Clinical Research in Cardiology (2023) 112:68–74
https://doi.org/10.1007/s00392-022-02031-0

ORIGINAL PAPER

Reliability of estimating left ventricular ejection fraction in clinical routine: a validation study of the SWEDEHEART registry

Joel Lenell[1] · Bertil Lindahl[1] · Per Karlsson[2] · Gorav Batra[1] · David Erlinge[3] · Tomas Jernberg[4] · Jonas Spaak[4] · Tomasz Baron[1]

Received: 16 September 2021 / Accepted: 28 April 2022 / Published online: 17 May 2022
© The Author(s) 2022

Abstract

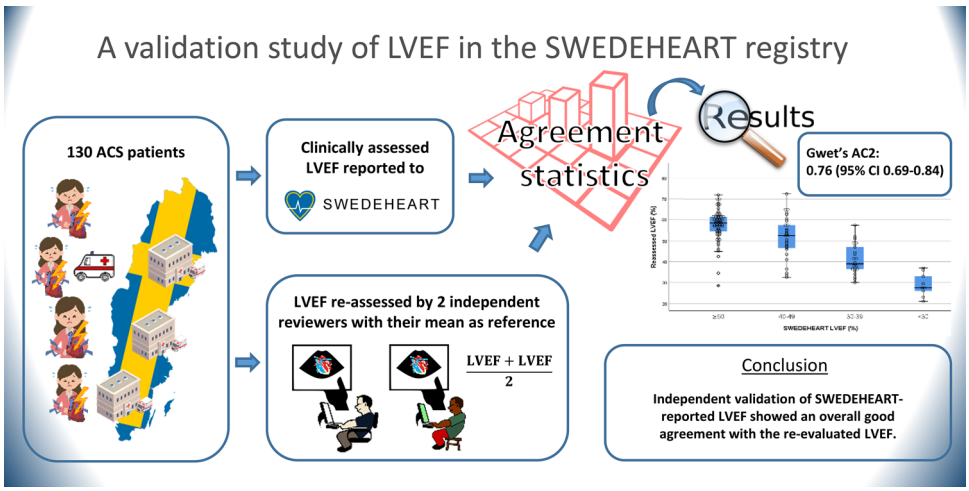
Objective Patients hospitalized with acute coronary syndrome (ACS) in Sweden routinely undergo an echocardiographic examination with assessment of left ventricular ejection fraction (LVEF). LVEF is a measurement widely used for outcome prediction and treatment guidance. The obtained LVEF is categorized as normal (> 50%) or mildly, moderately, or severely impaired (40–49, 30–39, and < 30%, respectively) and reported to the nationwide registry for ACS (SWEDEHEART). The purpose of this study was to determine the reliability of the reported LVEF values by validating them against an independent re-evaluation of LVEF.

Methods A random sample of 130 patients from three hospitals were included. LVEF re-evaluation was performed by two independent reviewers using the modified biplane Simpson method and their mean LVEF was compared to the LVEF reported to SWEDEHEART. Agreement between reported and re-evaluated LVEF was assessed using Gwet's AC2 statistics.

Results Analysis showed good agreement between reported and re-evaluated LVEF (AC2: 0.76 [95% CI 0.69–0.84]). The LVEF re-evaluations were in agreement with the registry reported LVEF categorization in 86 (66.0%) of the cases. In 33 (25.4%) of the cases the SWEDEHEART-reported LVEF was lower than re-evaluated LVEF. The opposite relation was found in 11 (8.5%) of the cases ($p < 0.005$).

Conclusion Independent validation of SWEDEHEART-reported LVEF shows an overall good agreement with the re-evaluated LVEF. However, a tendency towards underestimation of LVEF was observed, with the largest discrepancy between re-evaluated LVEF and registry LVEF in subjects with subnormal LV-function in whom the reported assessment of LVEF should be interpreted more cautiously.

Graphical abstract



Extended author information available on the last page of the article

Keywords LVEF · Registry · Validation · SWEDEHEART · Echocardiography

Background

Left ventricular ejection fraction (LVEF) is one of the most robust predictive parameters post-myocardial infarction offering guidance in both treatment and follow-up strategies [1]. However, despite its recognized value, LVEF by biplane Simpson is subject to a non-negligible inter-observer variability that increases with declining image quality [2–4]. Poor image quality may prevent reliable delineation of the endocardial border and hinder a quantitative assessment of LVEF [5]. In such cases, LVEF may be assessed using visual "eye-balling". This technique implicates an even greater uncertainty with a reported inter-observer variability of up to ± 14% LVEF [6].

The SWEDEHEART registry contains patient data from subjects admitted to Swedish hospitals nationwide due to acute coronary syndrome (ACS). The registry was established in 2009 by merging of four preexisting registries on ischemic heart disease into a larger, more comprehensive registry. The registry encompasses more than 450 variables including left ventricular ejection fraction (LVEF) by echocardiography [7]. SWEDEHEART has been a valuable source of insight in different aspects of ischemic heart disease and data from SWEDEHEART, including LVEF, has been used in a large number of studies [8].

Given the importance of LVEF and that the LVEF assessments in SWEDEHEART, and in several other registries, have not previously been validated we aimed to validate the LVEF assessments reported to SWEDEHEART.

Methods

The study cohort consisted of a random sample of 177 patients with ACS from three different Swedish hospitals [Uppsala University Hospital, Uppsala (site 1), Lund University Hospital, Lund (site 2) and Danderyd University Hospital, Stockholm (site 3)], in whom LVEF had been assessed according to local routine by 2D echocardiography during the index hospital stay. LVEF was then reported to SWEDEHEART in four categories: < 30%, 30–39%, 40–49%, and ≥ 50%. Missing data on LVEF in the registry prompted exclusion from the study (Fig. 1).

The echocardiographic raw data was collected from the imaging databases at the participating hospitals and LVEF was re-evaluated at the core lab in Uppsala by two independent reviewers according to the modified biplane Simpson method using TOMTEC-ARENA TTA2 software (TOMTEC IMAGING SYSTEMS, GMBH EDISONSTRASSE, Unterschleissheim DE) with manual delineation of the endocardial border [9]. Subjects with more than two untraceable myocardial segments due to insufficient image quality were excluded. The reviewers were experienced echocardiographers blinded to all patient related clinical data. The mean LVEF value of the two reassessments was

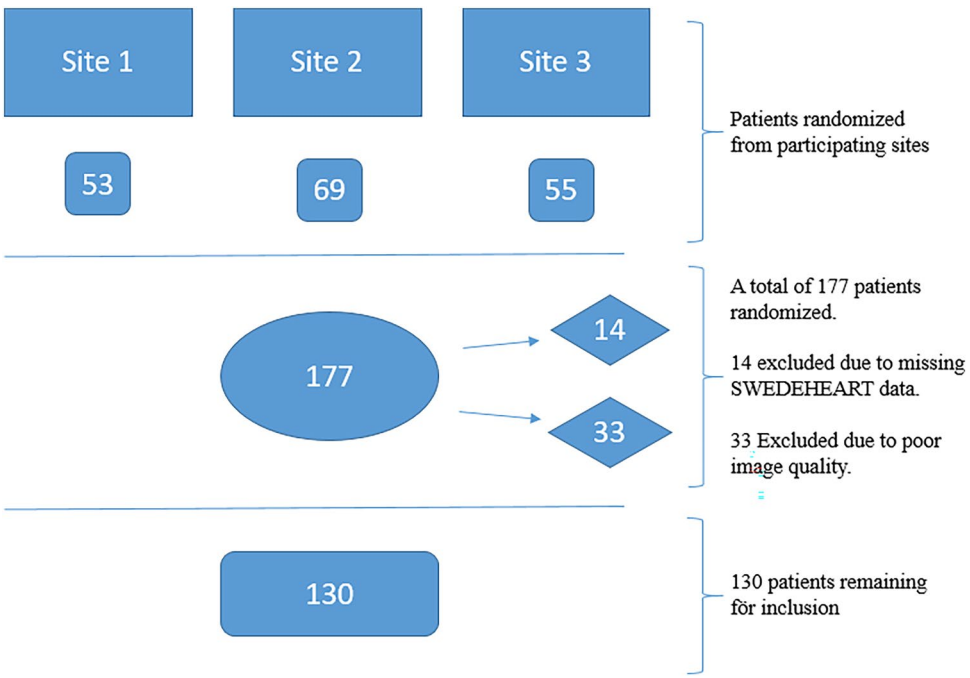


Fig. 1 Flow chart illustrating patient selection

used as reference when compared to SWEDEHEART data. Before comparison, the reassessed reference values were categorized from continuous into ordinal scale variables in accordance with the LVEF ranges found in SWEDEHEART.

Statistical analysis

Spearman's rank correlation was used to test correlation and Wilcoxon signed-rank test was used to assess bias. Gwet's AC1 and AC2 statistics were used to test agreement between SWEDEHEART data and the reassessments. The Gwet's AC2 analysis was performed with pre-specified linear weights. The Intraclass correlation coefficient (ICC) was used for inter-observer variability assessment between the two reference reviewers by a two-way mixed-effects model examining consistency in "single rater" type [10, 11]. The analysis was performed using SPSS software 26.0, SPSS Inc., Chicago, IL, USA and R (4.0.2) with statistical significance defined by $p < 0.05$. The results were presented in a tabular format with 95% confidence intervals. The study was approved by the Regional Committee for Medical Research Ethics (DNR 2017/759-31).

Results

The median age in the study population was 65 years, and 76% were men. See Table 1 for further baseline characteristics.

After LVEF was re-evaluated by the two reviewers with the modified biplane Simpson method, the mean LVEF was calculated for comparison with the SWEDEHEART data. The inter-observer variability between the two reference reviewers re-evaluating LVEF by biplane Simpson was close to excellent [ICC 0.87 (95% CI 0.82–0.92)]. Likewise, there was an excellent correlation with Spearman's *R* of 0.88 ($p < 0.001$). Figure 2a displays a scatterplot of the reassessments by the two reviewers and Fig. 2b illustrates their agreement in a Bland–Altman plot.

As displayed in Table 2 there is an overall good correlation between reassessed LVEF and SWEDEHEART LVEF (Spearman's $r = 0.69$, $p < 0.01$). Agreement as assessed by Gwet's AC1 was moderate [0.58 (95% CI 0.47–0.69)], however, when adjusted for misclassification errors (AC2), the agreement between SWEDEHEART LVEF and the gold standard was good [0.76 (95% CI 0.69–0.84)]. There was a trend towards lower LVEF values in the registry when compared to the reassessments. This bias was further explored in Table 3 by the Wilcoxon signed-rank test. The analysis showed a lower category LVEF in 25.4% ($p < 0.001$) of cases in SWEDEHEART and an absolute agreement in 66% of cases. Figure 3 displays a particularly wide distribution of reassessed LVEF among patients registered in the LVEF range between 40 and 49%, with a median above the 50% cut-off limit. There appears to be a slight bias towards registering a lower category LVEF in subjects with borderline function.

Table 1 Patient demographics and medical history

Characteristics	$n = 130$
Age, years median (IQR)	65 (58–72)
Male sex (%)	99 (76)
Active smoker, (%)	36 (28)
Diagnosis of ACS	
STEMI (%)	65 (50)
NSTEMI (%)	62 (48)
Other (UA, type 2 MI, Takotsubo) (%)	3 (2)
Medical history	
Hypertension (%)	72 (55)
Diabetes mellitus (%)	27 (21)
Atrial fibrillation (%)	6 (5)
Heart failure (%)	22 (17)
History of stroke (%)	10 (8)
History of MI (%)	29 (22)

IQR Inter quartile range, *STEMI* ST-elevation myocardial infarction, *NSTEMI* non-ST-elevation myocardial infarction, *BMI* Body mass index, *UA* Unstable angina pectoris

Discussion

Independent validation of LVEF in the SWEDEHEART registry shows an overall good agreement when re-evaluated by two independent reviewers using the modified biplane Simpson method. However, there was a tendency towards underestimation of LVEF in SWEDEHEART, with the largest discrepancy between re-evaluated LVEF and SWEDEHEART LVEF in subjects with LVEF < 50%. Only five patients had a greater than one category difference.

To our knowledge this is the first validation study of LVEF assessed by echocardiography in a quality registry by reassessment of the raw data. Consequently, it may be difficult to put these results in relation to previous registry validations. In 2016, Govatsmark et al. published a validation of the Norwegian myocardial infarction register with medical records as reference, without reinterpretation of the echocardiography raw data [12]. This approach, with medical records as gold standard, is common in registry validation studies, however, it fails to address potential flaws in index data acquisition [13, 14]. Publications on inter-observer variability in LVEF may be more feasible as a benchmark since they more closely reflect this SWEDEHEART validation study design. Such studies, covering different clinical settings, have presented good to excellent inter-observer

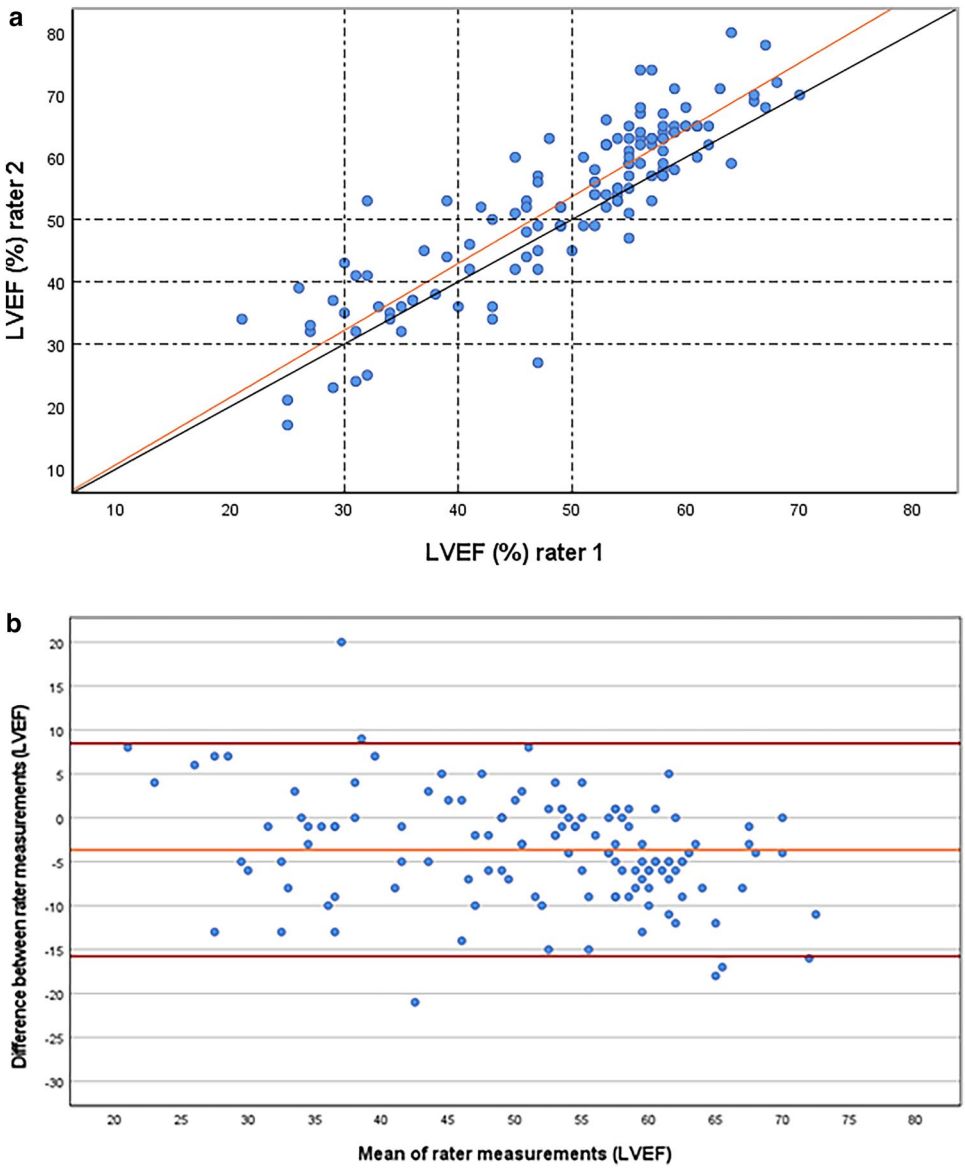


Fig. 2 **a** A scatter plot illustrating inter-observer variability between the two reviewers. The dotted lines mark the cut-off values defining LVEF categories in SWEDEHEART. **b** Bland–Altman plot illustrating the difference between the reviewers' measurements to their mean. Orange line = mean difference, Red lines = mean difference ± 1.96* standard deviation of the difference

Table 2 Crosstab displaying the overlap and distribution of the categorized re-evaluated LVEF (gold standard) and the LVEF registered in SWEDEHEART

SWEDEHEART	Reassessed LVEF				
	LVEF ≥ 50%	LVEF 40–49%	LVEF 30–39%	LVEF < 30%	Total
LVEF ≥ 50%	62	5	1	1	69
LVEF 40–49%	18	8	4	0	30
LVEF 30–39%	3	8	11	0	22
LVEF < 30%	0	0	4	5	9
Total	83	21	20	6	130

agreement of LVEF in populations with preserved ejection fraction, yet only moderate agreement in subjects with subnormal ejection fraction [15–17]. The findings from the SWEDEHEART registry are thus in accordance with previous results.

There were 47 subjects (36%) with LVEF < 50% in this study sample collected between 2008 and 2014. As of 2014, 35% of all the subjects registered to SWEDEHEART had an LVEF < 50%, indicating that the study sample was representative of the overall population [18].

Table 3 A Wilcoxon signed-rank test of bias between re-evaluated LVEF and LVEF estimates in the SWEDEHEART registry

Wilcoxon signed-rank test	*n* (%)
Reassessed LVEF > SWEDEHEART	33 (25.4)
Reassessed LVEF < SWEDEHEART	11 (8.6%)
Ties	86 (66.0)
Total	130

Assessment of LVEF by visual eye-balling was, at the time of enrollment, the most common method of LVEF assessment at all three participating centers. This approach has repeatedly shown to underestimate LVEF compared to quantitative assessment by the modified Simpson method [19–21]. Visually assessed LVEF has substantial inter-observer variability of up to ± 14% LVEF [6]. In such settings, grouping the values into larger categories might be reasonable. However, current guidelines on cardiac chamber quantification favor the more robust modified biplane Simpson approach that generates a continuous variable [22]. In such quantitative setting, it may be more reasonable to report the obtained LVEF value to avoid the obvious loss of granularity in categorized data. Categorization may also increase the perceived difference between two assessments that end up on different sides of a cut-off limit, despite being close in numeric value. The current categorization also fails to harmonize with important clinical cut-off limits such as the 35% LVEF limit in deciding on ICD and CRT implantation [23]. Algorithms enabling an automatic tracing of the endocardial border, generating an instant assessment of LVEF, are now common and validated tools in the echocardiography lab [24]. These techniques further lower the barriers to a quantitative assessment of a continuous LVEF variable. However, should quantitatively assessed LVEF come to be reported as a continuous variable it must still be interpreted with respect to a smaller, yet significant, inter-observer variation and a smallest detectable change of almost 7 % points [3, 25].

Kappa statistics has repeatedly been used in inter-observer variability studies on LVEF due to its property of adjusting for chance agreement [26]. However, Cohen's kappa is prone to underestimating agreement in populations with symmetrical imbalance [26] [27]. As the majority (63%) of the studied SWEDEHEART subjects presented with a preserved ejection fraction (LVEF > 50%), Cohen's kappa was rendered unfeasible in this setting. Instead, the weighed Gwet's AC2 method offered a more robust analysis, yet despite its advantages, it appears to be rarely performed in publications on assessments of LVEF agreement [12, 28, 29].

There was an agreement on categorized LVEF in roughly 50% of the cases with reduced LVEF (as determined by the reference method). This discrepancy is largest in the mid-range categories of 30–49% where important clinical cut-off limits are present. An improved precision of LVEF assessments in this subgroup would be most favorable. A Japanese study examined two different teaching interventions with regard to improved inter-rater variability in visually assessed LVEF [30]. The interventions significantly reduced the misclassification rates of LVEF regardless of operator experience and image quality. There are several other studies in support of similar teaching interventions [6, 31]. Such initiatives could possibly further improve the quality of LVEF assessments entered in SWEDEHEART.

Limitations

There is a theoretical risk that some LVEF assessments have been misclassified into the wrong category in SWEDEHEART since this process is performed manually. As we

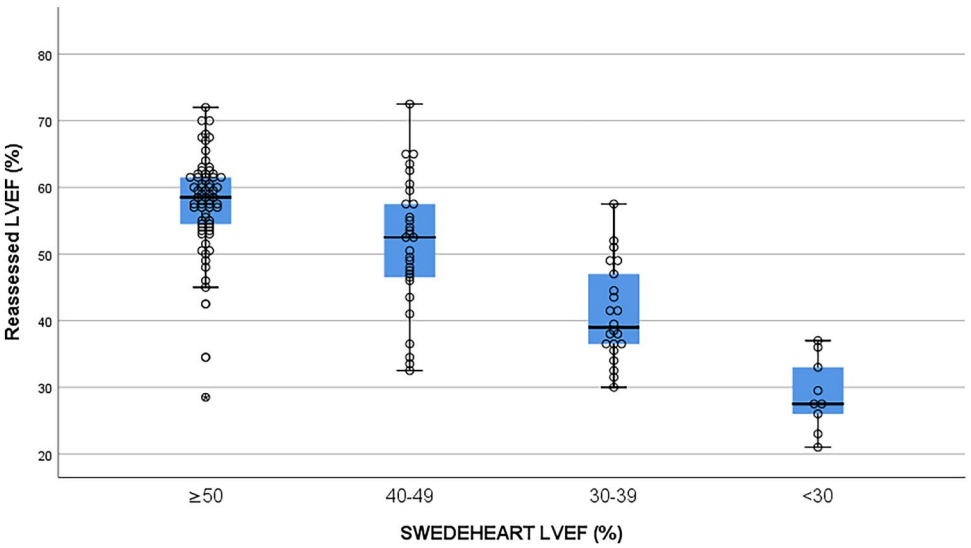


Fig. 3 Combined boxplot and dotplot with distribution of the reassessed LVEF in LVEF categories according to SWEDEHEART. Middle line = median; Box = interquartile range (IQR); Whiskers = lowest and maximum LVEF excluding outliers; Circle outside whiskers = mild outliers; asterisk in circle = severe outliers

have not validated the registry LVEF in relation to patient records such faults cannot be accounted for in this study which may be considered a limitation. The extreme outlier in Fig. 3 may be an example of such accidental misclassification. LVEF lacks a gold standard and, as a consequence, the proper reference method in LVEF validation studies is a subject open for debate. Given that alternative imaging modalities, such as cardiac magnetic resonance imaging (CMR) or cardiac computed tomography (CCT), have not been routinely performed in Swedish ACS subjects we deemed it appropriate that the reference in this study should be based on the 2D echocardiographic raw data. The use of a mean from two independent reassessments by the modified Simpson rule was proposed to strengthen the robustness in the reference method, however, there is scarce literature in support of this notion.

The patients were included at three university hospitals which may give rise to questions of sample representation since SWEDEHEART contains data from all Swedish hospitals treating ACS. A larger nationwide follow-up with a greater variation in hospital size may provide additional information on the data validity in SWEDEHEART. However, as previously mentioned, the representability of the material is supported by the concurring proportion of subjects categorized with reduced ejection fraction in this study and in the SWEDEHEART registry [18].

Conclusion

Independent validation of SWEDEHEART-reported LVEF shows an overall good agreement with the re-evaluated LVEF. However, a tendency towards underestimation of LVEF was observed, with the largest discrepancy between re-evaluated LVEF and registry LVEF found in subjects with subnormal LV-function in whom the reported assessment of LVEF should be interpreted more cautiously.

Funding Open access funding provided by Uppsala University. No funding was received for conducting this study.

Availability of data and material The data may be provided upon request.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical standards This study has been approved by the regional ethics committee and has, therefore, been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Consent to participate All study subjects gave their informed consent prior to their inclusion in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Collet J-P et al (2021) 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: the task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 42(14):1289–1367
2. Thavendiranathan P et al (2013) Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes. J Am Coll Cardiol 61(1):77–84
3. Baron T et al (2019) Test–retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. Clin Res Cardiol 108(4):355–365
4. Cole GD et al (2015) Defining the real-world reproducibility of visual grading of left ventricular function and visual estimation of left ventricular ejection fraction: impact of image quality, experience and accreditation. Int J Cardiovasc Imaging 31(7):1303–1314
5. McGowan JH, Cleland JGF (2003) Reliability of reporting left ventricular systolic function by echocardiography: a systematic review of 3 methods. Am Heart J 146(3):388–397
6. Johri AM et al (2011) Can a teaching intervention reduce interobserver variability in LVEF assessment: a quality control exercise in the echocardiography lab. JACC: Cardiovasc Imaging 4(8):821–829
7. Desta L et al (2016) Heart failure with normal ejection fraction is uncommon in acute myocardial infarction settings but associated with poor outcomes: a study of 91,360 patients admitted with index myocardial infarction between 1998 and 2010. Eur J Heart Fail 18(1):46–53
8. Hambraeus K et al (2014) SWEDEHEART annual report 2012. Scand Cardiovasc J 48:1
9. Foley T et al (2012) Measuring left ventricular ejection fraction—techniques and potential pitfalls. Eur Cardiol Rev 8:108
10. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155–163
11. Bunting KV et al (2019) A practical guide to assess the reproducibility of echocardiographic measurements. J Am Soc Echocardiogr 32(12):1505–1515
12. Govatsmark RE et al (2016) Interrater reliability of a national acute myocardial infarction register. Clin Epidemiol 8:305–312

13. Vikholm P et al (2018) Validity of the Swedish cardiac surgery registry. Interact Cardiovasc Thorac Surg 27(1):67–74
14. Ljung R et al (2019) The Swedish dental health register—validation study of remaining and intact teeth. BMC Oral Health 19(1):116
15. Gadsbøll N et al (1989) Interobserver agreement and accuracy of bedside estimation of right and left ventricular ejection fraction in acute myocardial infarction. Am J Cardiol 63(18):1301–1307
16. De Geer L, Oscarsson A, Engvall J (2015) Variability in echocardiographic measurements of left ventricular function in septic shock patients. Cardiovasc Ultrasound 13(1):19
17. Frikha Z et al (2015) Reproducibility in echocardiographic assessment of diastolic function in a population based study (The STANISLAS Cohort Study). PLoS ONE 10:e0122336
18. Szummer K et al (2018) Relations between implementation of new treatments and improved outcomes in patients with non-ST-elevation myocardial infarction during the last 20 years: experiences from SWEDEHEART registry 1995 to 2014. Eur Heart J 39(42):3766–3776
19. Sievers B et al (2005) Visual estimation versus quantitative assessment of left ventricular ejection fraction: a comparison by cardiovascular magnetic resonance imaging. Am Heart J 150(4):737–742
20. Rana S et al (2020) Left ventricle ejection fraction estimation by point of care echocardiography in patients admitted in intensive care unit. J Chitwan Med Coll 10:54–57
21. Jakobsen CJ, Torp P, Sloth E (2006) Assessment of left ventricular ejection fraction may invalidate the reliability of EuroSCORE. Eur J Cardiothorac Surg 29(6):978–982
22. Lang RM et al (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American society of echocardiography and the European association of cardiovascular imaging. Eur Heart J Cardiovasc Imaging 16(3):233–271
23. Priori SG et al (2015) 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 36(41):2793–2867
24. Knackstedt C et al (2015) Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: The FAST-EFs Multicenter Study. J Am Coll Cardiol 66(13):1456–1466
25. Thorstensen A et al (2010) Reproducibility in echocardiographic assessment of the left ventricular global and regional function, the HUNT study. Eur J Echocardiogr 11(2):149–156
26. Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. Fam Med 37(5):360–363
27. Feinstein AR, Cicchetti DV (1990) High agreement but low Kappa: i the problems of two paradoxes. J Clin Epidemiol 43(6):543–549
28. Wongpakaran N et al (2013) A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 13(1):61
29. Walsh P et al (2014) Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. PeerJ 2:e651
30. Kusunose K et al (2018) Reduced variability of visual left ventricular ejection fraction assessment with reference images: The Japanese Association of Young Echocardiography Fellows multicenter study. J Cardiol 72(1):74–80
31. Anilkumar S et al (2019) A teaching intervention increases the performance of handheld ultrasound devices for assessment of left ventricular ejection fraction. Heart Views 20(4):133–138

Authors and Affiliations

Joel Lenell[1] · Bertil Lindahl[1] · Per Karlsson[2] · Gorav Batra[1] · David Erlinge[3] · Tomas Jernberg[4] · Jonas Spaak[4] · Tomasz Baron[1]

✉ Joel Lenell
joel.lenell@medsci.uu.se

[1] Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden

[2] Department of Cardiology and Clinical Physiology, Uppsala University Hospital, Uppsala, Sweden

[3] Department of Clinical Sciences, Cardiology, Lund University, Lund, Sweden

[4] Division of Cardiovascular Medicine, Department of Clinical Sciences, Karolinska Institute, Danderyd Hospital, Stockholm, Sweden