Analysis of Interindividual Variance in Coronary Sinus Veins Anatomy Based on Computer Tomography Data

Aleksandr Sinitca[1], Arsenii Dokuchaev[1], Chiara Arduino[1], Stepan Zubarev[1,2], Margarita Budanova[1,2], Sergei Rud[2], Anastasia Bazhutina[1], Svyatoslav Khamzin[1], Mikhail Chmelevsky[1,2]

[1] XSpline Spa, Bolzano, Italy
[2] Almazov National Medical Research Centre, Saint-Petersburg, Russia

Abstract

The coronary sinus (CS) anatomy exhibits substantial inter-individual variance. Despite its clinical relevance, comprehensive studies of CS geometric variability are limited, with the majority investigations relying on autopsied human hearts. The purpose of current study was to propose an approach to analyze differences in CS anatomy. A cohort of 107 patients underwent contrast enhancement computed tomography, followed by manual segmentation of the CS and heart. Three-dimensional voxel CS models were normalized using the CS projection on the epicardial surface of the left ventricle and a modified universal coordinate system. Taking each distribution as a sample random variable, the distance was estimated using the Wasserstein distance (WD) and Kullback-Leibler (KL) divergence metric. These distances were then applied to numerically evaluate the variance of the set. To assess applicability, the set was randomly split into two equal subsets. Finally, intra- and intergroup and total variance were calculated based on distance, namely 0.071*,* 0.083*,* 0.077 *and* 0.154 *for WD-based values and* 0.037*,* 0.040*,* 0.041 *and* 0.080 *for KL-based, respectively. Intergroup and intragroup variance were comparable.*

1. Introduction

The high variability of coronary sinus (CS) veins is a well-known problem in the medical field [1]. At the same time, this structure is widely used for several types of invasive medical procedures. For this reason, it would be helpful to perform a reconstruction of CS prior to an intervention. Obtaining such information allows, for example, better prediction of left ventricle (LV) lead implantation and also reduces radiation exposure time for medical staff and the patient.

Difficulties in reconstruction and in making different measurements (length, angle, diameter) may be due to variability in CS veins geometry. Currently, this task can be accomplished using machine learning-based approaches, mainly deep learning neural networks, as previously proposed [2], combined with measurement algorithms. However, transferring the technology from research to clinical practice requires in-depth validation to mitigate risks. Thus, one of the main issues is the representativeness and equivalence of the datasets used for training and validation of data processing pipelines, as well as the consistency of the validation dataset with the target population. The traditional approach is based on statistical analysis of the patient population, which cannot demonstrate equivalence of CS anatomy. As a consequence, this leads to clinical trials to test the applicability of the method in each target population, which is not feasible due to the time-consuming, complex, large amount of data required and expert labeling of the collected medical imaging data. An alternative approach is to prove the equivalence of the distribution of anatomies between populations, which is a non-trivial issue given the three-dimensional nature of anatomies.

In current study, we propose an approach to quantify the variability in CS veins anatomy. This involves projecting anatomies into the same two-dimensional coordinate space and, further, analyzing anatomical variability based on the well-known Wasserstein distance (WD) [3] metric and the Kullback-Leibler (KL) [4] divergence in other scientific fields.

2. Materials and Methods

2.1. Dataset

In this study, we used cardiac computed tomography (CT) data from 107 candidates scheduled for conventional cardiac resynchronization therapy with LV lead implantation at the Almazov National Medical Research Center. The scans were performed according to the CT protocol that was proposed for the CRT-DRIVE clinical trial (ClinicalTrials.gov ID: NCT05327062). Written informed con-

sent was obtained from each patient.

Each Dicom CT image was manually labeled by our cardiologists according to the following classes: CS anatomy similar to the previous study [2] and segmentation of ventricular geometry [5]. Both manual segmentations were prepared using the same CT series for each case, ensuring similar spatial parameters of the masks such as origin, spacing, etc.

The manually obtained CS segmentation, that is, a three-dimensional image with a voxel mask of CS veins, was converted into a point cloud where each point corresponded to the center of the labeled voxel in world Cartesian coordinates. At the same time, manual ventricular segmentation was converted into a tetrahedral mesh, and then the mesh was refined using a modified universal ventricular coordinate (UVC) system prepared as proposed in [6]. UVC describes the position of each point in the biventricular model using 3 coordinates:

- the apicobasal coordinate represents the direction from the apex to the base of the ventricle (Fig. 1 A, left);
- the rotational coordinate (Circle) represents the circumferential rotation along the long axes of left and right ventricle (Fig. 1 B, right);
- the transmural coordinate (Radial) represents the distance from the endocardium to the epicardium.

2.2. CS Coordinates normalization

The first step in data processing is normalization. The necessity of this step is justified for the following reasons. First, all CT images are at different locations, and the hearts may be at different coordinates. The second reason is the variability in heart size, which makes a direct comparison of CSs not meaningful.

Thus, CS coordinates were projected onto homogeneous coordinates, for which a modified UVC [6] system was proposed. Given the proximity of the CS veins to the LV epicardial surface and the absence of transmural penetration into the LV walls, it is possible to substitute the Radial universal coordinate with only two coordinates: Circle and Apex-Base, which are sufficient. To determine the Circle coordinate on the CS point cloud, each point of the CS is projected onto the epicardial surface of the LV, followed by interpolation of the Circle coordinate from the LV surface onto each of the projected CS points using a radius basis function with a linear kernel.

Next, to determine the Apex-Base coordinate on CS, the CS points are projected into the central long axis of the LV. Then, a linear interpolation of the Apex-Base coordinate of the LV points is performed along the long axis, followed by interpolation onto the CS points. Since the CS points at the base of the LV extend beyond the geometry, the Apex-Base coordinate for these points is determined by linearly extrapolating the Apex-Base coordinate of the ventricles.

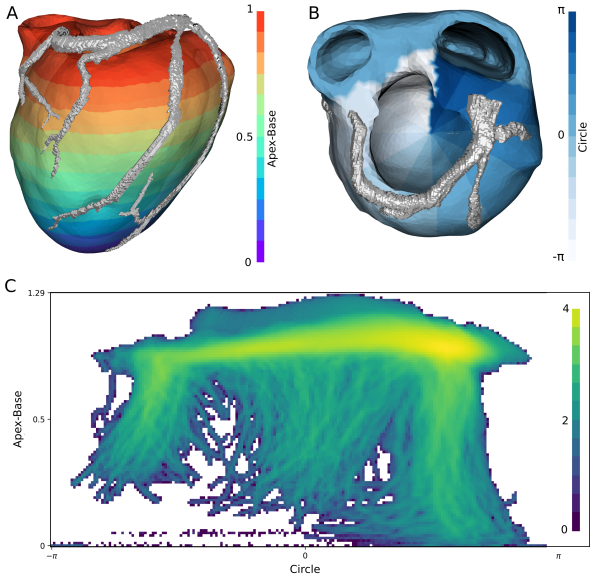

Figure 1. Top: an example of ventricle (colored by UVC coordinate values) geometry mesh with corresponding CS point cloud (white). **A.** Apex-Base UVC coordinate from 0 (blue color) to 1 (red color), **B.** Circle UVC coordinate in blue colors in range $-\pi$ to π radian. **C.** Heatmap of points density of CS point clouds, projected onto plane formed by Circle and Apex-Base UVC. Color scale indicates the probability density at each image point.

Hence, the corresponding UVC coordinate can be found for each point of CS, which allowed us to project the three-dimensional point cloud onto a two-dimensional plane as shown in Figure 1.

2.3. Variability Assessment

One of the main purposes of the variability assessment is a comparison of datasets. Therefore, we demonstrate the approach using random separation of the data into two groups. The random split of the dataset shall provide two sets that come from the same distribution by design.

Assumption 1. The CS anatomy is a random variable, and the CS representation made earlier is a sample from that random variable. And the variance of this random variable represents the variability of CS anatomy.

Next, according to [7], the variance of the random variable can be estimated using the distances between the samples as follows.

$$\sigma^2(X) = \frac{1}{n^2} \sum_j \sum_{j>i} (D_{ij})^2 \tag{1}$$

where, X is a random variable, D_{ij} is distance between samples i and j. Thus, it is possible to estimate the variance of the distribution of the CS anatomies using the same

approach but using the WD metric [3] and the KL divergence [4] instead of the Euclidean distance.

The WD, also known as the Earth Mover distance, is a measure of the distance between two probability distributions. In the theory of optimal transport, the WD is used to quantify the 'cost' of transporting the mass to convert one distribution to another. This cost is defined in terms of the amount of work it takes to rearrange the mass distributions, hence the name 'Earth Mover's Distance'. The mathematical definition of the WD between two probability measures P and Q on a metric space with metric d is given by:

$$W(P,Q) = (\inf_{\gamma \in \Gamma(P,Q)} \int d(x,y)^p d\gamma(x,y))^{1/p} \qquad (2)$$

where $W(P,Q)$ is the WD between P and Q, $\Gamma(P,Q)$ is the set of all joint distributions $\gamma(x,y)$ whose marginals are respectively P and Q, $d(x,y)$ is the distance between points x and y, $p = 20$ is WD order. In the scope of this work has been used an implementation proposed in [8].

The KL divergence, also known as relative entropy, is a measure of how a probability distribution P is different from a second reference probability distribution. The KL divergence is calculated using the following formula:

$$D_{KL}(P \parallel Q) = \sum_i P(x) \log \left(\frac{P(i)}{Q(i)} \right) \qquad (3)$$

where $D_{KL}(P \parallel Q)$ is the KL divergence of Q from P, $P(i)$ is the probability of event i according to distribution P, $Q(i)$ is the probability of event i according to distribution Q.

Using this approach, a total variance (σ^2_{tot}) and an intragroup variance (σ^2_i) can be estimated. Next, using the variance sum law, an intergroup variance (σ^2_{ig}) can be estimated as follows.

$$\sigma^2_{ig} = \sigma^2_{tot} - \frac{1}{N} \sum_i n_i \sigma^2_i \qquad (4)$$

where N is total number of samples, n_i is number of sample in i-th group.

3. Results

KL divergence and WD were estimated for the datasets used. 1378, 1431, and 11449 distances were estimated for the first, second, and entire dataset, respectively. The distributions of the corresponding distances are shown in Figure 2. As we can see, there is no expressed multimodality in the distributions, but some minor modes near 0.6 WD and 0.4 KL.

Variance	Wasserstein distance	KL-divergence
Intragroup-1	0.071	0.037
Intragroup-2	0.083	0.040
Total	0.154	0.080
Intergroup	0.077	0.041

Table 1. Estimations of intra- and intergroup variances based on Wasserstein and KL distances.

The values estimated according to equations 1 and 4 are presented in Table 1. Estimates of intergroup and intragroup variances have comparable values (0.077, 0.071, and 0.083 in terms of WD, respectively). Furthermore, under the assumption that the variability between different groups should be much larger than the variability within a group (according to the Fisher criterion proposed in [9]), we can predict that the subsets come from the same distribution.

4. Discussion

The WD measures the "cost" of transforming one distribution into another, capturing differences in both the location and the variance of CS anatomies. In contrast, KL divergence measures the difference in informational content (shape and entropy) between distributions. However, both metrics allow the assessment of anatomical variability without significant divergence from theoretical estimates.

The Distribution of Distances shown in Figure 2 has a weakly expressed multimodality, which is consistent with the conventional understanding of the variability of CS [1].

However, the current study has a number of limitations that need to be addressed in future works. First, it is necessary to perform an estimation of the variability between known classes of anatomies. Second, the proposed way to estimate the variance is challenging to interpret.

Nevertheless, the proposed approach can be used not only for variability estimation, but also for automatic classification of CSs or even anomaly detection. Being close to commonly used techniques for open-set classification and reidentification, namely distance-based approaches [10].

Moreover, the proposed approach can be extended. For example, using the same coordinate mapping when evaluating coronary artery anatomy or ventricular anatomy. Or by using a different coordinate normalization for other structures.

5. Conclusion

In the present study, an approach to estimate the numerical variability is proposed. This approach is based on the conversion of anatomy into two-dimensional coordinates and estimation of WD and KL divergence.

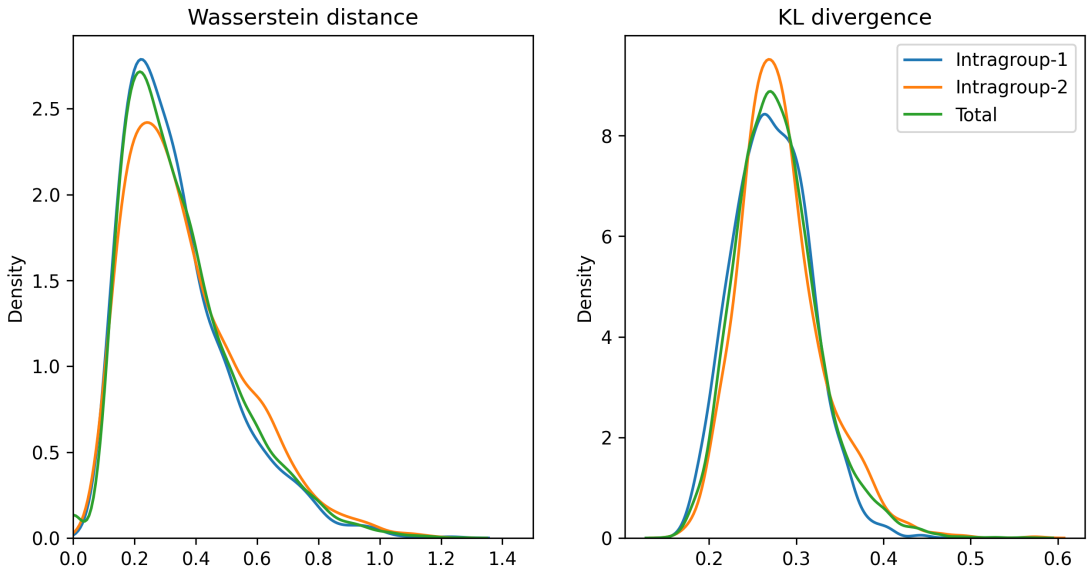

Figure 2. Histograms of intra-group and total distributions of WD (left) and KL divergences (right) between samples.

In the scope of the study, the proposed estimation has no significant discrepancy with the theoretically predicted results. Intergroup and intragroup variance are comparable. Moreover, based on the distribution of distances, it can be assumed that sufficient sensitivity is able to detect differences in anatomy. This may be the subject of further research.

Acknowledgments

Type of funding sources: Private company. Main funding source(s): XSpline Spa.

References

[1] von Lüdinghausen M. The Venous Drainage of the Human Myocardium. Springer Berlin Heidelberg, 2003. ISBN 9783642556234.

[2] Sinitca A, Chmelevsky M, Zubarev S, Shirshin A, Dokuchaev A, Budanova M, Arduino C, Khamzin S, Bazhutina A, Rud S, et al. A machine learning-based approach for automatic coronary sinus vein segmentation and anatomy reconstruction. In 2023 Computing in Cardiology (CinC), volume 50. IEEE, 2023; 1–4.

[3] Vaserstein LN. Markov processes over denumerable products of spaces, describing large systems of automata. Problemy Peredachi Informatsii 1969;5(3):64–72.

[4] Kullback S, Leibler RA. On information and sufficiency. The annals of mathematical statistics 1951;22(1):79–86.

[5] Arduino C, Zubarev S, Budanova M, Chmelevsky M, Rud S, Trufanov G, Sinitca A. Enhanced automatic segmentation of epi-endocardial ventricular anatomy using multi-architecture neural network ensemble. CinC 2024, in press.

[6] Bazhutina A, Khamzin S, Chmelevsky M, Zubarev S, Sinitca A, Budanova M, Rainer W. An automated algorithm for generating of aha model based on 3d heart geometry. In 2023 Computing in Cardiology (CinC), volume 50. IEEE, 2023; 1–4.

[7] Zhang Y, Wu H, Cheng L. Some new deformation formulas about variance and covariance. In 2012 Proceedings of International Conference on Modelling, Identification and Control. 2012; 987–992.

[8] Flamary R, Courty N, Gramfort A, Alaya MZ, Boisbunon A, Chambon S, Chapel L, Corenflos A, Fatras K, Fournier N, Gautheron L, Gayraud NT, Janati H, Rakotomamonjy A, Redko I, Rolet A, Schutz A, Seguy V, Sutherland DJ, Tavenard R, Tong A, Vayer T. Pot: Python optimal transport. Journal of Machine Learning Research 2021;22(78):1–8.

[9] Fisher RA. The use of multiple measurements in taxonomic problems. Annals of eugenics 1936;7(2):179–188.

[10] Sun J, Dong Q. A survey on open-set image recognition, 2023.

Address for correspondence:

Aleksandr Sinitca
XSpline Spa, Via Ressel 2F, 39100, Bolzano, Italy
sinitca@xspline.com