November Meeting

Last updated: 07 November, 2019

# CPRD Additional Analyses

## Opening notes

* All analyses used a complete case analysis.
* Effective sample size (number of participants with complete covariate information) is 773387, compared to the 1711679 participants who were eligible for the study.
* Covariates adjusted for in the full model: *age, sex, coronary artery disease, coronary bypass surgery, cardiovascular disease, Carlson index, IMD20-10, consultation rate, alcohol (ever), smoking (ever), BMI, peripheral arterial disease, & hypertension (+/- baseline total cholesterol)*.

## Adjusting for baseline cholesterol

**Results**

The original time-varying treatment analysis found an overall increased HR, driven mainly by the the large effect seen in the vascular dementia subgroup.

It was previously suggested that this was due to confounding by indication, where a higher baseline total cholesterol/LDL-c test result would result in someone being more likely to be prescribed statins and also more likely to develop vascular dementia.

In an attempt to deal with this, I adjusted each previous analysis to take account of baseline total cholesterol test results (analyses presented in grey on the forest plots below). The black lines are the results of analysis adjusted for all covariates listed above except for baseline total cholesterol. There was no noticeable effect on the probable Alzheimer’s results, but adjusting for baseline total cholesterol appears to increase the HR for vascular dementia (though the increase is consistent with chance).

See Section 1.3 for a possible explanation for these findings.

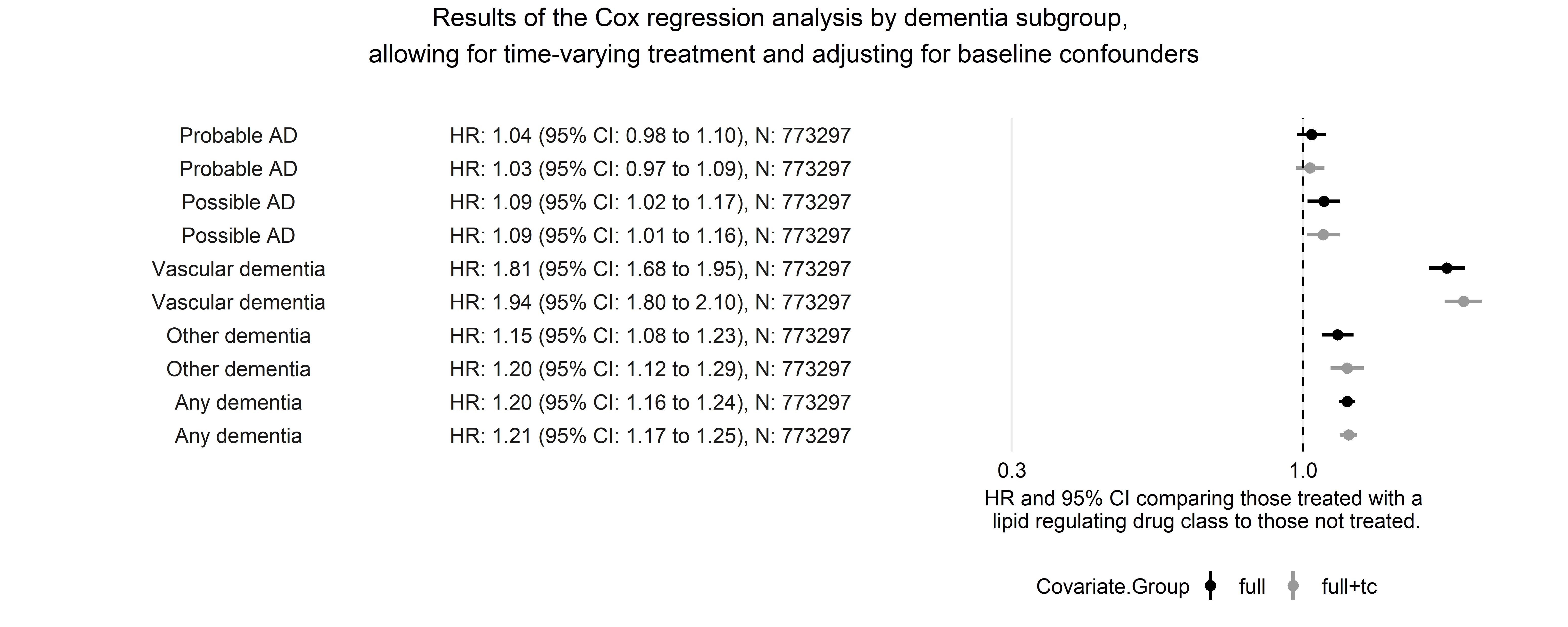


Figure 1: Dementia subgroup results, with and without adjustment for baseline total cholesterol

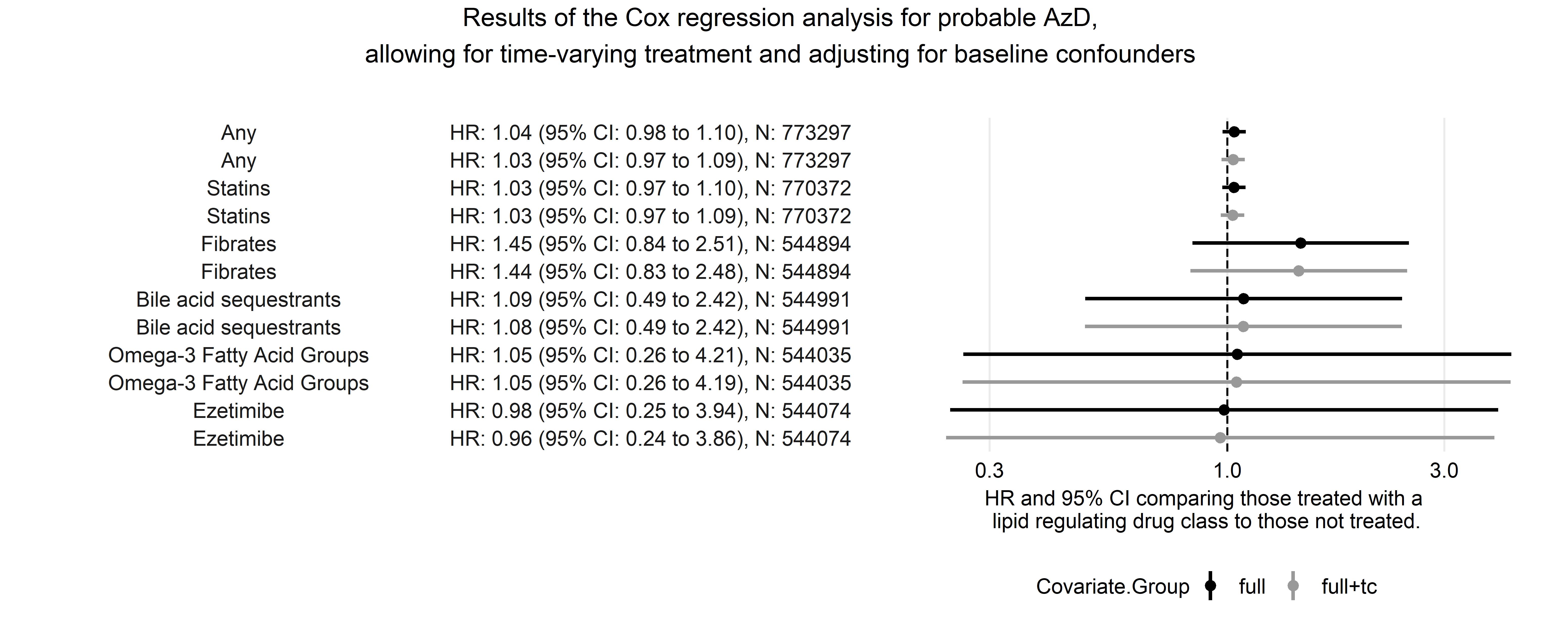


Figure 2: Probable Alzheimer’s disease results by drug class, with and without adjustment for baseline total cholesterol

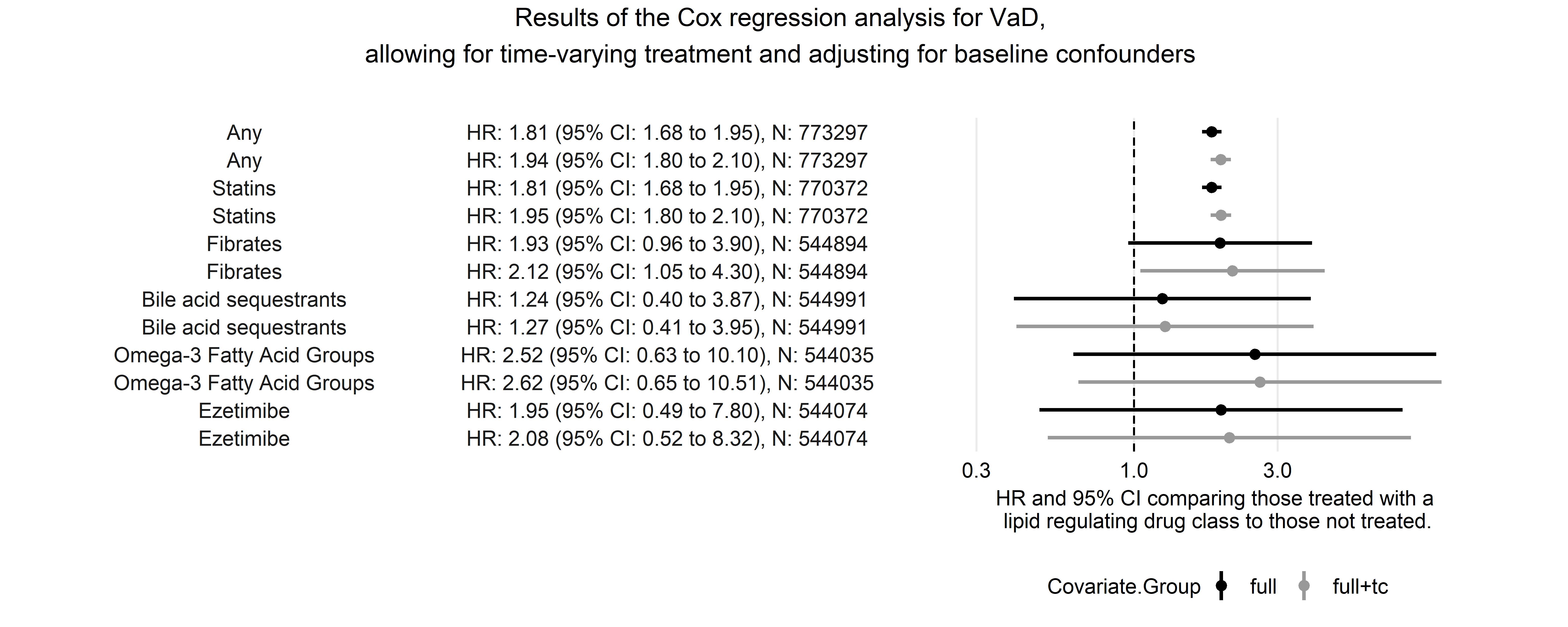


Figure 3: Vascular dementia results by drug class, with and without adjustment for baseline total cholesterol

**Questions**

* Is there any other important confounder that we may be missing, that could account for the increased HR, particular in the vascular dementia group?

## Regression discontinuity analysis

**Comments**

It was previously suggested that I perform a regression discontinuity analysis. In preparation for this, I spoke to Lauren Scott of the CLARHC who has performed a similar analysis in the CPRD using the QRISK2 score instead of a lipid test results. This analysis:

* found that QRISK2 is a poor instrument for statin prescription. Below NICE recommended cut-off, the probability of prescription of statin within 60 days was ~10%, and above the cutoff, the probability increased to only ~20% , even at quite high values of QRISK2 (Figure 4, below).
* had intended to look at side effects of statin use, but was unable to even demonstrate that statins lowered LDL-c levels 3-6 months after the date the QRISK2 score was recorded.

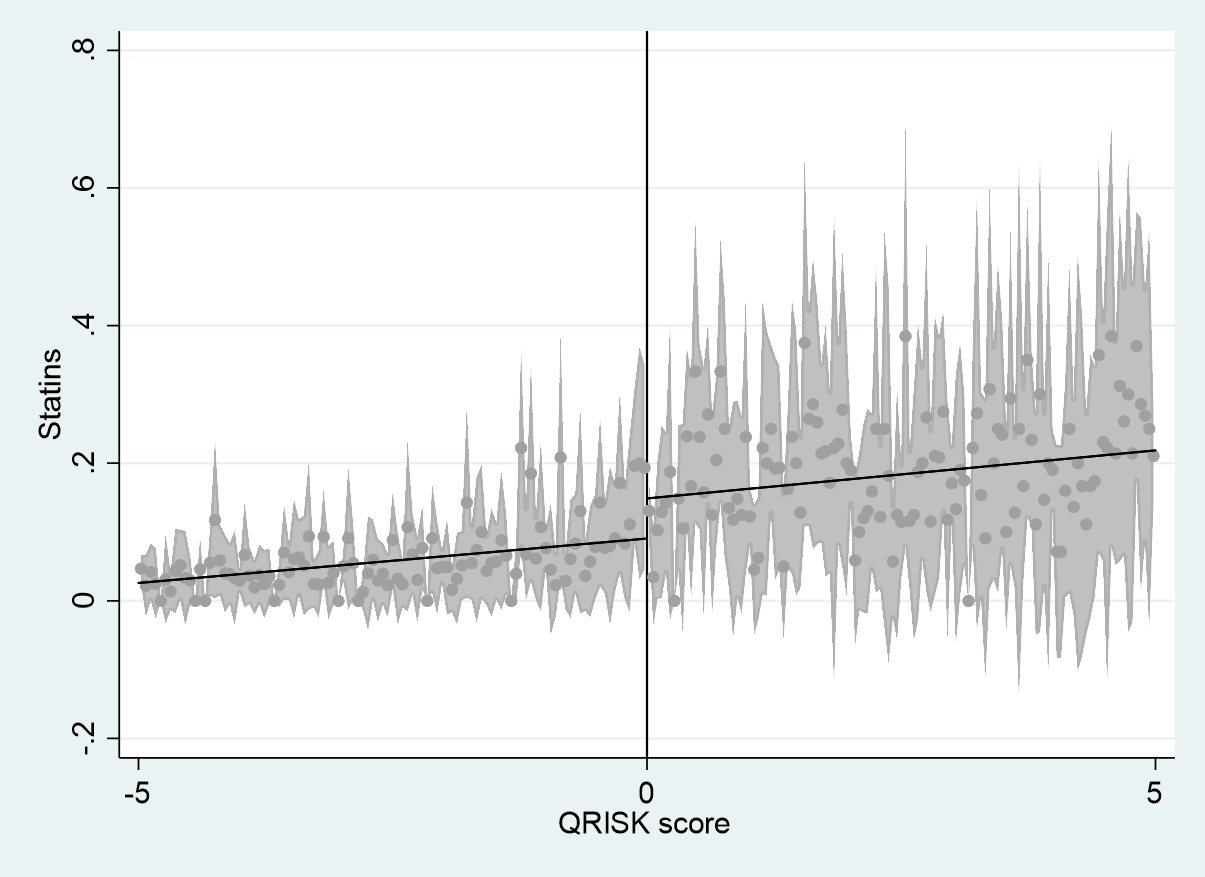


Figure 4: Results of a regression discontinuity analysis using QRISK2 score, showing a small discontinuity in statin prescription above and below NICE recommended cut-off. Graphic courtesy of Lauren Scott/Kate Tilling.

Based on the above analysis showing that even the QRISK2 score is a poor indicator of statin prescription, and assuming that it is still a better indicator than total cholesterol alone, I do not think it is worthwhile running a regression discontinuity analysis using total cholesterol.

These findings may also explain the lack of attenuation when including baseline total cholesterol in the models above in an effort to address confounding by indication, as it appears total cholesterol does not substantially influence whether someone is prescribed statins or not.

**Questions**

* Lauren has asked if we can replicate her analysis in our data-set, just to ensure that the findings were not a coding error on her part - I am assuming that this is okay for me to do?

## Results by age group at baseline

**Results**

At Yoav’s suggestion, I ran the fully adjusted models (including baseline cholesterol) in three age subgroups.

I had originally intended to break participants up into four categories: <65,65-74,75-84,>=85. However, the numbers were extremely low in the oldest age category, and so I split them into three groups instead: <65,65-74,>=75.

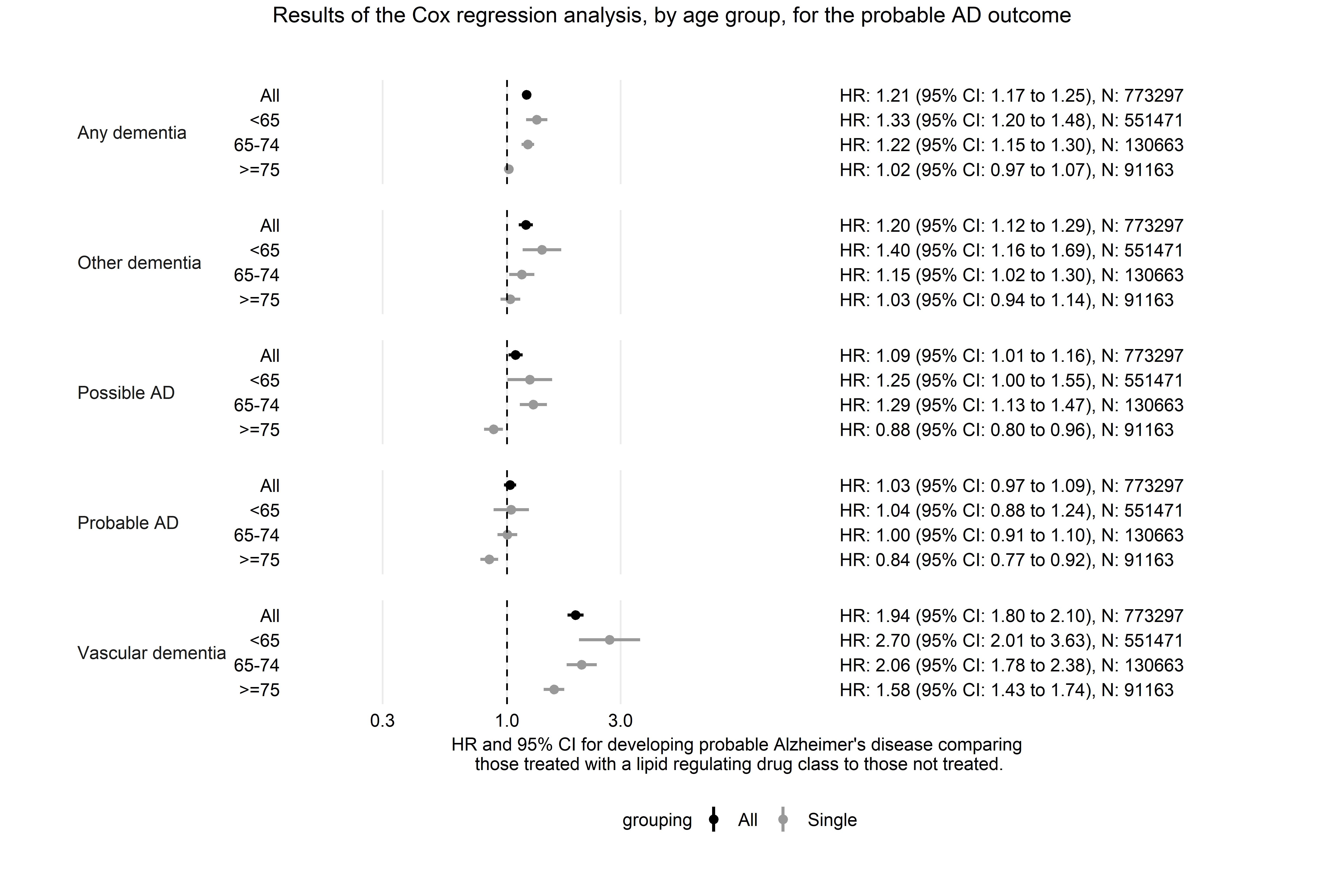


Figure 5: Effect of any lipid regulating agent on any dementia, by age group

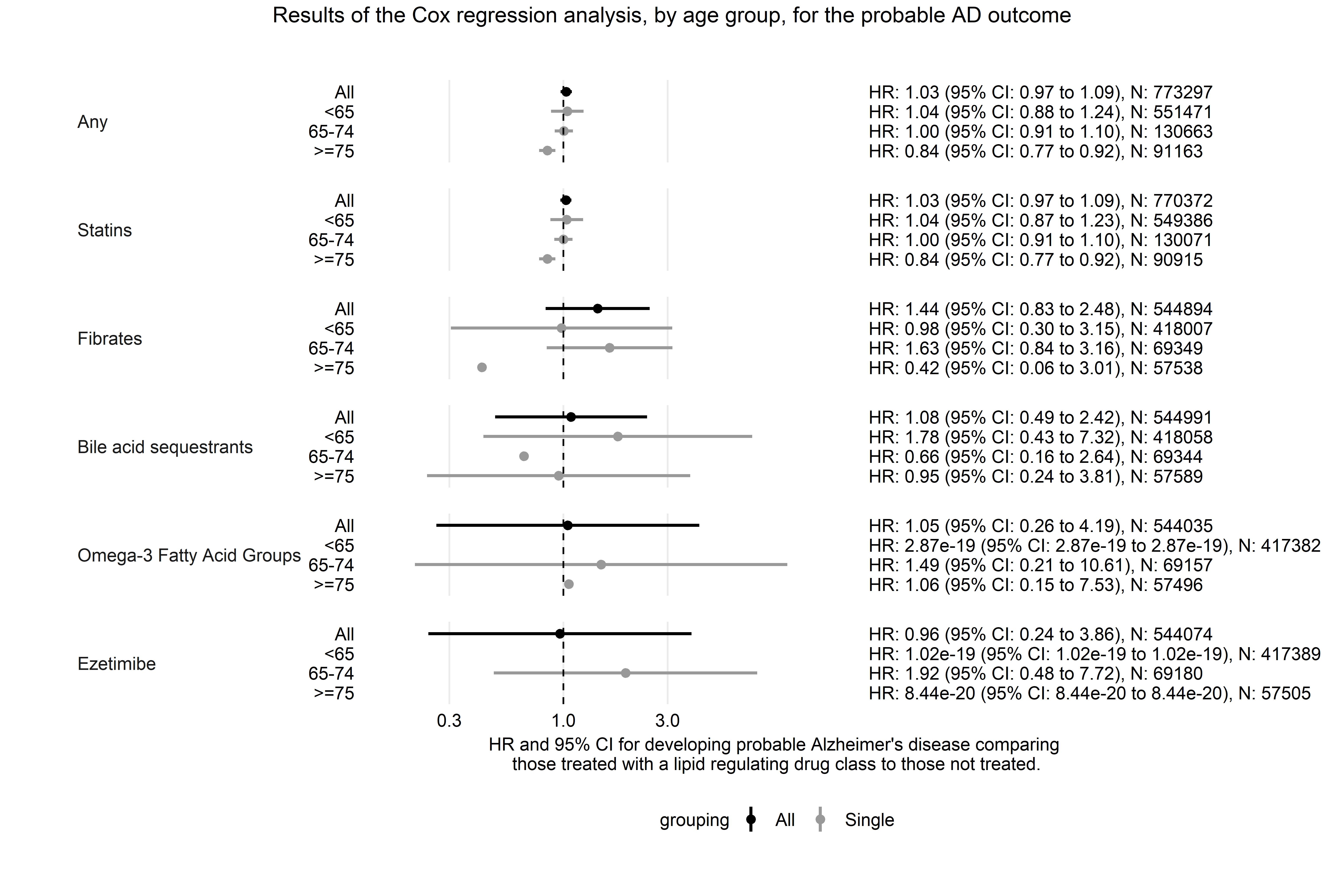


Figure 6: Effect of different drug classes on Probable AD, by age group. Note, some of the analyses give funny results due to the small number of both participants and events in that category.

It appears that the increased HRs seen are being driven by the youngest age groups, and that the effect is attenuated, or even qualitatively different (e.g. in the Possible AD subgroup) between the youngest and oldest categories.

**Questions**

* How to interpret these findings?

## Covariate selection

**Results**

From a series of post-estimation analyses, it appears that adjusting for many of the covariates included in the final model is unnecessary. The largest change in the effect comes from adjusting for age at baseline, following which there is very little further movement of the point estimate (an example is provided using the model for the effect of any drug class on any dementia outcome, Figure 7).

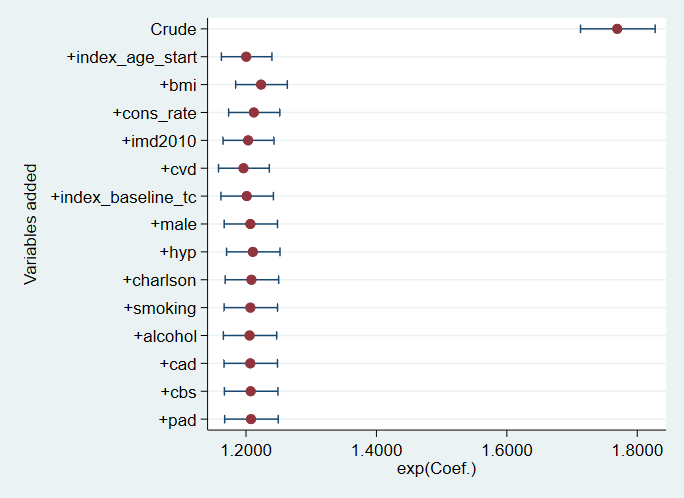


Figure 7: Postestimation analysis of a full adjusted model (including baseline total cholesterol) of the effect of any drug on any dementia diagnosis.

**Questions**

* Including all of the covariates significantly reduces the effective sample size of the complete case analysis from 773387 to 1711679 participants. Would it be better to exclude some of these covariates and therefore increase our effective sample size?

## Differential misclassification of Alzheimer’s disease vs vascular dementia

**Comments**

It seems likely that there is misclassification between Alzheimer’s disease and vascular dementia, as many of the methods to differentiate between the two conditions ask about cardiovascular history, including previous statin use.

To me, this means that it is likely that the vascular dementia results are overestimated, but also that the probable Alzheimer’s disease are underestimated (if you are on statins and you have a memory complaint, you are more likely to be classified as vascular dementia than Alzheimer’s). This could cause an issue as we had intended to make the Alzheimer’s disease results the core of the manuscript.

**Questions**

* Is this a reasonable interpretation?
* Is there any way to examine/tackle this potential misclassification in the analysis?

## Selection bias in the inclusion criteria

**Comments**

This was suggested by Yoav and Julian in my most recent supervisory meeting. It is possible that some selection bias was introduced in the way people are selected into the study if their index event is prescription of a drug. In this case, the analysis will not include those patients that same pre-selection testing/consultations but were not prescribed a drug.

This bias is not an issue in the current complete case analysis, as those with an index event of prescription of a drug are excluded as they do not have a baseline total cholesterol measurement.

However, this potential for bias will become important if we decide not to include baseline total cholesterol as a covariate (see Section 1.5, above).

**Questions**

* Is this a reasonable interpretation?
* Is there any way around this if we do decide to include those with a drug prescription as their index event?