UNIVERSITY OF WEST LONDON
The Career University

UWL REPOSITORY

repository.uwl.ac.uk

Origin of apparently-improved measurement error with artificial intelligence in Echocardiography

Almeida, G, Chawla, S, Ng, T S J, Stowell, C, Jevsikov, J, Manisty, C, Zolgharni, Massoud ORCID logoORCID: https://orcid.org/0000-0003-0904-2904, Shun-Shin, M and Francis, D P (2026) Origin of apparently-improved measurement error with artificial intelligence in Echocardiography. In: ESC Congress 2026, 28–31 August 2026, Munich, Germany.

https://doi.org/10.1093/eurheartj%2Fehag785.006

This is the Published Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/15734/

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

Imaging – Echocardiography, Technology

Origin of apparently-improved measurement error with artificial intelligence in Echocardiography

G. Almeida[1], S. Chawla[1], T.S.J. Ng[1], C. Stowell[1], J. Jevsikov[1], C. Manisty[2], M. Zolgharni[3], M. Shun-Shin[1], D.P. Francis[1]

[1]Imperial College London, National Heart and Lung Institute, London, United Kingdom of Great Britain & Northern Ireland
[2]University College London, London, United Kingdom of Great Britain & Northern Ireland
[3]University of West London, London, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: Type of funding sources: Foundation. Main funding source(s): British Heart Foundation

Downloaded from academic.oup.com/eurheartj/article/47/Supplement_1/ehag785.006/8997431 by guest on 31 October 2026

Background: Automated ("Artificial Intelligence", AI) methods seem to achieve better performance in matching human expert measurements of clinical variables, than experts themselves do. We explored the mechanism for this in a unique 6-observer, 100-patient, regulatory dataset. We examined the 4 contributors to ejection fraction and then 55 other directly measured variables.

Purpose: To determine why AI measurements in echocardiography outperform human expert measurements.

Method: Six experts each directly measured 59 variables in each of 100 echocardiograms. Each independently chose the video, beat, and frame, and then traced the distance or boundary curve. One AI system assessed every beat of every video. We quantified the measurement variance (expert-to-expert, and AI-to-expert) for all 59 variables. We then stratified these expert-expert or AI-expert pairings into whether the 2 measurements happened to be on:

- The same frame of the same video,
- A different frame of the same video, or
- A different video.

The variance in each stratum could then be expressed as a fraction of the variance overall.

To enable comparisons between different variables, we divided the differences by the population standard deviation of that variable. By squaring, we obtained measurement variance as a proportion of population variance.

Results: There were 5,963 pairings of 2 experts measuring the same patient's A4c LVEDV, of which 426 were same-frame, 2,665 were other-frame, and 2,862 were different-video. The disagreement variance in the same-frame pairings was only 18% of the overall disagreement variance ($p<0.001$ by F test). All 4 apical volumes showed this pattern: LVEDV and LVESV, in A4c and A2c views (Figure 1 left panels, $p<0.01$ by F-test). The other 55 directly measurable variables showed a similar pattern: the tracing variability on an identical frame being much smaller than the overall disagreement variability.

Strikingly, the AI-expert differences showed the same pattern (Figure 1 right).

Across all 59 variables both the expert-expert and AI-expert comparisons showed disagreement variance to be dominated by the accident of having chosen a different frame or different video. The results are shown in Figure 2, with same-frame variance and different-frame variance both being significantly smaller than the overall variance ($p<0.001$ and $p<0.001$ for expert-expert; $p<0.001$ and $p<0.001$ for AI-expert).

Conclusions: The true origin of reduced measurement error from AI measurement is not better accuracy in individual tracing but its comprehensive analysis of all available data.

Individual AI measurements suffer identically to expert measurements from biological variability between different beats (and different acquisitions). However, AI measurements routinely measure every beat and synthesise a summary. Although this has long been recognised as the gold standard for measurement accuracy (1), it has never been practical until the era of AI automation.

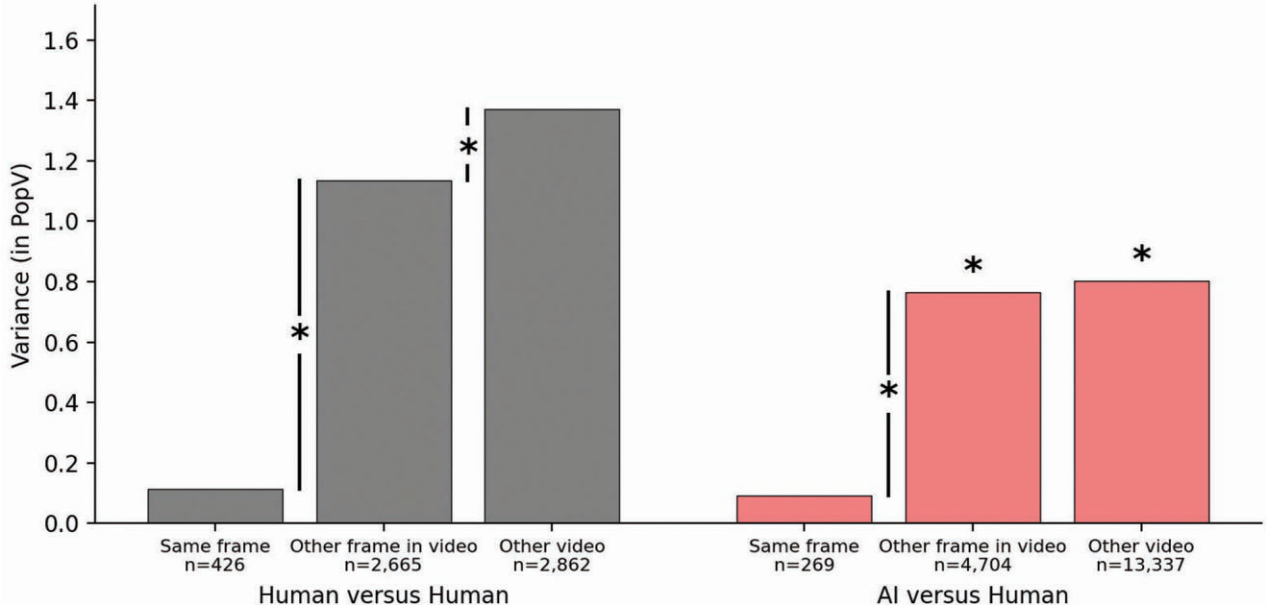


Figure 1

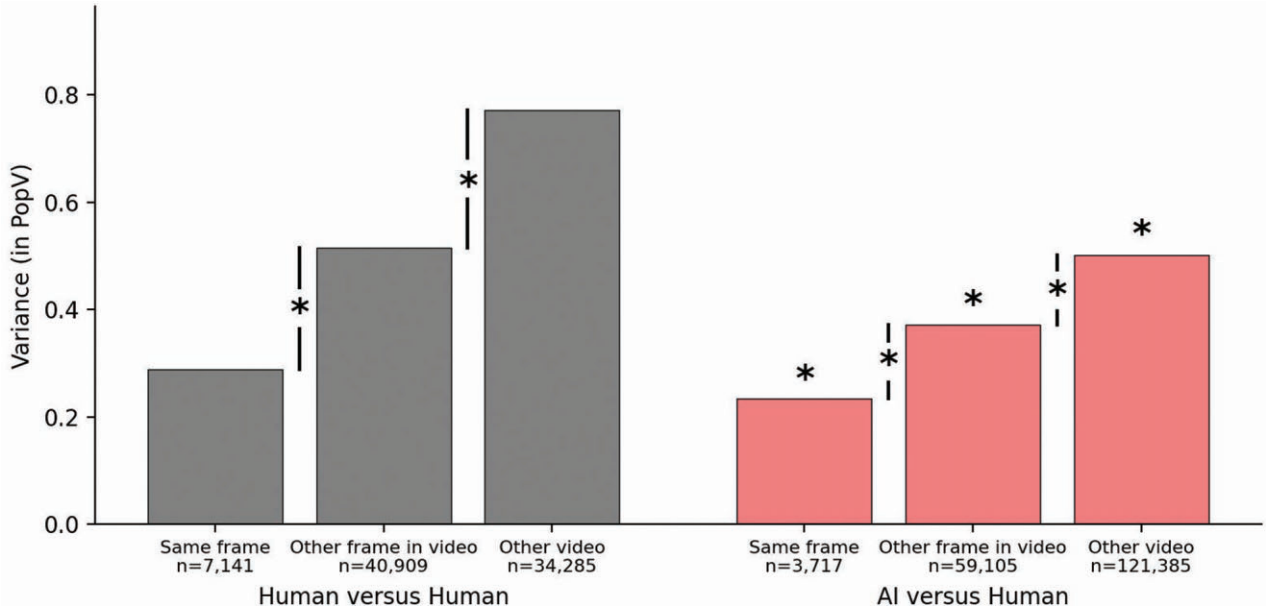


Figure 2

Downloaded from academic.oup.com/eurheartj/article/47/Supplement_1/ehag785.006/8997431 by guest on 31 October 2026