**Supplementary Material**

eAppendix A – p.2

eAppendix B – p.3

**eAppendix A**

The matrices for the three studies with complete data (Kungsholmen, Skåne, Nordanstig) are shown below.

The dimensions of the above matrices are reduced by one in the fourth study (Blekinge).

**eAppendix B**

We provide details of the three steps SNAC with four studies and a discrete predictor, gait speed, with three levels systematically missing in one of the studies. Of note, the steps can be followed for any number of studies and any levels of the missing predictor.

**Step I: The imputation model**

Consider the discrete predictor , observed in 4-1 studies, with levels. In our example, gait speed is coded as 0, 1, and 2 corresponding to the levels ≤0.8, 0.8-1.2, and >1.2 meters per second, respectively.

Let denote the probability that the predictor is equal to level conditionally on a set of explanatory variables in the *i*-th study with . We consider a multinomial regression model using the lowest level of gait speed (≤0.8 m/s), coded as 0, as referent for contrasting two logits of conditional probabilities:

with

The linear predictor for the *i*-th study includes all the predictors of 5-years mortality (ADL, IADL, MMSE, number of comorbidities, age, and sex) and the 5-years mortality indicator itself. Similar to other imputation approaches such as FCS, the rationale of including several covariates is to increase the plausibility of random missingness conditionally on everything that has been observed.

The linear predictor for the *i*-th study contrasting the conditional logit of the probability of a gait speed 0.8-1.2 m/s relative to the conditional logit of the probability of gait speed ≤0.8 m/s is

Similarly, the linear predictor for the *i*-th study contrasting the conditional logit of the probability of a gait speed >1.2 m/s relative to the conditional logit of the probability of gait speed ≤0.8 m/s is

Estimates of the parameters in and are obtained using the maximum likelihood method separately for each study (*i* = 1, 2, 3) with complete information (Kungsholmen, Skåne, Nordanstig).

**Step II: Predict conditional probabilities**

All estimates in of are combined across the three studies with complete information (Kungsholmen, Skåne, Nordanstig) using the inverse-variance method

with weight equal to one divided by the estimated variance to obtain two average linear predictors and as follows

The predicted conditional probability of falling into the three levels of the systematic missing variable, gait speed, for any individual in the fourth study (Blekinge) is obtained by using the observed variables in such study

where the sum of the predicted probabilities for each individual in the fourth study is

**Step III: Impute values for the predictor with systematic missing**

Let denote the imputed variable representing gait speed, a predictor of 5-years mortality, that is systematically missing in the fourth study. The inverse cumulative distribution function method is used to assign a value to . The predicted cumulative distribution function, varying across individuals according to the observed variables in , says that

A value for is assigned by inverting a cumulative distribution function of a uniform distribution over the interval 0 to 1. The imputed value is the quantile, , corresponding to the predicted cumulative conditional probabilities estimated in the previous step where is randomly drawn from a