Mtg w/ Patrick

Dempsey is “Physical activity volume, intensity, and incident cardiovascular disease” in the *European Heart Journal*

Good news:

**Sample of Results from Replicating Dempsey after fixing restricted cubic spline specification:**

|  |  |  |
| --- | --- | --- |
| **Physical Activity Energy Expenditure**  **Physical Activity Risk (Lower Risk = MORE Physical Activity)** | | |
| Physical Activity Energy Expenditure | *Adjusted HR (95% CI)* | *Dempsey Adjusted HR (95% CI)* |
| *15* | 1 (Reference) | 1 (Reference) |
| *20* | 0.91 (95% CI: 0.87, 0.95) | 0.88 (0.80-0.96) |
| *30* | 0.76 (95% CI: 0.67-0.87) | 0.73 (0.60-0.88) |
| *40* | 0.64 (95% CI: 0.54-0.77) | 0.69 (0.58-0.82) |
| *50* | 0.56 (95% CI: 0.47-0.67) | 0.64 (0.53-0.76) |
| *60* | 0.49 (95% CI: 0.40-0.61) | 0.60 (0.49-0.73) |

The Dempsey study on which much of this is modeled generally had stronger associations than we found. This makes sense given that we use coronary artery disease (a subset of CVD) as the outcome and they use cardiovascular disease as the outcome. Additionally, our sample is only about 75k individuals compared to 96k in Dempsey because we had to screen for individuals with high quality genetic data. Nevertheless, it does not appear that large statistical power differences exist (in part because of the similarly narrow confidence intervals). NOT A REPLICATION BUT NOT SUPPOSED TO BE.

**Thoughts & Questions:**

1. General Process:

Basically, I fit the function you shared in that Table and then do a Cox regression with PA and rqs.1 as the key exposure variables. I then figure out what rqs.1 is when PAEE = X and insert this value for rqs.1 in the delta.hrt function (baseline is before the first knot).

1. What does ‘3 evenly spaced knots’ mean in the context of a restricted cubic spline regression? And what really makes the most sense to do here?

My take is that it would mean having a spline with knots at the 10th/50th/90th, which would include the Kth knot and the K-1th knot, which is what I did. However, while Dempsey never mention the location of knots, they produce the following figure:

[LINK TO FIGURE](https://academic.oup.com/view-large/figure/385319909/ehac613f1.tif)

This suggests that there are knots at 25th/50th/75th and 95th and 10th as well. I guess I could see this logic too (that it would produce three variables compared to the single variable produced when using just 3 knots BEFORE restricting)…

1. After testing the proportionality assumption with cox.zph() in R, I found several variables in violation. I simply stratified if the covariate was discrete and interacted with time for continuous variables.
   1. HOWEVER, physical activity ALSO violates the PH assumption. What’s odd about this is that I would suspect that it would ALWAYS given the age range we’re focused on (50-mid 70s) and yet no one mentions correcting exposure for this.
   2. I have come across a paper in the literature “Why Test for Proportional Hazards?” that suggested that the pooled estimate should be treated as a WEIGHTED AVERAGE across time.
   3. If we DO want to get age-specific hazards, is there some way (a la Mantel Haenszel estimator) to POOL the results? Or should we keep it age-specific?
2. Worth experimenting with percentile cutoffs? For now in our study, I’m using quintiles of genetic and PA risk for reference. I fear this could be too narrow/being done slightlly differently here (modeled continuously, so we’re looking at it at the THRESHOLD of these quintiles - right idea?). Considering also tertiles to match a lot of the literature that does NOT focus on modeling exposures continuously.
3. Last big step in our analyses is to look at the effect of PA intensity on CAD risk HOLDING PHYSICAL ACTIVITY VOLUME CONSTANT
   1. Main issue here - Intensity and volume are naturally HIGHLY COLLINEAR (correlation = 0.69)
   2. Oddly enough, the VIF including PA and %MVPA was relatively minor… While it was higher than for other variables, ALWAYS below 5.. Which suggests it isn’t a massive issue
   3. How have others handled this question?
      1. Log and standardize PA in regression w/ MVPA and interact them (Strain et al)
      2. Interact the “four orthogonal spline variables” in a regression including PA and % MVPA (no transformations as far as I can tell) (Dempsey et al.)
      3. Compositional data analysis - Most rigorous and newest. One behavior can only increase at the expense of some other PA behaviors. Effective but a new rabbithole (Walmsley et al.)
   4. WEIRDNESS IN DEMPSEY RESULTS: (Simpson Paradox?)
      1. In **Table 3** of the paper, % MVPA has a consistent SIGNIFICANT effect at different PA levels on CVD risk.
      2. HOWEVER, in **Table S4** of the supplement, the authors perform same analysis sex stratified and get largely null results. Is this simply a statistical power issue? The effect sizes also look diminished.
      3. Why do I care? **BECAUSE MY ANALYSES ARE YIELDING RESULTS MORE IN LINE WITH TABLE S4**