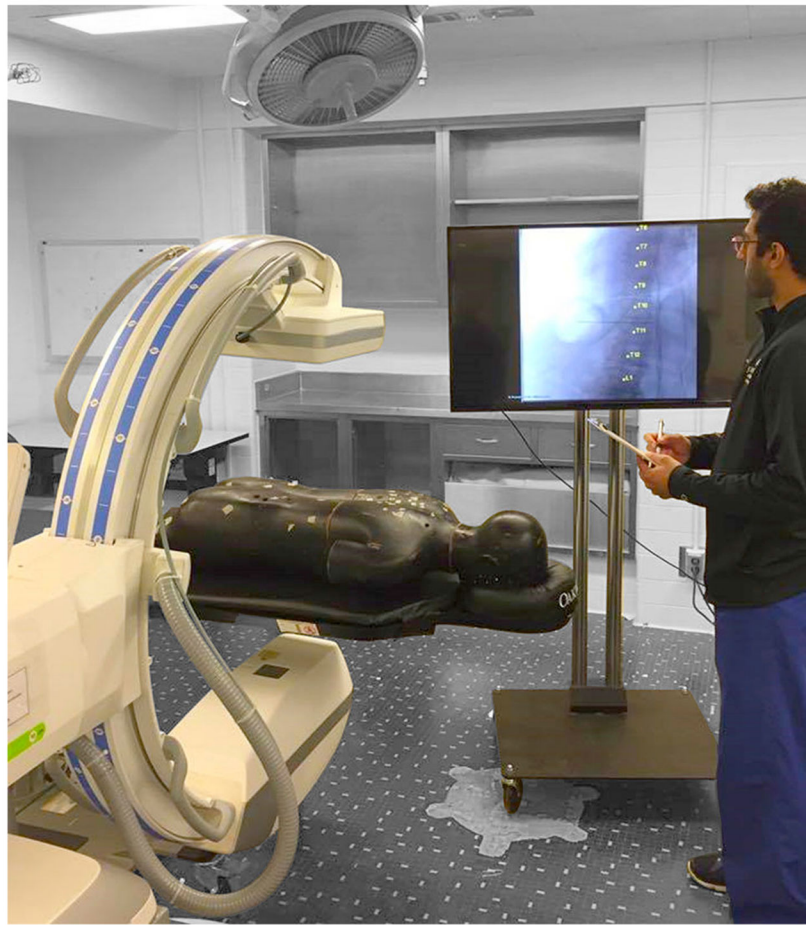


FIGURE 3. Experimental setup. Mock operating environment for the laboratory study using a mobile C-arm (Cios Alpha, Siemens Healthineers, Erlangen, Germany) and a spine phantom. Images (with or without *LevelCheck* labels) were drawn from a clinical study (not the spine phantom, which was present for illustrative/descriptive purposes).

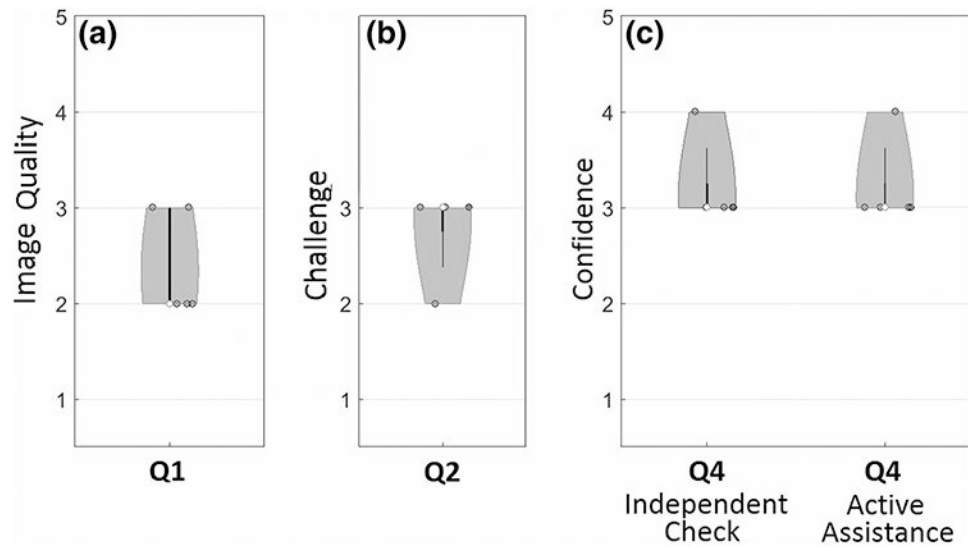


FIGURE 4. Violin plots showing responses to the laboratory study questionnaire (Table 1). The median value is marked as an open circle, and the range is marked by a gray region with width in proportion to the frequency of response.

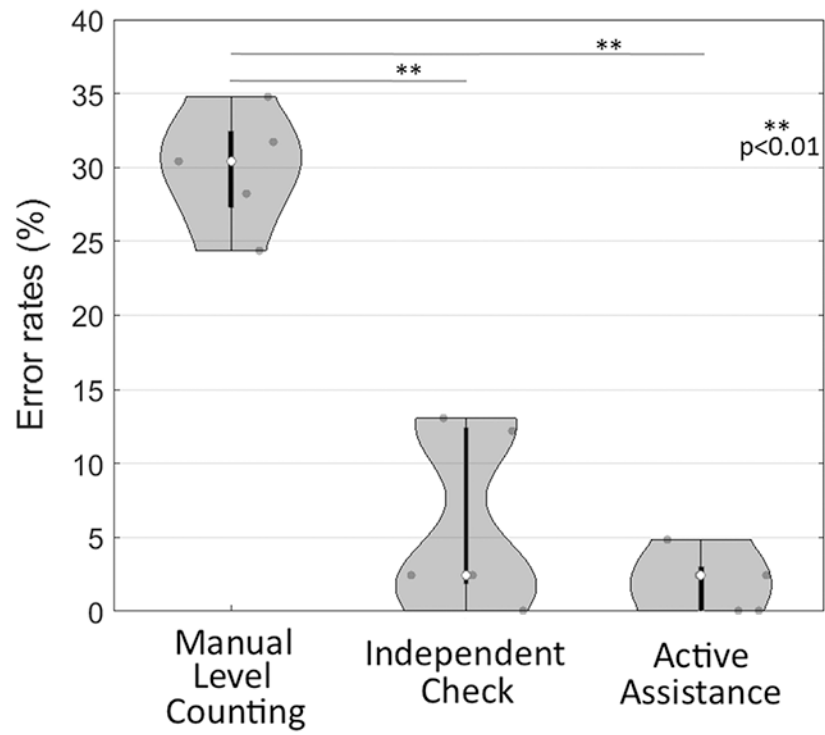


FIGURE 5. Error rate in vertebral localization in the laboratory study. The challenging nature of cases, evident in the high error rate for manual level counting, reduced significantly by decision support from *LevelCheck* in either mode of implementation.

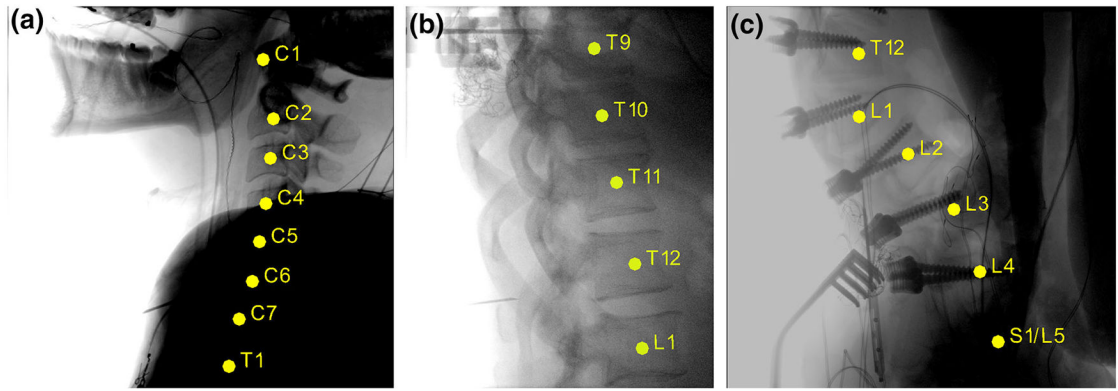


FIGURE 6.

Sample labeled radiographs from the clinical study. (a) Cervical spine example challenged by high attenuation in the shoulders. (b) Thoracic spine example with a 13th rib attached at L1. (c) Lumbar spine example with lumbarized S1-L5.

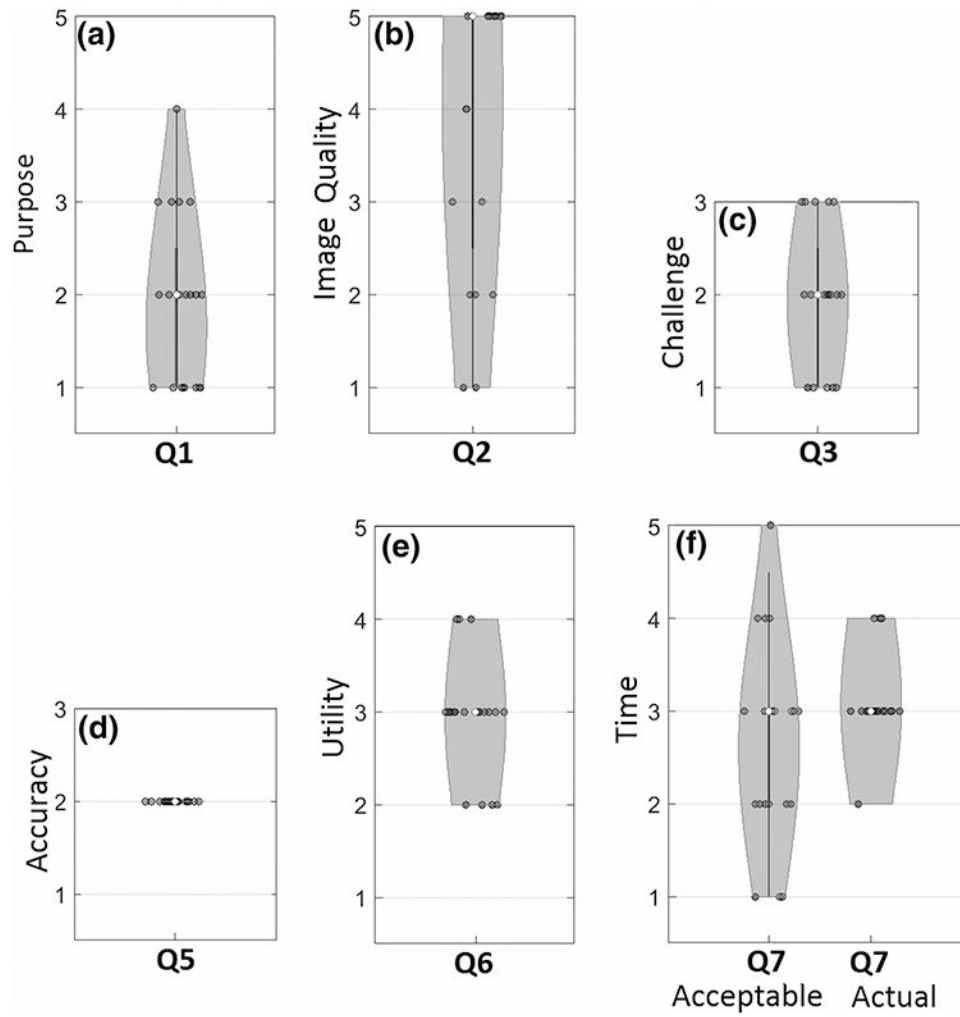


FIGURE 7. Violin plots showing responses to the clinical study questionnaire (Table 2). The median value is marked as an open circle, and the range is marked by a gray region with width in proportion to the frequency of response.

Questionnaire for the laboratory study of *LevelCheck* in two distinct modes of implementation: for the *Independent Check*, all questions shown in the table were posed; for the *Active Assistant*, only Q4 and Q5 were asked (since Q1–Q3 do not depend on the mode).

TABLE 1.

(1) Image quality: how do you rate the image quality, with respect to your ability to localize the target?

- [1] Very poor visibility/very challenging to assess
- [2] Poor visibility/challenging to assess
- [3] Fair visibility/ability to assess
- [4] Good visibility/ability to assess
- [5] Excellent visibility/ability to assess

(2) Potential utility: assess your confidence in localization based on the (unlabeled) radiograph and whether additional decision support would be helpful.

- [1] Localization is unambiguous, and verification/decision support would not be helpful
- [2] Localization is unambiguous, but verification/decision support would be helpful
- [3] Localization is challenging, additional verification/decision support would be helpful

(3) Localization challenges: describe particular challenges to localization for this case. (free response)

(4) Confidence: for this case, to what extent would *LevelCheck* affect your confidence in level localization?

- [1] Would degrade confidence by providing confounding information
- [2] Would not change confidence; provides confirmation of my original decision
- [3] Would improve confidence by providing reassurance of my original decision
- [4] Would improve confidence by providing additional information to reach my decision
- [5] It would have been extremely challenging without *LevelCheck*

(5) Localization challenges: are there particular challenges for this case for which the *LevelCheck* labels are particularly helpful? (free response)

TABLE 2.

Questionnaire for *Leve/Check* evaluation as implemented in the clinical study.

<p>Part I: before <i>Leve/Check</i> output</p> <p>(1) Clinical purpose of the radiograph acquisition</p> <p>[1] Preoperative patient positioning and localization</p> <p>[2] Intra-op localization</p> <p>[3] Intermediate-stage hardware confirmation</p> <p>[4] Final hardware confirmation</p> <p>[5] Other purpose</p> <p>(2) Image quality: how do you rate the intra-op image quality?</p> <p>[1] Very poor visibility/very challenging to assess</p> <p>[2] Poor visibility/challenging to assess</p> <p>[3] Fair visibility/ability to assess</p> <p>[4] Good visibility/ability to assess</p> <p>[5] Excellent visibility/ability to assess</p> <p>(3) Potential utility: assess your confidence in localization based on the (unlabeled) radiograph and whether additional decision support would be helpful</p> <p>[1] Localization is unambiguous, and verification/decision support would not be helpful.</p> <p>[2] Localization is unambiguous, but verification/decision support would be helpful</p> <p>[3] Localization is challenging, additional verification/decision support would be helpful</p> <p>(4) Localization challenges: describe particular challenges to localization for this CASE. (free response)</p> <p>Part II: after <i>Leve/Check</i> output</p> <p>(5) Geometric accuracy of <i>Leve/Check</i> output: The <i>Leve/Check</i> output was:</p> <p>[1] Different from surgeon's original decision and verified to be inaccurate</p> <p>[2] Confirmatory of surgeon's original decision</p> <p>[3] Different from surgeon's original decision and verified to be accurate</p> <p>(6) <i>Leve/Check</i> utility/confidence: for this case, to what extent would <i>Leve/Check</i> affect your confidence in level localization?</p> <p>[1] Would degrade confidence by providing confounding information</p> <p>[2] Would not change confidence; provides confirmation of my original decision</p> <p>[3] Would improve confidence by providing reassurance of my original decision</p> <p>[4] Would improve confidence by providing additional information to reach my decision</p> <p>[5] It would have been extremely challenging without <i>Leve/Check</i></p> <p>(7) Workflow adaptability: For this CLINICAL CONTEXT, what is the <i>maximum</i> amount of time that is acceptable within typical workflow for software to assist in level localization?</p>	
--	--

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

-
- [1] < 1 s
 - [2] 1–20 s
 - [3] 20 s–1 min
 - [4] 1–5 min
 - [5] 5–10 min