A systematic review of techniques used to validate the registration of AR images using a head-mounted device to navigate surgery

Abstract

Background

Augmented Reality (AR) Head Mounted Devices (HMD) allow the wearer to have digital images superposed onto their field of vision. It is being used to superpose annotations onto the surgical field akin to a navigation system. This review examines the published validation studies on HMD-AR systems, their reported protocols and outcomes. The aim is to establish commonalities and an acceptable registration outcome.

Materials and Methods

Multiple databases were systematically searched for relevant articles between January 2015 and January 2021. Studies examining registration of AR content, using a HMD to guide surgery were eligible for inclusion. The country of origin, year of publication, medical speciality, HMD device, software, method of registration were recorded. A meta-analysis of the mean registration error was calculated.

Results

4784 papers were identified, of which 23 met the inclusion criteria. This included studies using Hololens (22) and nVisor ST60 (1). 66% of studies were in hard tissue specialities. Eleven studies reported registration error using pattern markers (mean 2.6mm SD1.8mm). Four studies reported registration error in studies using surface markers (mean 3.8mm SD3.7mm). Three studies reported registration error using manual alignment (mean 2.2mm SD1.3mm).

Conclusions

The majority of studies in this review used in-house software with a variety of registration methods and reported errors. The mean registration error calculated in this study can be considered as a minimum acceptable standard. It should be taken into consideration when procedural applications are selected.

Introduction

Augmented reality (AR) is the term used for the superposition of digital images onto the real world view. In this way a person's vision is 'augmented'. AR was initially developed on computer screens. It has been used to add digital annotations to the video capture during endoscopy, robotic and microscopic procedures.

Head mounted devices (HMD) allow the augmentation to be directly overlaid onto the wearers vision. In this way the wearer can see the digital images superimposed onto their world view as they look around. Such information can include medical imaging from CT scans being overlaid onto the patient's body during surgery, or information from pre-operative planning such as a dissection path to access target structures.

The prospect of AR guided HMD is exciting. However, many technical questions exist before it can reliably be incorporated in the operating room. The important question is the ability of the HMD to superimpose AR images onto the target structure such as the patient. The process of aligning an image to the physical space is known as image registration. Registration would need to be accurate, comparable to current surgical standards, reproducible and acceptable to the end-user. Other questions include the learning curve, costs, time and space in the operating room to set-up the equipment.

IDEAL is a framework published in the Lancet describing the stages of adoption of new innovation in surgery. [1] Case report and series are widely being published on the use of HMD-AR to guide surgery. Competing hardware, software and registration methods are reported but clear superiority has not been established. The definition of superiority for a HMD-AR system is not in itself standardised either. This places HMD-AR technology in stage 2 of the IDEAL framework. In this stage, protocols for clinical application should be established with validated methods and defined outcome measures.

This study has examined all papers published within the last 5 years which have validated AR registration using HMD to guide surgery. The aim is to identify the most established software, hardware, clinical application and registration methods. Subsequently, how they have been validated and report the accuracy of registration achieved. This will allow the development of

standardised reporting of outcomes and agreement on criteria that should be met before introducing HMD-AR into clinical practice.

Method

A systematic review was carried out according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The National Centre for Biotechnology Information Pubmed (NCBI), Web of Science, Embase and Cochrane Central databases were searched for articles published between January 2015 and January 2021. The preceding 5 years was chosen to capture the period in which this technology has exponentially been reported in the literature. Limiting study period also avoids inclusion of discontinued technologies. The key terms used were augmented or mixed with real* and surg*. The asterisk denotes terms used with open suffix. These terms were chosen with input from a medical database specialist . Only full texts, in English language with primary data were included. Repeat articles were removed using EndNote software (www.myendnoteweb.com). Removing 'English' language as a limit produced an extra 65 studies in the initial search result. It is anticipated that this would represent an even smaller number of papers meeting the inclusion and exclusion criteria. This was not deemed to be a sufficient addition to warrant approaching investigators with academic fluency covering these languages. This review was registered with Research Registry [2]. The full search strategy used in Pubmed is stated in box 1. The reference list was manually checked to exclude studies which did not meet the study criteria. The database search was carried out by two authors independently. The two investigators discussed differences to reach agreement. The study selection for the metanalysis is shown in figure 1.

Inclusion criteria was any paper presenting the outcome of using HMD-AR device to register digital images onto a target object. Papers were excluded if they used an additional device to achieve registration or did not use this to guide a procedure. Authors of excluded papers were not contacted for missing information. This review does not include any grey literature from technology companies/manufacturers.

A standard proforma was used to collect data on included papers, medical speciality to which the study applied, HMD used, study design, software used, registration method, registration error and other outcomes reported. Any unclear or missing data was confirmed with a second

author and marked as ‘not specified’ in the results table. All data was collected and statistics performed on Excel with Microsoft 365.

18 studies reported registration error for Hololens only systems. Where multiple mean registration errors were reported, the highest value was used. These papers are quantitatively analysed in the meta-analysis.

The papers in this study have wide methodological heterogeneity. A modification of the CASP diagnostic test tool was applied to assess risk of bias. This tool was adjusted to consider ‘accurate registration’ as the ‘correct diagnosis’. The assessment was performed by discussion amongst two authors. Question 5: Is the disease status of the tested population clearly described? does not apply and was removed. Question 7: What are the results? Is presented in Table 1. Question 12: What would be the impact of using this test on your patients/population? Was also omitted as it is elaborated in the discussion section of the paper.

Results

The initial search identified 4784 unique papers. Following screening, 23 papers were included for qualitative analysis. [3-25] Three of these papers included 2 independent studies with results which were analysed separately. [24] This gave a final count of 26 included studies (Table 1).

Figure 2 shows more papers are included in this study from recent years and none in 2015. There is an unexpected reduction of papers published in 2021. This may be due to the COVID-19 pandemic.

Medical specialities represented

Figure 4 shows the medical specialities represented in the studies included in this review. Spinal procedures have been separated as they represent a combination of neurosurgical and orthopaedic teams. OMFS, spine and orthopaedics represent the largest specialities.

HMD and software

The majority of studies used Hololens (Microsoft®) as the HMD of choice (22/26). The second version (Hololens2) is widely commercially available costing £3500 (quoted April 2022).

One study used the nVisor ST60 (1/26). [23] This is produced by NVIS®. Their HMD device costs £25611 (quoted April 2022). Three studies did not clarify the HMD used. [23-25]

There is a paucity of off-the-shelf AR registration software. This review found that the majority of studies developed their own software (16/26) designed for research use. This included the use of platforms including Unity, ARToolKit, Vuforia and Vertostudio. The only off-the-shelf software used for registration was opensight. Used by 2 separate teams in the USA working on percutaneous spinal procedure applications. [11,23]

One paper developed HuaxiAR1.0 software system for OMFS facial fracture repair application which is not yet publicly available. [22]

Study design

The Hololens studies can be divided into 2 main types. In the first type, the virtual image is projected onto the target (11/26). This registration is thus evaluated. The most common evaluation is to quantify the registration error in distance and orientation. These studies evaluated registration on an anatomical phantom (10/11) or patient (1/11).

The second study type (14/26), asks the user, wearing the HMD, to employ the registration to perform a procedure. These studies evaluate the performed procedure. The majority of these studies were performed on an anatomical phantom (7/14), animal (2/14), abstract shape (1/14), cadaver (1/14) or on a patient (3/14).

The majority of performed procedures involved placement of a needle or screw (8/14). In 3 studies, bone osteotomy was performed. [12,17,20] This can be considered to be a more complex procedure. In 1 study, the projected virtual image was used to mark the position of ear reconstruction. [6]

Registration method

Registration is superposition of the digital image exactly over its real world target. For example the images from the CT scan are aligned with the patient anatomy (target). This is measured in studies by comparing the position of the projected virtual image with its real world counterpart. The method by which this position is calculated is included in table 1.

Registration was achieved using a pattern tracking marker (16/26), target surface markers (5/26) and manually (5/26).

Pattern tracking and object markers are positioned in the real world adjacent to the desired position of the digital image. Pattern tracking markers are printed patterns such as a QR code. The Hololens software is programmed to recognise this pattern. The virtual image is thus presented to the viewer at a set position relative to this pattern. Markers of the printed pattern type have been developed in a variety of shapes and styles. This includes 3D cubes, sheets of multiple patterns and single 2D squares. There are no studies which have directly compared the different patterns. The markers can be placed anywhere adjacent to the area of interest. It is important to place the marker in the exact position in the real world as it is planned in the Hololens program. This has been achieved by attaching it to a reproducible anatomical site such as dental prosthesis. [21]

In contrast, surface markers exist directly on the digital image and are aligned with recipricol surface markers on the real world target. This includes markers placed on the target during the medical imaging or using anatomical landmarks. The latter is more effective when working on a body surface with distinct topography such as the face. It is not possible for the Hololens to recognise a specific location within a bland surface such as the abdomen. All studies using surface topography were carried out on anatomical phantom model or cadaver.

Manual registration was used in 5 studies. This requires the user, wearing the Hololens, to move the digital image and real world target until they are coincident.

Registration Error

Registration error is reported in 18 of the studies. The largest reported mean registration error and standard deviation can be seen in table 1.

The meta-analysis synthesised mean registration error across all 18 studies was 2.8mm (SD 2.2mm). Of these, 11 studies reported registration error using pattern markers with a mean error of 2.6mm (SD 1.8mm SD). Four studies reported registration error in studies using surface markers, with a mean error of 3.8mm (SD 3.7mm). Three studies reported registration error using manual alignment to register with a mean error of 2.2mm (SD 1.34mm).

The heterogeneity in study design and varying number of studies in each subgroup is the likely cause of the relative large standard deviation. However, it is illustrative to see the scale of errors is within millimetres.

Other measured outcomes

Three studies had an endpoint of feasibility. [22,24] They all concluded the use of HMD-AR for their procedure was feasible and 1 found that the use of HMD-AR reduced surgical time. [22]

Four studies evaluated the outcome in clinical terms. All of these were percutaneous vertebral screw placement and reported a 94-97.5% screw placement acceptability compared to 100% in their gold standard. [5,11,19,23]

Assessment of risk of bias

The results of the assessment are presented in Table 2. No paper was excluded based on risk of bias assessment. All papers presented a clear aim. The reference standard and methodology was well explained in most papers. None of the papers discussed the outcomes important to the individual population or comprehensive discussion of alternatives. This reflects the pre-clinical stage of the studies. However, the protocols developed can all be applied to patients in the population of interest.

Two studies declared a conflict of interest. Gibby et al included an author employed by Novorad who produce Opensight, the software used in the study. [26] Muller et al included an author who founded Incremed AG, a Balgrist University Hospital start-up providing the software. [16]

Discussion

Performing AR guided medical procedure is increasing in popularity. This is shown by the increasing number of publications over the last few years. Hololens is by far the most popular device used. Hololens comes at a cost of £3500 compared to £25611 for nVisor ST60. The authors are not aware of any significant difference between the 2 devices. However, the larger scale production, wider availability and lower cost of Hololens may explain its popularity. Only 1 study reported registration for the nVisor device at 2.47mm SD 0.66mm. No statistical inference can be made but this is comparable to results from Hololens studies.

In contrast there is no established popular software with most studies developing their own in-house software. This is likely to contribute to the wide variation in reported registration error evidenced by the large Standard Deviation (SD) of the calculated mean registration error of all the studies (2.9mm SD2.1mm). The majority of software designs used pattern markers. This achieved equivocal registration error to manual alignment (2.7mm SD1.8mm vs 2.9mm SD0.06mm). The smaller SD for manual alignment is likely to represent the inclusion of only 2 studies. With pattern tracking achieving a similar registration error to manual alignment, the suggestion is that registration error is limited by a factor other than registration method. This may be the limitations of the Hololens device or AR perception of the user. [27]

However, this does set a standard minimum registration error that any HMD-AR system should achieve. Furthermore, an error of 2-3mm may well be within acceptable limits of many medical procedures. This is evidenced by the high vertebral screw placement acceptability achieved when HMD-AR systems. [5,11,19,23] The acceptable registration error would vary for each procedure and would need to be established and compared to what HMD-AR can achieve.

In this review, studies which incorporate an external device to achieve registration were excluded. These external devices, such as intra-operative imaging, are already in common surgical practice. The registration in these AR set-ups is dependent on the accuracy of the external device. External registration can only compare a position with that on a map. They cannot add annotations to the surgical field. To overcome this, AR can be added in combination. Studies reporting registration error using AR with infra-red navigation, magnetic

navigation and c-arm radiographs have reported mean registration errors of 3.9mm, 4.3mm and 5.7mm respectively. [28-30] These are commensurate to the values reported in this review. However, one such study combined fluoroscopy with the Hololens for registration. Fluoroscopy images were used to attain uptodate positional data on the patient's spine. The AR images were aligned using positional data of the spine. They report a mean registration error of 0.425mm SD 0.021mm. This is significantly less than any study included in this review. [31] When compared to such registration errors, the only advantage of HMD-AR systems are their ease of set-up and relative reduced cost. These parameters were not examined in this study and would be an area of great interest.

The majority of studies were in the surgical field of maxillofacial (8/40), orthopoedics (6/40) and spinal surgery (8/40). These specialities represent hard tissue (bone) procedures. Registration of digital images onto the target object requires pre-procedure imaging such as CT to be the same as the target object during the procedure. Hard tissue such as bone has the fortunate characteristic of not being easily deformed. Therefore, registration of hard tissue can be more readily achieved than soft tissue. Soft tissue deformation is difficult to predict. Currently the best way to overcome soft tissue deformation is to incorporate intra-procedure imaging using an external device such as c-arm or fluoroscopy. [28-30] This would further add to the advantage of using an external device to reduce registration error.

The main limitation of this review is the heterogeneity of the study methodologies. This varies from human trial to phantom models. Similarly, the reference used to measure registration accuracy differs. In most cases, HoloLens is used as the AR-HMD device. However, different studies have used either HoloLens generation 1 or 2. In addition, every study uses a different software. This is clear by the variation in accuracy even across studies applying the same registration method.

This review does not provide a final accuracy report on the use of the AR-HMD guided surgery. It is however a starting point. Whilst the accuracies reported vary, they show that a sub 5mm accuracy can be achieved. This is the standard of the current technology. This may further improve in the future as devices and software development continues. However, the accuracy we can achieve today may meet the requirement for many surgical applications. The authors hope that this review will act to motivate the surgical community to explore these applications.

References

1. McCulloch P. The IDEAL framework for ensuring safety and effectiveness of medical devices. BMJ. 2020;370:m3183.
2. Chegini S. REGISTRY OF SYSTEMATIC REVIEWS/META-ANALYSES DETAILS. 2022; https://www.researchregistry.com/browse-the-registry#registryofsystematicreviewsmeta-analyses/registryofsystematicreviewsmeta-analysesdetails/62d03ae8c2e109001e6995f4/. Accessed 14/07/2022.
3. Nguyen NQ, Cardinell J, Ramjist J, et al. Augmented reality systems for improved operating room workflow. Clinical Neurosurgery. 2019;66 (Supplement 1):101.
4. Mitsuno D, Ueda K, Itamiya T, et al. Intraoperative Evaluation of Body Surface Improvement by an Augmented Reality System That a Clinician Can Modify. Plastic and Reconstructive Surgery-Global Open. 2017;5.
5. Agten CA, Dennler C, Rosskopf AB, et al. Augmented Reality-Guided Lumbar Facet Joint Injections. Invest Radiol. 2018;53:495-498.
6. Nuri T, Mitsuno D, Otsuki Y, et al. Augmented Reality Technology for the Positioning of the Auricle in the Treatment of Microtia. Plastic and Reconstructive Surgery-Global Open. 2020;8.
7. Frantz T, Jansen B, Duerinck J, et al. Augmenting Microsoft's HoloLens with vuforia tracking for neuronavigation. Healthc Technol Lett. 2018;5:221-225.
8. Jiang T, Yu D, Wang Y, et al. HoloLens-Based Vascular Localization System: Precision Evaluation Study With a Three-Dimensional Printed Model. J Med Internet Res. 2020;22(4):e16852.
9. Moreta-Martinez R, García-Mato D, García-Sevilla M, et al. Augmented reality in computer-assisted interventions based on patient-specific 3D printed reference. Healthc Technol Lett. 2018;5:162-166.
10. Zhou ZY, Yang ZY, Jiang S, et al. Design and validation of a surgical navigation system for brachytherapy based on mixed reality. Medical Physics. 2019;46:3709-3718.
11. Gibby JT, Swenson SA, Cvetko S, et al. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. International Journal of Computer Assisted Radiology and Surgery. 2019;14:525-535.
12. Gao Y, Lin L, Chai G, et al. A feasibility study of a new method to enhance the augmented reality navigation effect in mandibular angle split osteotomy. J Craniomaxillofac Surg. 2019;47:1242-1248.
13. Liu H, Auvinet E, Giles J, et al. Augmented Reality Based Navigation for Computer Assisted Hip Resurfacing: A Proof of Concept Study. Ann Biomed Eng. 2018;46:1595-1605.
14. van Doormaal TPC, van Doormaal JAM, Mensink T. Clinical Accuracy of Holographic Navigation Using Point-Based Registration on Augmented-Reality Glasses. Oper Neurosurg (Hagerstown). 2019;17:588-593.
15. Liebmann F, Carrillo F, Farshad M, et al. First experiences of using mixed-reality for surgical navigation of corrective osteotomies. Swiss Medical Weekly. 2018;148:29S-29S.

16. Muller F, Roner S, Liebmann F, et al. Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging. Spine Journal. 2020;20:621-628.
17. Viehöfer AF, Wirth SH, Zimmermann SM, et al. Augmented reality guided osteotomy in hallux Valgus correction. BMC Musculoskelet Disord. 2020;21:438.
18. Jiang JG, Huang ZY, Qian W, et al. Registration Technology of Augmented Reality in Oral Medicine: A Review. Ieee Access. 2019;7:53566-53584.
19. Liu H, Wu J, Tang Y, et al. Percutaneous placement of lumbar pedicle screws via intraoperative CT image-based augmented reality-guided technology. Journal of Neurosurgery: Spine. 2020;32:542-547.
20. Zhu M, Liu F, Zhou C, et al. Does intraoperative navigation improve the accuracy of mandibular angle osteotomy: Comparison between augmented reality navigation, individualised templates and free-hand techniques. J Plast Reconstr Aesthet Surg. 2018;71:1188-1195.
21. Wang J, Shen Y, Yang S. A practical marker-less image registration method for augmented reality oral and maxillofacial surgery. International Journal of Computer Assisted Radiology and Surgery. 2019;14:763-773.
22. Chen G, Zeng W, Yin H, et al. The Preliminary Application of Augmented Reality in Unilateral Orbitozygomatic Maxillary Complex Fractures Treatment. J Craniofac Surg. 2020;31:542-548.
23. Urakov TM, Wang MY, Levi AD. Workflow Caveats in Augmented Reality-Assisted Pedicle Instrumentation: Cadaver Lab. World Neurosurg. 2019;126:e1449-e1455.
24. Rose AS, Kim H, Fuchs H, et al. Development of augmented-reality applications in otolaryngology-head and neck surgery. Laryngoscope. 2019;129:S1-S11.
25. Pepe A, Trotta GF, Mohr-Ziak P, et al. A Marker-Less Registration Approach for Mixed Reality-Aided Maxillofacial Surgery: a Pilot Evaluation. J Digit Imaging. 2019;32:1008-1018.
26. Gibby J, Cvetko S, Javan R, et al. Use of augmented reality for image-guided spine procedures. European Spine Journal. 2020;29:1823-1832.
27. Condino S, Carbone M, Piazza R, et al. Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. Ieee Transactions on Biomedical Engineering. 2020;67:411-419.
28. Meulstee JW, Nijsink J, Schreurs R, et al. Toward Holographic-Guided Surgery. Surgical Innovation. 2019;26:86-94.
29. Kuhlemann I, Kleemann M, Jauer P, et al. Towards X-ray free endovascular interventions - using HoloLens for on-line holographic visualisation. Healthc Technol Lett. 2017;4:184-187.
30. Andress S, Johnson A, Unberath M, et al. On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. Journal of Medical Imaging. 2018;5.
31. Liu J, Al'Aref SJ, Singh G, et al. An augmented reality system for image guidance of transcatheter procedures for structural heart disease. PLoS One. 2019;14:e0219174.

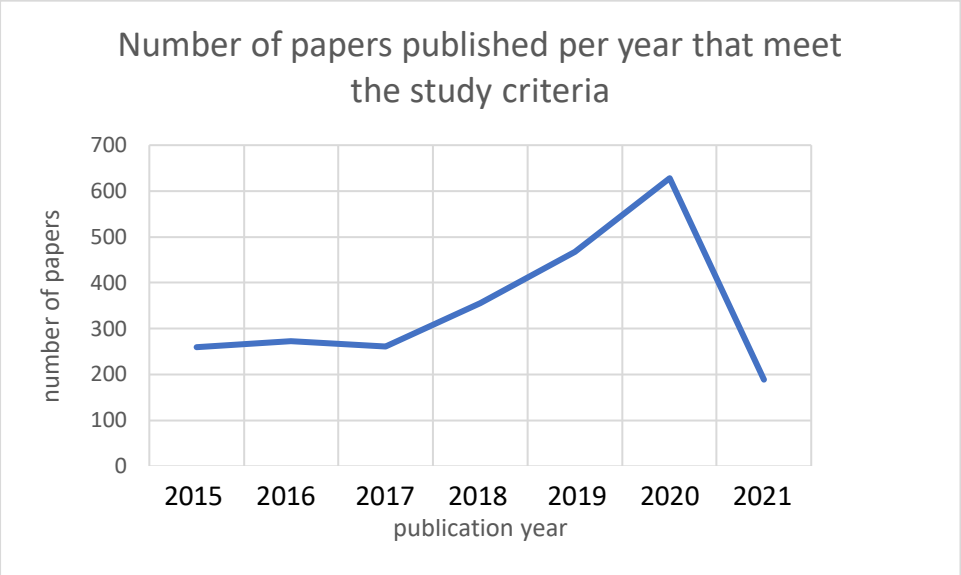

Figure 2: Number of papers published per year that meet the study criteria

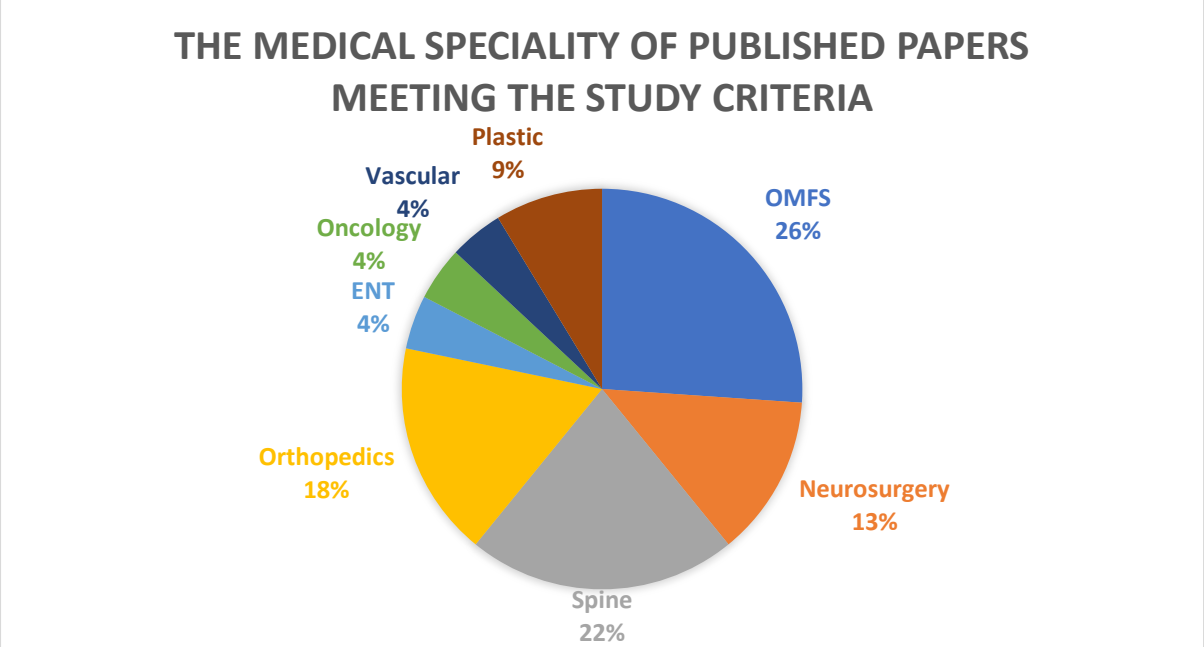

Figure 3: the medical speciality of published papers meeting the study criteria

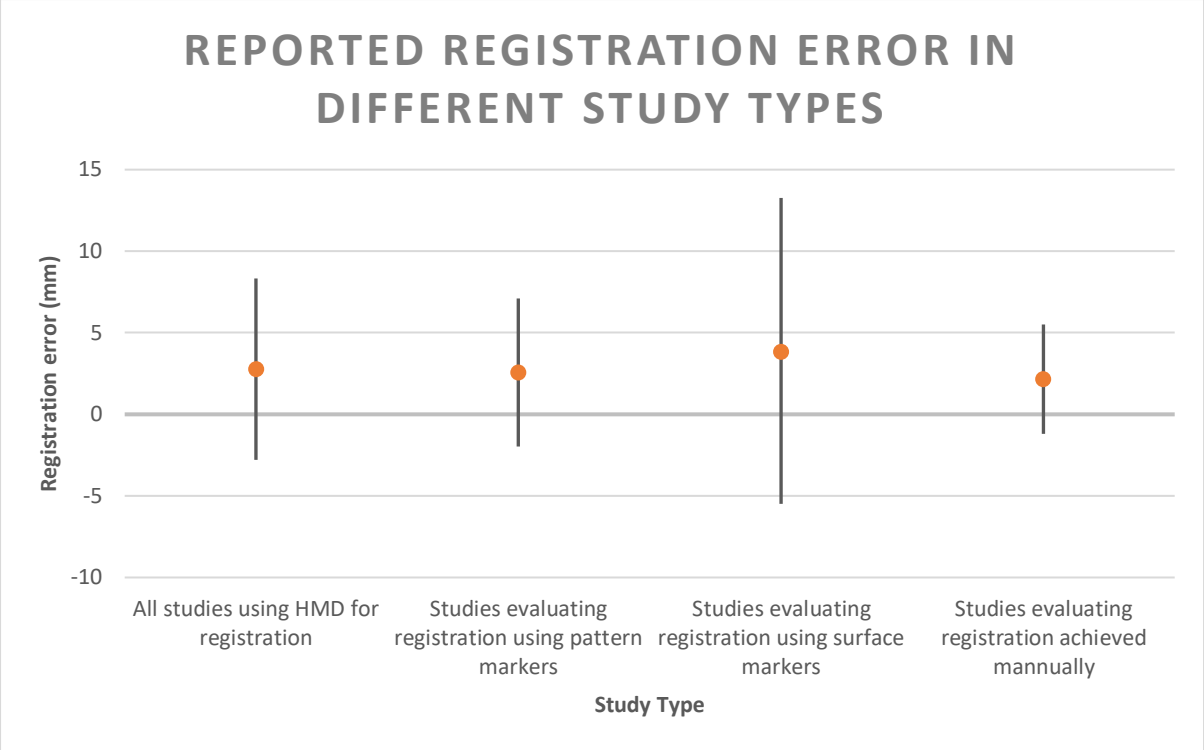

Figure 4: Reported registration error in different study types. The calculated mean is marked by the horizontal line. The vertical lines represent the mean +/- 2.5 SD of the means reported across all included studies.

Search terms: augmented, mixed, real*, surg*
Limit 2015-2020, English language, full text available

Figure 1. Study selection for the meta-analysis

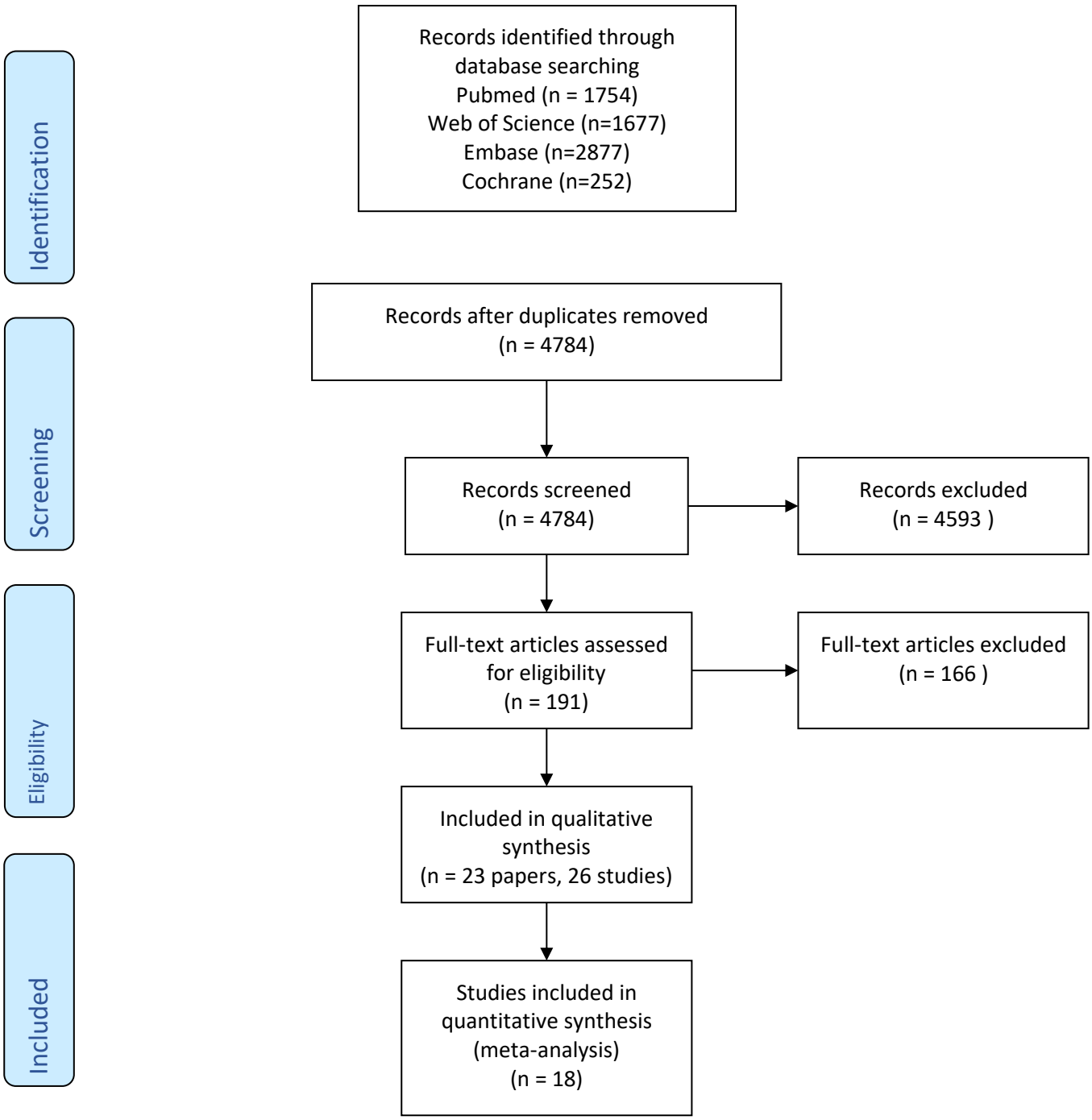

Table 1: Results of papers which meet the inclusion and exclusion criteria of this study

First Author, year	Medical Speciality	HMD	software	Registration method	Method: performing procedure or evaluating registration	Subject of study	Registration error mean (mm)	SD of Registration error mean (mm)	orientation error mean (deg)	orientation error SD (deg)	Other validation outcome
Nguyen, 2019	Neurosurgery	HoloLens	own	Manual	evaluating registration by measuring alignment of virtual object with respect to the real object on recorded views	phantom	2.9	1.8	6.8	5.9	
Mitsuno, 2018	Plastic	HoloLens	not specified	Manual	evaluating registration by measuring alignment of virtual object with respect to the real object on recorded views	phantom	2.98				
Nuri, 2019	Plastic	HoloLens	not specified	Manual	microtia repair	patient					Traced image by AR 1-2cm shorter than by paper tracing
Agten, 2018	Orthopaedics	HoloLens	Not specified	manual	percutaneous spinal procedures	phantom					97.5% of AR placed screws acceptable

Frantz, 2018	Neurosurgery	HoloLens	own	manual	evaluating registration by quantifying the location of a surface point on the phantom based on holographic markers	phantom	0.62	3.34			
Frantz, 2018	Neurosurgery	HoloLens	own	Pattern marker	evaluating registration by quantifying the location of a surface point on the phantom based on holographic markers	phantom	1.24	1.08			
Jiang, 2020	Vascular	HoloLens	own	pattern marker	evaluating registration by quantifying the location of points on the 3D printed model compared to the virtual model	phantom	1.35	0.43			
Moreta-Martinez, 2018	Orthopaedics	HoloLens	own	pattern marker	evaluating registration using an optical tracking system to measure the position of reference	phantom					Average root-mean error 2.9mm

					points on virtual image						
Gibby, 2018	Spine	HoloLens	Opensight	pattern marker	Percutaneous spine procedures	phantom	6.93				97% of AR placed screws acceptable
Zhou, 2019	oncology	HoloLens	own	pattern marker	brachytherapy needle	object	0.664		4.74		
Gao, 2019	OMFS	HoloLens	own	pattern marker	mandible angle osteotomy	phantom	1.89	0.51	2.03	1.15	
Zhou, 2019	oncology	HoloLens	own	pattern marker	brachytherapy needle	animal					
Liu, 2018	Orthopaedics	HoloLens	not specified	pattern marker	Percutaneous spine procedures	phantom	1.91	0.89	2.14	0.81	
Van Doormaal, 2018	Neurosurgery	HoloLens	own	pattern marker	evaluating registration by quantifying the location of points on patient head compared to the virtual image	patient	4.4				
Liebmann, 2019	Spine	HoloLens	own	pattern marker	percutaneous spinal procedures	phantom	2.77	1.46	3.38	1.73	
Muller, 2020	Spine	HoloLens	own	pattern marker	percutaneous spinal procedures	phantom	3.4	1.6	3.5	1.4	
Liu, 2020	Orthopaedics	HoloLens	own	pattern marker	percutaneous spinal procedures	phantom			4.9		94% of AR placed screws acceptable
Viehofer, 2020	Spine	HoloLens	not specified	pattern marker	hallux valgus osteotomy	phantom					
Jiang, 2019	OMFS	nVisor ST60	Own	pattern marker	Drill holes into mandible	animal	2.47	0.66	1.32	1.17	

Chen, 2020	OMFS	not specified	Huaxi AR	pattern marker	Facial Fracture Repair	patient					Feasibility study, showed shorter operation time
Zhu, 2018	OMFS	not specified	Own	pattern marker	Mandible angle osteotomy	patient	1.18	0.34			
Wang, 2017	OMFS	not named	own	surface anatomy (teeth)	evaluating registration by measuring alignment of virtual object with respect to the real object on recorded views	phantom	1	0.59			
Urakov, 2019	Spine	HoloLens	Opensight	surface anatomy	percutaneous spinal procedures	cadaver	2.5	0.44			7 screws placed in unacceptable position
Rose, 2019	ENT	HoloLens	own	surface markers	evaluating registration by quantifying the position of phantom compared to the virtual image	phantom	2.47	0.46			
Pepe, 2019	OMFS	HoloLens	not specified	surface markers	evaluating registration by quantifying the location of points on phantom compared to	phantom	9.3	6.1			

					the virtual image						
Rose et al, 2019	ENT	HoloLens	own	surface markers	foreign body removal	phantom					feasible

Table 2: CASP diagnostic test study

First Author, year	Was there a clear question for the study to address?	Was there a comparison with an appropriate reference standard?	Did all test subjects get the diagnostic test and reference standard?	Could the results of the test have been influenced by the result of the reference standard?	Were the methods for the test described in sufficient detail?	How sure are we about the results? Consequences and cost of alternatives performed?	Can the results be applied to your patients/the population of interest?	Can the test be applied to your patient or population of interest?	Were all outcomes important to the individual population considered?
Jiang, 2020	HIGH	HIGH	HIGH	HIGH	High	LOW	HIGH	HIGH	LOW
Frantz, 2018	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Rose, 2019	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW
Nguyen, 2019	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Moreta-Martinez, 2018	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW
Mitsuno, 2018	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW
Van Doormaal, 2018	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Pepe,	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW

2019									
Gibby, 2018	HIGH	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW
Liu, 2018	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Zhou, 2019	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Gao, 2019	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Liebmann, 2019	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW
Muller, 2020	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Urakov, 2019	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Nuri, 2019	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW
Viehofer, 2020	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Liu, 2020	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Agten, 2018	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Jiang, 2019	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Wang, 2016	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Chen, 2020	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW
Zhu, 2018	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW

Box1: Full search strategy used in Pubmed

((((((augmented) OR (mixed)) AND (real*))) AND (surg*))) AND (("2015/01/01"[Date - Publication] : "2021/01/01"[Date - Publication]))) AND (english[Language])