Supplementary file
**Evaluation of potential misclassification bias**

To determine whether the observed difference in survival between the earlier and later incident RA cohorts could be due to misclassification bias, whereby more prevalent cases of RA would be misclassified as incident RA cases in the earlier cohort due to shorter lead-in time, we have estimated how much misclassification of prevalent cases as incident cases of RA in the earlier cohort would be necessary to eliminate the difference in excess mortality (relative to population controls) observed between the earlier and later cohorts. To estimate this, we have varied the assumptions regarding (1) the hypothetical frequency of prevalent cases in the earlier cohort from 0% to 35% (using 5% increments). We have kept the frequency of prevalent cases in the later cohort at 0% (because adding prevalent cases in the later cohort would lead to a protective HR for death in RA relative to controls). In these calculations, we also varied the assumptions with regards to: (2) the duration of RA in the prevalent cases that were misclassified from 6 years (chosen because of the 6 year lead-in time) to 12 years; and (3) the HR representing the risk of death per year of RA duration from 1.05 to 1.20. The results of the analyses for the resulting 64 scenarios demonstrated that in order to eliminate the improvement in mortality between the earlier and later cohorts, highly unrealistic assumptions would be necessary regarding at least one (and often two) of the aforementioned parameters.

For example, when assuming an average RA duration of 6 years (among prevalent cases), a frequency of 35% prevalent cases in the earlier cohort are needed to eliminate the improvement in mortality assuming a HR of 1.20 per yr of RA duration; and the mortality improvement is not eliminated, even with a prevalence of 35% when the HR is < 1.20. When assuming an RA duration in prevalent cases of 12 years, 35 % prevalent cases are needed if the HR is 1.10 per year of RA duration; and 25% prevalent cases are needed if the HR is 1.15 per year of RA duration; and finally 20% prevalent cases are needed if the HR is 1.20 per year of RA duration.
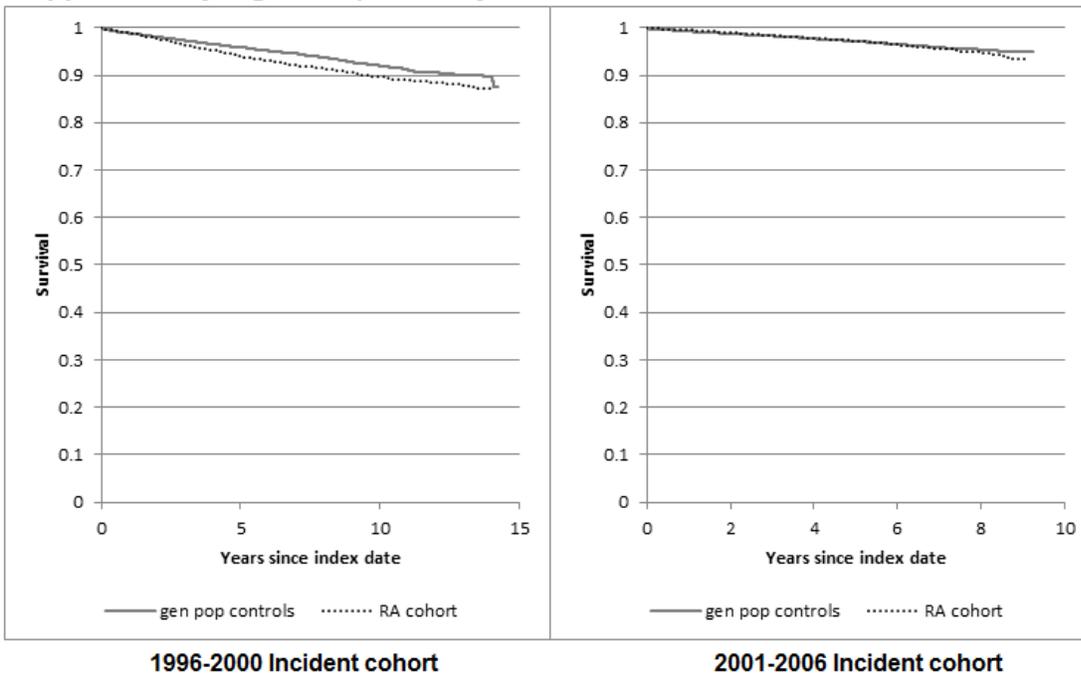
All of the above combinations of assumptions are highly improbable. It seems entirely unrealistic for 20% to 35% of RA patients to never have a medical visit coded as RA over a 6 year period, given the frequency distribution of RA visits in our cohort. Mortality HR in the range of 1.10 to 1.20 per year of RA duration are also unrealistic, as it would translate into a 50% - 100% increase in risk of mortality after five years of RA duration (among subjects of same age, sex and comorbidities). Therefore, we conclude, that misclassification cannot account for all of the observed difference in survival.

Supplementary Figure 1. Survival from all-causes, cardiovascular diseases, and cancer in RA and general population controls, calculated without censoring follow-up at 5 years
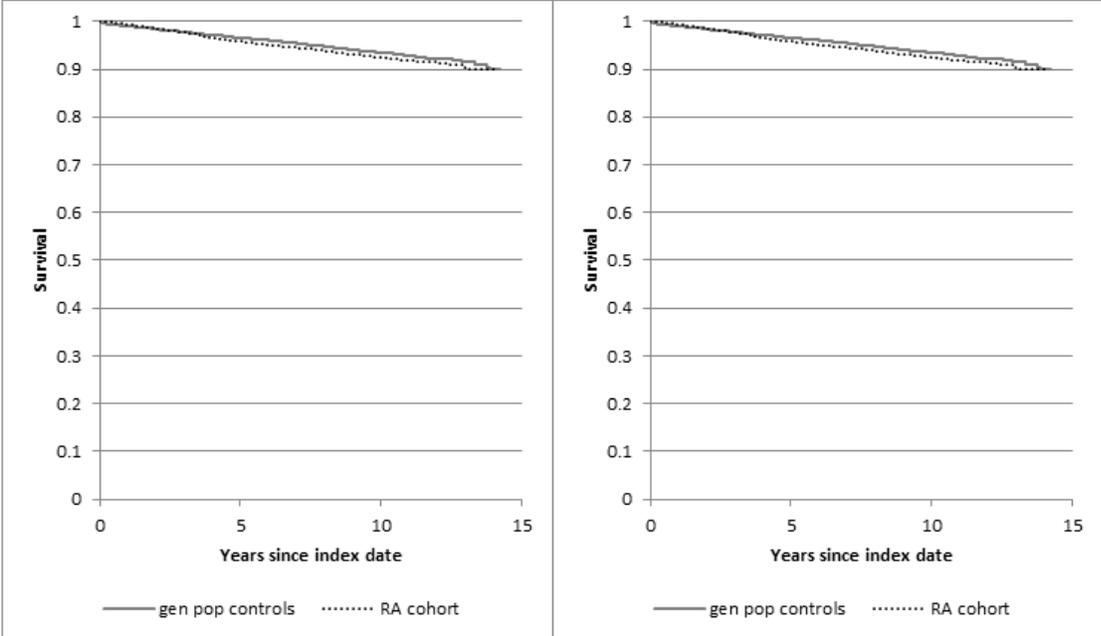
Supplementary Figure 1A) All-cause mortality

1996-2000 Incident cohort

2001-2006 Incident cohort
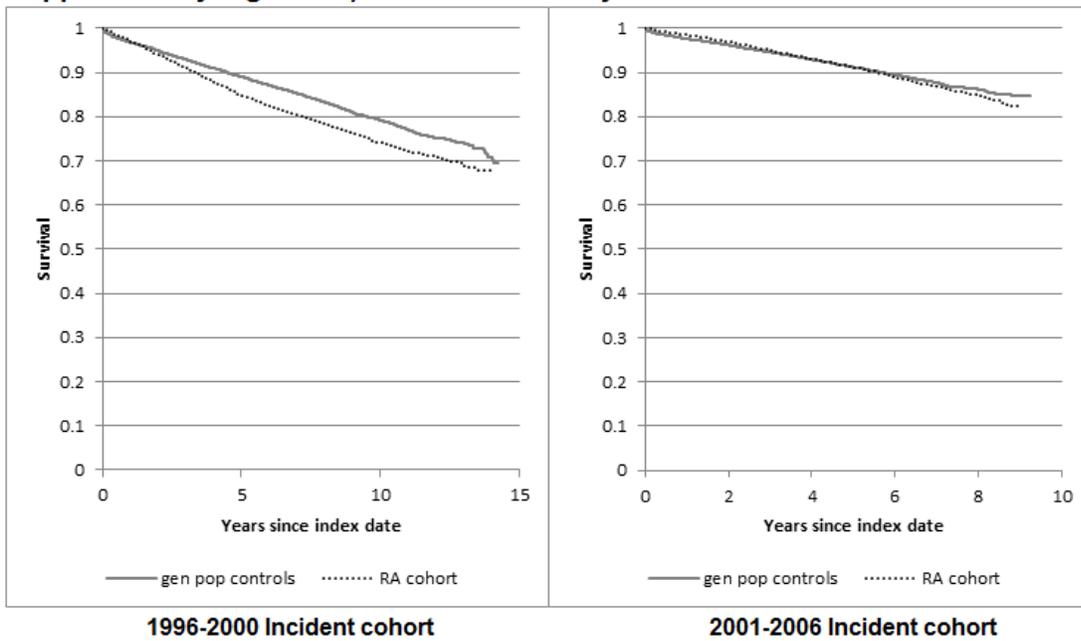
Supplementary Figure 1B) Mortality from cardiovascular diseases

1996-2000 Incident cohort

2001-2006 Incident cohort

Supplementary Figure 1C) Mortality from cancer

1996-2000 Incident cohort

2001-2006 Incident cohort