

Assessing Variability in Segmentation Algorithms for 3D Printing at the Point of Care

Magdalene Fogarasi

United States Food and Drug Administration

James Coburn (✉ james.coburn@fda.hhs.gov)

University of Maryland, College Park

Beth Ripley

Veterans Health Administration

Research Article

Keywords: Patient-specific anatomical model, 3D printing, geometric variation, medical image segmentation

Posted Date: October 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-841354/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: 3D printing (3DP) has enabled medical professionals to create patient-specific medical devices to assist in surgical planning. Anatomical models can be generated from patient scans using a wide array of software, but there are limited studies on the geometric variance that is introduced during the digital conversion of images to models. The final accuracy of the 3D printed model is a function of manufacturing hardware quality control and the variability introduced during the multiple digital steps that convert patient scans to a printable format. This study provides a brief summary of common algorithms used for segmentation and their principal features. We also identify critical parameters and steps in the workflow where geometric variation may be introduced. We then provide suggested methods to measure or reduce the variation and mitigate these risks.

Methods: Using a clinical head CT scan of a mandible containing a tumor, we performed segmentations in four separate programs using workflows optimized for each. Differences in segmentation were calculated using several techniques.

Results: Visual inspection of print-ready models showed distinct differences in the thickness of the medial wall of the mandible adjacent to the tumor. Residual volumes were calculated to generate pairwise agreement and disagreement percentages between each as the program's model. For the relevant ROIs, statistically significant differences were found globally in the volume and surface area comparisons between final bone and tumor models, as well locally between nerve centroid measurements – major variance introduced due to workflow is highlighted in difference heat maps. As with all clinical use cases, statistically significant results must be weighed against the clinical significance of any deviations found.

Conclusions: Statistically significant geometric variations can be introduced to patient-specific models from differences in software applications. The global and local variations should be evaluated for a full understanding of geometric variations. The clinical implications of these variations vary by anatomical location and should be evaluated on a case-by-case basis by certified clinicians. Understanding the basic functions of segmentation and 3D print preparation software is essential for users intending to adopt the use of patient-specific models for clinical intervention or decision making.

Contributions To The Literature

- The number of healthcare facilities using 3D printing to create patient-specific devices and anatomic models has dramatically increased in the last several years, but there is no systematic implementation strategy.
- Converting patient imaging to a 3D printable model is complex and can introduce geometric inaccuracies, with different software programs producing variable results.
- Description of key factors controlling the quality and accuracy of 3D printable meshes generated through segmentation and methods to evaluate and validate workflows

- Understanding these critical parameters will enable improvements to risk mitigation and quality control standards, thus improving patient care.

Introduction

Many clinicians are beginning to use 3D printing (3DP), a form of additive manufacturing (AM), to make anatomic models for surgical planning, patient education, and more. [1, 2] Over the last decade, traditional manufacturers have used 3DP to fabricate patient-specific medical devices [3]. However, the more recent trend is for health care systems to bring 3D printing capabilities within the walls of the hospital at the point of care. With increased implementation at the point of care, healthcare facilities need to develop methods to ensure that these devices are safe and do not increase risk to the patient. In August 2017, the FDA discussed several types of activities that could be undertaken at the point of care [4]. Common use cases include patient-specific implants, surgical cutting guides, and anatomic models [5]. Anatomic models may improve surgical outcomes and provide tactile stimulus during surgical planning [6]. Procedures to ensure patient-specific models meet clinical requirements can be developed through in-house experience, but a systematic approach developed from understanding critical to quality attributes and using established techniques can reduce patient risk and increase output consistency.

A basic workflow for generating a 3D printed model from patient anatomy begins with an appropriate patient imaging data set, progresses through several software steps to isolate the anatomy of interest, and then additional steps to convert volumetric segmentation data to surface data (Figure 1). Best practices for volumetric patient imaging (e.g. computed tomography (CT) and magnetic resonance imaging (MRI)) for designing patient specific models may be different than standard clinical imaging protocols used for diagnosis [1] and are optimized for all the equipment and software being used. Optimal imaging protocols for segmentation typically result in isotropic voxels with a small field of view in the XY plane and thin slices in the Z-axis – with potentially increased noise [7, 8]. However, the clinical needs include weighing the radiation dose and the benefits of additional imaging sequences [9].

Once patient imaging is acquired, software is used to segment a region of interest (ROI), all following a similar generalized workflow. Previous investigations into image acquisition and workflow found that the largest influences on model accuracy were insufficient scan quality and manual segmentation for complex soft tissue cases [10]. Recent algorithm improvements have increased the availability and degree of segmentation automation. As 3DP moves to point of care and the clinical environment, there is a need for users to validate software processes and measure their agreement with the true anatomy [11]. Each intended clinical use requires a specific level of accuracy.

For example, a model created to demonstrate relative anatomy may be functional with rough precision and accuracy, whereas a model intended for the sizing, placement, or templating of an implanted device may require submillimeter accuracy. The final accuracy of the 3D printed model is a function of manufacturing hardware quality control and the variability introduced during the workflow to convert patient scans to a printable format such as standard tessellation language (STL) files. STL surface

meshes are the most prevalent digital model files due to their history in computer graphics and their ability to minimize the storage and processing power needed for large volumes. While no standard imaging protocol for medical 3DP existed at the time of writing, most available protocols for FDA-approved devices and previous studies acknowledge that image slice thickness and slice interval are primary limiting factors in scan quality. Noise, beam hardening, patient movement, and metal artifacts can all contribute to inhomogeneities in gray values, negatively impacting segmentation precision. Similarly, for printing hardware/process output, maintenance, process controls, and material controls have all been identified as critical factors for accuracy. [12]

This study will focus on identifying key details of the digital workflow in segmentation software that will help a health care facility evaluate and implement the right solutions for their needs. Programmatic implementations and user options can affect the mesh either cosmetically or substantially, the effects of which are not always immediately evident to the user. This study uses several common programs with different automated approaches to illustrate the kind of variability that can exist between software programs and how to test for it. A single program may also have a variety of parameters that could lead to different outputs. We seek to identify and implement metrics to quantify potential geometric variation in 3D models arising from software implementation-associated workflows. Four programs were selected to represent a spectrum of those available, including both FDA cleared and noncleared software with proprietary or open-source implementations. Most programs operate in a similar fashion, but user control over smoothing options varies between programs. The comparison metrics used here can be extended and repeated with other software programs and workflows not described here.

Background: Representing Medical Imaging as Digital Solids or Segmented Anatomy

The first step in characterizing variability was to understand the underlying algorithms behind segmentation, mesh construction, and mesh smoothing and their typical implementation. Medical image volumes are – at their simplest – blocks of data stacked on top of each other into volumetric pixels or voxels. When regions of interest (ROIs) are segmented out and treated as separate datasets from the original imaging data, they are stored as digital solids. There are two ways to store these data. First, as solid volume information, all the voxels for an ROI are collected and named. While most representative of the patient data, volume models appear visually jagged as calculated only with voxel information. The second method is to create a surface mesh – essentially making a shell mold of the ROI by draping strings over it until they cover the entire region. This “surface mesh” is described in triangles. In the analogy, everywhere the strings intersect, it is called a vertex. The line between each vertex is called an edge, and then the space between any three nodes is called the face (Figure 1A). Similar to the real object, each face has an inside and an outside, and the outside direction is called its “normal”. All of these triangles together form a continuous mesh wrapped around the ROI.

Storing an ROI as a mesh is essential for 3DP, as most 3D printing software uses the coordinates of the vertexes and triangles – stored as stereolithography files (STLs) – to tell the printer motors where to

move and place material as the model is printed. While there are some limited advancements in printing technology allowing models to be stored with voxel information only [13], these are not the current standard.

Generating a uniform mesh – where triangles are even in size and distribution – is computationally complex but typically makes the digital object more accurate and easier to work with in the software (Figure 1 Bii). When increasing the resolution on a triangulated mesh, an increasing number of triangles are added to better approximate curved features. Dense meshes can be computationally complex and burdensome to work with. However, if the mesh density is not high enough, the solid may not accurately resemble the original object. In a simple case, a marble could be represented with 3 triangles and look like a pyramid or with thousands of triangles looking like a smooth sphere. In addition to misrepresenting a simple solid, areas with high curvature and small features can cause triangles to become very thin or create disconnected areas or holes. A balance must be struck between the complexity of adding triangle density and the need for dimensional fidelity. Many workflows mediate this by first creating a very dense original mesh and then reducing the number of triangles and optimizing their locations until an automatic criterion is met based on input parameters or the user manually determines that further modification is no longer suitable. Note that reducing the number of triangles does not always mean that the fidelity will be reduced (Figure 1 Bi). Many programs use constant density meshes, where the size of each triangle is relatively the same. Some programs can also produce variable density meshes, which will make the triangles smaller in regions that require higher resolution or fidelity. For most standard anatomic modeling applications, constant density meshes can accurately reproduce clinically relevant features.

Because of the complicated organic shapes of anatomic features, many algorithms and methods have been developed to help users identify boundaries between regions in medical images by automating segmentation and then increasing accuracy through refining processes. These processes can affect the final product in both clinically relevant and insignificant ways depending on the ROI's clinically relevant features and how the algorithm works. Knowledge of a few key parameters can help to determine what any software is doing, how it will affect the anatomic model, and ultimately how it will affect patient care and safety.

Background: Algorithms

Automated algorithms are used in multiple steps in the digital pipeline. The two steps with the greatest chance to impact the final print quality are automated segmentation and mesh smoothing. Segmenting begins with [14] thresholding to define the contour of the ROI. This selected area will be highlighted and color coded to visibly “mask” the ROI. Region growing can then be performed to refine the mask over the ROI [15]. Various methods exist, such as active contours [16, 17] and region competition [18], to semiautomatically refine segmentation masks using grayscale comparisons with weighting or probability calculations to decide which voxels to include in the refined contour. After refinement, the next step is to generate a 3D volume - generally implementing a version of the marching cubes algorithm[19]. This entails partitioning the mask into individual polygons representing each voxel and then fusing it into a

surface. The cube size is generally defined by the voxel size from the scan data and is set to a multiple of the slice thickness.

The newly created meshes can be smoothed and refined, terms that often overlap. *Refinement* may include any of several methods to increase the fidelity of the mesh in specific areas where feature resolution is needed. *Smoothing* will refer to decreasing noise in contours and flattening of surface features. Segmented meshes made from patient images tend to be large and very complex and can usually be decimated to reduce the computational load. *Decimation* can reduce the number of triangles that make up the mesh – decreasing file size and complexity while ideally preserving topology [20]. A summary of the most common smoothing algorithms with notes about implementations is summarized in Table 1.

Most mesh smoothing algorithms function by iterating through the mesh and relocating vertices according to mathematical restrictions to optimize the mesh to a given parameter. The most common smoothing algorithms are implementations of Laplacian smoothing [21, 22], a vertex-based technique that iteratively converges a curve toward a point. Not all smoothing algorithms are made equally. A pure Laplacian implementation does not correct for mesh shrinkage, but modifications such as Taubin smoothing [23] include a compensating inflation step after each mesh shrinkage step. Similar methods such as angle-based [24], bilateral [25], and curvature [26] have been implemented to optimize smoothing in a manner that preserves details (sharp points, small radius curvatures, or thin walls) in the mesh. Most programs include user-selectable options to preserve small features and boundaries.

Methods of optimally smoothing a mesh have been an important topic for a long time, and new implementations and optimization corrections are constantly being introduced. Open source and proprietary implementations can provide different options and benefits to the user, each with accompanying compromises. With open-source software, the user can directly view and sometimes modify the implementation but may sacrifice a user-friendly interface or extensive validation. Proprietary software programs generally include well-developed user interfaces but do not always publish complete descriptions of their algorithms, giving users less visibility and control over algorithmic behavior. However, that limited flexibility is often accompanied by additional software validation.

It is incumbent on the user to identify the characteristics of their software and determine which are most important for their application. Here, we use four commonly available segmentation programs to show differences that can arise from their implementation of algorithms. Then, we identify several methods of comparing algorithms, locating errors, and maintaining quality for clinical processes.

Materials & Methods

Materials

The test data set was a partial head CT scan used in an 2018 RSNA hands-on 3D printing training module. [27] (slice thickness 1.0 mm, no gantry tilt, 512x512 FOV, 0.3319 mm pixel spacing, 16-bit, 120

kVP, FC80 kernel, and 40 mA) The completely deidentified dataset was provided by Materialise NV (Leuven, Belgium) who obtained permission from the patient and institution for use of the deidentified data for non-diagnostic and research purposes. All methods were carried out in accordance with relevant guidelines and regulations. The data, available upon request, features a mandible with a large right mandibular tumor; the tumor is adjacent to both a nerve and an impacted wisdom tooth. This file was chosen because it contains several ROIs with varying degrees of segmentation complexity to challenge different aspects of the digital pipeline. All input images were in Digital Imaging and Communications in Medicine (DICOM) format, and output meshes were saved as STL files.

The four programs selected for this study will be referred to as Programs 1-4.

Surface comparisons between programs were performed using the final output STL meshes in Magics (23.0, Materialise). Linear dimensions and volumetric measurements were computed in 3-Matic (19.0, Materialise). Statistics were calculated using MATLAB (R2007a, Mathworks).

Methods

The process of obtaining an STL file from DICOM scans follows a similar workflow in all programs, generalized into five main steps: (1) DICOM import, (2) Segmentation, (3) Mesh/Model Generation, (4) Smoothing/Mesh Refinement, and (5) Exporting to STL.

All methods were carried out in accordance with relevant guidelines and regulations. The deidentified dataset was obtained with permission for use in this paper from Materialise NV and is available to the public from the company under reasonable request.

Global thresholding based on gray values/Hounsfield units was the principle starting point for bone volume segmentation. Thresholds identified the initial mask or created seed markers. Initial thresholds were made using the bone presets (if available) in each software and then refined in later steps. Mask refinement was performed manually for the nerves in all cases, and in some of the programs, it was necessary to manually refine the tumor segmentation.

Each of the segmentations was smoothed per program to three levels (Table 2): no smoothing 0, low smoothing 1, and high smoothing 2 to maximize how representative each final STL was to its respective program.

Several additional metrics were evaluated for their ability to discriminate between segmented regions of interest: volume, surface area, and linear measurements established by previous literature. Volume and surface area provide global comparisons between models of the same ROI generated by each software.

Linear measurements were taken on the mandibles at three smoothing levels for each program with digital calipers (Magics 23.0) using established fiduciary markers [28]. A residual volume comparison was used [29] to calculate the volumes of each of the tumors in the absence of a ground truth measurement. These residuals were used to calculate the agreement and disagreement volume and

percentages (Equations 1 and 2), and pairwise statistics were performed. [29, 30]. The agreement metric defines the space that is occupied by both models and can be used to assess accuracy and repeatability in segmentations either between operators or software [31]. Disagreement defines the space occupied by only one of the models and not the other.

$$\textbf{Agreement} \quad \textit{Volume} = A \cap B \quad \quad \quad \% = \frac{A \cap B}{\left(\frac{A+B}{2}\right)} \quad \quad \quad \textit{Equation 1}$$

$$\textbf{Disagreement} \quad \textit{Volume} = (A - B) + (B - A) \quad \quad \quad \% = \frac{(A-B)+(B-A)}{\left(\frac{A+B}{2}\right)} \quad \quad \quad \textit{Equation 2}$$

We also generated surface deviation heatmaps of the pair of models with the worst agreement to aid visual inspection and assist in identifying local areas where variation was more prominent.

To measure the tumor, standardized measurement planes were created by slicing each STL with identical, parallel datum planes (Figure 4A). Planar contours defined a centroid, used to measure the in-plane length and width of the tumor edges. Differences in alveolar nerve path segmentation were measured using the centroid coordinates of selected coronal slices.

Differences between programs for mandible and tumor surface area and volume were calculated using ANOVA ($\alpha = 0.05$) with Tukey post hoc testing.

Nerves were partitioned to create slices along their curve, and center points were extracted and compared between programs to assess local deviations. Distances between centroids were compared using ANOVA ($\alpha = 0.05$) to compare between programs with 3 models per program from different smoothing settings ($n_{\text{total}} = 12$).

Analysis of measurements was performed between programs using one-way ANOVA ($\alpha = 0.05$).

Results

Mandible Segmentation Differences

Initial visual inspection of print-ready models showed clear differences, especially at internal soft tissue boundaries. On the inner left wall of the mandible abutting the tumor, there were noticeable differences in wall thickness, as viewed in detail in Figure 2. Bones with a thin cortical layer are challenging to segment at baseline, and the disruption of the cortex caused by the adjacent tumor compounded the challenge such that most of the programs did not capture all of the bone contour along the inner wall.

Segmentation Complexity Differences

The time to complete each segmentation fully, including mask refinement, smoothing, and STL conversion, was relatively similar for each of the programs, and some manual editing by the user was

required in all cases.

Significant differences ($p < 0.0005$) were found between the volumes of the bone and tumor and the surface area of the bone.

Linear measurements were largely consistent across programs. Minimal significant differences were found but were not clinically relevant to the anatomic ROI.

Union and intersection figures (Figure 3) highlight variations in outer contour between software programs.

Signed difference heat maps of the pair of models with the worst agreement (Figure 4B) highlighted the areas of high deviation between programs in red (positive deviation) and dark blue (negative deviation).

Nerve Segmentation Differences

Visual inspection of stacked nerve segmented models showed differences in volumes (Figure 5B); however, no significant differences were found for the centroid comparisons per slice for either side. No significant differences were found on the healthy side, although a slight statistically significant difference between the center points of slices was found on the tumor side ($p < 0.05$) (Figure 5A).

Discussion

In this study, we set out to determine the extent to which different segmentation software programs produced similar 3D printable files (mandible, tumor and nerve) from a single imaging data set. The same metrics can be applied to a single software program to determine the differences made by changes to segmentation parameters and procedures. It is critical that software programs preserve anatomical accuracy during the conversion process when clinical decisions are made from these models. We discussed important points in the workflow where errors can be introduced. Segmentation, isosurface extraction, smoothing, and decimation algorithms each have specific effects on the final output. We then tested different metrics for measuring software performance.

Overall, it remains difficult to quantify variations in geometry between digital segmentations and model generation techniques. There is also a range in complexity dependent on ROIs. Generally, when using CT, soft tissue features can be difficult to segment without a contrast agent due to low variability in attenuation between adjacent structures. Here, segmentation automation did not appear to lend any significant impact to the final models when compared to manual editing processes. However, more complex cases may show differences in automated processing techniques.

To assess global differences between models generated from each program, we compared surface area and volume. The surface area was calculated along all sides of the mesh surfaces, while the volume was only measured on closed surfaces. Therefore, a small difference in the mesh, for example, a sharp point or a cluster of small triangles, affects the calculations of surface area more than those of volume. For this reason, it is also important to ensure that the final models are free from extra shells or floating

triangles. We expected that differences in surface area would be more sensitive than differences in volume because nonmanifold edges or small differences in masking would result in a large change in surface area because STLs contain only the surface meshes. Statistical differences in absolute volume or surface may not always be clinically relevant variances and must be assessed according to the ROI. The heat map view – constructed using Hausdroff distances – was useful to visualize differences between two models even without a gold standard.

We located regions of high geometric variation in agreement with other studies [32] in areas of high curvature, most notably in the channel formed by the tumor conforming around the nerve (Figure 5B). The differences are most likely caused by the geometry of the thin bone wall bordering the tumor. Visual inspection of the DICOM images showed a blurred boundary between the tumor and bone, implying that while the tumor had not breached the bone, it was unable to be segmented using global thresholding because of how few voxels the bone occupied in comparison to surrounding soft tissue (a scenario where partial volume averaging can significantly impact segmentation efforts). When altering thresholding to cover thin walls, care must be taken to maintain the boundaries of other similar tissues in the same area. Allowing automatic thresholding of thin walls and then smoothing these contours without specifying sufficient boundary preserving conditions may result in regions being erroneously removed by automated smoothing. While there are many benefits to automated methods, many default settings are not sufficient to preserve the integrity of the original data – in many cases, operators must manually interpret the anatomic regions and modify the segmentation masks.

Depending on the algorithm employed by each program, small differences in masking or contour can influence the level of variation in the final model depending on how each implementation traverses the surface mesh. Discrepancies in the plane of imaging or tangent to the plane may vary in prevalence depending on the algorithm handling of fringe condition regions. Laplace implementations without edge detection or preservation modifications will lose small features and may generate a model that is an underestimate of the original contour.

Linear measurements were the easiest way to assess the difference between specific points of interest. Multiple anatomical landmarks were tested for ease of repeatability, leading to the selection of those used in this study. Fiduciary markers should be selected for both clinical relevance and repeatability. If an anatomic region has a constant dimension or critical area for clinical intervention, then point dimensions are an easy way to assess accuracy (e.g., diameter of a vessel or length from a cut-plane to a nerve). Differences in linear measurements may also be attributed to the complexity of the anatomy at each location. As tumors are unique, linear measurements were not an effective method of comparison.

It was noteworthy that at higher smoothing levels, there was a larger deviation between the models, likely because of the difference in the ways each program controlled for mesh shrinkage. Linear measurements used here are dependent on the locations selected. A larger number of measurements may provide a better overall resolution of geometric deviations.

Risk Mitigation

Many studies have focused on dry bone models as their gold standard, but it has been established through this and other studies [33] that partial volume effects related to voxel size are a major cause of segmentation variation when soft tissue segmentation is necessary. Dry bone ground truth studies do not always translate to actual clinical results. Multiple prior studies have investigated the variations in printed anatomical models; however, with the rapid evolution of additional technologies such as virtual and augmented reality, we believe that digital-only models will dominate the industry. Once the system accuracy is established through a ground truth study, the accuracy of a patient-matching process requires more real-world data. The global and local geometric variation can be assessed more thoroughly due to multiple factors, which may assist in the validation of patient-matched models.

Minimizing the risk for geometric variation during the DICOM to 3D model conversion process can be done by understanding the software of choice and the anatomic ROI. This is best achieved by using staff trained well in both anatomy and segmentation processes to mitigate both software variability and the inherent variability of human operators. Models should always be assessed against patient data by relevantly trained engineers and physicians. A workflow checklist of potential risks and mitigations can be used and customized to fit users' needs (Figure 6).

Software may be cleared by the FDA for specific applications or intended uses such as segmenting medical images to generate 3D digital models for clinical use. Some of the programs selected for this study have been FDA cleared for specific indications for use (IFU). In that case, the software has shown the FDA that it can create accurate 3D representations of the anatomy listed in its IFU. It also includes sufficient instructions to let trained users replicate those results. While it is not necessary for clinicians to use FDA-cleared software to design anatomic models, they would then have to assume the burden of assessment and validation of the uncleared software. Validation of the software components may be done with the assistance of FDA guidance, but algorithm validation is generally left up to the specific vendor. For these reasons, there are benefits to both proprietary software whose algorithms have gone through a thorough review process but whose inner workings are private and open-source programs giving users the ability to see and control algorithm applications, but may not have gone through a use-specific review process. Regardless of the software choice, it appears that the most important aspect of designing models with minimal variability is the user's understanding of the smoothing parameters available and how they function within the software of choice. Knowing when to apply and tweak specific parameters is essential to minimizing geometric variability between DICOM sources and final STL models. The use of "cleared" software is not sufficient to mitigate risks without also understanding the mechanisms of the software and how certain parameters may negatively impact the fidelity of the final model.

Selecting parameters for metric assessment is always a challenge. Landmarks must be identifiable enough to be repeatable points of measurement between samples but also clinically relevant in their locations. Ideally, we would choose locations at critical regions – regions of high curvature, very thin

walls, etc. However, due to the nature of these regions being highly sensitive to variation due to workflows, they are often difficult to locate in models produced in different software programs. Maintaining the global coordinates of the original data set and performing analysis on models in a single separate software aids in reproducing measurement points between models but is still not a perfect failsafe.

Clinical Relevance

The clinical relevance of geometric deviations in digital models is case dependent and should be assessed by qualified clinicians in reference to the specific procedure and anatomical region of interest. Tumor resection surgeries with larger error margins may not be as affected by small surface geometry errors as other surgeries with narrower requirements. Clinical appropriateness criteria have been developed regarding which cases would benefit from patient-specific models [4]. The digital conversion workflow should attempt to maintain the integrity of the original patient imaging data as much as possible.

To minimize the risk for error during digital model creation, it is essential to start with a patient scan of adequate resolution for segmentation. Minimizing voxel size without subjecting the patient to excess radiation dosage is optimal for achieving high-quality segmentable data. Most diagnostic CT machines scan with slice thicknesses of less than 1 mm. Most patient-matched implants are currently accompanied by a specific scan protocol that is optimized for the anatomy around the implant location. Unique cases such as trauma, cancer, and congenital deformity that present some of the best potential for patient-matched technology do not yet have standardized requirements and should be evaluated based on clinical needs.

Conclusions

Quantifying the variations in model design will be essential for patient-matched technology to reach maximum potential. This study provides comparisons of several metrics that can be used to validate methods of preparing patient-matched 3D prints. The implications of these variations will still need to be assessed by a qualified clinician for each case. The critical parameters described here can compare segmentation quality and STL models regardless of workflow or program. Establishing best practices for evaluating variation between segmentation methods will allow users to develop optimized workflows - ideally accelerating the patient matched instrumentation implementation in industry and at point of care.

In the future, a standardized certification program for operator segmenting images for use in patient-matched devices would be beneficial. However, a basic understanding of the functionality of segmentation software is essential for ensuring patient safety as medical 3DP continues to expand.

Abbreviations

3DP 3D Printing

AM Additive Manufacturing

CT Computed Tomography

DICOM Digital Imaging and Communications in Medicine

FDA US Food and Drug Administration

IFU Indications for Use

MRI Magnetic Resonance Imaging

ROI Region of Interest

STL Standard Tessellation Language

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

DICOM data were used in a 2018 Radiological Society of North America training course and provided by Materialise NV. They are available upon request medical@materialise.be.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This project was supported in part by an appointment to the Research Participation Program at the Office of Science and Engineering Laboratories Center for Devices and Radiological Health Division of Biomedical Physics, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for

Science and Education through an interagency agreement between the U.S. Department of Energy and FDA.

Acknowledgements:

DICOM data set was provided by Materialise NV, Leuven, Belgium from their data library. 3D Printing resources were provided by the FDA Additive Manufacturing for Medical Products Laboratory.

Author contributions:

MF analyzed the segmentation data, organized algorithm background information and contributed to writing the manuscript. JC edited the manuscript and contributed to writing. BR contributed clinical perspectives and assisted in editing. All authors read and approved the final manuscript.

References

1. SME. Physicians as Manufacturers: The Rise of Point-of-care Manufacturing SME Report. 2018.
2. Christensen A, Wake N. Chapter 9 - Regulatory Perspectives for 3D Printing in Hospitals. In: Wake N, editor. 3D Printing for the Radiologist: Elsevier; 2022. p. 109-16.
3. SME. Medical Additive Manufacturing/3D Printing. SME Annual Report. 2018.
4. Chepelev L, Wake N, Ryan J, Althobaity W, Gupta A, Arribas E, et al. Radiological Society of North America (RSNA) 3D printing Special Interest Group (SIG): guidelines for medical 3D printing and appropriateness for clinical scenarios. 2018;4(1):11.
5. Bucking TM, Hill ER, Robertson JL, Maneas E, Plumb AA, Nikitichev DI. From medical imaging data to 3D printed anatomical models. PLoS One. 2017;12(5):e0178540.
6. Punyaratabandhu T, Liacouras PC, Pairojboriboon S. Using 3D models in orthopedic oncology: presenting personalized advantages in surgical planning and intraoperative outcomes. 3D printing in medicine. 2018;4(1):12.
7. Ibrahim N, Parsa A, Hassan B, van der Stelt P, Aartman IH, Wismeijer D. The effect of scan parameters on cone beam CT trabecular bone microstructural measurements of the human mandible. Dentomaxillofac Radiol. 2013;42(10):20130206.
8. Whyms BJ, Vorperian HK, Gentry LR, Schimek EM, Bersu ET, Chung MK. The effect of computed tomographic scanner parameters and 3-dimensional volume rendering techniques on the accuracy of linear, angular, and volumetric measurements of the mandible. Oral Surg Oral Med Oral Pathol Oral Radiol. 2013;115(5):682-91.
9. Wu XD, Xiang BY, Schotanus MGM, Liu ZH, Chen Y, Huang W. CT- versus MRI-based patient-specific instrumentation for total knee arthroplasty: A systematic review and meta-analysis. Surgeon. 2017;15(6):336-48.

10. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Medical engineering physics*. 2018;51:6-16.
11. Luo H, Meyer-Szary J, Wang Z, Sabiniewicz R, Liu Y. Three-dimensional printing in cardiology: Current applications and future challenges. *Cardiol J*. 2017;24(4):436-44.
12. Mitsouras D, Liacouras P, Imanzadeh A, Giannopoulos AA, Cai T, Kumamaru KK, et al. Medical 3D printing for the radiologist. 2015;35(7):1965-88.
13. Hosny A, Keating SJ, Dilley JD, Ripley B, Kelil T, Pieper S, et al. From improved diagnostics to presurgical planning: high-resolution functionally graded multimaterial 3D printing of biomedical tomographic data sets. *3D Printing and Additive Manufacturing*. 2018;5(2):103-13.
14. El-Baz A, Jiang X, Suri JS. *Biomedical image segmentation: advances and trends*: CRC Press; 2016.
15. Withey DJ, Koles ZJ, editors. *Medical image segmentation: Methods and software*. 2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging; 2007: IEEE.
16. Caselles V, Catté F, Coll T, Dibos FJNM. A geometric model for active contours in image processing. 1993;66(1):1-31.
17. Caselles V, Kimmel R, Sapiro GJlJocv. Geodesic active contours. 1997;22(1):61-79.
18. Zhu SC, Yuille AJIToPA, Intelligence M. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. 1996(9):884-900.
19. Lorensen WE, Cline HE, editors. *Marching cubes: A high resolution 3D surface construction algorithm*. ACM siggraph computer graphics; 1987: ACM.
20. Botsch M, Pauly M, Kobbelt L, Alliez P, Lévy B, Bischoff S, et al. *Geometric modeling based on polygonal meshes*. 2007.
21. Bade R, Jens Haase, Preim B. Comparison of Fundamental Mesh Smoothing. *SimVis*. 2006;6.
22. Field DA. Laplacian smoothing and Delaunay triangulations. *Communications in applied numerical methods*. 1988;4(6):709-12.
23. Taubin G. A Signal Processing Approach To Fair Surface Design. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995:351-8.
24. Zhou T, Shimada KJI. An Angle-Based Approach to Two-Dimensional Mesh Smoothing. *IMR*. 2000;2000:373-84.
25. Tomasi C, Manduchi R, editors. *Bilateral filtering for gray and color images*. *Iccv*; 1998.
26. Semenova I, Savchenko V, Hagiwara IJWToS. Curvature based mesh improvement. *WSEAS Transactions on Systems*. 2004;3(4):1346-54.
27. Wake N, Alexander AE, Christensen AM, Liacouras PC, Schickel M, Pietila T, et al. Creating patient-specific anatomical models for 3D printing and AR/VR: a supplement for the 2018 Radiological Society of North America (RSNA) hands-on course. 2019;5(1):1-10.

28. Ibrahim D, Broilo TL, Heitz C, de Oliveira MG, de Oliveira HW, Nobre SM, et al. Dimensional error of selective laser sintering, three-dimensional printing and PolyJet models in the reproduction of mandibular anatomy. *Craniomaxillofac Surg.* 2009;37(3):167-73.
29. George E, Liacouras P, Rybicki FJ, Mitsouras D. Measuring and Establishing the Accuracy and Reproducibility of 3D Printed Medical Models. *Radiographics.* 2017;37(5):1424-50.
30. Cai T, Rybicki FJ, Giannopoulos AA, Schultz K, Kumamaru KK, Liacouras P, et al. The residual STL volume as a metric to evaluate accuracy and reproducibility of anatomic models for 3D printing: application in the validation of 3D-printable models of maxillofacial bone from reduced radiation dose CT images. *3D Print Med.* 2015;1(1):2.
31. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. *Academic Radiology.* 2004;11(2):178-89.
32. Pinto JM, Arrieta C, Andia ME, Uribe S, Ramos-Grez J, Vargas A, et al. Sensitivity analysis of geometric errors in additive manufacturing medical models. *Medical Engineering & Physics.* 2015;37(3):328-34.
33. Rendón-Medina MA, Andrade-Delgado L, Telich-Tarriba JE, Fuente-del-Campo A, Altamirano-Arcos CA. Dimensional error in rapid prototyping with open source software and low-cost 3D-printer. *Plastic and Reconstructive Surgery Global Open.* 2018;6(1).

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

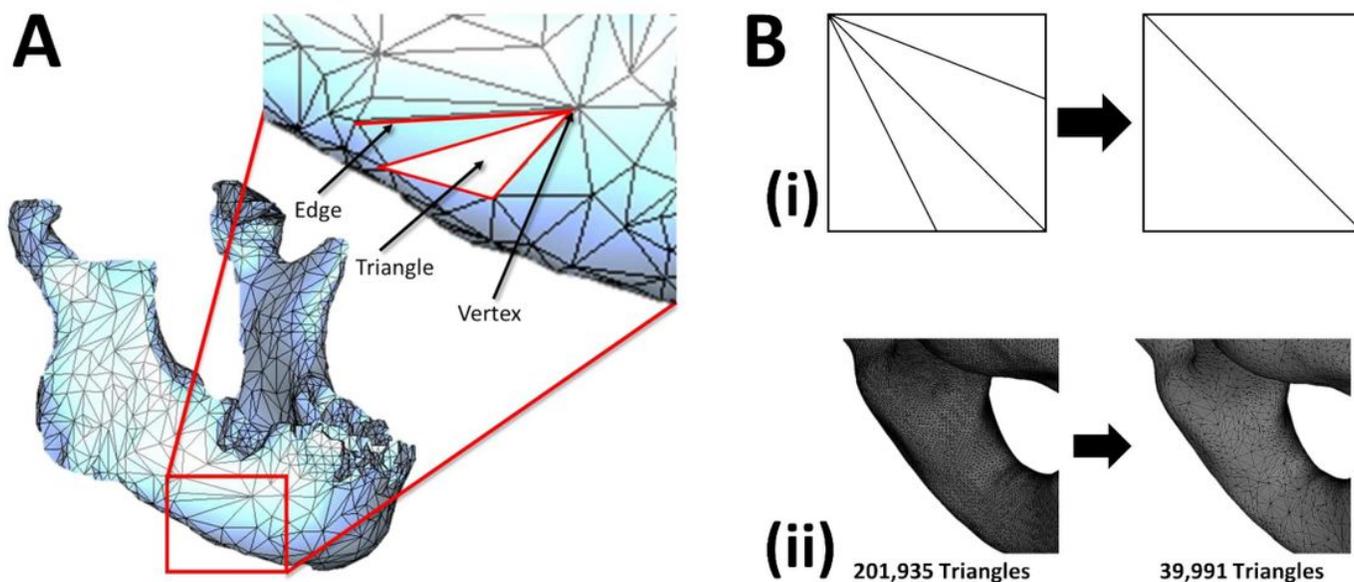


Figure 1

Meshing Basics: (A) Solid mesh of a mandible with major mesh features identified. (B) Decimation of features while maintaining fidelity of features in (i) a simple model and (ii) a complex model

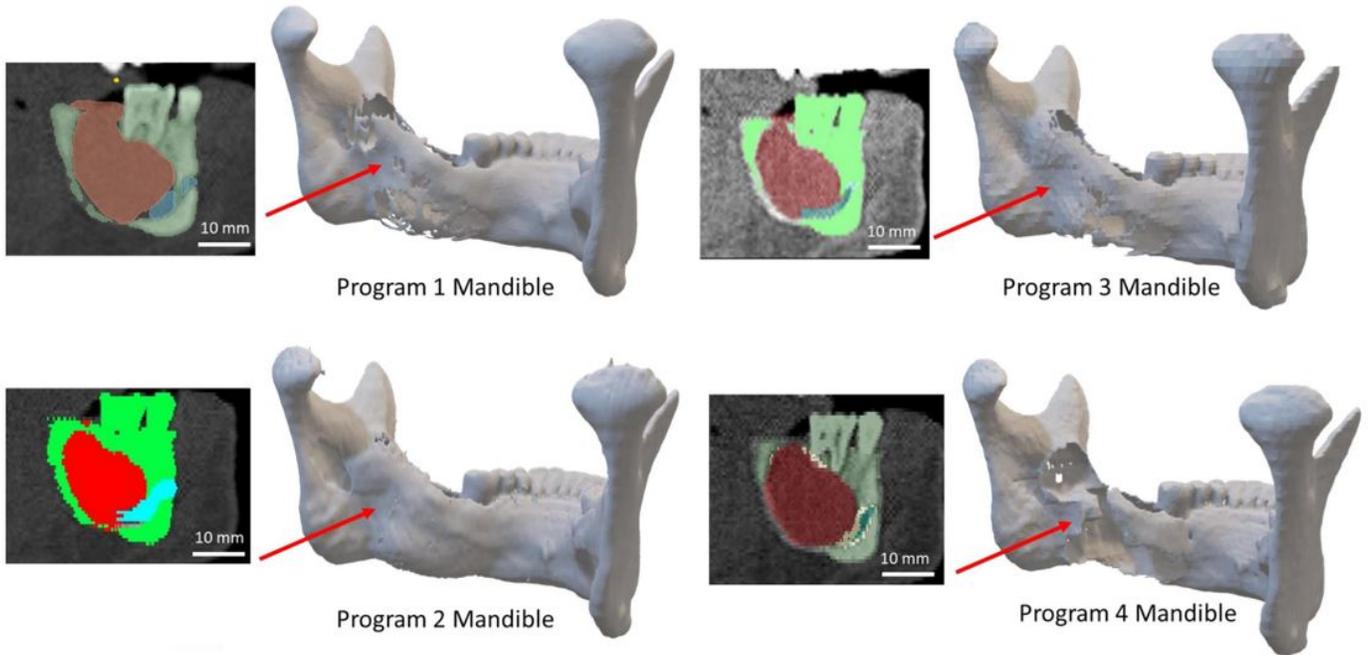


Figure 2

Volume representations of mandible segmentations with accompanying mask previews.

Intersection

-	Program 1	Program 2	Program 3	Program 4
Program 1	x			
Program 2		x		
Program 3			x	
Program 4				x

Union

+	Program 1	Program 3	Program 4
Program 2			
Program 1	x		
Program 3	x	x	

Figure 3

Tumor Difference Matrix. Intersection (L) and Union (R) matrices for the soft tumor used for calculating the agreement and disagreement statistics

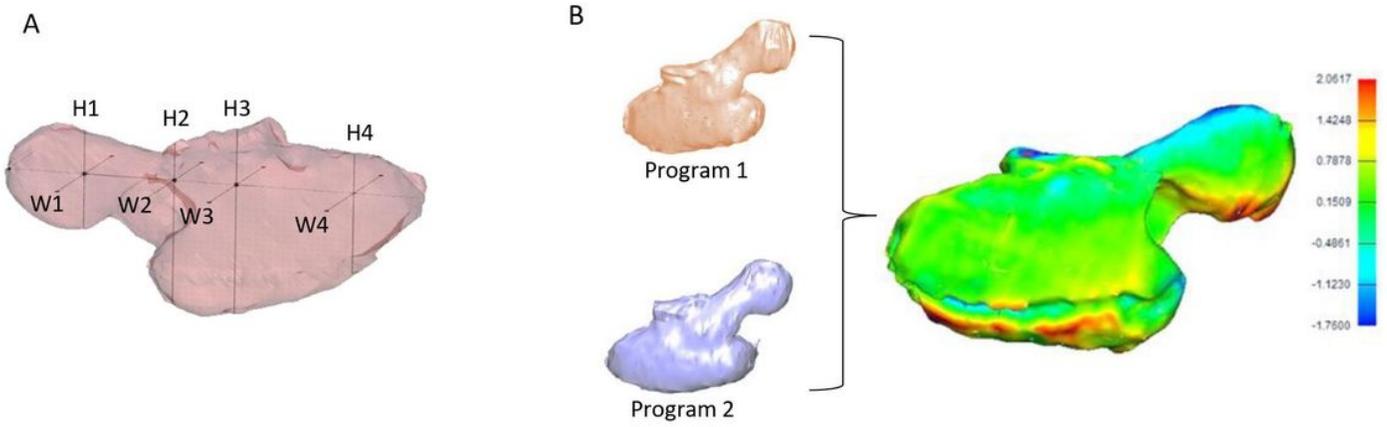


Figure 4

Arbitrary datum planes and locations of measurements on tumor (A). Heat Map Differences between Program 1 (top) and Program 2 (bottom) Tumor STLs, signed differences computed in 3-Matic, color bars in mm (B).

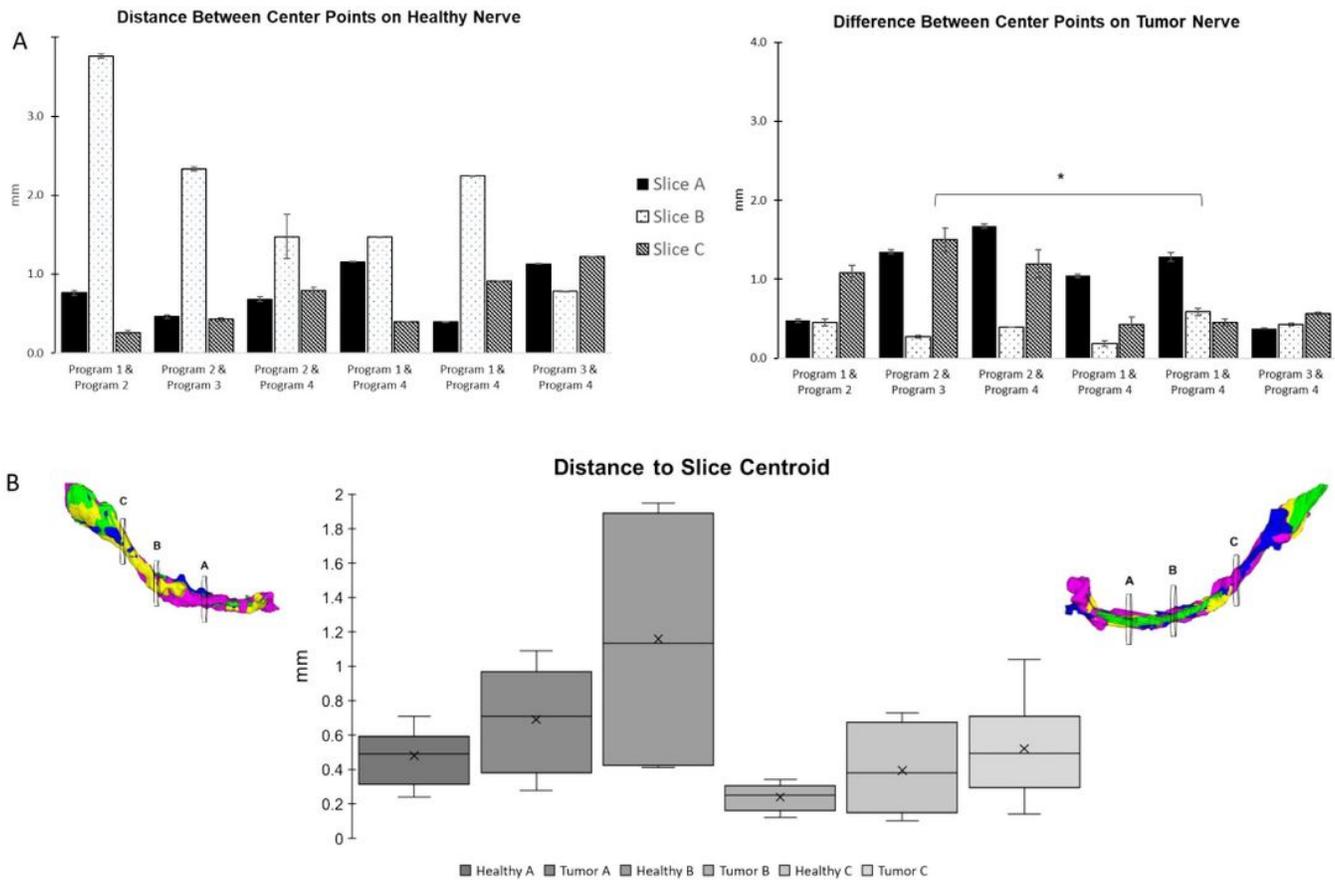


Figure 5

Summary of deviations in geometry between nerve segmentations. A) Differences between programs across all slices Healthy Nerve on left, Tumor Nerve on right. (* = $p < 0.05$, Tukey-Kramer, $Q_{6,39,0.05} = 4.237$) B) Differences between all programs per slice to centroid of each sliced based on the centroids of all models per slice alternating Healthy and Tumor Nerve centroid distance distributions for each slice. STL Stacks of Healthy (L) and Tumor (R) Nerves and location of slices on each.

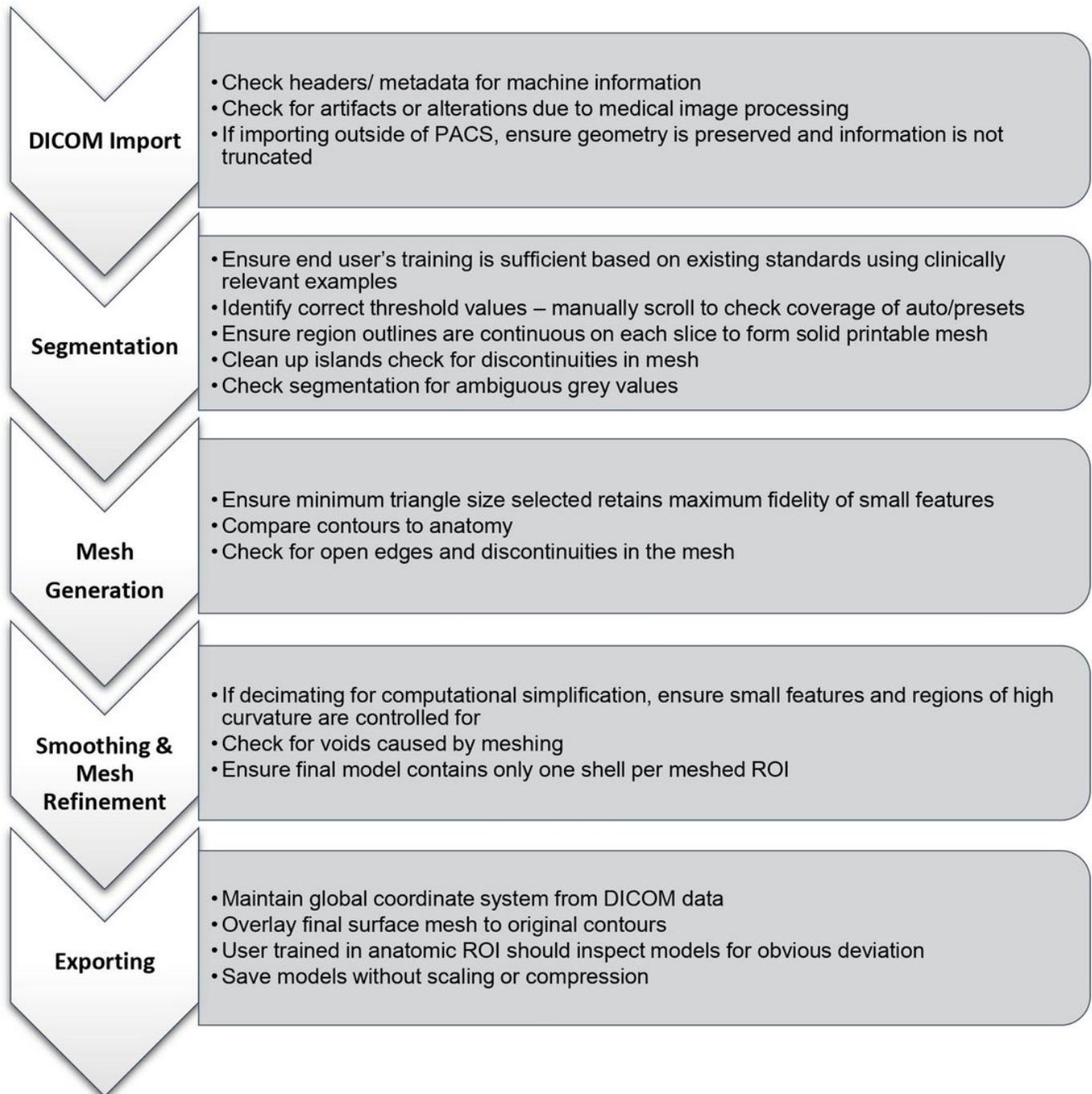


Figure 6

Check list to mitigate risk in digital workflow process. This is not an exhaustive list but provides basic checks for users to ensure that final STL files exhibit appropriate fidelity to original patient data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AssessingVariabilityinSegmentationAlgorithmsfor3DPrintingatthePointofCareTables.docx](#)