Thresholding Segmentation Errors and Uncertainty with Patient-Specific Geometries

Toma M.[1,2]*, Lu Y.[3], Zhou H.[3], Garcia J. D.[4]

ABSTRACT

Computer simulations provide virtual hands-on experience when actual hands-on experience is not possible. To use these simulations in medical science, they need to be able to predict the behavior of actual processes with actual patient-specific geometries. Many uncertainties enter in the process of developing these simulations, starting with creating the geometry. The actual patient-specific geometry is often complex and hard to process. Usually, simplifications to the geometry are introduced in exchange for faster results. However, when simplified, these simulations can no longer be considered patient-specific as they do not represent the actual patient they come from. The ultimate goal is to keep the geometries truly patient-specific without any simplification. However, even without simplifications, the patient-specific geometries are based on medical imaging modalities and consequent use of numerical algorithms to create and process the 3D surface. Multiple users are asked to process medical images of a complex geometry. Their resulting geometries are used to assess how the user's choices determine the resulting dimensions of the 3D model. It is shown that the resulting geometry heavily depends on user's choices.

Citation: Toma M, Lu Y, Zhou H, Garcia JD. Thresholding Segmentation Errors and Uncertainty with Patient-Specific Geometries. *J Biomed Phys Eng.* 2021;11(1):115-122. doi: 10.31661/jbpe.v0i0.2001-1062.

Keywords

Image Processing, Computer-Assisted, Errors; Uncertainty; Thresholding; Patient-Specific Modeling

[1]PhD, Department of Osteopathic Manipulative Medicine, College of Osteopathic Medicine, New York Institute of Technology, Old Westbury Campus, Northern Boulevard, Old Westbury, New York 11568-8000, USA

[2]PhD, Department of Mechanical Engineering, College of Engineering & Computing Sciences, New York Institute of Technology, Old Westbury Campus, Northern Boulevard, Old Westbury, New York 11568-8000, USA

[3]MSc, Department of Mechanical Engineering, College of Engineering & Computing Sciences, New York Institute of Technology, Old Westbury Campus, Northern Boulevard, Old Westbury, New York 11568-8000, USA

[4]DPT, Byrdine F. Lewis College of Nursing and Health Professions, Georgia State University, Urban Life Building, Suite 1280, Atlanta, GA, USA

*Corresponding author: M. Toma
Department of Osteopathic Manipulative Medicine, College of Osteopathic Medicine, New York Institute of Technology, Old Westbury Campus, Northern Boulevard, Old Westbury, New York 11568-8000, USA
E-mail: tomamil@tomamil.eu

Received: 30 January 2020
Accepted: 16 April 2020

Introduction

Medical imaging is mainly used to create visual representations of organs for clinical analysis and/or visual representation of their function. There are numerous software packages available to create 3D printable anatomic models from medical scans. With increasing frequency, patient-specific models are used to plan and optimize surgical procedures. Other uses include training simulators for educational purposes, verification and validation of new medical devices, and/or computational simulations and analyses.

Non-invasive MRI, CT or 3D ultrasound testing generates patient-specific imaging scans. Digital Imaging and Communications in Medicine (DICOM) is a standard for storing and transmitting medical images. The DICOM output provides a series of 2D images (slices) that are then reconstructed in 3D. The first step is to isolate the anatomy of interest (i.e. segmentation) and to generate a surface stereolithography (STL) file. The resulting STL file can be used for rapid prototyping, 3D printing, computer-aided manufacturing, and/or creating volume mesh. The volume mesh can then be used for computer simulations. A variety of commercial, freeware, and open-source software packages are available to perform the segmentation.

In segmentation, different schemes or algorithms have been developed to suit different conditions, i.e. multiseed traversal, threshold traversal and morphing-based volume splitting schemes. These schemes achieve relatively low bit rate overhead and high success rate (i.e. error resiliency), with meaning they are suitable for uniform pieces with small boundaries as patient-specific models.

Imaging artifacts and noise are always a big source of error affecting medical images with segmentation. Naturally, image acquisition artifact or artifact and noise is unable to avoid but only remove. The most used way to remove the artifact and noise is developing an algorithm or a tuning filter to reduce error and remove noises. However, it is beyond technology limit to completely achieve accurate imaging result bypass the artifact per noise influences. Once again, adding more complex method to make the results accurate will significantly increase the time and cost.

The first step in developing patient-specific models is to obtain DICOM data. All medical imaging devices, supporting the DICOM standard will export files readable by the editing software of user's choice. Human body is geometrically complex with multiple layers of varying densities. Thus, several imaging modalities are typically used for diagnostics and prognostics. One or more of the following modalities are used.

- MRI (magnetic resonance imaging) uses a strong magnetic field that excites hydrogen atoms in the body. The scanner then detects the radio frequency emitted by the hydrogen atoms. It is used for imaging soft structures in the body because hydrogen atoms in humans exist in large quantities especially in water and fat. Besides, it does not expose the patient to radiation.
- CT (computed tomography) uses a series of X-rays taken from different angles. In the resulting images, the brighter areas are denser than dark areas, e.g. a bone is brighter than the surrounding connective tissues. This method is quick and provides high resolution, but it is not best for soft tissues and results in radiation exposure.
- 3D ultrasound uses high-frequency sound waves sent into the body at different angles. As they reflect back, the receiving device displays them to produce a live 3D image of the internal organs. The method is cheap, provides live image, has no radiation, but it is of low resolution and does not show internal structures.

The software packages to display DICOM images vary from open-source software to enterprise-level solutions with the U.S. Food and Drug Administration's (FDA) clearance. There are dozens of options available. The data set must be segmented to separate the area of interest from the surrounding tissues. Segmentation is done by marking the relevant data and discarding the surrounding data.

Cutting-edge software solutions use advanced algorithms automatically or semi-automatically, separate specific organs from the surrounding tissues. Namely, some of methods used for automatic segmentation include region growing [1], region competition [2], digital subtraction [3], seed growing [4]. Recently, applying machine learning to medical imaging, especially deep learning, which has achieved state-of-the-art performance in image analysis and processing, is popular. In numerous instances, AI-based segmentation algorithms have successfully outperformed human experts. Such endeavors are actively transforming the field of medical image processing.

In scans where a certain anatomy has a very distinct set of pixel values, thresholding can be used, especially where high contrast is present. For example, in CT images, the pixel values represent the density. Hence, keeping only the brighter pixels leaves only the bone. The pixel values are set within a certain range with minimum value as black and maximum value as white. Values inside that range represent different shades of gray.

Moreover, it is important to keep in mind that thresholding segmentation is not the only algorithm applied in the process of acquiring the final geometry from DICOM images and segmentation is just one of the many sources of error. Each imaging modality has its shortcomings, e.g. image acquisition artifacts, corruption by noise. Moreover, various smoothing techniques, depending on the user, are used which determines the final geometry as well.

There has been a worldwide benchmark study to standardize CFD techniques used to assess the safety of medical devices [5], replicated also by our group with special focus on mesh sensitivity analysis [6]. The benchmark flow model used for this study consists of a nozzle with a concentrator and sudden expansion, as seen in Figure 1. Over 40 groups (self-ascribed as beginner, intermediate or expert) delivered their results. The graph in Figure 1 shows the best and worst fitting results from each group when compared to experimental results from 3 laboratories.

The results of the FDA study show that CFD results always need to be validated even when produced by experts. It can be seen, in Figure 1, that even the worst 'beginner' is better than the worst 'expert'. The study presented in this paper shows that the same logic can be applied to perform thresholding segmentation of patient-specific geometries where the threshold maximum and minimum values need to be set manually.

Many of the resulting geometries are subsequently used for CFD analyses. Hence, if both the CFD and geometry do not represent the actual flow and actual patient, respectively, the resulting analysis is effectively useless. However, every group's effort on validation is focused on validating the CFD only, if at all. Moreover, if some CFD analyses can produce wrong results even with such a simple geometry as the sudden expansion, using complex

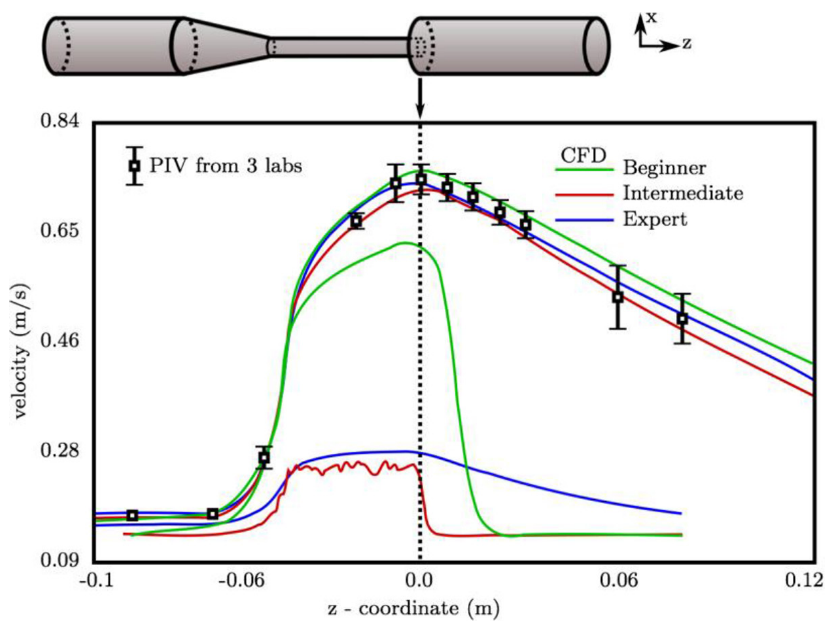

Figure 1: Results from the U.S. Food and Drug Administration's (FDA) "Critical Path" project to validate computational fluid dynamics (CFD) methods [5]. The best and worst fitting results are shown from each (self-ascribed) category.

patient-specific geometries even with a little variation from the actual patient's geometry can yield unrealistic results. The ultimate purpose of these simulations is to guide the decisions that physicians make to improve and save lives. Therefore, having confidence in the results is axiomatic. In this study, multiple users are asked to process the DICOM data of the same complex geometry. Their resulting geometries are used to assess how the user's choices determine the resulting dimensions of the 3D model.

Technical Presentation

To show that different users can produce different geometries when using the same set of medical images, an example of patient-specific geometry with high level of complexity ought to be considered. That requires a geometric model of human organ, containing both large and relatively small parts connected together. When smoothing techniques are applied, the smaller parts tend to suffer more, comparatively, from volume shrinkage. Volume shrinkage is a common issue with smoothing algorithms. Even when some of the shrinkage-free smoothing algorithms are used, the smaller parts of very complex geometric models tend to experience shrinkage as the algorithms focus on preserving the overall volume. Small, even seemingly negligible, the change of the overall volume of a complex geometry with both small and large parts connected together causes unacceptable shrinkage of the smallest parts of the geometric model. For this purpose, mitral valve (MV) geometry seems to be one of the best choices. The MV geometry is complex with large number of tendinous chords (chordae tendineae) with small diameters. When creating the MV geometry from the medical images with the focus on preserving all its components, all above mentioned smoothing-related issues arise. To preserve the chords in the patient-specific MV model, it is axiomatic, especially when that model is used to analyze chordae-related diseases, e.g. acute mitral regurgitation resulting from rupture of chordae tendineae [7]. In recent MV studies, thresholding segmentation is used to process the DICOM data resulting from μCT imaging [7, 8, 9]. Arguably, the user's choice of the threshold values determines the dimensions of resulting geometry. Especially with complex geometry, such as MV with multiple smaller parts, small variations in the threshold values depending on the user can lead to significantly different outputs.

A DICOM set of images from the MV studies using μCT imaging, see [7, 8, 9], is provided to 6 different users with their own preferences for the methods used. Except for the use of threshold segmentation, no other instructions are given. An example of geometry produced by the thresholding segmentation and smoothing by one of the users is shown in Figure 2(a).

The six users with different level of expertise processed the provided DICOM images and returned their resulting STL file for our analysis. Part of the MV geometry with largest concentration of chords is chosen for further analysis (Figure 2(b)). The diameters of these chords are measured for consequent comparative study to assess how the user's choices in processing the medical images of complex patient-specific geometries determine the resulting dimensions of the model.

All the measurements are summarized in Table 1 with corresponding mean values and standard deviations (SD) in the last two columns. The largest SD values are observed with chords of the largest diameters. A standard deviation close to 0 indicates that the diameter values are closer to the mean, while a high standard deviation indicates that the diameters are spread out over a wider range of values.

The resulting diameter values are further organized in Figure 3. For each chord, a graph is shown with the number of measurements, falling within given range of diameter values. It can be seen that for the chords with mean diameters of under 0.8 mm, all the measurements can be found within two or three ranges.

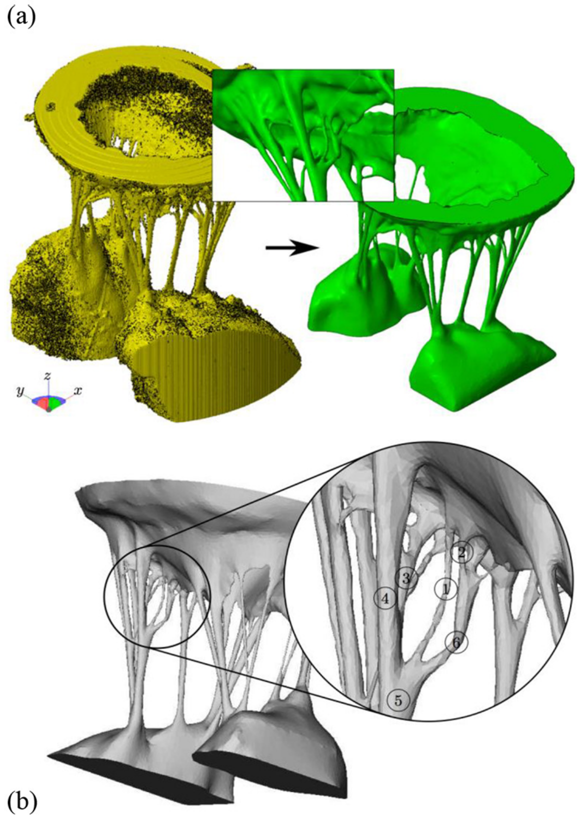

Figure 2: (a) Original (yellow) surface from the threshold segmentation compared to the final surface mesh (green) after smoothing techniques applied. Zoom-in to an attachment points between the chords and leaflets is shown to demonstrate the complexity of the geometry. (b) When the users returned their processed geometry the diameters of 6 chords (numbered) have been measured for comparison to analyze how user's choices determine the outcome.

For chords with mean diameters over 0.8 mm, the measurements have a wider spread.

Discussion

According to the worldwide benchmark study used to standardize CFD techniques, even experts can produce results that do not match the experimental measurements [5, 6]. It is worth to mention again that the referenced study used a simple geometry, while in patient-specific simulations, geometries of higher complexity are used. The ultimate purpose of computer simulations is to create methods and algorithms that can be trusted and, most of all, useful. For that, they need to produce results that accurately simulate real-life, i.e. not simplified, processes. A variety of techniques is used in the process to secure the accuracy of the simulations, e.g. mesh sensitivity studies are performed to assure the convergence of the results and the numerical methods used are validated against experimental (e.g. ex-vivo) measurements [10].

However, both the ex-vivo experiments (e.g. blood flow in arteries) and computer simulations using patient-specific geometries rely on accurate medical scanning and subsequent image processing techniques. If the 3D patient-specific model, that is also used in the ex-vivo experiments, does not accurately represent the patient, the computer simulation results might be validated and therefore considered accurate, but they cannot be used to assess the actual patient. Therefore, from this point

Table 1: Measured diameters [mm] of the 6 chords in geometries created by 6 different users working with the same Digital Imaging and Communications in Medicine (DICOM) set of images.

#	User 1	User 2	User 3	User 4	User 5	User 6	Mean	Standard Deviations (SD)
1	0.18668	0.15444	0.36169	0.13313	0 .15776	0.21982	0.20225	0.08368
2	0.45256	0.43443	0.53459	0.45575	0.45772	0.42153	0.45943	0.03943
3	0.44166	0.50039	0.48678	0.53943	0.50521	0.74251	0.53599	0.10602
4	0.74034	1.12607	0.93119	0.84165	0.83069	0.88540	0.89256	0.13088
5	0.95616	1.44648	1.34057	1.07457	1.41971	1.05813	1.21594	0.21099
6	0.51601	0.79046	0.58558	0.62163	0.60300	0.58943	0.61768	0.09192

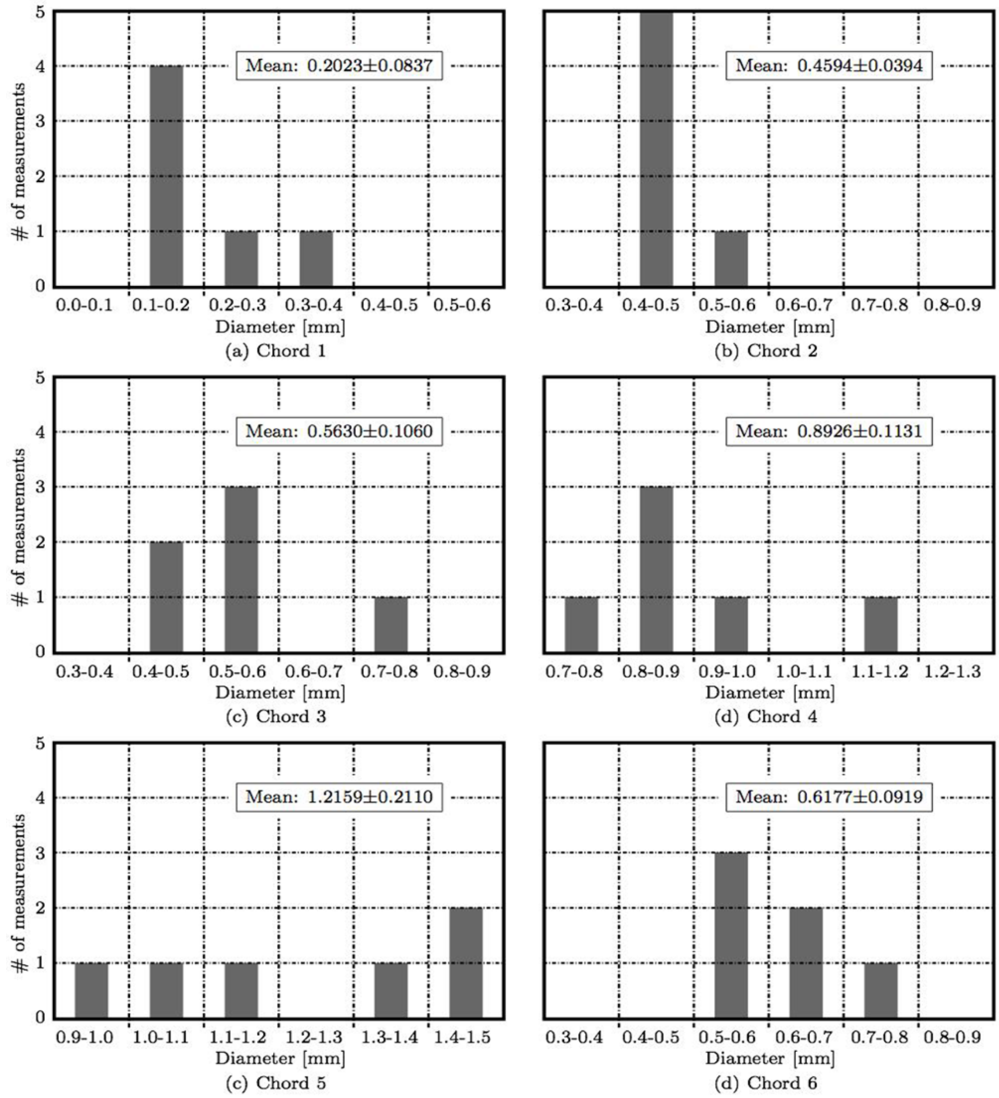

Figure 3: Number of diameter measurements, falling within a given range for each of the six chords.

of view, "accurate" does not mean "realistic". This work shows that different users processing medical images of complex organs can produce geometries with varying dimensions. This uncertainty can be avoided if the medical images are not used for both the ex-vivo experiments and computations at the same time. Considering all the new advancements in image segmentation techniques, the uncertainty discussed in this study will be gradually less damaging. In order to achieve more trustworthy validations of numerical models used, the numerical results should be compared against in-vivo measurement if possible.

Before constructing 3-D model for simulation analysis, uncertainty already existed from imaging the parts needed to be analyzed. Despite the fact that the modern technology saves time while providing accurate results, the patient-specific model will not be the same from the imaging. Therefore, the simulation results from different users will be even more apart from each other. Nevertheless, a wide variety of approaches have been proposed to reduce the error and uncertainty in segmentation with imaging and 3-D modeling. These studies helped to select proper values of threshold to improve the quality of analyzed model. In addition, artificial intelligence tools are being developed to help the study.

Conclusion

In engineering mostly, computational simulations of high fidelity are increasingly becoming an indispensable tool designed to replace laboratory experiments. However, when the simulations are too complex, they can be computationally expensive (i.e. have long runtime) and have convergence issues. To overcome these obstacles, users tend to simplify the geometries used. Needless to say, it is more desirable to work with more manageable geometries instead of computationally costly fully comprehensive patient-specific models. Even converting the medical images to geometric 3D models would be faster if the resulting geometries could be kept simplified. However, to keep the model patient-specific, it has to truly represent the patient from whom the model was extracted. If a simplified model is used, arguably it does not represent that patient anymore. Even though, modelers are generally aware of the fact that their models have a limited scope and are predictive for a certain range of applications. For example, the MV model used in this study does not always need to include the 3D representation of its chordae tendineae. For most purposes, e.g. when simulating healthy closure, the chordal part of MV is usually replaced by 1D elements.

The main limitation of this study is that it includes only two sources of error causing geometric models to deviate from the patient-specific anatomy, i.e. manual segmentation and the use of smoothing techniques. Such interpretation would be inaccurate as there are other sources of error to be considered, e.g. image acquisition artifacts, corruption by noise, and inherent shortcomings of each imaging modality. However, the subsequent threshold segmentation is expected to be partially influenced by some of these shortcomings.

Conflict of Interest

None

References

1. Zucker SW. Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*. 1976;**5**(3):382-99. doi: 10.1016/S0146-664X(76)80014-7.
2. Zhu SC, Yuille A. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1996;**18**(9):884-900. doi: 10.1109/34.537343.
3. Balter S, Ergun D, Tscholl E, Buchmann F, Verhoeven L. Digital subtraction angiography: fundamental noise characteristics. *Radiology*. 1984;**152**(1):195-8. doi: 10.1148/radiology.152.1.6374758. PubMed PMID: 6374758.
4. Adams R, Bischof L. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1994;**16**(6):641-7. doi: 10.1109/34.295913.
5. Stewart SFC, Day S, Burgreen GW, Paterson EG, Manning KB, Hariharan P, et al. Preliminary results of FDA's 'Critical Path'project to validate computational fluid dynamic methods used in medical device evaluation. *Asaio J.* 2009;**55**(2):173.
6. Toma M. Computational Fluid Dynamics Simulations Using FDA's Idealized Medical Device Demonstrating the Importance of Model Validation. *Biomed Res Rev.* 2018;**1**(1):104. doi: 10.31021/brr.20181104.
7. Toma M, Bloodworth CH, Pierce EL, Einstein DR, Cochran RP, Yoganathan AP, Kunzelman KS. Fluid-structure interaction analysis of ruptured mitral chordae tendineae. *Annals of Biomedical Engineering*. 2017;**45**(3):619-31. doi: 10.1007/s10439-016-1727-y.
8. Bloodworth CH, Pierce EL, Easley TF, Drach A, Khalighi AH, Toma M, Jensen MO, Sacks MS, Yoganathan AP. Ex vivo methods for informing computational models of the mi-

tral valve. *Annals of Biomedical Engineering*. 2017;**45**(2):496-507. doi: 10.1007/s10439-016-1734-z. PubMed PMID: 27699507. PubMed PMCID: PMC5300906.

9. Toma M, Bloodworth CH, Einstein DR, Pierce EL, Cochran RP, Yoganathan AP, Kunzelman KS. High-resolution subject-specific mitral valve imaging and modeling: experimental and computational methods. *Biomech Model Mechanobiol.* 2016;**15**(6):1619-30. doi: 10.1007/s10237-016-0786-1. PubMed PMID: 27094182.
10. Wei ZA, Sonntag SJ, Toma M, Singh-Gryzbon S, Sun W. Computational fluid dynamics assessment associated with transcatheter heart valve prostheses: a position paper of the ISO Working Group. *Cardiovasc Eng Technol.* 2018;**9**(3):289-99. doi: 10.1007/s13239-018-0349-y. PubMed PMID: 29675697. PubMed PMCID: PMC6095731.