

**Supplementary Material to accompany:**

**Derivation and assessment of risk prediction models using case-cohort data**

Jean Sanderson<sup>1</sup>, Simon G Thompson<sup>1</sup>, Ian R. White<sup>2</sup>, Thor Aspelund<sup>3,4</sup>, Lisa Pennells<sup>1</sup>

<sup>1</sup> Department of Public Health and Primary Care, University of Cambridge, UK

<sup>2</sup> MRC Biostatistics Unit, Cambridge, UK

<sup>3</sup> University of Iceland

<sup>4</sup> Icelandic Heart Association

### Supplementary Figure 1: Mean error in 10-year cumulative baseline survival

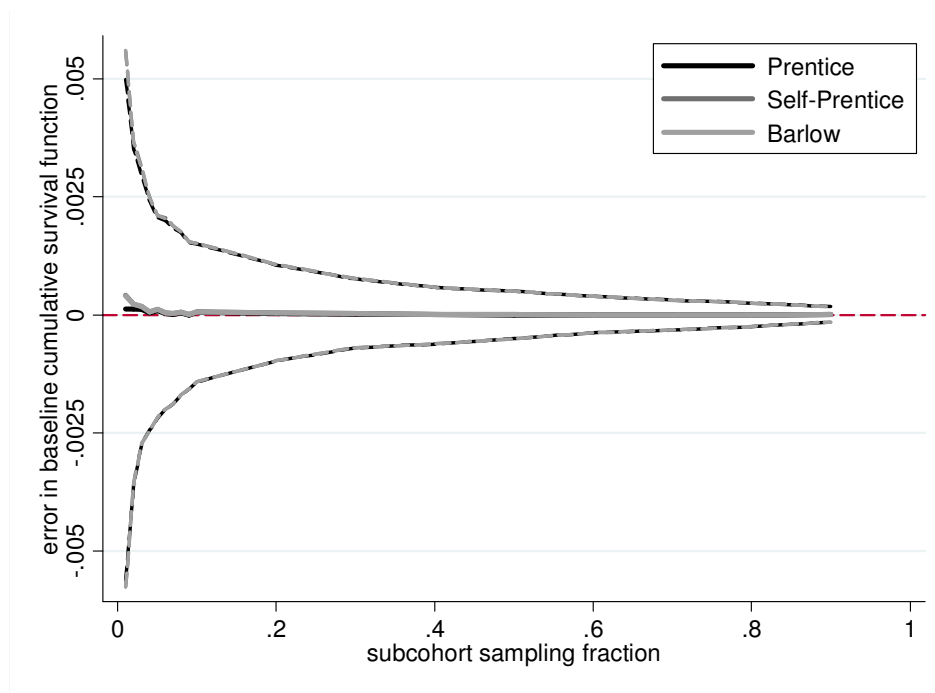


Figure shows the difference between cumulative baseline survival obtained using a prediction model derived on the case-cohort set, and that obtained using the model derived on the full cohort. Solid lines show mean error, and dashed lines show empirical upper and lower fifth percentiles, averaged over all individuals in the case-cohort set, and averaged over simulations, for three case-cohort weighting methods.

## Supplementary Figure 2: Error in log hazard ratios

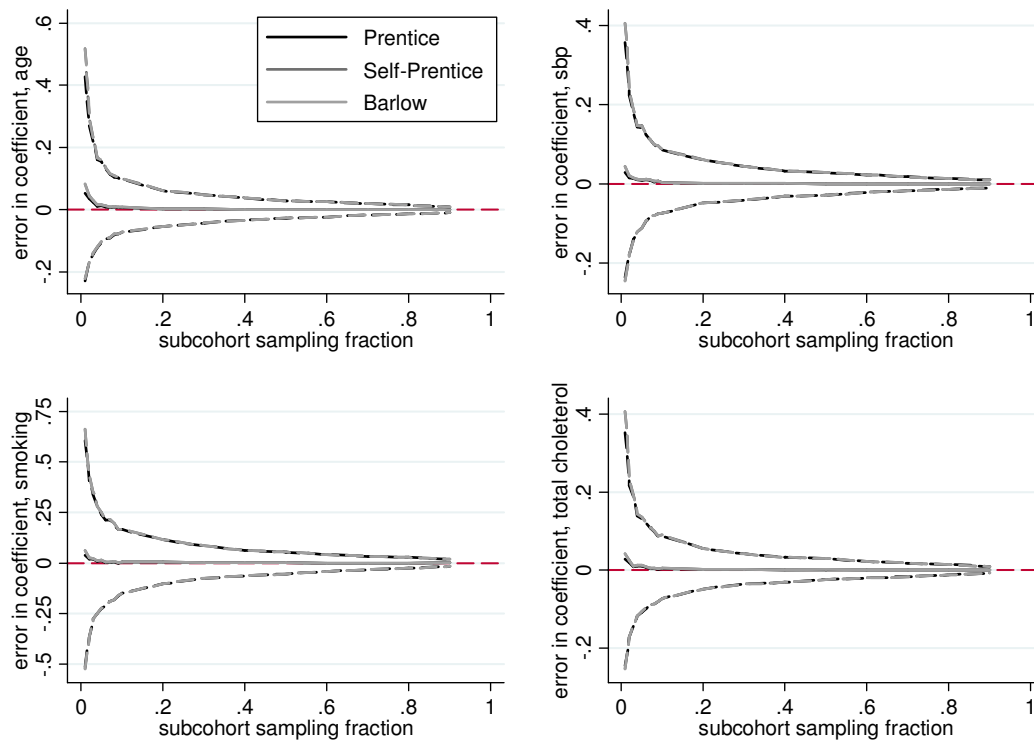
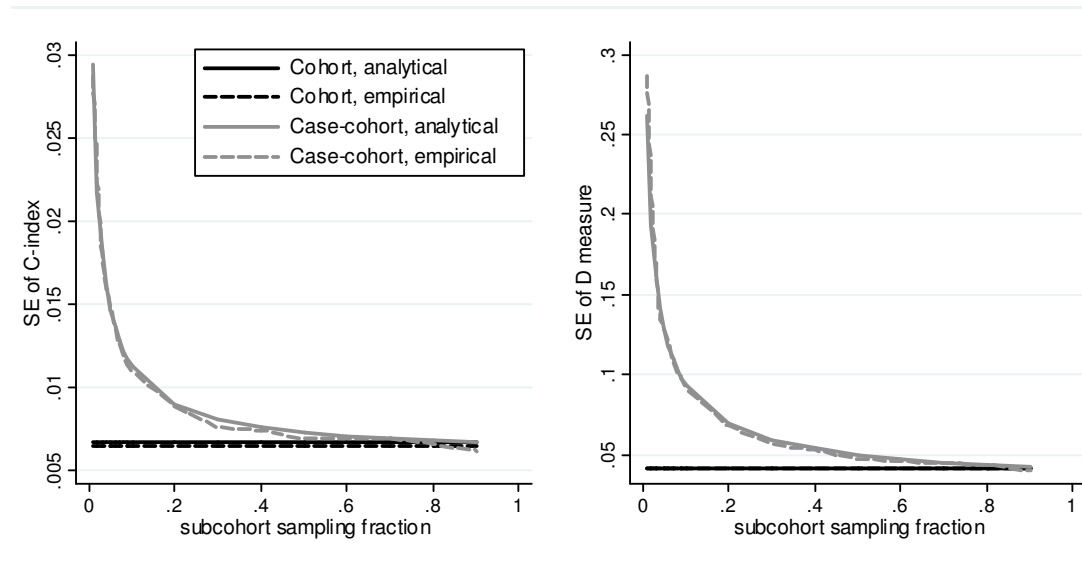
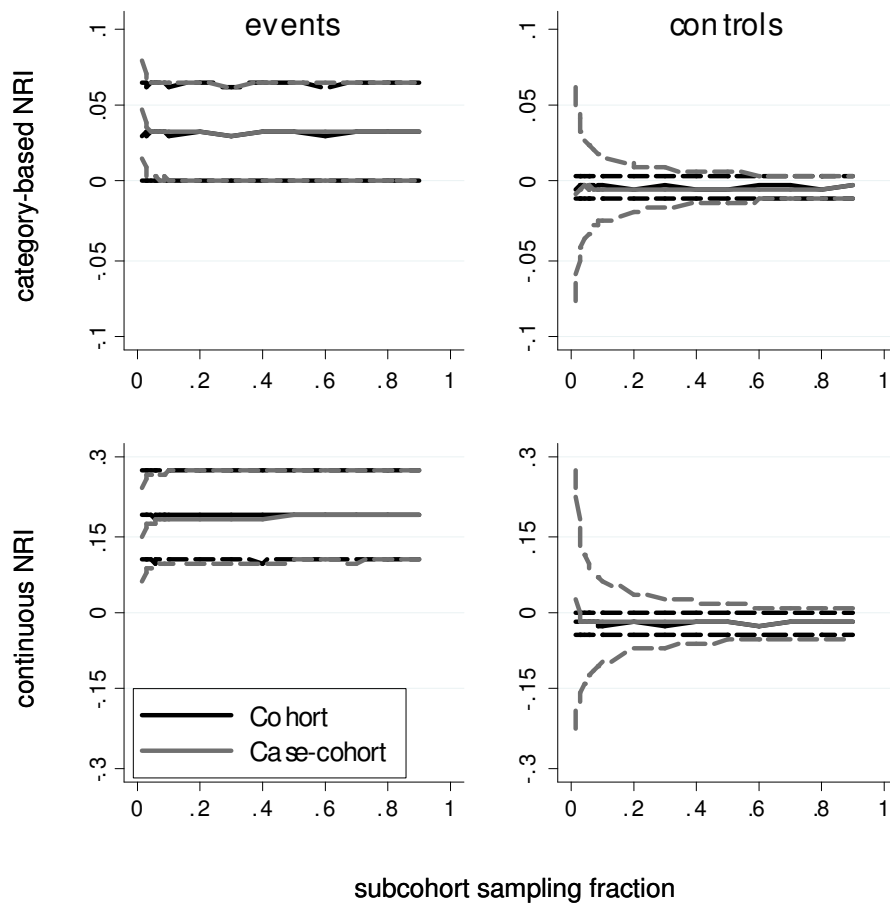


Figure shows the difference between log hazard ratios obtained using a model derived on the case-cohort set, and that obtained using the model derived on the full cohort. Solid lines show mean error, and dashed lines show empirical upper and lower fifth percentiles, averaged over all individuals in the case-cohort set, and averaged over simulations, for three case-cohort weighting methods.

**Supplementary Figure 3: Empirical and mean analytical SE for the C-index and D measure: full cohort and case-cohort estimates.**



**Supplementary Figure 4: Category-based and continuous NRI for events and non-events: full cohort and case-cohort estimates**



Dotted lines indicate empirical 95% CI

**Supplementary Figure 5: Empirical and mean analytical standard errors for category-based and continuous NRI: full cohort and case-cohort estimates.**

