Online Supplement
Biomarkers of Inflammation and Longitudinal Evaluation of Lung Function, Physical Activity, and Grip Strength: A Secondary Analysis in CASCADE

David M. MacDonald, MD, MS[1,2] Sarah Samorodnitsky,[2] Eric F. Lock,[2] Vincent Fan,[3] Zijing Chen,[4] Huong Q. Nguyen,[5] Chris H. Wendt[1,2]

[1]Minneapolis VA Health Care System, Minneapolis, Minnesota, United States

[2]University of Minnesota, Minneapolis, Minnesota, United States

[3]VA Puget Sound Health Care System, Seattle, Washington, United States

[4]Tsinghua University, Beijing, China

[5]Kaiser Permanente Southern California, Pasadena, California, United States

Full description of Methods

Data Cleaning: The CASCADE data contained measurements of 14 biomarkers, IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-13, IFN, GMCSF, TNF, and CRP, measured on 302 individuals at baseline, year 1, and year 2. Biomarkers at each timepoint were transformed using a log(1+x) transformation. After transformation, all biomarkers were centered and scaled to mean 0, standard deviation 1.

Statistical Analysis: We sought to identify biomarkers at baseline associated with a change in seven outcomes: forced expiratory volume in 1-second (FEV_1), total distance walked, average step total, average minutes active, grip strength, and percentage of time spent inactive. We collectively refer to total distance walked, average step total, average minutes active, grip strength, and percentage of time spent inactive as our *activity outcomes.*

We measured change as both a raw difference between baseline and year 2, i.e.,

$$FEV1_{year\ 2} - FEV1_{baseline},$$

and as the proportional change from baseline to year 2, i.e.,

$$\left(FEV1_{year\ 2} - FEV1_{baseline}\right)/FEV1_{baseline}.$$

We calculated the Pearson correlation between each baseline biomarker and change in each outcome (raw and percent change) and tested the null hypothesis that this correlation is equal to 0. To control the overall false discovery rate (OFDR) (Benjamini & Heller, 2008) across the biomarkers, we used a hierarchical hypothesis testing procedure to test the significance of the correlation between each biomarker and each outcome, as described in Li & Ghosh (2014). We treated each biomarker as a "set" of hypotheses in which 14 hypotheses (raw and percent change in seven outcomes) were tested. By controlling the OFDR, we control the expected number of biomarker sets falsely rejected. The steps to our hierarchical hypothesis testing procedure were as follows:

1. Treat each biomarker as a group or set of null hypotheses. These null hypotheses are that a given biomarker is not correlated with raw change or percentage change in the

seven outcomes. We index each of the $q = 14$ hypotheses within a set using $j, j = 1, \dots, 14$.

Using Procedure 4 defined in (Li & Ghosh, 2014), first test the screening hypothesis for each biomarker, which is whether each biomarker is significantly associated with any outcome. To do this test, calculate the p-value for the association between biomarker $i, i = 1, \dots, 14$ with each of the outcomes. Then obtain the screening p-values, $p_{(0)}(i)$, by calculating $\min_{1 \le j \le q} \left((q + 1 - j) p_{(j)}(i) \right)$ where $p_{(j)}(i)$ is the p-value corresponding to hypothesis *j* for biomarker set *i.*

2. Apply the Benjamini-Hochberg FDR correction (Benjamini & Hochberg, 1995) to the screening p-values, $p_{(0)}(1), \dots, p_{(0)}(14)$. Let R be the number of rejected screening hypotheses at the 0.05 level.
3. For biomarker i, define $R_i = \max\{ 1 \le j \le q : p_{(j)}(i) \le R\alpha \left(m(q + 1 - j) \right)^{-1} \}$ or $R_i = 0$ if the maximum does not exist.
4. For every i and j such that $p_{(j)}(i) \le p_{(R_i)}(i)$, reject the corresponding null hypothesis for outcome *j* in biomarker set *i.*

We report if we can reject any hypotheses within a biomarker set and if so, which hypotheses within that set that were rejected.

We also tested if baseline biomarkers can predict if an individual experienced a rapid decline in FEV_1 from baseline to year 2 using Fisher's linear discriminant analysis (LDA). We defined rapid decline in FEV_1 as an average drop of 40 milliliters in FEV_1 from baseline to year 1 and year 1 to year 2. We used 10-fold cross validation where we trained the LDA model on 9 out of 10 training folds and predicted rapid decline in FEV_1 on the held-out test fold. We used the area under the receiver operating curve (AUROC) to evaluate the predictive performance. We used a permutation testing framework to assess the significance of this AUROC in which we permuted whether each individual experienced a rapid decline in FEV_1 across the sample. We then fit the LDA model to predict rapid decline in FEV_1 on the permuted samples using 10-fold cross validation. We repeated this 100 times and calculated the permutation p-value as:

$$Permutation\ P - Value\ = \frac{\left(\sum_{i=1}^{100} 1(AUC_i^{perm} > AUC^{true}) \right) + 1}{101}$$

Samples were required to have at least two consecutive FEV1 measurements (baseline and year 1 and/or year 1 and year 2) to be included in this analysis. Samples missing any baseline biomarker measurements were not included.

Lastly, we used a linear mixed modeling analysis to investigate the relationship between baseline biomarkers over time and eight outcomes: FEV_1, FEV_1-percent-predicted (FEV_1pp), total distance walked, average step total, average minutes active, grip strength, and percentage of time spent inactive. We considered baseline biomarker, time, and smoking status at each year of follow-up as fixed effects. We also adjusted for FEV_1 in the models for total distance walked, average step total, average minutes active, grip strength, and percentage of time spent inactive. We included an interaction between baseline biomarker and time to capture any change over time of the baseline biomarker's effect on the outcome. We included a random intercept, b_{0i}, for each subject i to account for subject-specific variation in the observations. The model for FEV1 and FEV1pp at time $t, t = baseline, year\ 1, year\ 2$ was as follows:

$$y_{it} = \beta_0 + \beta_1 biomarker_{i0} + \beta_2 1(t = year\ 1) + \beta_3 1(t = year\ 2) + \beta_4 1(t = year\ 1) * biomarker_{i0} + \beta_5 1(t = year\ 2) * biomarker_{i0} + \beta_6 smoking\ status_{it} + b_{i0} + \epsilon_{it}$$

where $1(\cdot)$ represents an indicator function.

The model for our activity outcomes at time $t, t = baseline, year\ 1, year\ 2$ was as follows:

$$y_{it} = \beta_0 + \beta_1 biomarker_{i0} + \beta_2 1(t = year\ 1) + \beta_3 (t = year\ 2) + \beta_4 1(t = year\ 1) * biomarker_{i0} + \beta_5 1(t = year\ 2) * biomarker_{i0} + \beta_6 smoking\ status_{it} + \beta_7 FEV1_{it} + b_{i0} + \epsilon_{it}$$

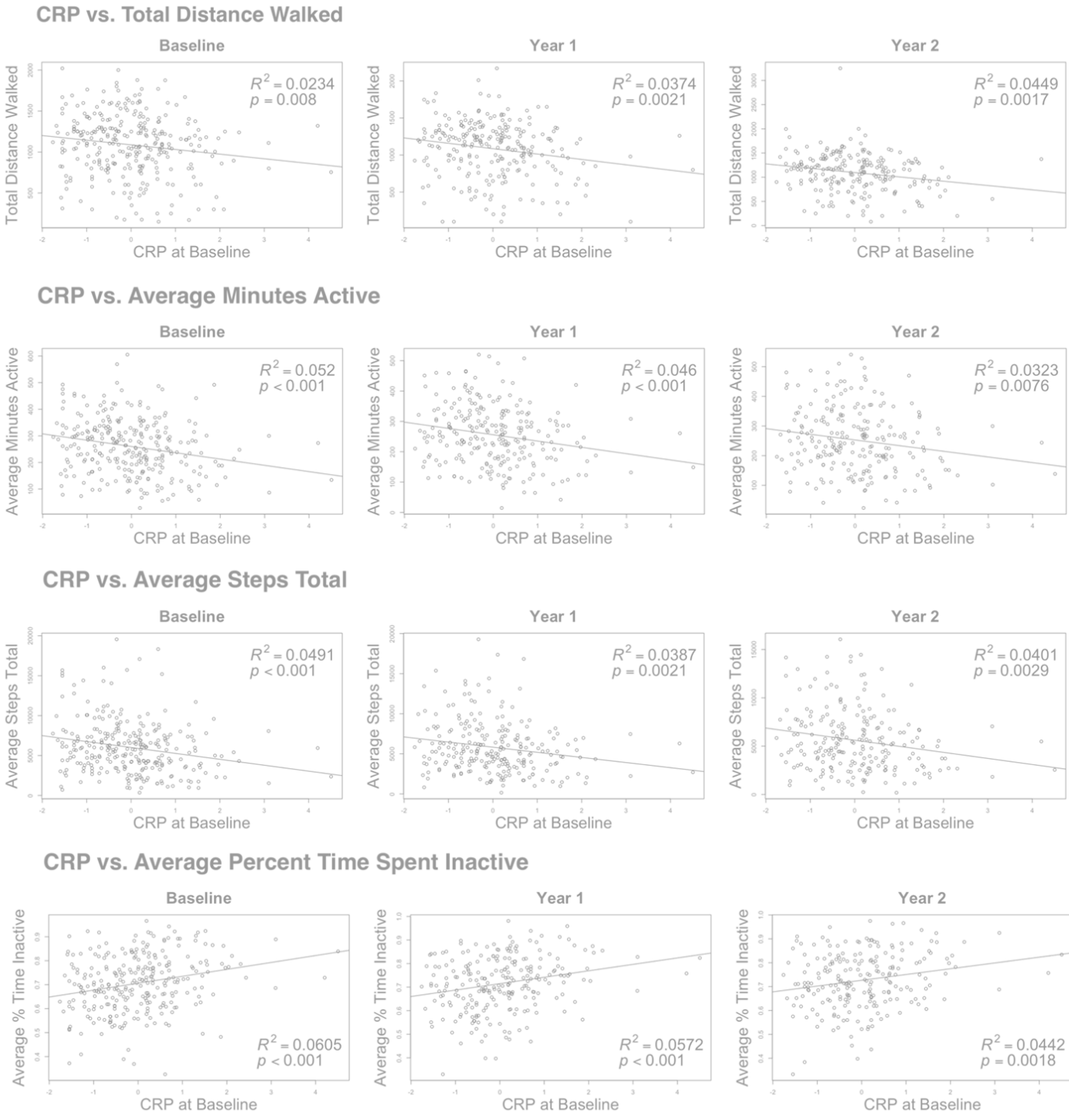
We were interested in the significance of the baseline biomarker effect on the outcome. We applied an analogous hierarchical hypothesis testing framework for each biomarker to assess significance, where each biomarker was treated as a set. The null hypotheses within each set were that each baseline biomarker has no effect on the seven outcomes considered in our analysis.

Supplementary Table 1: Mean annual changes in outcomes for all participants (Total) and stratified by rapid FEV_1 decline.

Characteristic	Total (n=302)	Rapid FEV_1 decline* (n=122)	No rapid FEV_1 decline* (n=131)
FEV_1	-32.8 (141)	-137 (98.9)	64.5 (98.4)
Average minutes active	-9.23 (43.0)	-14.3 (46.0)	-3.84 (38.6)
Average steps total	-335 (1310)	-474 (1540)	-172 (1020)
6-minute walk total distance (ft)	-16.1 (164)	-52.7 (147)	17.6 (171)
% time spent inactive	1.75 (4.40)	2.34 (4.66)	1.17 (4.00)
Grip strength (kg)	-1.25 (3.06)	-1.14 (2.57)	-1.37 (3.45)

* Rapid FEV_1 decline was defined as an average annual drop of 40 mL or higher. Forty-nine samples were missing two consecutive (baseline and year 1 or year 1 and year 2) measurements of FEV1.
FEV1, forced expiratory volume in 1-second

Supplementary Figure 1: Scatterplots of associations between baseline CRP and total distance walked (first row), average minutes active (second row), average steps total (3rd row), and average percent time spent inactive (fourth row). Associations between baseline CRP and activity outcomes are shown at baseline (first column), year 1 (second column), and year 2 (3rd column).



Supplementary References

Benjamini, Y., & Heller, R. (2008). Screening for Partial Conjunction Hypotheses. *Biometrics*, *64*(4), 1215–1222. https://doi.org/10.1111/j.1541-0420.2007.00984.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
Kawagoshi, A., Iwakura, M., Furukawa, Y., Sugawara, K., Takahashi, H., & Shioya, T. (2020). The association between systemic inflammation and the time spent in posture and movement during daytime in patients with chronic obstructive pulmonary disease and lower weight. *Journal of Physical Therapy Science*, *32*(12), 804–809. https://doi.org/10.1589/jpts.32.804
Li, Y., & Ghosh, D. (2014). A two-step hierarchical hypothesis set testing framework, with applications to gene expression data on ordered categories. *BMC Bioinformatics*, *15*(1), 108. https://doi.org/10.1186/1471-2105-15-108