COMPARISONS OF ORTHODONTIC RESIDENTS' PERFORMANCE AND ATTITUDES USING 2D, 3D, AND VIRTUAL REALITY SURGICAL SIMULATION METHODS

Lauren E. Sytek[1], Marita R. Inglehart[2], Vidya Ramaswamy[3], Sharon Aronovich [4], Sean Edwards[5], and Hera Kim-Berman[6]

[1]Department of Orthodontics and Pediatric Dentistry, School of Dentistry, The University of Michigan, Ann Arbor, Michigan, USA
[2]University Diversity and Social Transformation Professor, Professor of Dentistry, Department of Periodontics and Oral Medicine, School of Dentistry & Adjunct Professor of Psychology, College of Literature, Science & Arts (LS&A), The University of Michigan, Ann Arbor, Michigan, USA
[3]Associate Director of Curriculum and Program Evaluation, School of Dentistry, The University of Michigan, Ann Arbor, Michigan, USA
[4]Clinical Associate Professor, Department of Oral and Maxillofacial Surgery, School of Dentistry, The University of Michigan, Ann Arbor, Michigan, USA
[5]Clinical Professor, Department of Oral and Maxillofacial Surgery, School of Dentistry, The University of Michigan, Ann Arbor, Michigan, USA
[6]Clinical Associate Professor, Program Director - Orthodontics, Department of Orthodontics and Pediatric Dentistry, School of Dentistry, The University of Michigan, Ann Arbor, Michigan, USA

Funding Sources: The University of Michigan Le Gro Fund; Oculus Education Group

Email Addresses:

Lauren Sytek	bauerlau@umich.edu
Marita R. Inglehart	mri@umich.edu
Vidya Ramaswamy	ramaswav@umich.edu
Sharon Aronovich	saronovi@umich.edu
Sean Edwards	seanedwa@umich.edu
Hera Kim-Berman	bermanh@umich.edu

Corresponding author:
Hera Kim-Berman, DDS, MMSc
Department of Orthodontics and Pediatric Dentistry
The University of Michigan School of Dentistry
1101 N. University Avenue
Ann Arbor, MI 48109-1078
Tel: (734) 764-1520 Fax: (734) 763-8100
bermanh@umich.edu

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jdd.12598.

Abstract

Purpose: Advances in virtual reality technology for surgical simulation methods may improve diagnosis and treatment planning of complex orthognathic surgery cases. The objective was to assess orthodontic residents' performance and attitudes when treatment planning orthognathic surgery cases using 2D digital, 3D digital, and virtual reality surgical simulations.

Methods: The study had a mixed methods study design involving 20 graduate orthodontic residents. Their previous experiences, confidence, and competence with orthodontic diagnosis and surgical treatment planning was assessed with a baseline survey. Each resident completed 2D, 3D, and VR treatment planning and simulation tasks in a randomized order and recorded their diagnosis, objectives, treatment plan, and special surgical concerns for each case using a treatment planning worksheet. The worksheets were scored and quantitative data were analyzed. Attitudinal responses to the simulation experience were captured with a post-survey and interview.

Results: The number of total prescribed surgical movements was greater for 3D and VR simulation methods ($p = 0.001$). There were no differences in the overall total written treatment plan analysis score among the three surgical simulation tasks. Participants took longer to complete the VR and 3D tasks ($p < 0.001$) and asked more questions regarding manipulation ($p < 0.001$) and software features ($p < 0.002$) for higher fidelity tools. Analysis of qualitative feedback showed positive attitudes towards higher fidelity tools in regards to visualization, manipulation, and enjoyment of the task.

Conclusions: The results demonstrate that simulation methods of increased fidelity (3D and VR) are appropriate alternatives to 2D conventional orthognathic surgical simulation methods when combined with traditional records. Qualitative feedback confirms residents' readiness to adopt VR simulation. However, comprehensive training is needed to increase familiarity and comfort with using the new technology.

MeSH key words:

Virtual Reality	Simulation	Resident Education	Orthodontics
Diagnosis	Treatment Planning	Orthognathic Surgery	

INTRODUCTION

Historically, oral maxillofacial surgeons and orthodontists planned orthognathic surgery based on acetate tracings of a patient's lateral cephalogram. This technique allowed to rearrange the structures to simulate surgical outcomes.[1,2] With the introduction of digital radiographs and software analysis tools, surgical simulations can now be completed digitally without the need for manual tracings.[1] Like acetate tracings, the digital method falls short in that only sagittal and vertical corrections can be simulated.[1]

The introduction of cone beam computed tomography (CBCT) has paved the way for significant improvements in surgical planning with the use of computer-assisted surgical simulation (CASS) and three-dimensional (3D) patient data.[3] The CBCT images capture both hard and soft tissue which can be used to assess resultant facial changes after skeletal manipulation. Although similar to 2D simulation, the accuracy for lip and chin changes needs improvement.[4,5] Planning orthognathic surgeries using CASS has positively impacted patient care by lessening the ambiguity of 2D planning and simplifying surgical procedures by use of customized 3D printed surgical guides and splints.[6,7] Previously, when 2D treatment plans were executed intraoperatively, unexpected surgical complications could arise such as bony collisions, rotational axis discrepancies, or residual chin inadequacies.[8] CASS and 3D imaging have helped to revolutionize orthognathic surgical planning by minimizing these unexpected outcomes.[8] Although CASS is highly regarded as a treatment modality, the propriety software is too cost prohibitive to allow for multiple simulations for diagnostic purposes or as an educational tool.

More recently, emerging technology platforms such as virtual reality (VR) are being integrated in medical and dental education.[9] VR allows users to interact in an entirely computer generated virtual world. VR can be fully immersive using a head mounted device (HMD) or non-immersive using 2D monitors or stereoscopic glasses. Applications in medical virtual reality include simulation surgery for tendon transplants, abdominal surgery and virtual endoscopy among others.[10,11,12] Other medical applications include interpersonal communication training, anatomy instruction, exposure therapy and pain management.[13-17] Virtual reality educational tools for maxillofacial surgery planning have been developed.[18-22] However, the available tools are few in number and the use of virtual reality technology in dentistry is still in its infancy.

The need for a high-fidelity patient model that simulates orthognathic procedures for educational purposes has helped drive the development of the Virtual Reality Patient Model (VRPM).[23] The VRPM was developed at the University of Michigan, Duderstadt Center, using Jugular software, the university's in-house VR platform. Contrary to previous computer-based scenario assessments, the VRPM is a relatively low-cost, portable system.[24] The portable workstation includes a computer with the appropriate hardware specifications to support the Oculus Rift HMD, sensors, and Touch controllers (Oculus VR Inc., Menlo Park, CA). The Oculus Rift HMD has a microcontroller, gyroscope, accelerometer, magnetometer, and infrared LED sensors that provides sensor based head and positional tracking.[25] Touch controllers feature capacitive sensors that detect how the controllers are being held and enables the user to have virtual hands that are responsive in the immersive environment.[26]

Recent studies report that compared to other medical surgical disciplines, there is limited use, underuse and lack of validated studies of simulated clinical teaching for maxillofacial surgery education.[27, 28] The objective of this study was to assess orthodontic residents' performance and attitudes when treatment planning orthognathic surgery cases using 2D digital, 3D digital, and virtual reality surgical simulations.

METHODS

Institutional Review Board (IRB) approval was obtained from the University of Michigan IRB (#HUM00144492, #HUM00145046) to include graduate level residents as study participants and grant access to patient data from January 1, 2007 through March 31, 2018, for the creation of VRPMs of varying complexity from clinical cases.

The study used a convergent mixed methods design with twenty graduate level orthodontic residents (n=20) (Figure 1). Each session began with completion of a baseline survey to assess general demographic data and previous surgical simulation experiences.

Three potential cases were identified and determined to be of similar surgical complexity by evaluating the case records, the American Board of Orthodontics' (ABO) Board Case Oral Examination (BCOE) worksheet keys, and the Grummons cephalometric analysis for each case.[29, 30] The cases were then randomly assigned to a treatment simulation method (2D, 3D, and VR) (Figure 2). The patient case assigned to each method remained consistent across all participants. However, the order in which the participants completed the 2D, 3D, and VR tasks was randomized.[31]

The 2D task was completed using Dolphin Imaging: Treatment Simulation software (Dolphin Imaging and Management Solutions, Version 11.8, Chatsworth, CA) on a desktop computer. Participants used the mouse and keyboard to manipulate the digital cephalometric tracing. The 3D simulation task was also completed on a desktop computer using the Dolphin Imaging: 3D Surgery treatment module (Dolphin Imaging and Management Solutions, Version 11.8, Chatsworth, CA). Participants used the mouse and keyboard to translate and rotate the CBCT segmented skeletal structures. The VR simulation was completed in a seated position within an immersive environment using the Oculus Rift HMD and Touch controllers to view and manipulate the VRPM.

Before completing each of the simulation tasks, the subjects participated in a training module to familiarize them with each software's functionality. The participants listened to a standardized training script with instructions for viewing the patient data and performing the treatment simulation. The total time spent in the training module was recorded and the frequency and content of the participants' questions during the training were recorded. All sessions were recorded using Open Broadcaster Software (OBS https://obsproject.com/).

For each simulation task, participants received a lateral cephalogram, a limited set of cephalometric measurements, de-identified facial photos, digital model photos, and a panoramic radiograph. Participants completed a modified version of the ABO's Board Case Oral Examination (BCOE) worksheet to document their diagnosis, objectives, plan, and any surgical considerations for the case (i.e. bony overlap, residual asymmetry). The total time to complete each task and the frequency and content of the participants' questions were recorded.

The BCOE worksheet analysis was reported as the percentage of correctly identified worksheet key items. Where applicable, participants received one point for identifying a problem and a second point for identifying its severity. For severity, a response was accepted if it fell within the defined ranges of mild (1-3mm), moderate (>3-6mm), and severe (>6mm).[32] The surgical simulation was analyzed by tallying the number of prescribed linear and rotational movements for each jaw. Retest intra-rater reliability testing was determined three weeks after the initial scoring. Inter-rater reliability of the BCOE worksheet scoring method was previously validated as a testing instrument by Sakowitz *et al*.[33]

A semi-structured interview was conducted and audio taped post-intervention to collect open-ended feedback about the residents' experiences. The recorded text was transcribed using the Otter application (Otter.ai Version 2.1.7-1613, Los Altos, CA).

Quantitative and Qualitative Data Analysis

The quantitative data were analyzed using SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.). Descriptive statistics such as means, standard deviations and frequency distributions were calculated to provide an overview of the survey responses and performances on each simulation task. Repeated measurement analyses of variance were used to compare the respondents' performances on the three simulation tasks that were measured with continuous variables. Tukey's post hoc tests were used when the univariate analyses showed that the three simulation task responses differed significantly. Chi-square analyses were used to analyze categorical outcome variables. Intra-class correlation coefficients (ICC) were calculated for intra-rater reliability. A significance level of $p < 0.05$ was accepted for all analyses.

The qualitative free response data were coded by first categorizing each response as either positive or negative about each method and then assigning a code based on the remark's content. The codes used for analysis were the same as those identified in a previous study by Kim-Berman *et al.*[34] An additional code, "information" was added to the codes defined by Kim-Berman *et al.* for statements such as, "I have limited information with this tool." The final coding for all interview statements was reviewed by a second investigator. Discrepancies were discussed and resolved.

RESULTS

Baseline Survey Results

The sample had nearly equal proportions of male (55%) and female (45%) participants in each of the three residency years (1st year: 35%; 2nd year: 35%; 3rd year: 30%). At baseline, participants had the most experience with 2D simulation and reported being most competent using 2D surgical simulation tools (on a scale from 1 = lowest to 5 = highest: mean 2D = 3.30, mean 3D = 1.60, mean VR = 1.55; $p \leq 0.001$).

Training and Test Time

There was a difference among the three methods in the amount of time spent in the training modules (mean 2D = 2.13 minutes; mean 3D = 3.95 minutes; mean VR = 6.59 minutes; $p \leq 0.001$) as well as time spent in minutes to complete each test task (mean 2D = 19.25, mean 3D = 24.71, mean VR 29.29; $p \leq 0.001$) (Table 1).

Training and Test Questions

Participants asked more questions during the VR training module compared to the 2D module (mean 2D = 0.45, mean VR = 1.70, p = 0.012). Participants also asked more questions regarding manipulation in the VR training module compared to both the 2D and 3D training modules (mean 2D = 0.25, mean 3D = 0.30, mean VR = 1.10; p = 0.006) (Table 1). During the test session, participants asked more questions during the 3D and VR tasks compared to the 2D task (mean 2D = 1.80, mean 3D = 3.90, mean VR 4.70; p = 0.002). Participants asked more questions regarding software features during the 3D and VR tasks (mean 2D = 0.15, mean 3D 1.15, mean VR 1.30; p = 0.001) (Table 1). There were differences between all tasks in the number of test questions regarding manipulation (mean 2D = 0.00, mean 3D = 1.35, mean VR = 0.55l; $p \leq 0.001$) (Table 1).

BCOE Worksheet Written Case Analysis

The scoring of the BCOE worksheet showed differences in the skeletal diagnosis, dental diagnosis, and special surgical considerations worksheet categories. However, the post hoc tests did not detect significant differences for pairwise comparisons (Table 2). Skeletal diagnosis scores (mean 2D = 56.88%, mean 3D = 66.25%, mean VR =58.89%; p = 0.038) and surgical considerations scores (mean 2D = 30.00%, mean 3D = 42.50%, mean VR =27.50%; p = 0.028) were higher for the 3D task compared to the 2D and VR tasks (Table 2). Dental diagnoses scores were higher for the 2D task than the 3D and VR tasks (mean 2D = 61.58%, mean 3D = 53.61%, mean VR =57.06%; p = 0.041). Participants scored higher on the skeletal objectives for the VR task compared to the 2D task (mean 2D = 46.88%, mean 3D = 53.13%, mean VR = 58.13%; p = 0.008).

A higher percentage of participants was able to identify the mandibular asymmetry severity using the 3D simulation tool compared to the 2D and VR tools (2D = 40%, 3D=80%, VR =50%; p = 0.029) (Table 2). All participants correctly identified the presence of a mandibular asymmetry and its direction every time. There were no differences in the participants' ability to identify the transverse discrepancy or either of the special surgical considerations (i.e. residual soft tissue asymmetry, bony interferences/gaps) (Table 2).

Intra-class correlations (ICC) for the intra-rater reliability of the BCOE worksheet scoring ranged from 0.757-1.000. All ICCs were significant at p <0.001.

2D, 3D, and VR Simulation Analysis

There were differences in the number of total movements prescribed for the 2D, 3D and VR simulations (mean 2D = 4.00, mean 3D = 5.15, mean VR = 5.85; p = 0.001) (Table 3). A higher percentage of participants treatment planned maxillary pitch movements for the VR case compared to the 2D and 3D case (2D = 30%, 3D = 45%, VR = 80%; p = 0.005) (Table 3). Participants planned more mandibular sagittal movements for the 3D case compared to the 2D and VR case (2D = 40%, 3D = 90%, VR = 65%; p =0.004). A higher percentage of participants treatment planned mandibular pitch movements for the VR case (2D = 25%, 3D = 30%, VR = 70%; p = 0.007) (Table 3).

Participant Interviews

When the participants were asked about the value and accuracy of each simulation method, the majority of responses were positive towards the VR tool and centered on themes of enhanced visualization and manipulation. When asked about the difficulty of forming a diagnosis, the majority of responses were negative towards 2D, with nearly all comments attributing the difficulty to having a limited view. For the difficulty in forming a treatment plan based on simulation, participants responded negatively towards the VR tool, with comments centered on the sensitivity of manipulation. An equal number of positive comments for both 3D and VR were documented for the most preferred treatment tools. The positive 3D comments focused on enhanced soft tissue visualization and the positive VR comments centered on its perceived accuracy. For their least preferred method selection, the majority of comments were negative towards the limitations of the 2D simulation. For the most enjoyable method, a majority of the participants responded positively to VR, describing it as a "cool", "interactive", and "immersive" environment.

The significant quantitative findings were merged and integrated with the qualitative remarks to provide insights into how user experiences may potentially impact written analysis performance (Table 4).

DISCUSSION

With the enhanced features of higher fidelity tools, we expected that residents' written analysis performance would be higher with increased fidelity. It was observed that participants tended to rely on the provided records rather than using the simulation tool for their diagnosis, treatment objectives and treatment plan. If participants had not be supplied with a full set of records, there may have been greater observed difference in the worksheet scores. This emphasizes the importance of having complete patient records for proper diagnosis and treatment planning.

Participant interview responses provided little support for the 3D tool. However, the participants scored significantly higher on the skeletal diagnosis and special surgical considerations for the 3D case. They were also significantly better at identifying the asymmetry severity using the 3D tool. These findings suggest that even though participants did not favor 3D as a treatment simulation tool, it may provide superior diagnostic value compared to 2D simulation. This conclusion was further supported by positive 3D comments in the qualitative feedback about superior visualization.

A specific point of interest was to look at the residents' ability to recognize special surgical considerations for asymmetric cases including residual soft tissue asymmetry after completing bony changes and issues with bony overlap that can occur when correcting yaw deformities. We expected that with high fidelity these issues would become clear. However, even with standardized prompting, a low percentage of participants correctly recognized either of the considerations. One potential explanation could be the lack of movement restrictions in all of the simulation software programs allowing participants to simulate "ideal" outcomes, regardless of the surgical feasibility. Participants cited they had "too much free reign". This statement might indicate that they may have understood what they were simulating may not be realistic. However, they lacked the surgical experience to make appropriate adjustments. A future study could investigate oral surgery residents' performance using the VRPM to see if they would be better able to recognize these surgical considerations.

The intra-rater reliability ICC values were all well above 0.75 which is considered a good indication of reliability.[35] The BCOE worksheet keys leave little room for ambiguity and the grading was well calibrated.

Training and test time trended upwards with increasing fidelity likely because residents were most familiar with the 2D method at baseline. Despite having similar exposure to 3D and VR simulation methods at baseline, the participants spent significantly more time in the VR training module and on the VR task. The qualitative feedback suggests that the time increase could be due to the immersive nature of VR as well as the novelty of exploring a new environment with a new technology. It may also be due to the steeper learning curve associated with VR.[36]

Another important consideration is the time taken to simulate skeletal movements vs. dental movements. For the 2D and VR tasks, participants had the ability to manipulate both the dental and skeletal structures, whereas with the 3D software, the teeth were static. Although the ability to simulate dental movements could have impacted the time spent on each simulation, it does not entirely explain the significant difference between 2D and VR test time because dental movements could be simulated to some extent in both modules. The significant differences in time are more likely attributable to baseline differences in familiarity with the three simulation tools as well as the immersive qualities of VR. The training format may not have been sufficient as evidenced by the higher frequency of software feature and manipulation questions asked during the testing sessions for the 3D and VR tools. Incorporating an assessment at the end of each training module that must be passed in order to move on to the test task could minimize the differences between tools that are attributed to unfamiliarity and novel equipment.

Looking at baseline familiarity, it was expected that participants would ask more questions during the 3D and VR testing sessions regarding the software features and manipulation. Interestingly, even though the 3D software manipulation with mouse and keyboard was similar to 2D, the number of manipulation questions for the 3D tool were significantly higher compared to both the 2D and VR tools. Difficulty navigating the 3D software interface and manipulating the model within standardized views were negative user themes that may explain the higher frequency of software feature and manipulation questions asked during the 3D task. A lower frequency of manipulation questions asked during the VR task suggests that an advantage of VR systems is the natural user interface.[37] By incorporating gestures that have physical meaning and mimic real-world behavior, the VR interface becomes more intuitive and more readily adoptable by the user.[37]

For the treatment plan simulations, we observed an increasing number of average total movements prescribed that trended with increasing fidelity. This suggests that with high fidelity, residents are correcting discrepancies in more spatial planes and are therefore eliciting more complex and thorough treatment plans. The participants' treatment planned significantly more maxillary and mandibular pitch changes for the VR patient case. This may be a true fidelity-based observation. The use of different patient cases, however well matched, may also explain the observed differences in the number of pitch movements. When pitch movements were excluded from the statistical analysis, a significant difference in the number of total movements for higher fidelity methods still remained. The confounding variation that comes from using three different patient cases could be resolved by using one patient case for all tasks and then randomly assigning the participants to complete one of the tasks. With this methodology, the number of participants would have to be significantly higher. However, a revision of the methodology in this way would not allow a comparison of the qualitative responses concerning the three approaches.

This study had several limitations. The small sample size made it harder to discern meaningful trends for some variables. Given the limited sample size, a mixed methods study design was used to allow for integration of the quantitative and qualitative data. Mixed methods research functions on the premise that the use of both quantitative and qualitative data and analysis methods provides a better understanding of research problems than using either approach alone.[38] For this study, integration of the quantitative and qualitative datasets was used to identify the advantages of higher fidelity surgical simulation methods as well as to highlight potential areas of improvement.

Additionally, only a small number of patient cases fit the inclusion criteria for the patient modules. Although, the selected patient cases were matched as closely as possible, there were still some discrepancies in the malocclusions. Collaborating with other institutions could increase the number of patient modules and increase the likelihood of finding an even more closely matched trio of cases. Despite these limitations, the study findings support continued evaluation of VR methods for orthognathic simulation in conjunction with qualitative assessments in order to improve upon the educational and clinical applications of higher fidelity simulation tools.

CONCLUSIONS

A comparison of 2D, 3D, and VR surgical simulation tools showed significant differences and an increase in the number of prescribed surgical movements during simulation that tracked with increased fidelity. This suggests that higher fidelity tools may facilitate the creation of more thorough and complex treatment plans - which has clinical and educational relevance. However, the study results also showed that there was no significant difference in the residents' overall performance on the written analysis exercise that involved diagnosis, treatment objectives and surgical considerations.

Our findings suggest that the highest fidelity simulation tools may be applied best to treatment planning tasks or simulations rather than diagnostic tasks. Acting under the assumption that clinicians will have access to traditional diagnostic records in addition to surgical simulation tools, future research with the VRPM should shift its focus to elaborating on the impact of fidelity on surgical treatment simulations specifically. Although more training for the residents is needed to increase familiarity and comfort with using a new technology, qualitative feedback confirms residents' readiness to adopt virtual reality tools for orthodontic surgical simulation.

ACKNOWLEDGMENTS

We thank Theodore Hall, Sean Petty and Stephanie O'Malley and the entire Emerging Technologies Group at the University of Michigan, Duderstadt Center , for their contributions in the creation of the virtual reality patient model for this study. We want to thank The University of Michigan Le Gro Fund and the Oculus Education Group for their financial support for this research.

REFERENCES

1. Kim-Berman H, Suh E. Embracing novel technologies in dentistry and orthodontics. In: Craniofacial Growth Series,56th vol.Ann Arbor, MI: Department of Orthodontics and Pediatric Dentistry, 2019: 110-24.
2. Proffit WR, White RP. Surgical orthodontic treatment. St. Louis, MO: Mosby Year Book. ix, 1991.
3. Hua J, Aziz S, Shum JW. Virtual surgical planning in oral and maxillofacial surgery. Oral Maxillofac Surg Clin North Am 2019;31(4):519- 30.
4. Lo LJ, Weng JL, Ho CT, Lin HH. Three-dimensional region-based study on the relationship between soft and hard tissue changes after orthognathic surgery in patients with prognathism. PLoS One 2018;13(8).
5. Larson BE. Orthodontic preparation for orthognathic surgery. Oral Max Surg Clinics 2014;26(4): 441-458..

6. Gateno J, Xia J, Teichgraeber JF *et al.* The precision of computer-generated surgical splints. J Oral Maxillofac Surg 2003;61(7):814-7. doi: 10.1016/s0278-2391(03)00240-4.
7. Becker O, Scolari N, Santos MF *et al.* Three-dimensional planning in orthognathic surgery using cone-beam computed tomography and computer software. J Comput Sci Syst Biol 2013;6(6):311-6. DOI: 10.4172/jcsb.100012.
8. Ho CT, Lin HH, Liou EJ, Lo LJ. Three-dimensional surgical simulation improves the planning for correction of facial prognathism and asymmetry: A qualitative and quantitative study. Sci Rep 2017;7:40423. doi.org/10.1038/srep40423.
9. Shu Y, Huang YZ, Chang SH, Chen MY: Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. Virtual Real 2018;23(4):437-446.
10. Zheng C, Li J, Zeng G, et al.. Development of a virtual reality preoperative planning system for postlateral endoscopic lumbar discectomy surgery and its clinical application. World Neurosurg 2019;123:e1-e8.
11. Wong DT, Mehta A, Singh KP, et al. The effect of virtual reality bronchoscopy simulator training on performance of bronchoscopic-guided intubation in patients: A randomised controlled trial. Eur J Anaesthesiol 2019;36(3):227-233.
12. Khan R, Plahouras J, Johnston BC, et al.. Virtual reality simulation training in endoscopy: A cochrane review and meta-analysis. Endoscopy 2019;51(7):653-664.
13. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. J Med Libr Assoc 2018;106(4):498-500.
14. Izard SG, Juanes Mendez JA, Palomera PR. Virtual Reality Educational Tool for Human Anatomy. J Med Syst 2017;41(5):76.
15. Carl E, Stein AT, Levihn-Coon A, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019;61:27-36.
16. Minns S, Levihn-Coon A, Carl E, et al..Immersive 3d exposure-based treatment for spider fear: A randomized controlled trial. J Anxiety Disord 2018;58:1-7.
17. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: A systematic review and meta-analysis. J Pain Res 2019; 12:2053-2085.
18. Arikatla VS, Tyagi M, Enquobahrie A, et al. High fidelity virtual reality orthognathic surgery simulator. Proc SPIE Int Soc Opt Eng 2018;10576.
19. Grabowski HA, Krempien R, Munchenberg JB, et al. Simulation of frontal orbital advancement. Virtual Reality 1999;4:235-240.
20. Neumann P DS, Schuiz A, Fauikner G, et al. Using virtual reality techniques in maxillofacial surgery planning. Virtual Reality 1999;4:213-222.
21. Xia J, Samman N, Yeung R, et al. Three dimensional virtual reality surgical planning and simulation workbench for orthognathic surgery. Int J Adult Ortho and Orthog Surg 2000;15(4):265-282.

22. Zaragoza-Siqueiros J, Medellin-Castillo HI, de la Garza-Camargo H, et al. An integrated haptic-enabled virtual reality system for orthognathic surgery planning. Comput Methods Biomech Biomed Engin 2019;22(5):499-517.
23. UMichDent. Dentistry and VR—Jaw surgery simulation [Video file]. Aug 2019. Retrieved from https://www.youtube.com/watch?v=YCyrqmrMwPQ&t=3s. Accessed May 2020.
24. Kramer GA, Albino JE, Andrieu SC et al. Dental student assessment toolbox. J Dent Educ 2009;73(1):12-35.
25. Desai PR, Desai PN, Ajmera KD, Mehta K: A review paper on oculus rift-a virtual reality headset. International journal of engineering trends and technology *(IJETT)* 2014;13(4):175-179.
26. Oculus Touch Review. Available from https://www.theverge.com/2016/12/5/13811232/oculus-touch-rift-vr-motion-controller-review Accessed February 2021.
27. Maliha Samantha G, Rodrigo D-S, Plana Natalie M, et al.. Haptic, physical, and web-based simulators: are they underused in maxillofacial surgery training? J Oral Maxillofac Surg 2018;76(11):2424.e1–2424.e11. https://doi.org/10.1016/j.joms.2018.06.177
28. Ahmed N, McVicar IH, Mitchell DA. Simulation-based training in maxillofacial surgery: are we going to be left behind? Br J Oral Maxillofac Surg 2019;57(1):67–71. https://doi.org/10.1016/j.bjoms .2018.11.009
29. American Board of Orthodontics. Board case oral exam examinee worksheet [PDF]. Available from https://www.americanboardortho.com/media/1164/bcoe-examinee-worksheet.pdf Accessed January 2020.
30. Grummons DC, KappeyneVan De Coppello MA. A frontal asymmetry analysis. JCO 1987;21(7):448-465.
31. List Randomizer. Available from www.random.org/lists/ Accessed January 2020.
32. Sayin MO, Turkkahraman H. Malocclusion and crowding in an orthodontically referred Turkish population. Angle Ortho 2004;74:635-9.
33. Sakowitz S, Inglehart M, Ramaswamy V et al. Comparison of 2-dimensional prediction tracing and a 3-dimensional virtual reality method for treatment planning of orthognathic cases - a preliminary study. Virtual Reality 2020;24:399-409.
34. Kim-Berman H, Karl E, Sherbel J et al. Validity and user experience in an augmented reality virtual tooth identification test. J Dent Educ 2019;83(11):1345-52.
35. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276- 82.
36. Sung GK, Kyung SY, Young HK et al. A study on the learning curve of the robotic virtual reality simulator. J Laparoendosc Adv Surg Tech A 2012;22(5):438-42.

37. Peterson N, Stricker D. Continuous natural user interface: reducing the gap between real and digital world. ISMAR 2009;1:23-26. DOI:10.1109/ISMAR.2009.5336502.
38. Creswell JW, Creswell JD. Mixed methods research: Developments, debates, and dilemmas. Research in organizations: Foundations and methods of inquiry. San Francisco, CA: Berrett-Koehler Publishers, Inc., 2005: 315-27.

Table 1: Average time to completion and differences in number of questions asked by type in training and test situations

Time to Completion (in minutes)	**2D Mean SD**	**3D Mean SD**	**VR Mean SD**	**p**
Training Time	2.13 0.956	3.95 1.676	6.59 1.830	<0.001[a,b,c]
Test Time	19.25 3.893	24.71 4.789	29.29 5.823	<0.001[a,b,c]
Question Frequency	**2D Mean SD**	**3D Mean SD**	**VR Mean SD**	**p**
Total training question frequency	0.45 0.686	0.95 0.999	1.70 1.593	0.012[b]
- task related	0.10 0.308	0.20 0.523	0.05 0.224	0.427
- case details	0.00 0.000	0.00 0.000	0.00 0.000	n/a
- software features	0.10 0.308	0.45 0.605	0.45 0.686	0.073
- manipulation	0.25 0.639	0.30 0.571	1.10 1.294	0.006[b,c]

- other	0.00 0.000	0.00 0.000	0.10 0.308	0.135
Total test question frequency	1.80 1.795	3.90 1.651	4.70 3.629	0.004[a,b]
- task related	0.45 0.826	1.05 0.945	1.50 1.933	0.059
- case details	0.90 1.165	0.35 0.587	0.95 0.826	0.066
- software features	0.15 0.366	1.15 0.933	1.30 1.380	0.002[a,b]
- manipulation	0.00 0.000	1.35 0.933	0.55 0.826	<0.001[a,b,c]
- other	0.30 0.733	0.10 0.308	0.40 0.598	0.287

Note:
Post hoc comparisons: a = 2D vs. 3D = $p<0.05$; b = 2D vs. VR = $p<0.05$; c = 3D vs. VR = $p<0.05$.

Table 2: Average percentages of correct responses by type of simulation and percentages of participants who correctly identified specific points of interest

BCOE Worksheet Scores **% Correct**	**2D** **Mean** **SD**	**3D** **Mean** **SD**	**VR** **Mean** **SD**	**p**
Skeletal Diagnosis	56.88 11.088	66.25 13.512	58.89 14.008	0.038[a,c]
Dental Diagnosis	61.58 15.091	53.61 10.705	57.06 8.556	0.041[a,b]

Skeletal Objectives	46.88 12.082	53.13 13.374	58.13 14.776	0.008[b]
Dental Objectives	55.00 15.460	61.50 18.432	57.27 17.224	0.332
Special Surgical Considerations	30.00 37.697	42.50 37.258	27.50 34.317	0.028[a,c]
Total Score	55.21 9.7131	56.96 9.158	56.38 6.567	0.632
Specific Points of Interest	**2D** **N (Yes)** **% (Yes)**	**3D** **N (Yes)** **% (Yes)**	**VR** **N (Yes)** **% (Yes)**	**p**
Correctly identified transverse discrepancy	17 85%	14 70%	17 85%	0.392
Correctly identified mandibular asymmetry direction	20 100%	20 100%	20 100%	n/a
Correctly identified mandibular asymmetry severity	8 40%	16 80%	10 50%	0.029[a,c]
Identified possible bony interferences/iatrogenic gaps	6 30%	10 50%	6 30%	0.614
Identified possible residual soft tissue asymmetry/genio to finalize esthetics	6 30%	7 35%	5 25%	0.314

Note:
Post hoc comparisons: a = 2D vs. 3D = $p<0.05$; b = 2D vs. VR = $p<0.05$; c = 3D vs. VR = $p<0.05$.

Table 3: Number of linear and rotational movements planned using each simulation method by percentage of participants that prescribed each type of movement

Treatment Planning: **Maxillary linear movements:**	**2D** **N / %**	**3D** **N / %**	**VR** **N / %**	**p**
- sagittal movements	7 35%	12 60%	11 55%	0.247
- vertical movements	2 10%	3 15%	2 10%	0.851
- transverse movements	11 55%	13 65%	15 75%	0.766
Maxillary rotational movements:	**2D** **N / %**	**3D** **N / %**	**VR** **N / %**	**p**
- yaw movements	0 0%	0 0%	2 10%	0.126
- pitch movements	6 30%	9 45%	16 80%	0.005[b]
- roll movements	3 15%	1 5%	3 15%	0.524
Mandibular linear movements:	**2D** **N / %**	**3D** **N / %**	**VR** **N / %**	**p**
- sagittal movements	8 40%	18 90%	13 65%	0.004[a]
- vertical movements	0 0%	3 15%	2 10%	0.217
Mandibular rotational movements:	**2D** **N / %**	**3D** **N / %**	**VR** **N / %**	**p**

- yaw movements	17 85%	19 95%	18 90%	0.574
- pitch movements	5 25%	6 30%	14 70%	0.007[b,c]
- roll movements	2 10%	1 5%	6 30%	0.064
Total number of movements	**2D** **Mean** **SD**	**3D** **Mean** **SD**	**VR** **Mean** **SD**	**p**
	4.00 1.686	5.15 1.182	5.85 1.387	0.001[a,b]

Note: Post hoc comparisons: a = 2D vs. 3D = $p<0.05$; b = 2D vs. VR = $p<0.05$; c = 3D vs. VR = $p<0.05$.

Table 4: Integration matrix merging significant quantitative findings with qualitative remarks

	Quantitative Finding	**Qualitative Feedback**	**Integration and Interpretation**
Time	Significantly more time taken to complete task with higher fidelity tools	"I felt like I got lost [in VR]." "Tough to focus while trying to navigate new controls and new environment [in VR]" "2D was most familiar; quick and easy"	Residents spend less time completing familiar tasks. Increased time in VR may be attributed to immersion and novelty of equipment.
Interactions	Significantly more total interactions for higher fidelity tools	"[VR is] unfamiliar, frustrating for not knowing how to use it properly."	Residents ask more questions about unfamiliar tools.
	Significantly more questions about software features with higher fidelity tools (3D & VR vs. 2D)	"[3D is] limited by the interface; had to find the right button to press, confined to presets" "Did not like the [3D] software, felt dumb using it.	Differences in software interface have significant impact on the user experience.
	Significantly more questions about manipulation with higher fidelity tools (3D > VR	"[VR is] intuitive and the power to freely move it vs. clicking in increments	VR natural user interface facilitates easier manipulation.

	> 2D)	made a big difference” “Difficult to navigate [3D] with the mouse and two separate views.”	
Worksheet Scores (Diagnosis)	No significant difference in worksheet total scores.	“Don’t need to be able to see from all perspectives to diagnose a case.”	Residents tend to rely heavily on what is most familiar; residents are used to treatment planning with standard records.
	Significantly better at identify mandibular asymmetry severity using 3D	“I can see soft tissue changing in a dynamic way; I can adjust opacity too”	3D may be a better diagnostic tool compared low fidelity methods due to superior soft tissue visualization
	Significantly higher scores on skeletal objectives using VR	n/a	Higher skeletal objectives score may be indicative of VR’s advantages for treatment planning.
	No significant differences in identifying surgical considerations.	“Too much control over the pieces; may be unrealistic since so freeform” “Too much free reign, can make anything look good”	Residents may understand that what they are simulating is unrealistic but lack the surgical competence to make appropriate adjustments
Treatment Simulation (Tx. Planning)	Significantly more total movements prescribed with higher fidelity tools	“[VR] facilitates trying multiple plans” “I can do more with VR”	Higher fidelity tools may facilitate the creation of more thorough and complex treatment plans.
	Significantly more maxillary pitch movements prescribed with VR	“Gives accurate representation of surgical movements and relationship to surrounding structures”	Pitch movements may be more readily manipulated in VR indicating a fidelity-based difference.
	Significantly more mandibular pitch movements prescribed with VR	“Gives accurate representation of surgical movements and relationship to surrounding structures”	Pitch movements may be more readily manipulated in VR indicating a fidelity-based difference.
	Significantly more mandibular sagittal movements prescribed with 3D	“More control over manipulation, one plane, step-wise”	Sagittal discrepancies may be more readily manipulated in 3D indicating a fidelity-based difference

Table 4 continued:

Post Survey	Significantly higher value ratings for diagnosing with 3D and VR	“I can see all the hard tissue and I can see in all three planes of space”	Residents find higher fidelity tools more helpful for forming an accurate diagnosis.

	Significantly higher value ratings for treatment planning with 3D and VR	"In VR you get all dimensions and you can see exact measurements for your movements in all planes."	Residents find higher fidelity tools more helpful for forming a treatment plan.
	Significantly higher difficulty ratings for diagnosis with 2D	"I was limited in what planes I had so I couldn't even diagnose in the transverse."	Residents find low fidelity tools too limiting to form an accurate diagnosis.
	Majority of participants selected VR as most accurate	"I can move VR the most freely and I like that you are able to move the teeth" "Can see all the planes of space and contours of structures"	Enhanced visualization and manipulation contribute significantly to perception of accuracy.
	Significantly higher enjoyment rating for VR vs. 2D	"Cool, interactive, felt real, immersive"	Immersion and interaction are key factors contributing to enjoyment.
Baseline vs. Post-Survey	Significant decrease in value ratings for 3D for forming an accurate treatment plan	"Very hard to work with" "Couldn't get things to go where I wanted"	Difficult interface negatively affected perception of value of 3D tool.
	Significant increase in selections for VR as most preferred"	"Prefer VR because can be more complete" "More interactive and feels more fluid"	Residents exhibit a readiness to adopt VR due its enhanced visualization and natural user interface

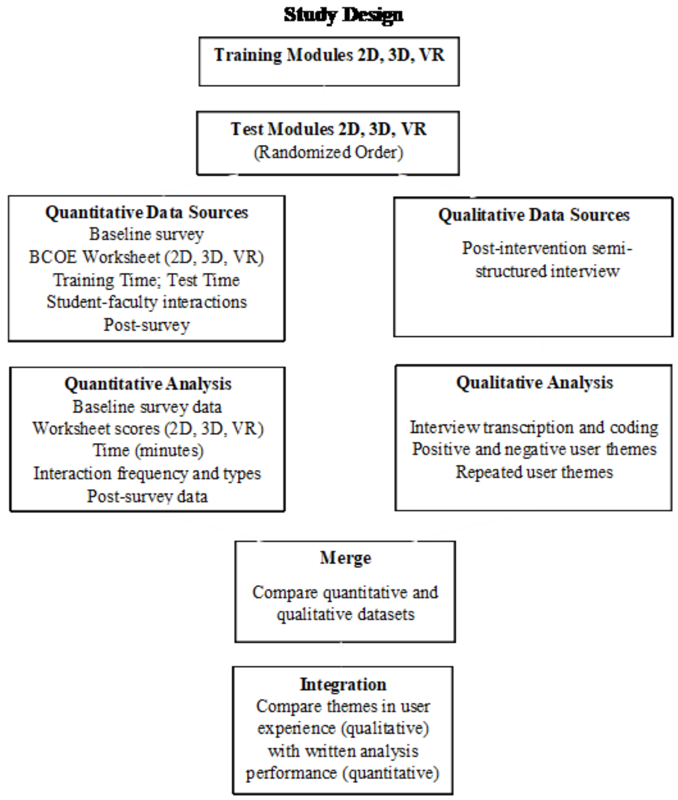

Figure 1. Study design flowchart where participants (n=20) diagnosed and treatment planned orthognathic surgery cases using 2 dimensional (2D), 3 dimensional (3D) and virtual reality (VR) methods. Qualitative and quantitative data were analyzed and integrated.