Concordance of Pericardial Effusion Size Between Computed Tomography and Echocardiography

Robert S. Zhang, MD[a], Eric Alter, MD[b], Samuel Kozloff, MD[c], Alana Choy-Shan, MD[a,d], Yuhe Xia, MS[e], Kunal Patel, MD[f], Elliott K. Gozansky, MD, PhD[g], Muhamed Saric, MD, PhD[a], Jadranka Stojanovska, MD, MS[g], and Robert Donnino, MD[a,d,g,*]

Discrepancy between computed tomography (CT) and transthoracic echocardiography (TTE) regarding pericardial effusion (PEff) size is common, but there is limited data regarding the correlation between these 2 imaging methods. The aim of this study is to examine the real-world concordance of observed PEff size between CT and TTE. We performed a retrospective analysis of all imaging reports available from 2013 to 2019 and identified patients with a PEff who underwent both a chest CT and TTE within a 24-hour period. We evaluated the agreement between CT and TTE in assessing PEff size. Of 1,118 patients included in the study, mean age was 66 (±17 years) and 54% were female. The median time interval between the 2 studies was 9.4 hours (interquartile range 3.5 to 16.6). Patients within a half-grade or full-grade of agreement were 71.9% and 97.2%, respectively. The mean difference in grade of agreement (TTE minus CT) between the 2 imaging methods was −0.1 (±0.6, p <0.0001). CT was more likely to report a higher grade (i.e. larger PEff size) when compared with TTE (261 patients vs 157 patients, p <0.001). The weighted kappa was 0.73 (95% confidence interval 0.69 to 0.76). After excluding patients with trace/no effusion, 42.3% and 94.1% of patients' studies were within a half-grade or full-grade of agreement, respectively. Of the 18 patients who had large discrepancies, 9 patients had loculated effusions, 2 patients had large pleural effusions, and 6 patients had suboptimal TTEs images. In conclusion, TTE and CT showed relatively strong agreement in estimation of PEff size, with CT sizes larger than TTE, on average. Large discrepancies in size may be related to reduced image quality, large pleural effusions, and loculated PEff. Published by Elsevier Inc. (Am J Cardiol 2023;203:92−97)

Pericardial effusion (PEff) is a common clinical entity, with implications that range from a benign incidental imaging finding to life-threatening disease.[1] Transthoracic echocardiography (TTE) remains the primary imaging modality to assess PEff, although the presence of effusion is often detected initially by computed tomography (CT), which may offer additional information regarding the nature of the effusion and associated thoracic findings beyond that of TTE.[1] Despite their widespread clinical use, there has only been limited data published regarding the correlation between CT and TTE in evaluation of PEff size. Previous studies comparing the 2 imaging methods have been limited to small sample sizes with long intervals between imaging studies.[2,3] Appropriate classification of PEff size is pivotal to guide management and decrease the need for unnecessary testing. Although guidelines exist on how to report PEff size on TTE,[1] there remains a significant amount of subjectivity, especially in asymmetric and/or loculated effusions. There are no standardized, quantitative guidelines on how to report PEff size on CT. Furthermore, most routine CT studies are not electrocardiogram-gated and as a result, PEff size cannot be measured at a consistent point in the cardiac cycle. Finally, chest CT scans are typically read by chest radiologists and echocardiograms by cardiologists, another potential source of discrepancy between methods. Given these and other potential limitations of quantifying PEff size with both imaging methods, our aim was to examine the real-world concordance of clinically reported PEff size between CT and TTE with a large sample size of patients with imaging studies that had been performed within a 24-hour period of each other.

Methods

We performed a retrospective analysis of all clinical imaging reports available in the electronic medical record from 2013 to 2019 at a large academic center to identify patients who underwent both a chest CT and echocardiogram within a 24-hour period, with a PEff reported on

[a]Leon H. Charney Division of Cardiology, Department of Medicine, New York University Grossman School of Medicine, New York; [b]Hartford Healthcare Heart and Vascular Institute, St. Vincent's Medical Center, Bridgeport, Connecticut; [c]Department of Medicine, University of Utah, Salt Lake City, Utah; [d]Veterans Affairs Medical Center, New York, New York; [e]Department of Population Health, New York University Grossman School of Medicine, New York; [f]Division of Hospital Medicine, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia; and [g]Department of Radiology, New York University Grossman School of Medicine, New York, New York. Manuscript received March 12, 2023; revised manuscript received and accepted June 29, 2023.

Drs. Zhang and Alter contributed equally to this manuscript.

Funding: none.

See page 96 for Declaration of Competing Interest.

*Corresponding author

E-mail address: robert.donnino@nyulangone.org (R. Donnino).

0002-9149/Published by Elsevier Inc.
https://doi.org/10.1016/j.amjcard.2023.06.098

either chest CT, TTE, or both. Chest CT was acceptable if performed with or without contrast and/or gated or non-gated to the cardiac cycle. Institutional review board approval was obtained with a waiver of informed consent for this Health Insurance Portability and Accountability Act-compliant retrospective cohort study.

The electronic medical record generated a dataset of all the patient encounters in which a patient received a TTE and a CT scan of the chest within the same 24-hour period. This dataset was then filtered to a subset where, for a given patient, a PEff was present in either the TTE or CT narrative report. Patients were excluded if they had any surgical/procedural intervention affecting the PEff (e.g., pericardiocentesis) between the CT and TTE studies. Patients were also excluded if they had any tumor filling the pericardial space, any surgical material (e.g., draining tubes) in the pericardial space, or if the assessment of PEff was not reported in either study. Based on the TTE or CT report description of the PEff size, a numerical value on a scale of 0 to 3 was assigned: "trace/none" = 0, "trace to small" = 0.5, "small" = 1, "small to moderate" = 1.5, "moderate" = 2, "moderate to large" = 2.5, and "large/very large" = 3. For example, if the TTE report described a PEff as "small" and the CT report described it as "small to moderate," we assigned a numerical value of "1" to that patient's TTE effusion size and a "1.5" to the patient's CT effusion size. A manual chart review of any missing data or ambiguous report findings was performed by a physician, and any reports that contained uncertainty regarding the size of effusion were eliminated. When an individual patient had undergone multiple imaging studies, the 2 performed closest in time were used. Individual patients could only be used once, even if they had multiple paired studies. Patients with a large difference in PEff size (i.e., a difference in grades of $\geq$2 on the ordinal scale) between CT and echo underwent manual chart review and review of both CT and echocardiographic images. As part of a prespecified analysis, we also examined the concordance between the 2 imaging methods after removing the trace/none effusion group.

Continuous variables were presented as mean $\pm$ SD and compared using two-sided *t* test. Categorical data were expressed as frequency and proportions and compared using chi-square test or Fisher's exact test, when appropriate. A weighted kappa coefficient and Bland-Altman plot was used to examine inter-test agreement. A weighted kappa using the Fleiss-Cohen weights was chosen to reflect the degree of agreement so that it attaches a greater emphasis to larger differences between ratings than to small differences. McNemar's test was performed to test if CT is more likely to report a higher grade than TTE. A two-sided $p < 0.05$ was considered to be statistically significant. All analyses were performed using STATA software (StataCorp, College Station, Texas).

Table 1
Baseline characteristics

Age (years)	66 $\pm$ 17
BMI (kg/m^2)	26.9 $\pm$ 6.6
Female, n (%)	607 (54%)
Mean time between TTE & CT (hours)	9.4 (IQR 3.5, 16.6)
TTE performed before CT, n (%)	355 (32%)
ECG gated, n (%)	3/1118 (0.26%)

Results

A total of 1,118 patients met inclusion/exclusion criteria and were included in the final study analyses. Mean age was 66 ($\pm$17 years), and 54% were female (Table 1). The mean body mass index was 26.9 $\pm$ 6.6 kg/m^2. The median time interval between the 2 imaging modality reports was 9.4 hours (interquartile range 3.5 to 16.6). TTE was performed before CT in 355 (32%) of the patients. Frequency of effusion size between both imaging methods is listed in Tables 2 and 3.

The percentage of patients that were within a half-grade or full-grade of agreement were 71.9% and 97.2%, respectively (Figure 1). The mean difference in grade (TTE minus

Table 2
Frequency of pericardial effusion size on TTE and CT

Grade	Pericardial Effusion Size	TTE, n (%)	CT, n (%)
0	No effusion	196 (17.5%)	444 (39.7%)
	Trace	536 (48%)	225 (20.1%)
0.5	Trace to small	3 (0.27%)	8 (0.7%)
1	Small	217 (19.4%)	233 (20.8%)
1.5	Small to moderate	62 (5.5%)	54 (4.8%)
2	Moderate	57 (5.1%)	93 (8.3%)
2.5	Moderate to large	14 (1.25%)	31 (2.8%)
3	Large	33 (3%)	30 (2.7%)

Table 3
Frequency of Pericardial Effusion Size on CT and TTE

TTE	CT							
	Trace/none	Trace to small	Small	Small to moderate	Moderate	Moderate to large	Large	Total
Trace/none	572	7	134	12	5	0	2	732
Trace to small	2	0	0	0	1	0	0	3
Small	90	1	73	18	31	3	1	217
Small to moderate	4	0	18	15	19	3	3	62
Moderate	1	0	8	7	24	11	6	57
Moderate to large	0	0	0	2	4	3	5	14
Large	0	0	0	0	9	11	13	33
Total	669	8	233	54	93	31	30	1118

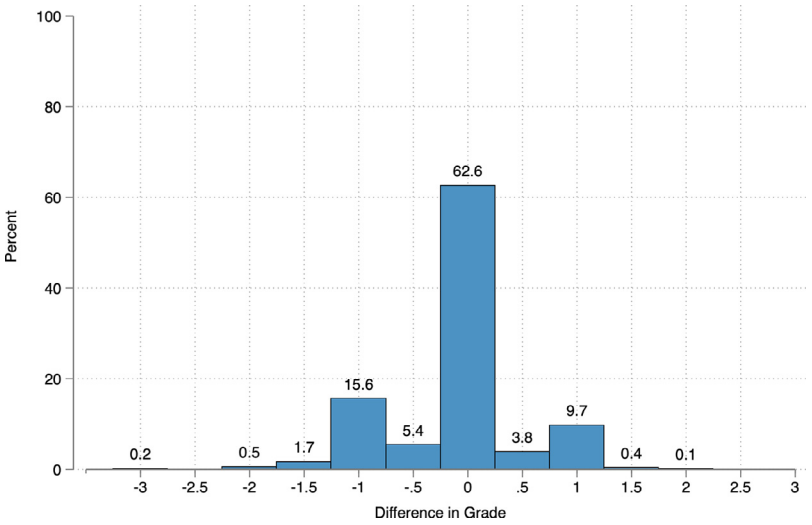

Figure 1. Difference in grade between CT and TTE is shown on the *x*-axis, and the frequency in which each difference in grade was observed between the 2 methods on the *y*-axis.

CT) between the 2 imaging methods was −0.1 (±0.6, p <0.0001). Bland-Altman comparisons between methods for determining PEff size are shown in Figure 2. The Bland-Altman plot demonstrated that 32 of 1,118 of studies (2.86%) fell outside the limits of agreement. Based on the Bland-Altman plot, there is no specific PEff size where TTE and CT are more likely to become discordant. In patients who had a grade difference that fell outside of the limits of agreement, CT chest tended to give the PEff a higher grade/large size. Similarly, in studies that were not in perfect agreement (n = 418, 38%), CT was more likely to report a higher grade (i.e., larger effusion) when compared with TTE (261 patients vs 157 patients, p <0.001). Pearson's correlation coefficient between the 2 imaging methods in assessment of PEff size was 0.73. The percent agreement between the 2 methods was 62.6% and the weighted kappa was 0.73 (95% confidence interval 0.69 to 0.76). This value suggests substantial agreement between the 2 imaging methods based on the Fleiss-Cohen weights.[4] TTE was more likely to report a trace effusion versus CT scan (48% vs 20.1%). Of the 536 patients reported as having a trace PEff on TTE, 378 (70.5%) were reported as having no effusion on chest CT.

In a secondary analysis, after excluding patients with trace or no effusion identified by both imaging methods, the percentage of patients that were within a half-grade or full-grade of agreement was 42.3% and 94.1% (Supplementary Figure 1), respectively, with the mean difference in grade (TTE minus CT) between the 2 imaging methods −0.2 (±0.83, p <0.0001). Frequency of effusion size between both imaging methods in this sensitivity analysis is shown in Supplementary Table 1.

There was a total of 18 patients who had a grade difference of 2 or more (e.g., small vs large) between the 2 imaging methods. All 18 patients underwent manual review of both study images. In patients with a grade difference of 2

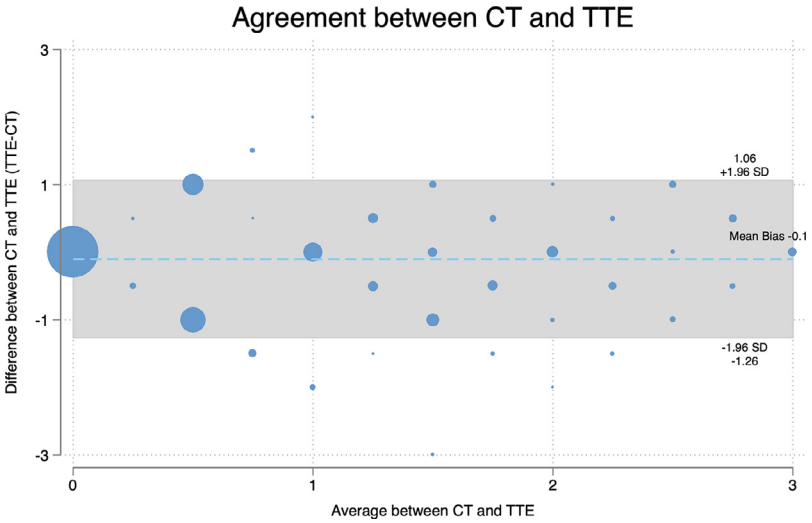

Figure 2. Bland-Altman Plot demonstrating the average agreement between CT and TTE on the *x*-axis and the difference between CT and TTE (TTE-CT). The dashed line represents mean bias. The top and the bottom of the box represent 95% levels of agreement (1.96 SD; 2.1 to −2.5). The size of the dot on the plot represents the percentage of patients. A larger dot represents a larger percentage of patients.

or more between the 2 imaging methods, CT reported larger sizes, with the mean grade for TTE and CT 0.53 and 1.5 respectively, with a mean difference of 1 (±1.76, p = 0.03). Nine of these patients had loculated effusions and 2 patients had large pleural effusions that made PEff difficult to assess on CT. Six of these patients had TTEs that were technically difficult studies with reports tending to underestimate the size of the effusion. One patient had a small PEff that was difficult to differentiate from pericardial fat.

Discussion

To the best of our knowledge, this is the largest reported comparison of CT and TTE for the evaluation of PEff size. Our primary finding is that there is a high degree of concordance in the assessment of PEff size between CT and TTE with 71.9% of the studies decreasing within a half-grade of each other and 97.2% within 1 full-grade. After removing patients with trace/none identified by both imaging methods, the concordance was weaker at 42.3% within a half-grade, but concordance within 1 grade was still high at 94.2%. It is important to note that our study compared clinical reporting of PEff size to reflect real-world practice, rather than a standardized, direct-image comparison.

Despite their widespread clinical use, there has been a paucity of previous data regarding the correlation between CT and TTE regarding PEff size. One previous single-center retrospective study examined the size of PEff on CT compared with TTE based on reports of 96 subjects.[2] Similar to our study, the size of a PEff was found to be larger when evaluated by CT than by TTE. In this study, however, agreement between the 2 methods was significantly worse with only 50% of the studies decreasing within a half-grade of each other. In addition to the relatively small sample size, an important limitation of this study was the inclusion of TTE and CT performed up to 14 days apart. PEff may show substantial growth (or regression) over that time period, limiting the generalizability of the conclusions. Our study only included TTE and CT studies completed within 24 hours (median time of 9.4 hours, interquartile range 3.5 to 16.6). Other small studies have also observed that PEff size tends to be reported as larger on CT compared with TTE, similar to our study.[3,5]

We found that trace PEffs were more frequently reported on TTE than CT. This may suggest that TTE is more sensitive for trace effusions, although because only image reports were analyzed, it may be because of, at least in part, how radiologists report (or choose not to comment on) such a small amount of effusion. Regardless, because PEffs are often presumed to be a physiologic/normal finding, it is unlikely that this difference would have a significant clinical impact.

It is important to note that our results do not suggest superiority of one modality over another for measuring PEff size, as we had no gold standard for comparison. In a previous study, Leibowitz et al[3] compared the quantification of PEffs by TTE and CT of the chest to that drained at pericardiocentesis in a retrospective study that included 19 patients. Calculated TTE volumes correlated relatively well (r^2 = 0.73) with the amount of fluid drained, whereas the correlation with CT was weaker (r^2 = 0.4). It should be noted, however, that the length of time permitted between studies was up to 13 days (mean 2 days). In addition, the amount of PEff drained during pericardiocentesis may not be a perfectly accurate gold standard, as the entire effusion may not be fully drained, especially in cases of loculated effusion.

Both CT and TTE have advantages and disadvantages in the evaluation of PEff, and these should be viewed as complementary imaging methods. TTE remains the primary imaging modality to assess PEff because of its noninvasive nature, lack of radiation, bedside availability, and, perhaps most importantly, its ability to evaluate physiology in addition to anatomy. Although TTE is typically the initial modality performed to evaluate suspected (or known) PEff, they are discovered incidentally by CT scans of the chest in up to 5% of studies.[6] The most significant disadvantage of CT is the inability to evaluate for physiologic consequences (e.g., cardiac tamponade). In some circumstances, CT may have advantages over TTE in evaluating PEff, such as when acoustic windows are limited on TTE and/or for loculated effusions in an anatomic location not well seen by TTE. In our manual review, the images in the 18 patients with large size discrepancies between the 2 methods, the discrepancies were attributed to large pleural effusions, technically difficult TTE studies, and/or a loculated PEff

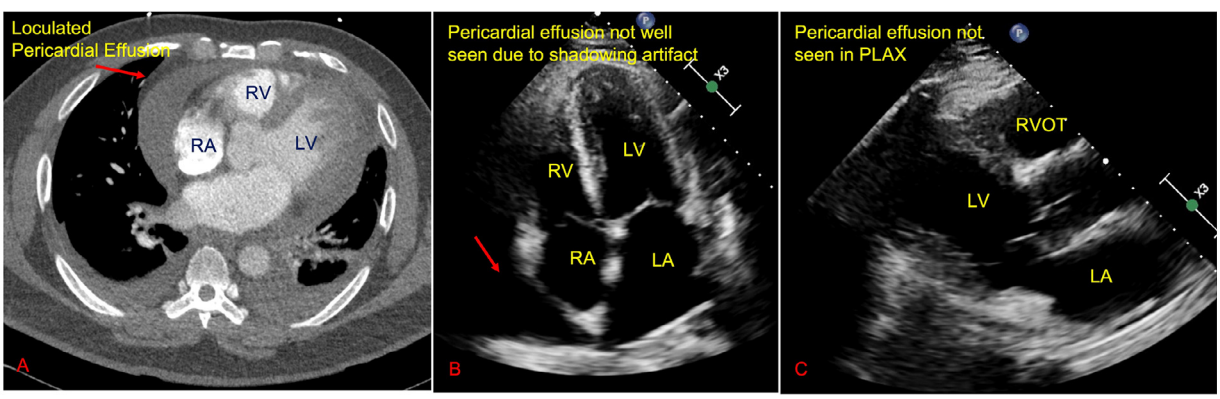

Figure 3. Loculated effusion seen on CT but not well seen on transthoracic echocardiogram apical 4 chamber view because of shadowing artifact near the right atrium. The effusion is not present in the parasternal long-axis view. LA = left atrial; LV = left ventricular; RA = right atrial; RV = right ventricular; RVOT = right ventricular outflow tract

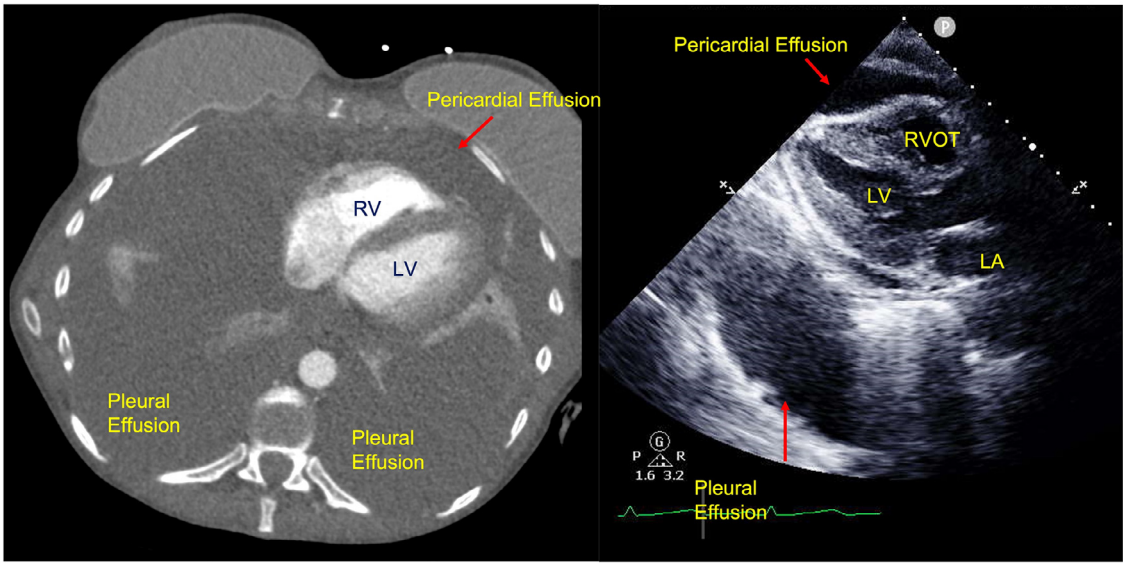

Figure 4. Demonstration of large pleural effusion obscuring the heart border making it difficult to assess pericardial effusion size. LA = left atrial; LV = left ventricular; RV = right ventricular; RVOT = right ventricular outflow tract

(Figures 3 and 4). Another advantage of CT is its ability to characterize the effusion as complicated or simple and differentiate effusion from clot and epicardial fat using the degree of attenuation (i.e., Hounsfield units).[1,7] Additionally, CT may also identify extracardiac findings that may provide information regarding the etiology of the effusion such as mediastinal lymphadenopathy or neighboring malignancies. Thus, especially for larger effusions, both TTE and CT are often important complementary methods for a comprehensive clinical evaluation.

Finally, our study represents a real-world representation of how TTE and CT chest compare with each other regarding PEff size. Although there are quantitative methods to measure PEff size, the assessment of PEff size on TTE and CT is often done by qualitative visual assessment, which inevitably leads to variability between readers. Despite this, our study suggests that the correlation overall is strong. Clinicians should note, however, that in certain settings it may be prudent to employ more than a single imaging modality, such as in the presence of large pleural effusions, unevenly distributed (i.e., loculated) PEff and limited image quality (e.g., because of poor acoustic windows).

Our study has several limitations. First, as a single-center study, generalizability to other sites may be reduced. Second, the retrospective nature of the study raises the potential for selection bias. For example, because we examined PEffs that had reports of both imaging methods within 24 hours without any procedures in between the 2 studies, it is possible that patients with larger effusions were less likely to obtain a second imaging modality (e.g., if TTE was performed first) before pericardiocentesis if there is clinical instability with evidence of cardiac tamponade. This may have led to a lower representation of large PEffs in our study. Our study does not include clinical outcome information (e.g., patients with tamponade or those who underwent pericardiocentesis) which may have provided additional insight. To investigate the "real-world" interpretation of these imaging methods, our study is based upon review of study reports and not review of the images themselves. This may introduce significant interpretation bias, however, as interpreting physicians were not blinded to the other imaging methods, nor to clinical data. Finally, we were not able to identify whether one imaging test was linked to the other, which could provide important clinical implications and insights into healthcare utilization.

In conclusions, in our large dataset, clinically reported TTE and CT completed within 24 hours of each other showed relatively strong agreement in estimation of PEff size between both methods, with size estimated by CT generally larger than TTE, on average. In the relatively small number of patients with large discrepancies in size between TTE and CT, several factors played a role, including reduced image quality (i.e., technically difficult studies), the presence of large pleural effusions, and loculated PEff. Both TTE and CT are important and complementary imaging methods in the evaluation of PEff.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.amjcard.2023.06.098.

1. Klein AL, Abbara S, Agler DA, Appleton CP, Asher CR, Hoit B, Hung J, Garcia MJ, Kronzon I, Oh JK, Rodriguez ER, Schaff HV, Schoenhagen P, Tan CD, White RD. American Society of Echocardiography clinical recommendations for multimodality cardiovascular imaging of patients with pericardial disease: endorsed by the Society for Cardiovascular Magnetic Resonance and Society of Cardiovascular Computed Tomography. *J Am Soc Echocardiogr* 2013;26:965–1012. e15.
2. Moustafa A, Liu X, Ye F, Khuder S, Alhazmi L, Youssef E, Alim H, Alsamman MA, Khan MS, Eltahawy E. Computed tomography scan

overestimates the size of pericardial effusion compared to echocardiography. *Cureus* 2019;11:e5182.
3. Leibowitz D, Perlman G, Planer D, Gilon D, Berman P, Bogot N. Quantification of pericardial effusions by echocardiography and computed tomography. *Am J Cardiol* 2011;107:331–335.
4. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–282.
5. Katsanos S, van Rosendael P, Kamperidis V, van der Kley F, Regeer M, Al-Amri I, Karalis I, Palmen M, de Weger A, Marsan NA, Bax JJ, Delgado V. Pericardial effusion following transcatheter aortic valve implantation: echocardiography and multi-detector row computed tomography evaluation. *Int J Cardiovasc Imaging* 2015;31:37–43.
6. Hall WB, Truitt SG, Scheunemann LP, Shah SA, Rivera MP, Parker LA, Carson SS. The prevalence of clinically relevant incidental findings on chest computed tomographic angiograms ordered to diagnose pulmonary embolism. *Arch Intern Med* 2009;169:1961–1965.
7. Verhaert D, Gabriel RS, Johnston D, Lytle BW, Desai MY, Klein AL. The role of multimodality imaging in the management of pericardial disease. *Circ Cardiovasc Imaging* 2010;3:333–343.