Replicated Blood-based Biomarkers for Myalgic Encephalomyelitis Not Explicable by Inactivity

Sjoerd Viktor Beentjes[1,2], **Artur Miralles Méharon**[1], Julia Kaczmarczyk[1], Amanda Cassar[1], Gemma Louise Samms[1], Nima S Hejazi[3], Ava Khamseh[1,2], Chris P Ponting[1]

[1]University of Edinburgh, [2]University of California, Berkeley, [3]Harvard T.H. Chan School of Public Health

UKRI AI Centre for Doctoral Training in Biomedical Innovation
The University of Edinburgh
UK Research and Innovation

Background & Motivation

What is Myalgic Encephalomyelitis / Chronic Fatigue Syndrome (ME/CFS)?

- Debilitating multi-system disease affecting $\approx 0.6\%$ of the UK population (predominantly women)
- Key symptom: post-exertional malaise (worsening after activity)
- **No diagnostic biomarker**; biology poorly understood
- No cure and no widely effective therapy

The Problem: Confounding by Activity

- ME/CFS $\rightarrow$ reduced physical activity
- Many biomarkers are activity-dependent
- **Key question:** Are biomarker differences disease-driven, or explained indirectly through reduced activity?

Data

UK Biobank (UKB), *UK*

- 1,455 ME/CFS cases and 131,303 controls
- Blood traits (63), NMR metabolites (251), Proteins (2923)
- Activity measures: walking duration, days of moderate activity, minutes of moderate activity

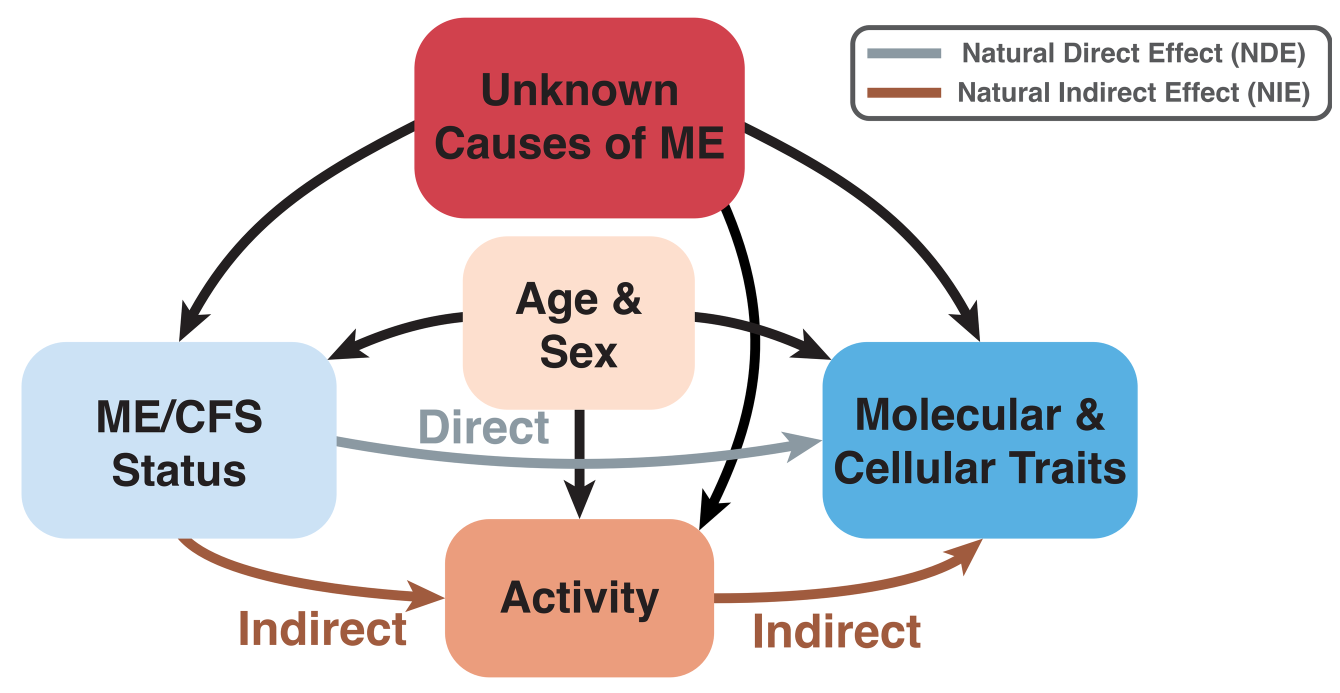
PEM subgroup (UK Biobank)

- Defined using the UKB Pain Questionnaire (fatigue 6 months, post-exertional worsening, not relieved by rest).
- 297 identified; 239 used after matching to non-PEM ME/CFS cases.

All of Us (AoU), *USA* (Replication)

- 903 cases and 75,943 controls
- 14 overlapping blood traits

Methods



Note: Full causal identifiability is not guaranteed; mediation effects are interpreted as statistical quantities adjusted for age and sex.

Semi-parametric mediation analysis

We used ensemble machine learning (Super Learner) to flexibly estimate all nuisance functions required for the one-step estimation of:

- Total Effect (TE)
- Natural Direct Effect (NDE)
- Natural Indirect Effect (NIE)

$$\underbrace{\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]}_{\text{TE}} = \underbrace{\mathbb{E}[Y(1, M(0))] - \mathbb{E}[Y(0, M(0))]}_{\text{NDE}} + \underbrace{\mathbb{E}[Y(1, M(1))] - \mathbb{E}[Y(1, M(0))]}_{\text{NIE}}$$

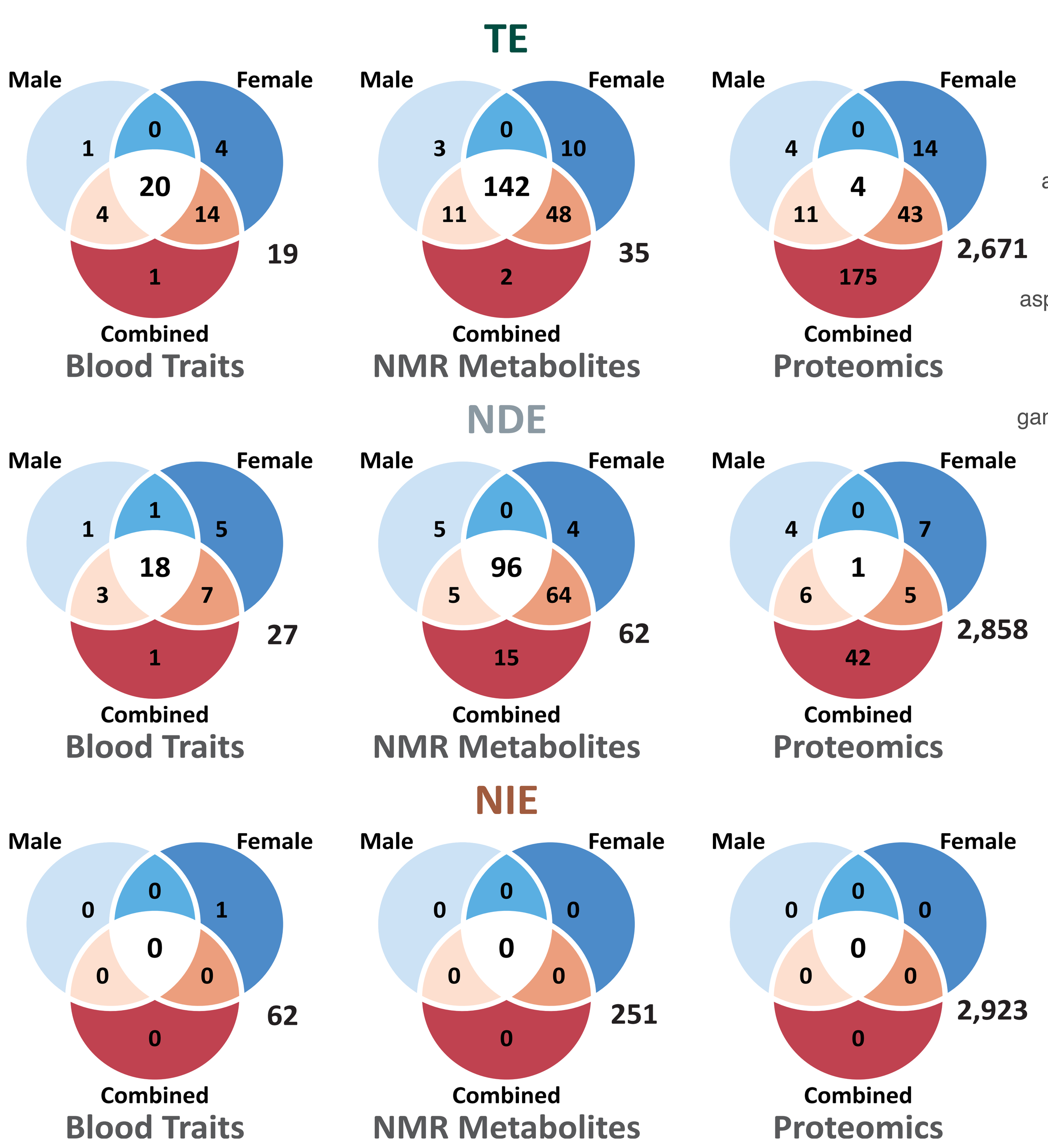
Super Learner Library

- Earth (MARS), GLMnet (LASSO)
- GLM with interactions
- XGBoost/LightGBM
- Highly Adaptive Lasso (HAL)

Software (R packages)

- TE: `npcausal` (Kennedy, 2021) + `SuperLearner` (Polley et al, 2011)
- NDE & NIE: `medoutcon` (Hejazi et al, 2022) + `sl3` (Coyle et al, 2021).

Results



	AoU ME	AoU Ctrl	AoU Sig	UKB ME	UKB Ctrl	UKB Sig
alanine_aminotransferase (U/L)	540	19539		1387	125055	+
alkaline_phosphatase (U/L)	594	22078	+	1389	125086	+
aspartate_aminotransferase (U/L)	583	22094		1381	124675	+
c_reactive_protein (mg/L)	323	4917		1386	124859	+
gamma_glutamyltransferase (U/L)	72	1811		1386	125038	+
glucose (mmol/L)	751	30314	+	1251	114080	+
hba1c (mmol/mol)	450	14571	+	1393	124688	+
hdl_cholesterol (mmol/L)	579	23194	-	1253	114161	-
leukocyte_cnt (10^9 cells/L)	617	22343	+	1413	127347	+
neutrophil_cnt (10^9 cells/L)	494	16389	+	1408	127109	+
TG.to.HDL.C (ratio)	567	22684	+	1253	114064	+
triglycerides (mmol/L)	589	23221	+	1388	124988	+
TyG (index)	480	20480	+	1251	113997	+
urea (mmol/L)	754	30147	+	1387	125002	-

Key findings

- **511 biomarkers** differ between ME/CFS cases and controls across blood traits, NMR metabolites, and proteomics.
- Effects are **consistent across males and females**: 166 biomarkers are significant in both sexes.
- **Mediation via physical activity is negligible**: almost no NIEs are significant despite strong TEs.
- **External replication** in All of Us: 9 of 14 blood traits tested show consistent effects (only urea differs).
- **Sensitivity analysis**: individuals with PEM-like symptoms show stronger biomarker shifts.

References

I Díaz, N S Hejazi, K E Rudolph, and M J van Der Laan. Nonparametric efficient causal mediation with intermediate confounders. *Biometrika*, 108(3):627–641, August 2021.

Nima Hejazi, Kara Rudolph, and Iván Díaz. medoutcon: Nonparametric efficient causal mediation analysis with machine learning in R. *J. Open Source Softw.*, 7(69):3979, January 2022.

Edward H Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *J. Am. Stat. Assoc.*, 114(526):645–656, April 2019.

Mark J. Van Der Laan and Sherri Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Series in Statistics. Springer, New York, NY, 2011. ISBN 978-1-4419-9781-4. doi: 10.1007/978-1-4419-9782-1.

Mark J van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Stat. Appl. Genet. Mol. Biol.*, 6(1), January 2007.