# A Appendix

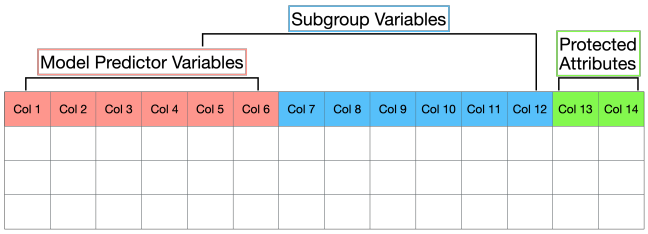


Figure A.1: A diagram illustrating the description of datasets by their columns. The columns can be categorized into the three types: i) predictor variables; ii) subgroup variables; iii) protected attributes.

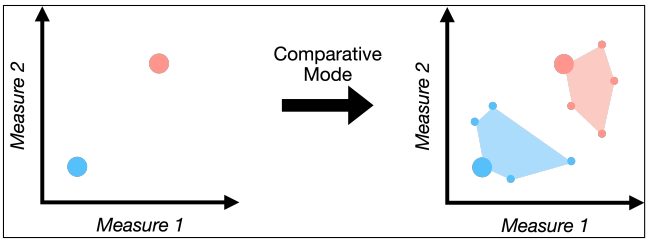


Figure A.2: Illustration of the comparative mode of scatter plots used for the performance and fairness of subgroups in Subgroup Summary. Initially, each scatter plot shows dots colored by subgroups on two continuous measures. As users switch to the comparative mode, the chart introduces other risk models to the chart by drawing a polygon per subgroup (color). Each vertex of the polygon represents a different risk model.

Table A.1: Population selection from the UK Biobank (incident AF).

| Number of patients | Selection criteria |
|---|---|
| 502,521 | All registered patients. |
| 456,793 | Age $\geq$ 45 at baseline ("date of enrollment"). |
| 453,573 | Complete values: height, weight, SBP, and DBP at baseline. |
| 445,357 | No history of AF on or prior to baseline. |
| 445,329 | Exclude patients who asked to opt out from the study. |

## A.1 Individual Fairness

For individual fairness, we adopted the auditing procedure of [20]. This procedure aims to find individuals that are similar to those in the original dataset while experiencing different model outcomes. Different decisions on a pair of similar individuals constitute an individual fairness violation. We report a fraction of data points where individual fairness is violated, i.e. the auditing algorithm [20] manages to find similar individuals with predicted risk scores on the opposite sides of the user-specified threshold. The similarity of individuals is quantified using the fair distance learned from the data following the sensitive subspace approach described in Appendix B.1 of [30]. Their method measures the similarity of two individuals with covariates $x$ and $x'$ with a fair distance $d(x,x')$ which ignores all differences in a sensitive subspace and only accounts for the differences in the directions orthogonal to this subspace. The subspace consists of the directions which are most informative for the sensitive attributes and are obtained from logistic regression coefficients which predict a sensitive attribute from the other covariates. This fair distance treats two individuals to be similar if they are similar in all respect, except for the sensitive attributes. For an individual with covariates $x$, the auditing method [20] aims to find a similar individual $x'$ which is close to $x$ in terms of the fair distance and maximizes the prediction loss. Let $p(x)$ denote the predicted risk score for an individual $(x)$, and $y(x)$ to be the indicator of whether an individual belongs to a high-risk group. Then $x'$ is obtained as

$$x' = \begin{cases} \arg\min_u p(u) + \lambda d(x,u) & \text{if } y(x) = 1, \\ \arg\max_u p(u) - \lambda d(x,u) & \text{if } y(x) = 0, \end{cases}$$

and, the metric is calculated as

$$\text{individual fairness violation rate} = \frac{\#\text{ individuals with } y(x) \neq y(x')}{\text{sample size}}.$$


Figure A.3: The model behavior explanation plots, the SHAP plot and the feature distribution summary, for the five income subgroups from low (left) to high (right). The feature distributions show some differences in age of patients at the baseline (dotted areas).

Table A.2: Patient population characteristics of the study cohort.

| Characteristics | % or Mean (Standard Deviation) |
|---|---|
| Patient Demographics at baseline | |
| Female (%) | 55.0 |
| Age (years) | 58.4 (7.0) |
| White race (%) | 94.7 |
| Smoking (%) | 10.7 |
| SBP (mmHg) | 138.9 (18.6) |
| DBP (mmHg) | 82.5 (10.1) |
| Height (cm) | 168.2 (9.2) |
| Weight (kg) | 77.9 (15.8) |
| Hypertension (%) | 30.5 |
| Diabetes (%) | 2.5 |
| Hyperlipidemia (%) | 15.7 |
| Heart failure (%) | 0.4 |
| Annual Income Level (%): | |
| 1. Less than £18,000 | 19.5 |
| 2. £18,000 to £30,999 | 22.1 |
| 3. £31,000 to £51,999 | 21.9 |
| 4. £52,000 to £100,000 | 16.9 |
| 5. Greater than £100,000 | 4.5 |
| Do not know or prefer not to answer | 15.1 |

## A.2 Calculation of Cardiovascular Risk Models

Each score is calculated by a sum of the product of the coefficient and value across the $n$ covariates for each patient:

$$Score = \sum_{k=1}^{n} \beta_k x_k \tag{1}$$

where $\beta$ is the coefficient for predictor covariate $x$. Then, using the baseline hazard and mean covariate estimates, we can compute 5-year estimated risk:

$$5\text{-}Year\,Estimated\,Risk = 1 - c^{\exp(Score - bias)} \tag{2}$$

where c is a constant, representing average AF-free survival probability at 5 years.

Table A.3: Multivariate model coefficients (row) for the three scores (columns): EHR-AF, CHARGE-AF, and $C_2$HEST. Each risk model comprises different combinations of predictor variables. - indicates that the corresponding variable (row) was not used to predict the corresponding score (column).

| Variables | EHR-AF | CHARGE-AF | $C_2$HEST |
|---|---|---|---|
| Male | 0.137 | - | - |
| Age, per 10-yr increase | 1.494 | 1.016 | - |
| Age $\geq$ 75 | - | - | 2 |
| Squared age, per 10-yr increase | -0.048 | - | - |
| White | -0.208 | 0.465 | - |
| Smoking | 0.152 | 0.359 | - |
| Height, per 10-cm increase | -0.231 | 0.248 | - |
| Squared height, per 10-cm increase | 0.012 | - | - |
| Weight, per 15-kg increase | -0.050 | 0.115 | - |
| Squared weight, per 15-kg increase | 0.021 | - | - |
| Systolic BP, per 20 mm Hg increase | - | 0.197 | - |
| Diastolic BP, per 10 mm Hg increase | - | -0.101 | - |
| Diastolic blood pressure $\geq$ 80 mm Hg | -0.104 | - | - |
| Diabetes | - | 0.237 | - |
| Myocardial infarction | - | 0.496 | - |
| Hypertension | 0.106 | - | 1 |
| Hyperlipidemia | -0.156 | - | - |
| Heart failure | 0.563 | 0.701 | 2 |
| Coronary heart disease | 0.210 | 0.349 | 1 |
| Valvular disease | 0.487 | - | - |
| Previous stroke/TIA | 0.132 | - | - |
| Peripheral artery disease | 0.126 | - | - |
| Pulmonary disease | - | - | 1 |
| Chronic kidney disease | 0.279 | - | - |
| Hypothyroidism | -0.138 | - | 1 |
| 5-year estimated risk variables: | | | |
| c | 0.971 | 0.972 | 0.975 |
| bias | 6.454 | 12.582 | 0.370 |

The predictor variable coefficients for the three risk score models are shown in Table A.3. To learn more about the model details, readers are advised to read the respective papers for EHR-AF [11], CHARGE-AF [2], and $C_2$HEST [17].