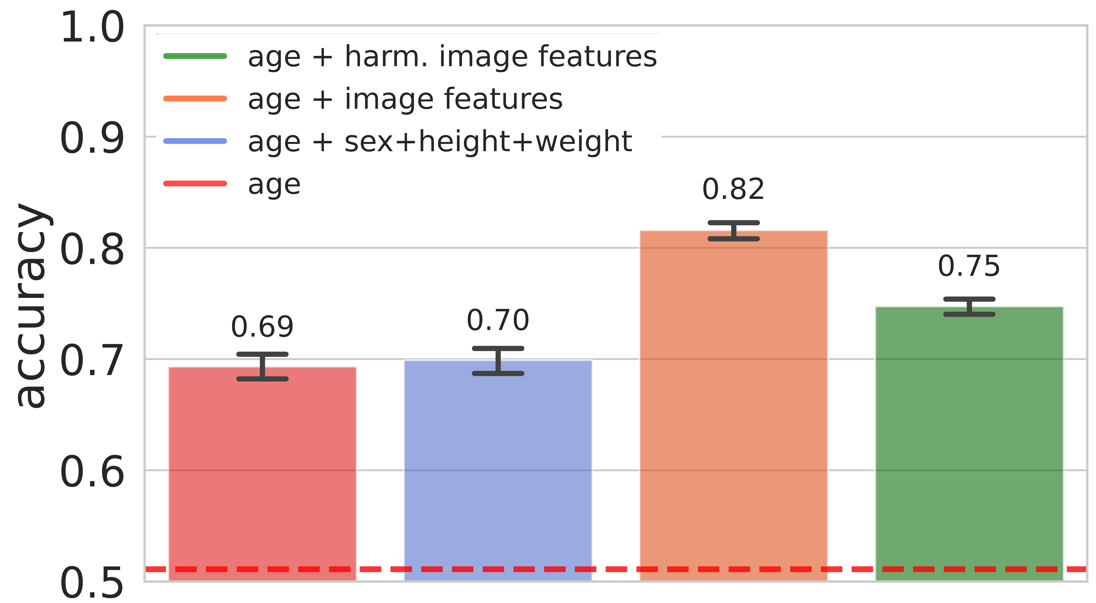
**Supplemental Material 1**

****

**Figure S1:** Accuracy for predicting the data source (UKBB vs. NAKO) using different feature combinations in addition to age as the main known source of variation. The red line represents the study proportion (random classification). The use of image features in addition to age increased prediction accuracy of data source identification (orange bar); this increase in accuracy was markedly reduced after feature harmonization across studies (green bar). The features sex, height and weight did not provide relevant information regarding the data source beyond the feature age.

**Methodology and Results:** To assess the success of data harmonization, we trained random forest classifiers models to predict the data source (UKBB vs. NAKO) from image-derived features. The core idea is that the possibility of predicting the data source from observed data implies the existence of source-dependent bias – successful data harmonization should reduce this bias and thus decrease data source classification accuracy.

We used random forest classifiers to predict the data source (UKBB vs. NAKO) from different combinations of image-derived features and epidemiological data (age, sex, height, weight). Knowing that age was an existing source of bias in this study, we assessed the predictive capacity of the remaining features beyond age information by always adding age as an input feature. All random forest models were trained with 100 trees and five-fold cross-validation.

As expected, the variables sex, height and weight, which were balanced across studies, had no relevant predictive capacity regarding the data source (Fig. S1). In contrast and not surprisingly, age alone allowed to predict the source of a data sample with 69% accuracy. When using extracted image features in addition to age, source classification accuracy further increased substantially to 82% pointing towards additional, age-independent source-related bias in image-derived features. Data harmonization resulted in a marked decrease in the predictive capacity of image features from 82 % using original image features to 75 % using harmonized image features as input.