Replicating Results from Neale Lab BMI GWAS

6/15/2022

<https://github.com/Nealelab/UK_Biobank_GWAS#imputed-v3-association-model>

<http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas>

<https://pan.ukbb.broadinstitute.org/downloads>

<https://pan.ukbb.broadinstitute.org/docs/per-phenotype-files/index.html>

<https://docs.google.com/spreadsheets/d/1AeeADtT0U1AukliiNyiVzVRdLYPkTbruQSk38DeutU8/edit#gid=268241601>

Line 4110 in Pan-UK Biobank phenotype manifest (first try)

Betas correlate VERY closely – GREAT!

BUT the coefficients are pretty different… Is this just based on what we control for???

Trying again using JUST sex and PCA to control:

Okay pretty much identical… WHY are my estimates just like 3 times higher???

Almost EXACTLY 3 times higher… (Don’t think this would really affect anything.. But why is it like that?)

Okay… Cut it into a third the size and see if this actually affects things…

MAIN TROUBLE is I don’t know WHY it’s inflated by factor of 3… Or whether this also affects second stage regressions..

SO reducing BOTH by a third doesn’t affect things at all…. BUT ONLY REDUCING Bx and Bxse and NOT By and Byse by a third results in a far larger effect estimate

LOOKING INTO HOW THEY ALL CODED BMI… If it’s different but scaled, could explain why correlation is nearly perfect but 3 times larger…

OH SHIT!! THIS IS BECAUSE THEY BOTH \*DO\* STANDARDIZE THIS BMI MEASURE BEFORE RUNNING IT!!!!

SO IT’S REALLY THE SAME THING BUT SCALED DOWN BY SD, WHICH EXPLAINS THE DISCREPANCY!

NOW it all works! Only 96 because one variant didn’t copy over… Doesn’t much matter

VERY close to each other and highly correlated! So we’re good!

Saved as “BetaComparisonsFileNealeLockeUs.csv”