

# openheart Enhanced detection of severe aortic stenosis via artificial intelligence: a clinical cohort study

Geoff Strange <sup>1,2</sup>, Simon Stewart <sup>3,4</sup>, Andrew Watts <sup>5</sup>, David Playford <sup>6</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/openhrt-2023-002265>).

**To cite:** Strange G, Stewart S, Watts A, *et al.* Enhanced detection of severe aortic stenosis via artificial intelligence: a clinical cohort study. *Open Heart* 2023;**10**:e002265. doi:10.1136/openhrt-2023-002265

Received 18 January 2023  
Accepted 30 June 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Cardiology, Heart Research Institute Ltd, Newtown, New South Wales, Australia

<sup>2</sup>The University of Notre Dame Australia, School of Medicine, Fremantle, Western Australia, Australia

<sup>3</sup>Institute for Health Research, The University of Notre Dame Australia, Fremantle, Western Australia, Australia

<sup>4</sup>School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow, UK

<sup>5</sup>Echo IQ Pty Ltd, Sydney, New South Wales, Australia

<sup>6</sup>School of Medicine, The University of Notre Dame Australia, Fremantle, Western Australia, Australia

**Correspondence to**  
Dr Geoff Strange; [gstrange@neda.net.au](mailto:gstrange@neda.net.au)

## ABSTRACT

**Objective** We developed an artificial intelligence decision support algorithm (AI-DSA) that uses routine echocardiographic measurements to identify severe aortic stenosis (AS) phenotypes associated with high mortality.

**Methods** 631 824 individuals with 1.08 million echocardiograms were randomly split into two groups. Data from 442 276 individuals (70%) entered a Mixture Density Network (MDN) model to train an AI-DSA to predict an aortic valve area <1 cm<sup>2</sup>, excluding all left ventricular outflow tract velocity or dimension measurements and then using the remainder of echocardiographic measurement data. The optimal probability threshold for severe AS detection was identified at the f1 score probability of 0.235. An automated feature also ensured detection of guideline-defined severe AS. The AI-DSA's performance was independently evaluated in 184 301 (30%) individuals.

**Results** The area under receiver operating characteristic curve for the AI-DSA to detect severe AS was 0.986 (95% CI 0.985 to 0.987) with 4622/88 199 (5.2%) individuals (79.0±11.9 years, 52.4% women) categorised as 'high-probability' severe AS. Of these, 3566 (77.2%) met guideline-defined severe AS. Compared with the AI-derived low-probability AS group (19.2% mortality), the age-adjusted and sex-adjusted OR for actual 5-year mortality was 2.41 (95% CI 2.13 to 2.73) in the high probability AS group (67.9% mortality)—5-year mortality being slightly higher in those with guideline-defined severe AS (69.1% vs 64.4%; age-adjusted and sex-adjusted OR 1.26 (95% CI 1.04 to 1.53), p=0.021).

**Conclusions** An AI-DSA can identify the echocardiographic measurement characteristics of AS associated with poor survival (with not all cases guideline defined). Deployment of this tool in routine clinical practice could improve expedited identification of severe AS cases and more timely referral for therapy.

## INTRODUCTION

Affecting millions worldwide, aortic stenosis (AS) is the most common, acquired form of valvular heart disease managed in clinical practice.<sup>1</sup> When left untreated, there are substantial societal costs attributable to high rates of premature mortality and quality-adjusted life years lost across the entire spectrum of disease.<sup>2</sup> This is important when

### WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Many people with moderate-to-severe aortic stenosis (AS) are at risk of dying without timely evidence-based care.

### WHAT THIS STUDY ADDS

⇒ We developed and tested an artificial intelligence decision support algorithm to detect the phenotype associated with severe AS, in addition to clinical guideline-defined cases of severe AS from their routine echocardiographic report. The algorithm rapidly identified these high-risk cases with excellent performance and identified patients with a high risk of death.

### HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This algorithm has the capacity to be uniformly applied as an automated alert system for high-risk patients with AS in routine clinical practice.

considering the long-held concept that only severe symptomatic cases of AS should be referred to a specialist heart care team for aortic valve replacement.<sup>3–5</sup> Consistent with a broader phenotype of 'high-risk' AS (encompassing individuals undergoing mal-adaptive changes to their cardiac structure/function in response to their failing aortic valve),<sup>6</sup> there is increasing evidence that both asymptomatic severe and moderate AS are also associated with high rates of mortality.<sup>3,4,7</sup> Accordingly, the capacity to improve survival rates in such individuals via interventional strategies is being tested in prospective randomised trials.<sup>8</sup>

Beyond recognising the prognostic implications of all forms of AS, there is increasing clinical tension to apply optimal and timely strategies to rapidly detect and definitively treat high-risk individuals.<sup>9</sup> However, even within the most well-resourced healthcare settings, this is problematic.<sup>10</sup> Consequently, the potential for machine learning/artificial intelligence (AI) systems to systematise the detection, treatment and prognostication

of AS is being explored via the synthesis of directly acquired and reported clinical data (including from cardiac auscultation,<sup>11</sup> mechanical sensors,<sup>12</sup> echocardiography,<sup>13–14</sup> computed axial tomography/MRI<sup>15</sup> and electronic medical records.<sup>16</sup>

We evaluated the performance of a newly refined artificial intelligence decision support system (AI-DISA)<sup>13</sup> to identify adults with more severe forms of AS associated with high-rates of mortality. The reference standard for the performance of the AI-DISA was severe AS as defined by current guidelines.<sup>17–18</sup> We further sought to explore the capacity of the AI-DISA to identify those with characteristic moderate-to-severe AS and elevated risk of mortality.<sup>7</sup> We also sought to determine the ‘minimum’ echocardiographic parameters routinely reported in clinical practice required by the AI-DISA to operate efficiently.

## METHODS

The data that support the findings of this study are available from the corresponding author on reasonable request.

### Study design

In this retrospective, patient cohort and outcome analysis, we evaluated the performance of a stand-alone AI-DISA to detect individuals with high probability of the severe AS phenotype.<sup>17–18</sup> The AI-DISA scrutinised the same echocardiographic information used to evaluate potential heart disease as part of the routine clinical practice. The subsequent pattern of (actual) 5-year all-cause mortality according to the AI-DISA outputs (including a feature to automatically identify guideline-defined severe AS within in the AI-DISA-identified population) was then compared independently. Where appropriate, this study conforms to the ‘Standards for Reporting Diagnostic accuracy studies’ guidelines.<sup>19</sup>

### Data sources

As described in greater detail previously,<sup>2–7,9</sup> this unique resource is a vendor agnostic source neutral database containing measurement and text outputs from multiple echocardiographic laboratories Australia-wide. Source data are derived from individuals being routinely investigated/managed with heart disease within Australia’s multicultural population via its well-resourced, universal healthcare system. Only those records with insufficient demographic details to enable highly secure/anonymised individual data linkage to the well-validated National Death Index<sup>20</sup> were excluded.

For these analyses, we used the latest version of the mortality-linked database containing 1 077 145 studies from 631 824 individuals aged  $\geq 18$  years during the period from 29 May 1985 to 26 June 2019. Individual all-cause mortality was established during a median (IQR) of 4.3 (2.3–7.3) years follow-up from last echocardiogram to a census date of 21 May 2019). A total of 280 individual echocardiographic variables (online supplemental material pp.1–11), representing base measurements

and calculations as part of a standard echocardiography examination, were provided for the AI-DISA training and validation. Mortality status was not provided and did not form part of the AI-DISA training. Individuals with prior aortic valve replacement were excluded.

### Development of the AI model

The data set was split (ratio 70:30) into two separate groups, one for model training and the other for test/validation of the trained model (online supplemental material pp.12–18). As part of a six-step process, a modified Mixture Density Network was used to train on the 442 276 individuals (70%) and their 754 503 echocardiograms randomly assigned to the training set. As part of this development process, we were able to directly address the sparsely filled data sets typical in clinical echocardiography. Critically, the left ventricular outflow tract (LVOT) data relevant to the continuity equation (specifically, all velocity, gradient, velocity time integral, dimension and area) were *withheld* from the AI-DISA test set model. The MDN model was then used to predict the probability of severe AS (after being trained on the entire echo) defined by an aortic valve area (AVA)  $< 1 \text{ cm}^2$ . The trained model was designed to be general purpose and can perform inference using arbitrary sets of available measurements.

### Development of the guideline-quarantined patients

Subsequently, a guideline analysis was developed using the current AS diagnostic guidelines, to quarantine severe patients with AS from the AI-DISA-derived phenotype, if the peak velocity was  $\geq 4.0 \text{ m/s}$ , the mean gradient was  $\geq 40 \text{ mm Hg}$  and/or the AVA  $\leq 1.0 \text{ cm}^2$ .

### Evaluating the model

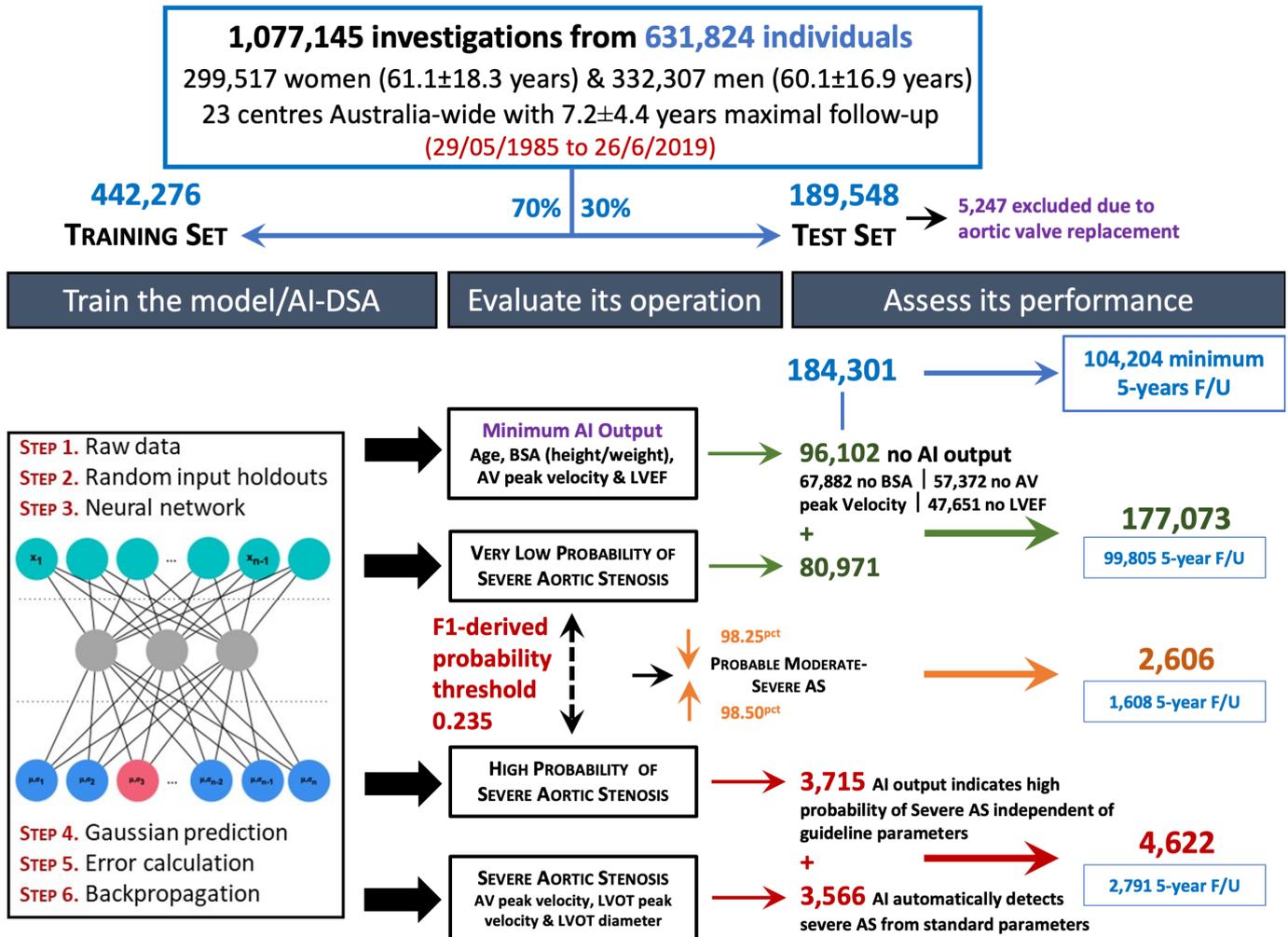
Once developed, the model’s performance was independently evaluated using data from the remaining 189 548 (30%) individuals and their 322 642 echocardiograms (online supplemental material pp.19–21) that had never been seen by the AI-DISA during training and were linked to mortality outcomes. As an initial diagnostic, selected groups of related measurements were withheld, and the AI predictions were evaluated against known values. These results indicated that the predicted measurements had minimal bias and surprisingly low error bounds considering the heterogeneous nature of the data and that key information (ie, LVOT data) had been intentionally removed.

### Testing the AI-DISA

Using the Mixture Density Network (MDN)-predicted output for AVA:

$$\text{AVA} = (\pi(\text{LVOTd})^2/4) \times (\text{LVOT}_{\text{VTI}}/\text{AV}_{\text{VTI}})$$

where LVOT=left ventricular outflow tract, AV=aortic valve, d=dimension and VTI=velocitytime integral and using the cumulative distribution function to determine what percentage of the distribution falls above the severe AS threshold, we calculated the probability of an AVA  $< 1 \text{ cm}^2$ —online supplemental material pp.22–27.



**Figure 1** Flow chart of the training, evaluation, and testing of the AI-DSA to detect severe forms of AS. This schema shows the distribution of cases/investigations from the NEDA cohort used to train and evaluate the AI-DSA (investigation-based, no mortality data) and then test/assess its performance in accurately detecting the severe form of aortic stenosis associated with high 5-year mortality (individual-based). The highest F1 score was chosen as the probability at the peak of the precision/recall relationship, corresponding to a probability of 0.235. The moderate-to-severe aortic stenosis group corresponds to a probability output of >0.0625 (98.25 to 98.5 percentile of probability spectrum below the f1-derived threshold). AI-DSA, Artificial Intelligence Decision Support Algorithm; AS, aortic stenosis; AV, aortic valve; BSA, body surface area, F/U, follow-up; LVEF, left ventricular ejection fraction; LVOT, left ventricular outflow tract; NEDA, National Echo Database of Australia; pct, percentile.

Curves for receiver operating characteristic (ROC) and areas under the ROC (AUROC) were then generated to evaluate the performance of the severe AS classifier, and the probability threshold value was adjusted for a maximum f1-score calculation. Analyses were conducted on the entire test group and then left ventricular ejection fraction (EF) thresholds of <30% (4203 cases) and <50% (18 799 cases)). For testing the performance of the AI-DSA, mortality outcome data were added to the analysis database.

#### Individual classification according to AI-DSA outputs

As shown in figure 1, after excluding those with an AV replacement, the overall performance of the AI-DSA, both in terms of determining the severity of AS<sup>17 18</sup> at their last echocardiogram and subsequent survival, was assessed in 184 301 individuals aged >18 years. Three main groups (incorporating five subgroups) were identified from the

AI-DSA outputs. The first subgroup comprised those with insufficient data (age, body surface area, aortic valve peak velocity and EF) to enable the AI-DSA to produce a probability output. To reflect real-world clinical practice, these were automatically defaulted into the main 'low-probability' group. Those individuals with sufficient data, according to the f1-derived threshold of 0.235 (below and above), were initially categorised as 'low' or 'high' probability of the severe AS phenotype identified by the AI. Within the 'low' probability group, a third main group was prospectively derived from those with sufficient data and a probability score below the f1-derived threshold. Specifically, applying 98.25th to 98.50th<sup>h</sup> percentile of probability distribution below the f1-derived threshold, a probability range of patients with 'moderate-to-severe' AS was identified (probability >0.0625 to <0.235). The final safety function then automatically identified

guideline-based severe AS.<sup>17 18</sup> The AI – DSA was then tested against the 58 170 individuals with a native AV, with guideline-applicable AS data reported.

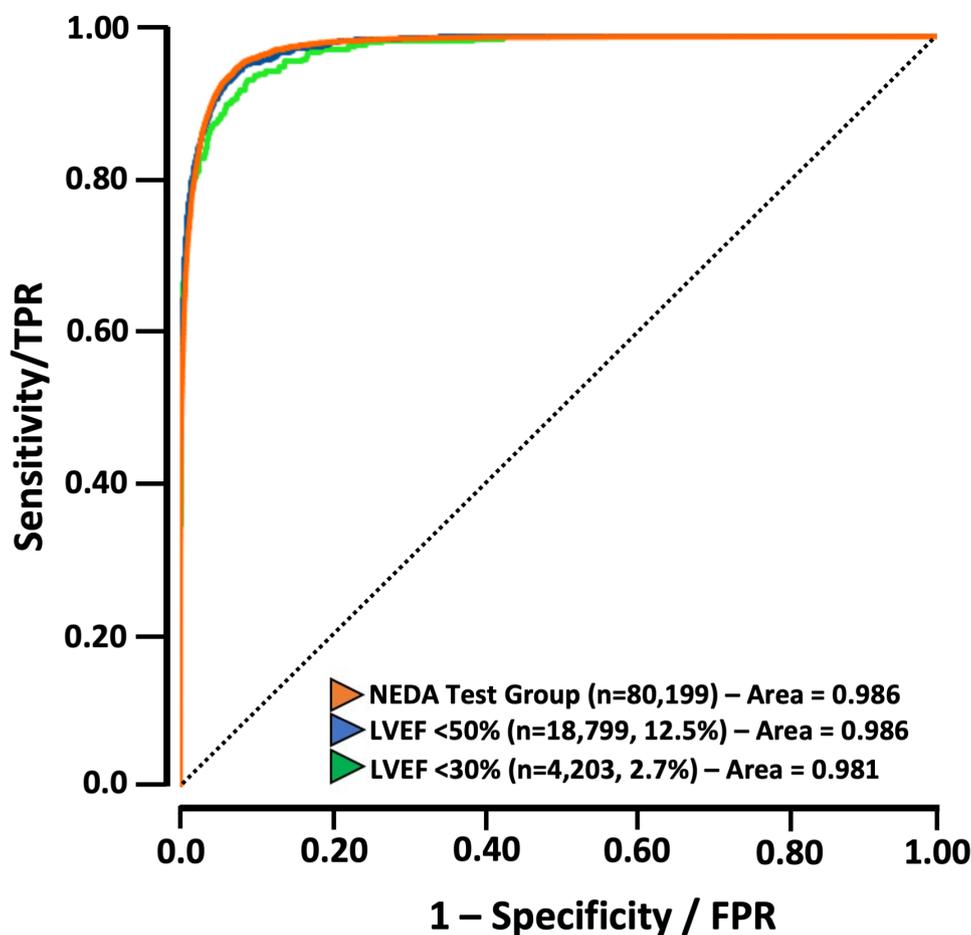
### Statistical analyses

Given the size and scope of National Echo Database of Australia (NEDA) cohort (>5 000 000 individuals with >1 million echocardiograms) collated on a consecutive basis, no formal power calculations were performed. Beyond those analyses described above, standard methods for describing grouped data, including means ( $\pm$ SD), median (IQR) and proportions (with 95% CIs) were performed. Between-group comparisons included ANOVA, Student's t-test and  $\chi^2$  analyses where appropriate. Actual 5-year mortality was calculable in 104 204 cases with complete 5-year follow-up. For each predetermined group identified, multiple logistic regression (entry model) was used to generate age-adjusted and sex-adjusted OR with 95% CI for 5-year mortality with the lowest probability group (as determined by the AI-DSA) set as the reference group. The same method was repeated for age-specific OR for men and women separately. All descriptive and survival analyses were performed with SPSS V.28.0 (IBM Corporation, Chicago, Illinois) and

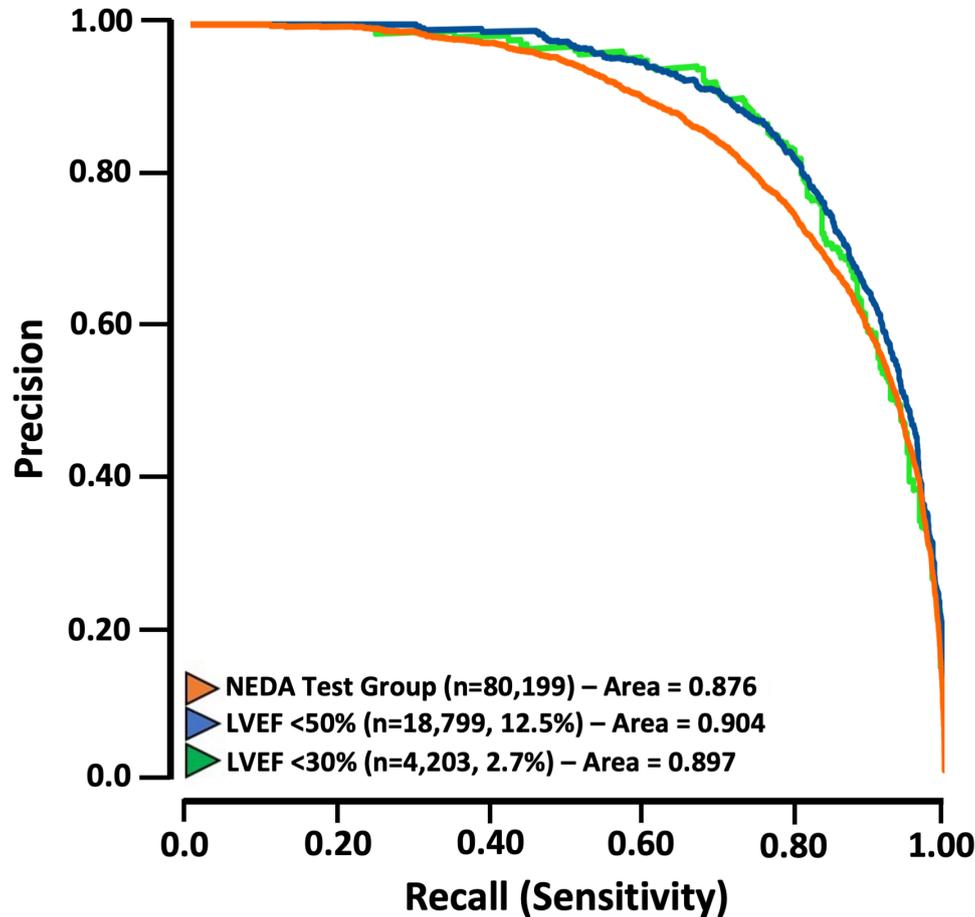
statistical significance accepted at a two-sided alpha of <0.05.

### RESULTS

The performance of our AI-DSA to detect severe AS is shown in figure 2—the AUROC (95% CI) being close to one overall (0.986 (0.985 to 0.987)) and among those with impaired (LVEF <50%–0.986% (0.984 to 0.988)) to severely impaired (LVEF <30%–0.981% (0.975 to 0.986)) left ventricular systolic function. The subsequent precision-recall of the AI-DSA to detect severe AS in those with more complete echocardiographic reports is shown in figure 3—the AUPR (95% CI) being close to 0.9 overall (0.876 (0.869 to 0.883)) and among those with impaired (LVEF <50%–0.904% (0.892 to 0.915)) to severe impaired (LVEF <30%–0.897% (0.871 to 0.920)) left ventricular systolic function. An f1-derived threshold (the harmonic mean of precision/recall) based on the probability output of the AI-DSA was identified at 0.235—figure 4. Among the 184 301 individuals comprising the test cohort, 80 971 (52.3%) did not have the minimum data to produce a probability output. Minimum data to allow for AS guideline application were also not available



**Figure 2** Performance of the model to detect severe AS. This graph shows the performance of the model underpinning the AI-DSA to identify an aortic valve area of <1.0 cm<sup>2</sup>. AI-DSA, Artificial Intelligence Decision Support Algorithm; AS, aortic stenosis; FPR, false positive rate; LVEF, left ventricular ejection fraction; NEDA, National Echo Database of Australia; TPR, true positive rate.



**Figure 3** Precision-recall performance of the model to detect severe AS. This graph shows the precision values (true positives/(true positives+false negatives)) on the y-axis and recall values (true positives/(true positives+false positives)) on the x-axis derived from the AI-DISA output. AI-DISA, Artificial Intelligence Decision Support Algorithm; AS, aortic stenosis; LVEF, left ventricular ejection fraction; NEDA, National Echo Database of Australia.

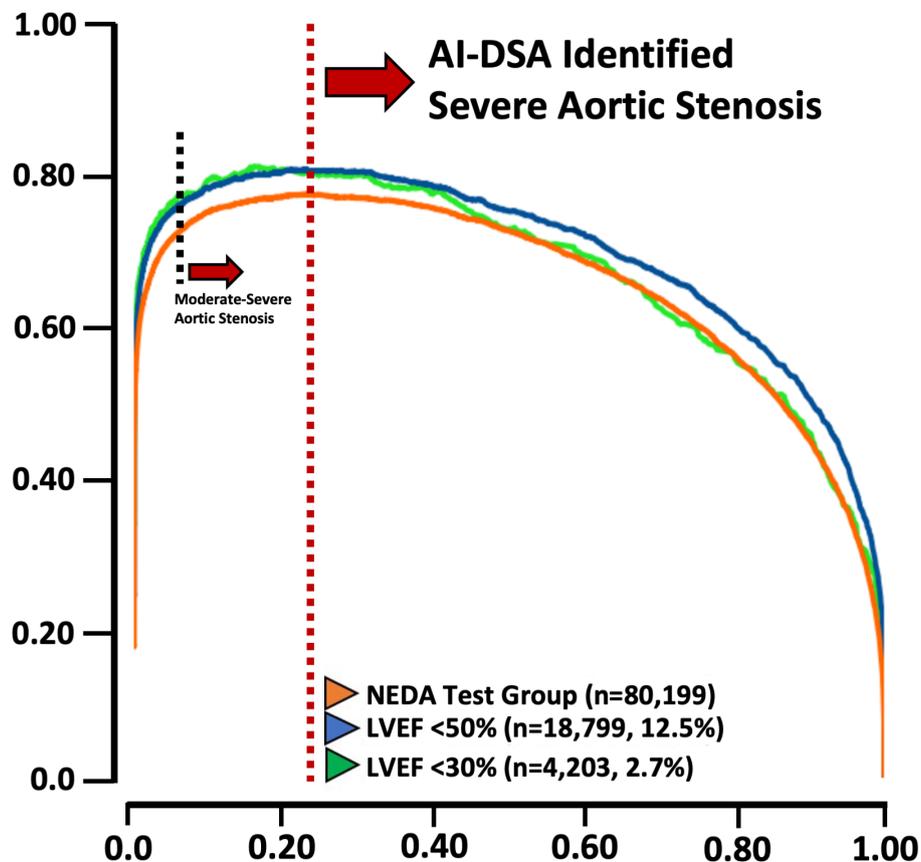
in 128 228 individuals (66.9%), noting that this predominantly comprised echocardiograms without AS—Most echos with insufficient data to calculate the AVA had AV peak velocity  $\leq 2$  m/s (in 62 537 individuals, 55.9%). Notably, however, there was a small but important proportion of individuals with elevated AV peak velocities ( $\geq 3$  m/s, 1356 individuals, 28.1%) had insufficient data to calculate the AVA using the continuity equation (n=1264 (26.2%) without LVOT peak velocity and/or LVOT diameter and n=1984 (41.1%) without AV VTI and/or LVOT VTI and/or LVOT diameter). All of these individuals had an AS probability determined by the AI.

Table 1 summarises the demographic and echocardiographic characteristics of the three main groups identified within the test cohort; comprising 177 073 individuals (96.1% designated as low probability of severe AS, including 80 971 with a definitive AI-DISA probability score), 2606 (1.4% overall) identified as ‘increased risk’ of the moderate-to-severe AS phenotype and 4622 (2.5% overall) identified as severe AS. Of the latter, 3566 (77.2%) had severe AS according to clinical guidelines—the AI-DISA identifying all such cases when possible.

Overall, there were statistically significant differences ( $p < 0.001$  all comparisons) between the three main

groups, with exception of heart rate and left ventricular diastolic/systolic diameter. Low probability AS cases identified by the AI-DISA were significantly younger and comprised more women than those with a high probability of moderate-to-severe and severe AS, while demonstrating more normal cardiac parameters. Conversely, the individuals the AI-DISA identified as high probability severe AS (the oldest, male predominant group) had appropriately high levels of AV, left ventricular and right ventricular dysfunction. Those identified with high probability moderate-to-severe AS also had high levels of cardiac dysfunction, but to a less extent. The specific comparison between those identified by the AI-DISA as a high probability of severe AS, versus those who met guideline criteria for severe AS, revealed minor, but statistically significant differences between the two in respect to AV function, indices of diastolic dysfunction and evidence of left ventricular remodelling (all worse in the guideline group). Overall, those identified as high probability of the severe AS phenotype by AI, but outside guidelines for severe AS, had findings typically found at the higher spectrum of currently reported moderate AS.

Actual 5 year mortality was 67.9% and 56.2% among the 1896 and 903 individuals identified as high probability



**Figure 4** Probability threshold of the model to detect severe AS. This graph shows the plots used to determine the F1-derived threshold based on the average of precision and recall of the AI-DSA to detect severe aortic stenosis (main red dotted line) overall (F1-derived probability threshold 0.235) and in those with a left ventricular ejection fraction <50% and <30%. It also shows (short black dotted line)—the 0.0625 probability threshold for identifying ‘moderate aortic stenosis’ group. AI-DSA, Artificial Intelligence Decision Support Algorithm; AS, aortic stenosis; LVEF, left ventricular ejection fraction; NEDA, National Echo Database of Australia.

severe and moderate-to-severe AS, respectively. This compared with a mortality rate of 22.9% in the low probability group—the age-adjusted and sex-adjusted OR (95% CI) for all-cause mortality being 1.82 (1.63 to 2.02) and 2.80 (2.57 to 3.06) for the moderate-to-severe and severe AS groups (figure 5). Within the low probability group, 5-year mortality was significantly lower in comparison to the specific AI-DSA identified group (9068/47 345 (19.2%)) than those with insufficient data (13 452/51 578 (26.3%)); age-adjusted and sex-adjusted OR 1.60 (95% 1.55 to 1.65,  $p<0.001$ ). Within the high probability group, 5-year mortality was significantly higher in those who met guideline criteria for severe AS (1438/2081 (69.1%)) compared with those identified by the AI-DSA (458/711 (64.4%)); age-adjusted and sex-adjusted OR 1.26 (95% 1.04 to 1.53,  $p=0.021$ )—figure 6. On a sex-specific basis, relative to the low probability group, the age-adjusted OR (95% CI) for 5-year all-cause mortality was consistently higher for the 49 120 women versus 54 160 men categorised as moderate-to-severe AS (2.03 (1.74 to 2.36) vs 1.63 (1.40 to 1.91)) and severe AS (3.00 (95% CI 2.68 to 3.40) vs 2.56 (95% CI 2.32 to 2.91)).

We examined the potential importance of a simple measure of the AVA compared with the probability

assessment in a series of sensitivity analyses. Of the 29 102 individuals with actual 5-year mortality follow-up data available, 1566 individuals had a calculated AVA <1 cm<sup>2</sup>, 898 had an AVA 1.0 cm<sup>2</sup> to <1.2 cm<sup>2</sup>, 1738 had AVA 1.2 cm<sup>2</sup> to <1.5 cm<sup>2</sup>, 4348 had an AVA 1.5 cm<sup>2</sup> to <2 cm<sup>2</sup> and 20 552 had an AVA >2 cm<sup>2</sup>. As expected, each AVA group had an elevated 5-year mortality, with the highest risk in the lowest AVA group. However, the probability score transcended the AVA by continuing to predict mortality beyond each AVA group, providing an additional age-adjusted and sex-adjusted hazard of 1.40 (1.06 to 1.86),  $p=0.019$ .

In a separate sensitivity analysis, we examined whether the AI probability could predict mortality risk in low-flow AS beyond the AVA alone. Of 341 patients with a recorded Stroke Volume Index below 35 mL/m<sup>2</sup> and AVA <1 cm<sup>2</sup>, 259 individuals (76.0%) had died within 5 years. After adjustment for age and sex, the AI probability continued to be independently highly associated with mortality (HR 1.98, 95% CI 1.13 to 3.46,  $p=0.016$ ), whereas low-flow low-gradient severe AS was not independently associated with mortality (HR 1.29, 95% CI 0.89 to 1.87,  $p=0.17$ ). These data confirm that the probability score is not confounded by the severity or type of AS.

**Table 1** Distribution of the test cohort according to the AI-DSA outputs

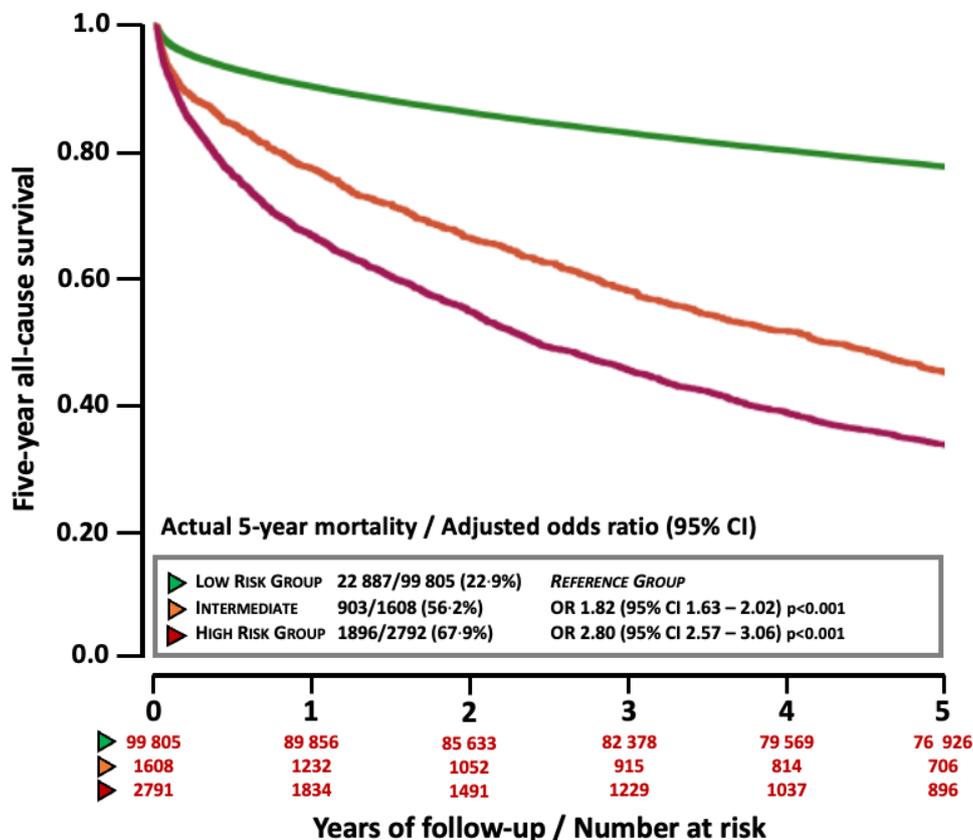
	Aortic Stenosis Cohort (n=4622)			Non-Severe AS Cohort (n=179679)	
	ALL (n=1 843 011)	High probability severe AS (n=4622)	Severe AS guideline (n=3566)	High probability moderate AS (n=2606)	Low probability severe AS (n=1 77 073)
Demographic profile					
Age, years	61.8±17.8	79.0±11.9	78.9±12.1	77.1±12.3	61.1±17.6
Women, %	87 987 (47.7%)	2423 (52.4%)	1782 (50.0%)	1306 (50.1%)	84 258 (47.6%)
Clinical profile					
Body mass index, kg/m <sup>2</sup> (n=67 882)	28.1±6.4	26.7±5.8	26.8±5.8	28.3±6.5	28.1±6.4
Systolic   diastolic BP, mm Hg (n=21 005)	134±22   77±12	137±25   74±13	136±25   73±13	140±24   78±11	133±22   77±11
Heart rate, bpm (n=66 997)	72.4±15.5	73.1±15.6	73.3±15.7	72.4±15.3	73.0±15.5
Aortic valve profile					
Peak velocity, m/s (n=1 269 29)	1.40 (1.20–1.70)	3.70 (3.10–4.29)	3.99 (3.10–4.44)**	2.80 (2.38–3.20)	1.40 (1.20–1.62)
Mean gradient, mm Hg (n=69 359)	4.50 (3.13–7.56)	31.00 (21.00–43.50)	35.50 (22.00–47.00)**	19.00 (14.70–24.00)	4.10 (3.00–6.06)
Aortic valve area, cm <sup>2</sup> (n=61 073)	2.52 (1.93–3.16)	0.84 (0.69–0.98)	0.81 (0.67–0.93)**	1.21 (1.10–1.36)	2.60 (2.08–3.22)
≥ Mild aortic regurgitation (n=66 796)	17 514 (26.2%)	1410 (53.5%)	1518 (42.6%)**	915 (57.1%)	15 189 (24.3%)
Right ventricular function and dimensions					
eRVSP, mm Hg (n=90 529)	37.3±11.9	45.7±14.2	45.7±14.3	46.2±12.5	36.7±11.5
TR peak velocity, m/s (n=87 199)	2.60±0.49	2.95±0.54	2.95±0.54	2.99±0.47	2.57±0.48
Left ventricular function and dimensions					
Left atrial volume index, mL/m <sup>2</sup> (n=54 305)	41.6±28.8	61.0±44.7	57.2±39.2**	94.3±56.2	39.9±25.9
LV diastolic dimension, cm (n=1 35 063)	4.68±0.73	4.61±0.80	4.59±0.79	4.91±0.95	4.68±0.72
LV systolic dimension, cm (n=1 06 935)	3.04±0.76	3.05±0.90	3.04±0.89	3.06±0.90	3.04±0.75
LV mass index, g/m <sup>2</sup> (n=41 872)	82.8±25.1	106.0±30.7	115.0±34.7**	90.8±28.3	81.9±24.5
LV ejection fraction, % (n=1 36 650)	62±13	59±16	58±16	63±18	62±12
Septal E:e' ratio (n=64 626)	10.60±5.03	17.30±8.49	17.20±8.46	15.10±5.76	10.40±4.77
Transmitral E/A ratio (n=1 16 143)	1.15±0.67	1.18±1.02	1.16±1.06*	1.29±1.23	1.15±0.65
Stroke volume index, mL/m <sup>2</sup> (n=40 947)	40.1±11.7	36.5±14.6	35.8±14.8**	39.5±10.7	40.2±11.5

This table shows the demographic and echocardiographic profile of each group (including the key reference group—severe aortic stenosis according to guideline criteria (identified separately in blue) within the test cohort with a native aortic valve. \*p<0.05 and \*\*p<0.001 for the comparison between the two severe aortic stenosis groups. LV mass index calculated using the 2D ASE formula. Stroke volume index was derived from left ventricular outflow tract (LVOT) velocity time integral and the LVOT dimension.<sup>17,18</sup> eRVSP, estimated right ventricular systolic pressure; LV, left ventricular; TR, tricuspid regurgitation.

As a final analysis to determine whether the AI-DSA is simply mimicking expanded thresholds of risk based on AVA and Vmax (when available), we generated five new moderate-to-severe AS subgroups. These groups demonstrated the expected small increments in 5-year mortality, with the severe AS subgroup of AVA <1.0 cm<sup>2</sup> (and Vmax >3.2 m/s) demonstrating the worst survival (see online supplemental table S1, figure S1–S3). The moderate (55.4%) and severe AS (65.4%) subgroups demonstrated the expected adverse 5-year mortality, with the comparator group experiencing 23.4% 5-year mortality (online supplemental figure S2). We also confirmed the high 5-year mortality for the two guideline-derived severe AS subgroups in addition to those without an AVA available (AVA <1.2 cm<sup>2</sup> and Vmax >4.0 m/s, AVA <1.0 cm<sup>2</sup> and Vmax >3.2 m/s, and no AVA and Vmax >3.2 m/s, see online supplemental figure S3).

## DISCUSSION

To our knowledge, this is the largest study to train and test the performance of an AI-DSA to interpret routinely acquired echocardiographic reporting data to transcend the more traditional AVA calculation and identify the phenotype of more severe AS associated with high mortality. In almost 90 000 individuals with age, BSA, peak aortic velocity and LVEF recorded, the AI-DSA identified 5.2% as 'high-probability' for severe AS. Overall, 23.8% of these cases were not guideline-defined severe AS. AUROC and precision-recall analyses consistently demonstrated excellent performance values (close to one) for all individuals, including those with impaired left ventricular systolic function. Compared with the low-probability group, the age-adjusted and sex-adjusted odds of 5-year mortality was 2.41-fold higher among those categorised as as high-probability severe AS (67.9% mortality), respectively. Our methodology is novel and



**Figure 5** Actual 5-year all-cause mortality according to three main outputs from the AI-DSA. This graph shows the 5-year actual mortality curves (all-cause and with no censoring of cases) for the three main output groups from the AI-DSA—with low probability individuals being the reference group). The ORs for age and sex were 1.08 (95% CI 1.08 to 1.08 per annum) and 1.50 (95% CI 1.46 to 1.55 for men vs women). AI-DSA, Artificial Intelligence Decision Support Algorithm.

meets the growing calls for automated echocardiographic reporting, to identify patients reliably and consistently with AS phenotypes that may benefit from early clinical review and consideration for AV intervention.<sup>21 22</sup>

Recently, a range of AI techniques have been applied to expedite the diagnosis of (predominantly severe) AS and improve its prognostication following AV replacement. In routine clinical practice, simple cardiac auscultation augmented by a digital AI-assisted stethoscopes has potential to identify those requiring cardiac imaging.<sup>11</sup> A similar approach has been explored in relation to AI detection of characteristic T wave changes on surface electrocardiograms.<sup>23</sup> Novel technology involving non-invasive inertial sensors<sup>12</sup> has also shown distinct phenotypes can be identified using invasive haemodynamics and cardiac imaging, which strongly predict future clinical outcomes.<sup>6 22</sup> Previously, we have demonstrated that the relationship between echocardiographic measurement variables could be automatically associated together to predict severe AS.<sup>13</sup> This approach has been subsequently refined by applying newer machine learning techniques and clinical practice guideline criteria.<sup>17 18</sup> Collectively, these studies have revealed the progressive nature of AV disease (involving valve calcification and myocardial fibrosis) as well as the normalisation of AV function following surgical intervention.<sup>24</sup> After trans-aortic

valve intervention, a machine learning system involving multiple clinical, CT, electrocardiographic and echocardiographic inputs was able to outperform standard clinical risk scores 1 year in predicting outcomes.<sup>25</sup> Thus, there is increasing potential to apply machine learning techniques to streamline the identification of high-risk cases with AS in routine clinical care.<sup>26</sup>

Why should an AI system like ours add value to clinical practice, when conventionally, there is only a dichotomy of treatable cases in AS (symptomatic severe vs the rest)? Our primary goal is to highlight at-risk patients for timely clinical review, closer monitoring and guideline-directed therapy where appropriate. Typical changes associated with progressive AS include AV calcification, left ventricular hypertrophy, diastolic dysfunction, increased left atrial pressure and pulmonary hypertension.<sup>15</sup> These pathophysiological changes have been used to improve timely identification of individuals in need of AV replacement, with the goal of improving clinical outcomes and minimising mortality risk.<sup>26</sup> Although direct visualisation of the multiple interconnected layers inherent to the neural network is not possible, our AI-DSA automatically identifies the typical phenotypic AS features associated with progressive disease, including a gradient of mortality with increasing AS probability that remains independently predictive after adjustment for the traditionally measured

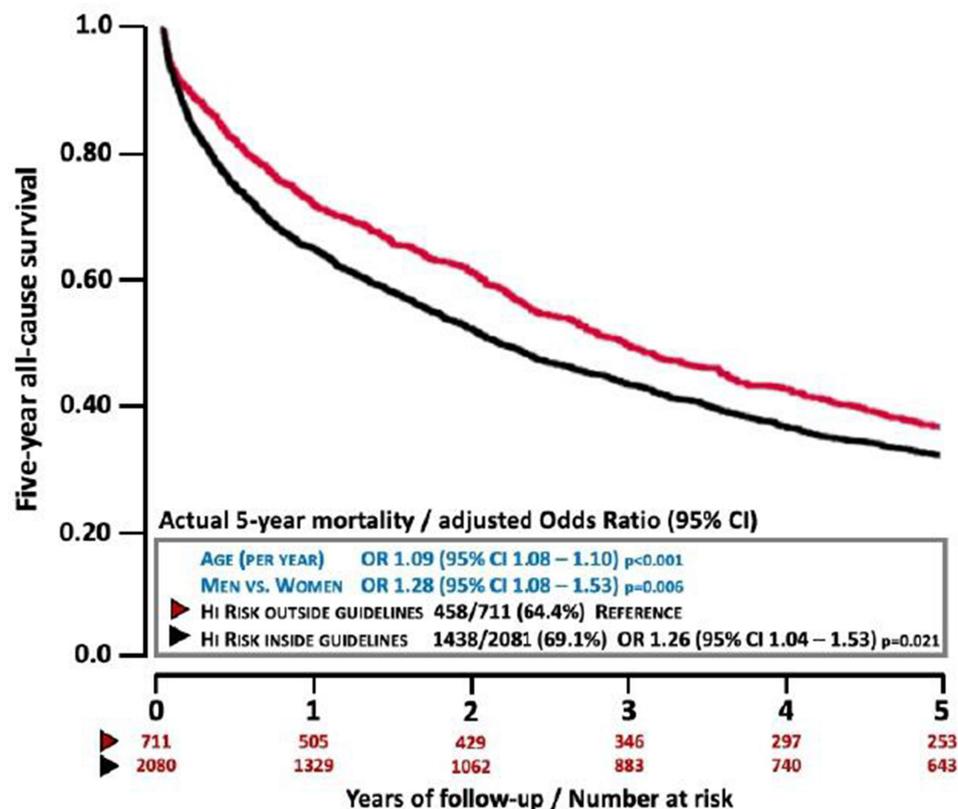
AVA. Using a progression of probability of severe AS from the AI-DISA, it is possible to discriminate differing risk thresholds, as we have shown with identification of the severe phenotype (inside and outside of traditional guidelines) as well as the moderate-to-severe group of individuals who are also at increased mortality risk, and thereby highlighting at-risk individuals in need of clinical review.

Timely recognition of those at risk of dying from progressively worse AS remains a high-priority clinical issue.<sup>2</sup> Studies from a range of health systems have consistently demonstrated the under-treatment of severe AS with consequently high mortality rates.<sup>27</sup> Low-gradient AS, which can be technically challenging to diagnose, is particularly subjected to variation in clinical interpretation of severity and under-represented in AV intervention.<sup>28</sup> Our AI-DISA performs equally well in high-gradient AS and in low-gradient AS, with AUCROC and AUPR almost identical in normal EF, mildly impaired and severely impaired systolic function. Recent observations of a significant detection and treatment gap in AS have prompted ‘urgent calls to action’ by major cardiology and echocardiography societies.<sup>21–29</sup> Our AI-DISA can be directly incorporated into echocardiographic reporting systems, or applied to echo labs directly, thereby triggering an automated alert to the presence of

conventional severe AS, in addition, those who require further consideration for action. When combined, these represented ~7% (ratio of 1:1) of all test cases with sufficiently detailed reporting data. Critically, the AI-DISA output requires no modification of standard echocardiographic imaging or acquisition workflow. As recently reported by rECHOmend Investigators, it is possible that similar machine learning technology with a broader relevance to all forms of structural valvular disease has the potential facilitate meaningful recommendations for echocardiography in clinical practice.<sup>30</sup>

### LIMITATIONS

The AI-DISA is not intended to replace human clinical decision-making, given some patients will have conditions (eg, dementia, frailty or systemic disease) precluding surgical management. Furthermore, it is currently incapable of identifying other cardiac diseases that may adversely affect prognosis (eg, amyloid cardiomyopathy). Moreover, the AI-DISA is dependent on minimum echocardiographic measurements being performed and could not deliver a definitive output in just over 50% of echocardiograms, although most missing aortic valve data were in patients without AS consistent with common echocardiography practice. While affected individuals (who also had



**Figure 6** Actual 5-year all-cause mortality in the two severe AS groups (AI-DISA identified vs guidelines). This graph shows the 5-year actual mortality curves (all-cause and with no censoring of cases) for the two output groups identified by the AI-DISA as severe aortic stenosis—according to clinical guideline criteria (black line) or otherwise (red line—reference group). The ORs for age and sex were 1.09 (95% CI 1.08 to 1.10 per annum) and 1.28 (95% CI 1.08 to 1.53 for men vs women). AI-DISA, Artificial Intelligence Decision Support Algorithm; AS, aortic stenosis.

insufficient data to apply AS guideline criteria) had lower mortality than positively identified AS cases, it was higher than those more definitively identified as 'low probability AS'. Although the AI-DISA reliably identifies low-gradient severe AS, this form of AS is typically associated with other cardiac diseases. Given that data from a large multi-cultural Australian cohort were used to train and test the AI-DISA, it has yet to be tested in other geographic regions/health systems or specific ethnic groups. Critically, we were unable to discriminate between symptomatic and non-symptomatic severe AS and NEDA does not (currently) capture potentially confounding data on comorbidities, hospital episodes and/or pharmacotherapies. The AI-DISA has yet to be compared with the routine clinical detection and management of AS in respect to the cost-effective management of severe forms of AS.

## CONCLUSIONS

In summary, we have demonstrated that a readily deployable AI-DISA has the potential to identify individuals with the phenotype of more severe AS associated with poor survival (if left untreated). Consistent with calls from major cardiac societies, the AI-DISA raises an automatic alert when the results of routine echocardiography are being reported. It can do so without increasing the clinical workload for the sonographer, echocardiography laboratory or the referring clinician. With further integration into current workflow, this AI-DISA could improve expedited identification of severe AS cases and more timely referral for therapy.

**Twitter** David Playford @PlayfordDavid

**Acknowledgements** Dr Stewart is supported by the NHMRC of Australia (GNT1135894).

**Contributors** The study design was conceived by SS and DP. ECHO IQ Pty Ltd developed the AI-DISA (EchoSolv®) and provided the optimal f1 probability (Dr Watts) but played no role in clinical data collection, data analysis, data outcome interpretation, or writing of the report. The data were independently analysed in a blinded fashion by Profs Stewart, Strange and Playford. SS drafted the first draft of the manuscript and did the sensitivity analyses. All authors contributed to the final version of the submitted manuscript. The corresponding author (Prof Strange) had full access to all the data (derived from the training and test groups) with the addition of mortality data, and the final responsibility to submit for publication.

**Funding** National Health & Medical Research Council of Australia and ECHO IQ Pty Ltd. Drs Strange and Playford are the Co-Principal Investigators and Directors of NEDA Ltd (a not-for-profit research entity). NEDA has received investigator-initiated funding support from Novartis Pharmaceuticals, Pfizer Pharmaceuticals, ECHO IQ and Edward Lifesciences in the past 3 years. Dr Stewart has received consultancy fees from NEDA and ECHO IQ. The funders played no role in the overall study design.

**Competing interests** Profs Stewart, Playford and Strange have previously received consultancy/speaking fees from Edwards Lifesciences. Profs Playford and Strange have received consultancy fees from Medtronic, Edwards Lifesciences, Abbott Laboratories and ECHO IQ Pty Ltd. Dr Watts is employed by ECHO IQ Pty Ltd.

**Patient consent for publication** Not applicable.

**Ethics approval** All data were derived from the National Echo Database of Australia (NEDA). NEDA has obtained ethical approvals across Australia from all relevant Human Research Ethics Committees. A patient consent waiver was authorised for all retrospectively acquired data used in these analyses.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Geoff Strange <http://orcid.org/0000-0001-6800-7119>

Simon Stewart <http://orcid.org/0000-0001-9032-8998>

Andrew Watts <http://orcid.org/0000-0003-3300-9127>

David Playford <http://orcid.org/0000-0003-4492-1103>

## REFERENCES

- 1 Tsao CW, Aday AW, Almarzoq ZI, *et al*. Heart disease and stroke Statistics-2022 update: A report from the American heart Association. *Circulation* 2022;145:e153–639.
- 2 Stewart S, Afoakwah C, Chan Y-K, *et al*. Counting the cost of premature mortality with progressively worse aortic stenosis: A clinical cohort study. *The Lancet Healthy Longevity* 2022;3:e599–606.
- 3 Postolache A, Tridetti J, Nguyen Trung ML, *et al*. Moderate aortic stenosis: a new actor has come into stage. *J Thorac Dis* 2020;12:7064–8.
- 4 Badiani S, Bhattacharyya S, Azimnia N, *et al*. Moderate aortic stenosis: what is it and when should we intervene *Interv Cardiol* 2021;16:e09.
- 5 San Román JA, Vilacosta I, Antunes MJ, *et al*. The 'wait for symptoms' strategy in asymptomatic severe aortic stenosis. *Heart* 2020;106:1792–7.
- 6 Kwak S, Everett RJ, Treibel TA, *et al*. Markers of myocardial damage predict mortality in patients with aortic stenosis. *J Am Coll Cardiol* 2021;78:545–58.
- 7 Strange G, Stewart S, Celermajer D, *et al*. Poor long-term survival in patients with moderate aortic stenosis. *J Am Coll Cardiol* 2019;74:1851–63.
- 8 Popma JJ, Adams DH, Reardon MJ, *et al*. Transcatheter aortic valve replacement using a self-expanding Bioprosthesis in patients with severe aortic stenosis at extreme risk for surgery. *J Am Coll Cardiol* 2014;63:1972–81.
- 9 Stewart S, Chan YK, Playford D, *et al*. Incident aortic stenosis in 49 449 men and 42 229 women investigated with routine echocardiography. *Heart* 2022;108:875–81.
- 10 Li SX, Patel NK, Flannery LD, *et al*. Trends in utilization of aortic valve replacement for severe aortic stenosis. *J Am Coll Cardiol* 2022;79:864–77.
- 11 Chorba JS, Shapiro AM, Le L, *et al*. Deep learning algorithm for automated cardiac murmur detection via a Digital stethoscope platform. *J Am Heart Assoc* 2021;10:e019905.
- 12 Shokouhmand A, Aranoff ND, Driggin E, *et al*. Efficient detection of aortic stenosis using morphological characteristics of Cardiomechanical signals and heart rate variability parameters. *Sci Rep* 2021;11:23817.
- 13 Playford D, Bordin E, Mohamad R, *et al*. Enhanced diagnosis of severe aortic stenosis using artificial intelligence: A proof-of-concept study of 530,871 echocardiograms. *JACC Cardiovasc Imaging* 2020;13:1087–90.
- 14 Sengupta PP, Shrestha S, Kagiya N, *et al*. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *JACC Cardiovasc Imaging* 2021;14:1707–20.
- 15 Kang N gyu, Suh YJ, Han K, *et al*. Performance of prediction models for diagnosing severe aortic stenosis based on aortic valve calcium on cardiac computed tomography: incorporation of Radiomics and machine learning. *Korean J Radiol* 2021;22:334.
- 16 Solomon MD, Tabada G, Allen A, *et al*. Large-scale identification of aortic stenosis and its severity using natural language processing on electronic health records. *Cardiovasc Digit Health J* 2021;2:156–63.

- 17 Otto CM, Nishimura RA, Bonow RO, *et al.* ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American college of cardiology/American heart Association joint committee on clinical practice guidelines. *Circulation* 2021;143:e72–227.
- 18 Vahanian A, Beyersdorf F, Praz F, *et al.* ESC/EACTS guidelines for the management of valvular heart disease. *Eur J Cardiothorac Surg* 2021;60:727–800.
- 19 O'Connor MK, Dai H, Fraga GR. PRAME immunohistochemistry for Melanoma diagnosis: A STARD-compliant diagnostic accuracy study. *J Cutan Pathol* 2022;49:780–6.
- 20 Powers J, Ball J, Adamson L, *et al.* Effectiveness of the National death index for establishing the vital status of older women in the Australian longitudinal study on women's health. *Aust N Z J Public Health* 2000;24:526–8.
- 21 Lindman BR, Lowenstern A. The alarm Blares for Undertreatment of aortic stenosis: how will we respond? *J Am Coll Cardiol* 2022;79:878–81.
- 22 Lachmann M, Rippen E, Schuster T, *et al.* Subphenotyping of patients with aortic stenosis by Unsupervised Agglomerative clustering of echocardiographic and hemodynamic data. *JACC Cardiovasc Interv* 2021;14:2127–40.
- 23 Jentzer JC, Kashou AH, Lopez-Jimenez F, *et al.* Mortality risk stratification using artificial intelligence-augmented electrocardiogram in cardiac intensive care unit patients. *Eur Heart J Acute Cardiovasc Care* 2021;10:532–41.
- 24 Casacalang-Verzosa G, Shrestha S, Khalil MJ, *et al.* Network tomography for understanding Phenotypic presentations in aortic stenosis. *JACC Cardiovasc Imaging* 2019;12:236–48.
- 25 Hernandez-Suarez DF, Kim Y, Villablanca P, *et al.* Machine learning prediction models for in-hospital mortality after Transcatheter aortic valve replacement. *JACC Cardiovasc Interv* 2019;12:1328–38.
- 26 Namasivayam M, Myers PD, Gutttag JV, *et al.* Predicting outcomes in patients with aortic stenosis using machine learning: the aortic stenosis risk (asterisk) score. *Open Heart* 2022;9:e001990:9:..
- 27 Strom JB, Playford D, Stewart S, *et al.* Increasing risk of mortality across the spectrum of aortic stenosis is independent of Comorbidity & treatment: an international, parallel cohort study of 248,464 patients. *PLoS One* 2022;17:e0268580. 10.1371/journal.pone.0268580 Available: <https://doi.org/10.1371/journal.pone.0268580>
- 28 Snir AD, Ng MK, Strange G, *et al.* Prevalence and outcomes of low-gradient severe aortic stenosis-from the National echo database of Australia. *J Am Heart Assoc* 2021;10:e021126.
- 29 Barnhart GR, Martin RP, Thomas JD, *et al.* The need for echocardiography alerts for aortic stenosis: the time has come. *J Am Soc Echocardiogr* 2020;33:355–7.
- 30 Ulloa-Cerna AE, Jing L, Pfeifer JM, *et al.* rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of Undiagnosed structural heart disease detectable by echocardiography. *Circulation* 2022;146:36–47.