

## Supplementary Appendix

### Supplementary methods

#### A. Long COVID QALDs calculation

Long COVID QALDs were computed as the complement of the sum of the relevant utility weights for each participant, transformed by 365 days, which assumes symptoms persist for a year:

$$\text{Long COVID QALDs}_i = (1 - \sum_{k=1}^5 \text{QALY.weight.post}_{ki}) * \text{duration}$$

Where  $i$  denotes subject,  $k$  denotes each of the five items of the EQ-5D-5L survey,  $\text{QALY.weight.post}_{ki}$  is the weight from the appropriate value set corresponding to the chosen response for item  $k$  for subject  $i$ , and  $\text{duration}$  is the assumed duration of long-term COVID symptoms (1 year for all)

#### B. Variable preparation, missing data imputation, and post-stratification weighting

In the UK and Russian cohorts, COVID-19 disease severity was defined following the categorization in Reyes et al.<sup>1</sup> That is, subjects who reported any of the following outcomes (high-flow nasal cannula, ventilation, or the use of inotropes or vasopressors) during their hospital admission were classified as having severe acute COVID-19 disease.<sup>1</sup> Severity indicators were classified as unknown among subjects who had missing entries for all five possible indicators. In the Norway cohort, there was little information available on any of the aforementioned outcomes and another measure of acute COVID-19 severity, hospitalization for acute COVID-19 infection (any or ICU admission), was very sparsely recorded. Therefore, we excluded this variable among our factors under consideration in this cohort.

Vaccination data was obtained using a search for the following key terms within the field for immunization against COVID-19: Moderna, Pfizer, AstraZeneca, and Janssen, as well as “COVID vaccination”, “COVID-19 vaccination”, and “COVID-19 vaccine type”, which were the general categories recorded when no specific brand was noted.

Antiviral and additional treatment data was obtained using a search for the following key terms within the field for treatment: darunavir, remdesivir, acyclovir, valganciclovir, lopinavir/ritonavir, metformin, as well as “Antiviral agent”, which was a general category recorded when no specific type was noted. We selected these drugs as those explicitly listed as possible options in the ISARIC follow-up surveys, as well as a diabetes drug that has been identified to be protective against long COVID (metformin).<sup>2</sup> Data on the two most currently supported therapies nirmatrelvir and molnupiravir was not available in any cohort.

Missingness was reported for the following variables: Norway: educational attainment (17.6%), race/ethnicity (2.3%), sex (1.2%), age (0.006%), vaccination status (0.90%), antiviral treatment during illness (83.5%), and the severity indicator (75.7%); UK: employment status (11.2%), race/ethnicity (1.6%), vaccination status (66.5%), and antiviral treatment during illness (22.3%); and Russia: employment status (4.8%), race/ethnicity (100%), vaccination status (100%), and antiviral treatment during illness (100%).

Missing variable imputation was performed via Multiple Imputation by Chained Equations (MICE) using the *mice* package.<sup>3</sup> We cannot make a definitive conclusion on the missingness mechanisms for sex at birth. That is, we have no reason to believe that females compared to males are less likely to self-report their sex, i.e. we do not have any evidence in favor of a significant difference or missing not at random (MNAR) mechanism. Thus, the use of MICE, which assumes a missing (completely) at random mechanism, appears justified. However, vaccination status and the socio-economic indicators of educational attainment and employment status are more likely to be MNAR due to social stigmas and fear of judgment. Nonetheless, a relatively low fraction of most of these values were missing, with the exception of the vaccination status indicator in the UK and Russia cohorts, antiviral treatment in the Norway and Russia cohorts, and the severity indicator in the Norway cohort. Consequently, to avoid imposing significant bias, we chose not to impute these specific variables and excluded them from their respective variable sets. For all cohorts, we specified 10 imputations and 10 iterations following standard practice, where, for each iteration, missing variables were expressed as a function of all other variables via polytomous regression (for categorical variables), logistic regression (for binary variables), and predictive mean matching (for continuous variables). Missing values were then assigned the majority imputation across runs for the corresponding variable.

The distributions of sex at birth in our sampled populations non-negligibly deviated from those in the underlying Norwegian and British populations. Thus, to correct for this binary mis-representativeness, we conducted sensitivity analyses where subjects were assigned post-stratification weights according to the *raking* method (Supplementary Appendix: Methods and Results).<sup>4</sup> In brief, this approach adjusts the distribution of each of the sampled populations on the basis of sex to correspond to its distribution in the overall population.<sup>5-7</sup>

Finally, we note that in both Norway and the UK, most participants self-identified as white (93% and 96%, respectively), which is a profound deviation from the underlying race/ethnicity distribution in both countries, with some key minority ethnic groups being very sparsely captured. Due to these multi-group imbalances, we could not validly apply population corrections and concluded that any assessment of the role of race/ethnicity in shaping long COVID QALDs would not be justified, thereby omitting it from our social variables under consideration.

### C. Quantifying variable importance

For all analyses, comorbidities were filtered to only include conditions with at least ten subjects reporting having them. All of our random forest regressions were trained on 80% of the dataset using the *caret* package.<sup>8</sup> For the individual RF and pre-grouped RF, variables were ranked according to their associated % increase in mean squared error, averaged over 100 runs. To run these RFs, we used the *randomForest* package.<sup>9</sup> While we averaged our results across 100 runs for the individual and pre-grouped analyses, we considered a fewer number of runs (50) for the model-grouped analysis to avoid exorbitant model costs, due to the resource intensiveness of each CoV-VSURF run, using the *CoVVSURF* package.<sup>10</sup> For Norway, we used dummy variables representing belonging to each quintile of educational attainment compared to the referent category of quintile 1 for RF #3, as CoV-VSURF does not accept ordered factors as inputs. Finally, for the population adjustment sensitivity analysis, weights were incorporated in all random forest implementations using the `sample_weight` parameter in the `randomForest`<sup>9</sup> and `covsurf`<sup>10</sup> functions.

For our pre-grouped random forest regression implementation, we applied multiple factor analysis to each cluster, which corresponds to a principal component analysis for numeric variables and multiple correspondence analysis for categorical variables, considering the first and second principal components of each. For binary variables, we replaced subjects' values with the loading obtained for that variable, with variable coordinates defined as the square root of the component's eigenvalue<sup>11</sup>, if present, and with 0, if absent. For categorical variables with more than one level, we replaced subjects' values with the loading obtained for the relevant level of that variable. For numeric variables, we multiplied subjects' values by the loading obtained for that variable. For each cluster, we then summed the resulting weighted variables by subject and by component, reducing the number of variables for each cluster  $i$  from  $p_{cluster.i}$  to two. The resulting reduced number of covariates were then used as features for our random forest model runs. We used the *FactoMineR* package<sup>12</sup> to conduct this cluster summary analysis.

### D. Estimating NDEs, NIEs, and proportions non-mediated

#### *Defining exposures*

For Norway, our binary proxy of SES was high/low educational attainment, where 'high' encompassed quintiles 3-5 and 'low' encompassed quintiles 1 and 2. For the UK and Russia, our binary proxy of SES was high/low employment status. For the UK, 'high' encompassed full-time employment and 'low' encompassed all other employment status categories (furloughed, part-time employment, student, and the unemployed), with the exception of retirement. For Russia, 'high' encompassed full-time employment and 'low' encompassed all other employment status categories, where retirees are included due to notable disparities in representation of full-time employees versus all other categories in this cohort.

### *Defining mediators*

Mediators were selected following a literature search of studies distinguishing key long COVID predictors<sup>11,12</sup>, consistent with the data available for each country, resulting in the following final mediator sets: asthma, chronic cardiac disease (not hypertension), hypertension, chronic pulmonary disease (not asthma), type 1 diabetes (T1D), type 2 diabetes (T2D), diabetes (type not specified), psychological disorder, smoking, and vaccination status (Norway), asthma, chronic cardiac disease (not hypertension), hypertension, chronic pulmonary disease (not asthma), T2D, diabetes (type not specified), psychological disorder, ischemic heart disease, smoking, obesity, and antiviral treatment (UK) and asthma, chronic cardiac disease (not hypertension), hypertension, chronic pulmonary disease (not asthma), T2D, diabetes (type not specified), smoking, and obesity (Russia).

### *Running mediation analyses*

Estimating the NDE, NIE, and proportion mediated via targeted maximum likelihood estimation (TMLE) requires two key steps.<sup>13,14</sup> First, the data is divided into  $k$  cross-validation folds (subsets of the data), to which we train and test our algorithms of choice, where the response measure is the expected mean of long COVID QALDs dependent on each of our binary SES proxies or female sex, the confounders, and mediators.<sup>13,14</sup> The algorithms we consider include a simple intercept regression model and generalized additive models (GAMs)<sup>15</sup>, the latter equipped to accommodate non-linear forms for continuous covariates, in addition to gradient boosted regression trees<sup>16</sup>, using the *lightgbm* package<sup>17</sup>, and feed-forward neural networks<sup>18</sup> using the *nnet* package<sup>19</sup>, both of which can better capture underlying nonlinear trends non-parametrically. The super learner ensemble modeling procedure<sup>20</sup> (via the *s/3* package<sup>21</sup>) then pinpoints the combination of weights to assign to each algorithm that minimizes the (cross-validated) mean squared error (MSE) across all folds in a theoretically optimal manner, through a non-negative least squares meta-learner (via the *nnls* package<sup>22</sup>). The expected means of long COVID QALDs dependent on each of our binary SES proxies or female sex, the confounders, and mediators under the two levels of the exposure are then estimated using the weighted sum of each algorithm's fit to the original data.<sup>13,14</sup> The difference in these fitted means of long COVID QALDs dependent on each of our binary SES proxies or female sex, the confounders, and mediators for each level of the exposure, alongside the conditional probability of each of our binary SES proxies or female sex, dependent on confounders and the conditional probability of each of our binary SES proxies or female sex, dependent on confounders and mediators are subsequently used as ingredients to compute estimates of the NDE and NIE.<sup>13,14</sup>

In all analyses, 95% confidence intervals (CIs) were provided for our measures of interest estimated in *medoutcon*<sup>23</sup>, i.e., the NDE, NIE, and proportion mediated. Proportions non-mediated are calculated as the complement of estimated proportions mediated. For CIs of the proportion mediated with upper bounds falling below 0 (i.e., upper bound of the CI of the proportion non-mediated exceeding 1), we reported CIs bounded by 1. For the population adjustment sensitivity analysis, weights were incorporated using the *survey\_weight* parameter in the *medoutcon* package.<sup>23</sup>

Directed acyclic graphs (DAGs) were drawn using the *dagitty*<sup>24</sup> and *ggdag*<sup>25</sup> packages.

#### E. Additional analyses

To compute the crude association between long COVID QALDs and sex (female vs male), we conducted a one-sided Wilcoxon rank sum test which assumes independence and equal variance for the two groups, but does not assume normality of long COVID QALDs.

All analyses were run using the R programming language and environment for statistical computing and graphics (version 4.2.2).<sup>26</sup>

### Supplementary results

*Comparison of EQ-5D-5L responses by demographic groups (in the final study populations for each cohort, post-missing data imputation)*

#### A. Norway

We found minimal differences across educational attainment quintiles in reported responses for the dimensions anxiety/depression, mobility, self-care, and usual activities (Supplementary Figure S3B). However, quintiles 3-5 skewed towards lower responses for pain/discomfort, indicating fewer challenges for that dimension (Supplementary Figure S3B). Males and females reported similar distributions in responses for the EQ-5D-5L dimensions anxiety/depression, mobility, self-care, and usual activities (Supplementary Figure S4B). However, males skewed towards lower values for pain/discomfort (Supplementary Figure S4B).

#### B. UK

There was considerable heterogeneity across employment status categories in responses to each of the five EQ-5D-5L dimensions (Supplementary Figure S3C).

Males and females reported similar distributions in responses for the EQ-5D-5L dimensions for mobility, pain/discomfort, and usual activities (Supplementary Figure S4C). However, males skewed towards lower values for anxiety/depression and self-care (Supplementary Figure S4C).

#### C. Russia

As with the UK, there was considerable heterogeneity across employment status categories in responses to each of the five EQ-5D-5L dimensions (Supplementary Figure S3C).

The distribution of EQ-5D-5L responses was nearly equivalent for both sexes on every dimension except mobility, where females skewed towards greater challenges (Supplementary Figure S4A).

Tables

**Supplementary Table S1. Selected studies for extracting utility weights**

<u>Cohort</u>	<u>Study (Country)</u>
Norway	Sun et al., 2022 <sup>27</sup> (Sweden)
UK	Devlin et al., 2018 <sup>28</sup> (England)
Russian cohort	Golicki et al., 2019 <sup>29</sup> (Poland)

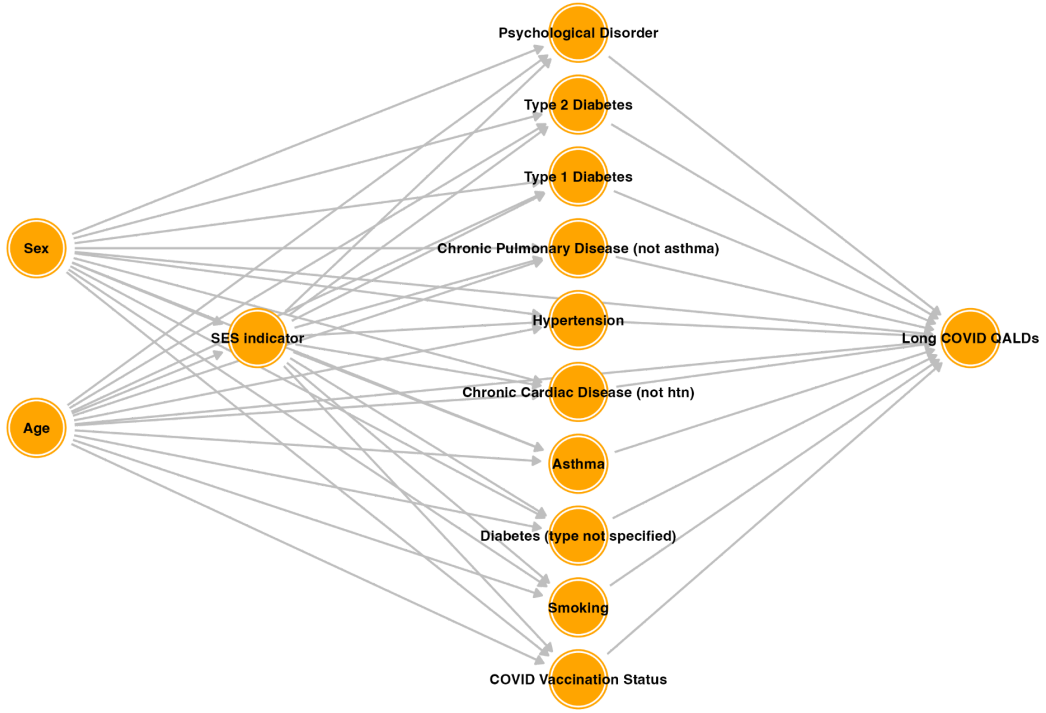
**Supplementary Table S2. Final employment status groupings**

<b>Group category</b>	<b>Norway</b>	<b>Russia</b>
Retired	“Medically retired”, “Retired”, “Retired_Medically retired”, “Unable to work due to chronic illness_Retired”	“Early Retirement Due to Illness”, “Retired”
Unemployed	“Unemployed”, “Unable to work due to chronic illness”	“Unemployed”, “Unable to Work Due to Chronic Illness”
Full-time employment	“Full-time employment”, “Full time carer (children or other)”, “Full-time employment_Prefer not to say”, “Full-time employment_Full time carer (children or other)”, “Working full-time”, “Working Full-time”	“Working Full-Time”
Carer	NA (small n, so grouped w/ full-time employment above)	“Full Time Carer (Children or Others)”,
Furloughed	“Furloughed”, “Full-time employment_Furloughed”	NA
Student	“Student”	Student
Part-time employment	“Part-time employment”, “Working Part-time”, “Working part-time”	“Working Part-Time”

**Supplementary Table S3. Predetermined groupings of variables, by country, for RF #2**

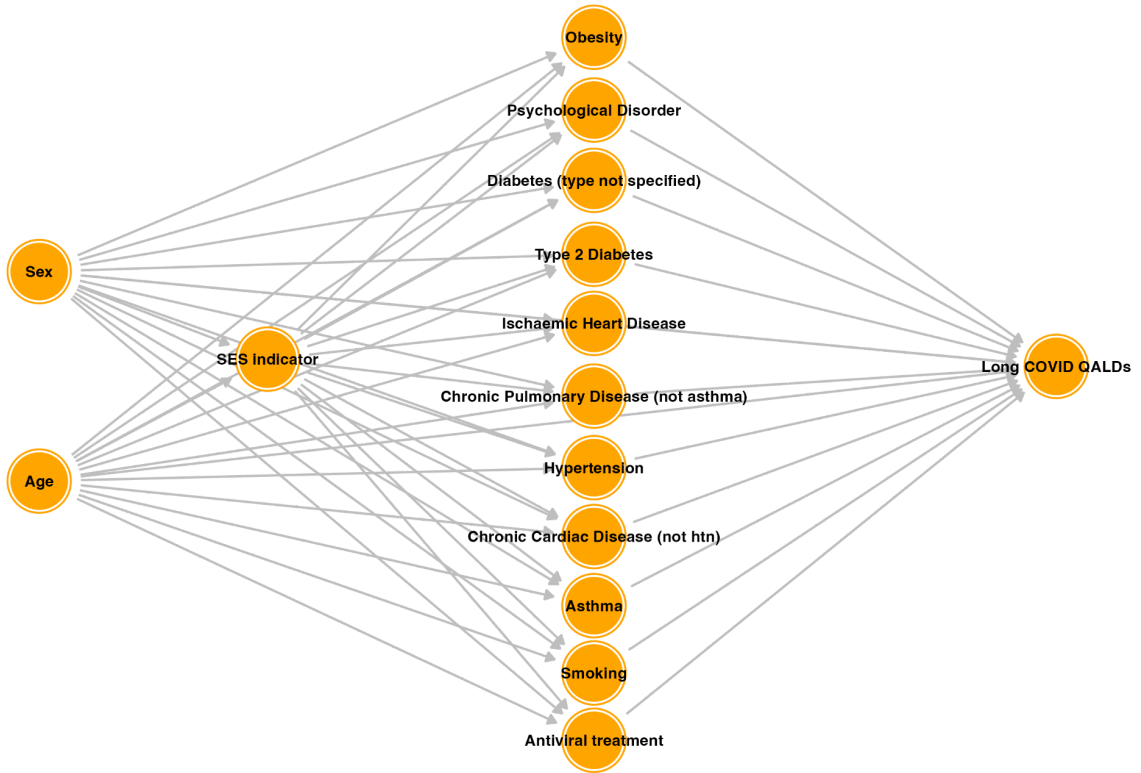
<b>Group</b>	<b>Norway</b>	<b>UK</b>	<b>Russia</b>
1	Age	Age	Age
2	Educational attainment (years), educational attainment quintile, sex	Employment status, sex	Employment status, sex
3	Asthma, chronic pulmonary disease (not asthma), smoking	Asthma, chronic pulmonary disease (not asthma), bronchiectasis, smoking	Asthma, chronic pulmonary disease (not asthma), smoking
4	Obesity, type 1 diabetes, type 2 diabetes, diabetes (type not specified)	Obesity, type 1 diabetes, type 2 diabetes, diabetes (type not specified)	Obesity, type 2 diabetes, diabetes (type not specified)
5	Chronic cardiac disease (not hypertension), hypertension	Congestive heart failure, ischemic heart disease, atrial fibrillation, chronic cardiac disease (not hypertension), hypertension, peripheral vascular disease	Chronic cardiac disease (not hypertension), hypertension
6	Chronic kidney disease, liver disease (severity not specified)	Gastrointestinal (GI) disease, chronic metabolic endocrine disease, GI reflux disease, hypothyroidism, lipid disorder	Chronic kidney disease, mild liver disease
7	Chronic hematological disease, rheumatological disorder	Chronic hematological disease, rheumatological disorder	Rheumatological disorder
8	Psychological disorder, chronic neurological disorder	Psychological disorder, chronic neurological disorder	Dementia, chronic neurological disorder
9	Malignant neoplasm	Malignant neoplasm	Malignant neoplasm
10	Other	Myocardial infarction (MI)	Other
11	Vaccination status	Chronic infection	Country
12	N/A	Other	COVID-19 disease severity indicator
13	N/A	Antiviral treatment	N/A
14	N/A	COVID-19 disease severity indicator	N/A

(A)

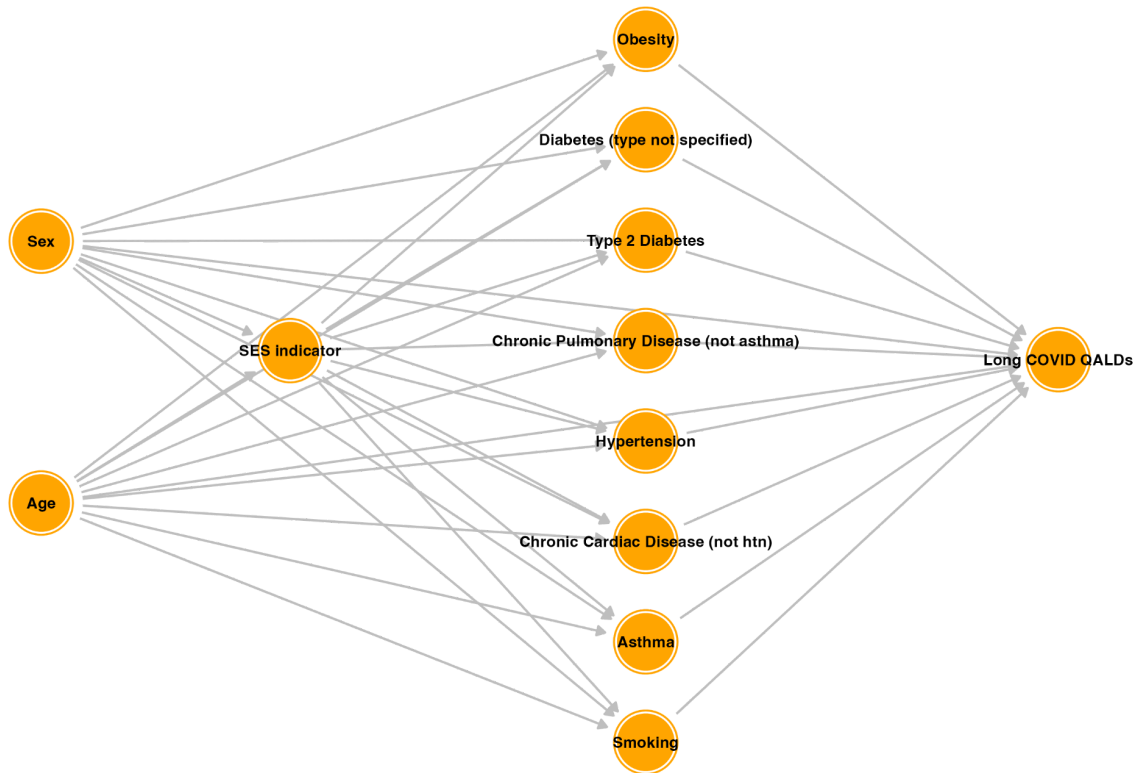




(B)

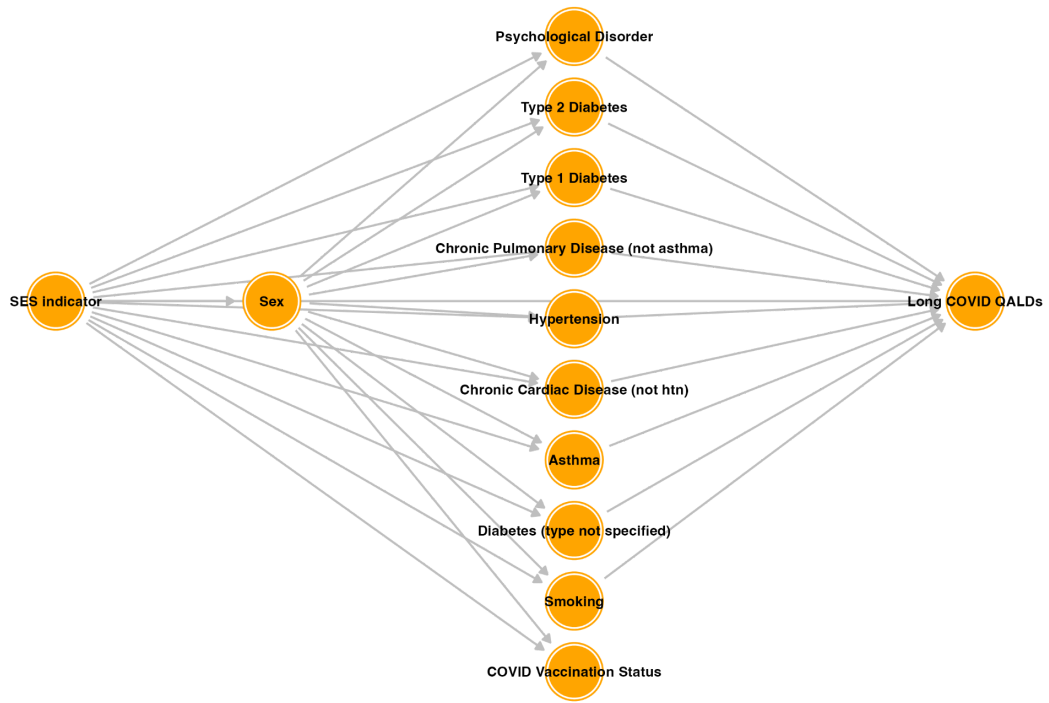


(C)

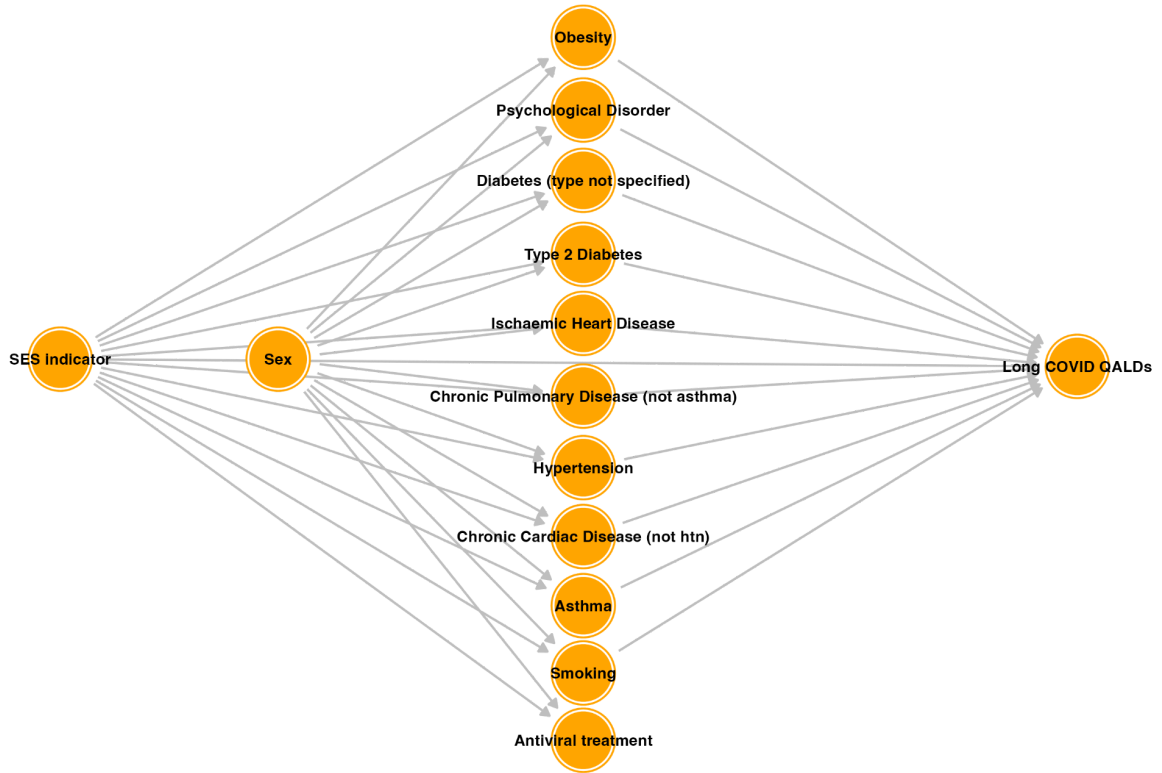


**Supplementary Figure S1. Assumed DAGs for the Norway, UK, and Russia cohorts with SES indicators as exposure. DAG for educational attainment binary variable in Norway (A), DAG for employment status binary variable in the UK (B), and DAG for employment status binary variable in the Russia cohort (C). Note: htn = hypertension, in “Chronic Cardiac Disease (not htn).”**

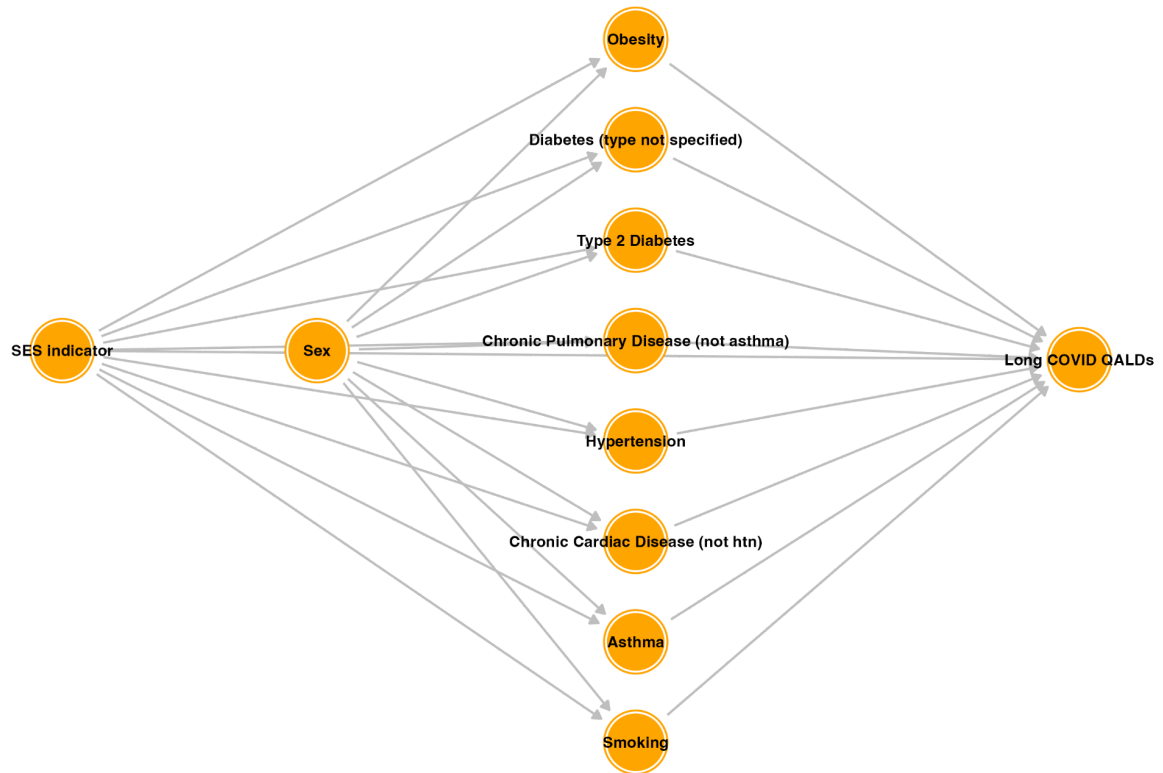
(A)



(B)

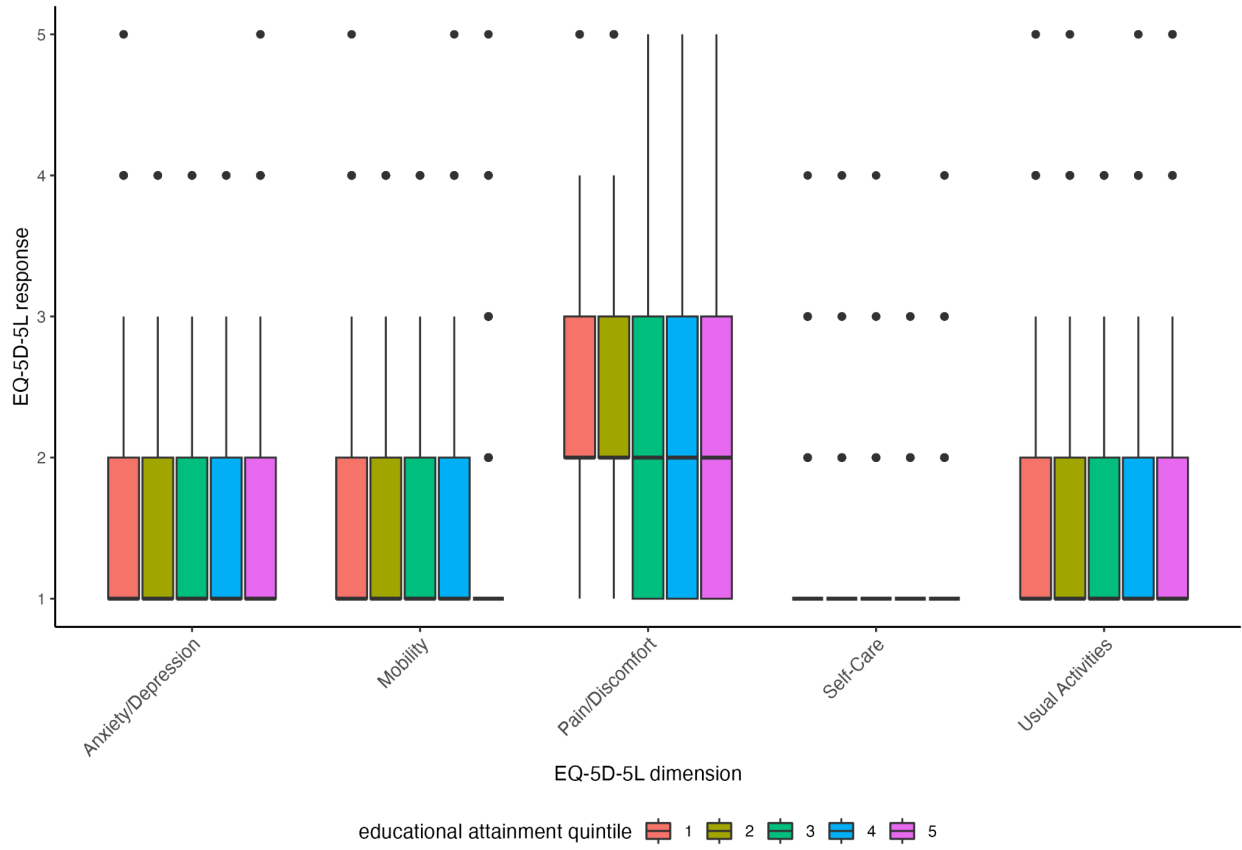


(C)

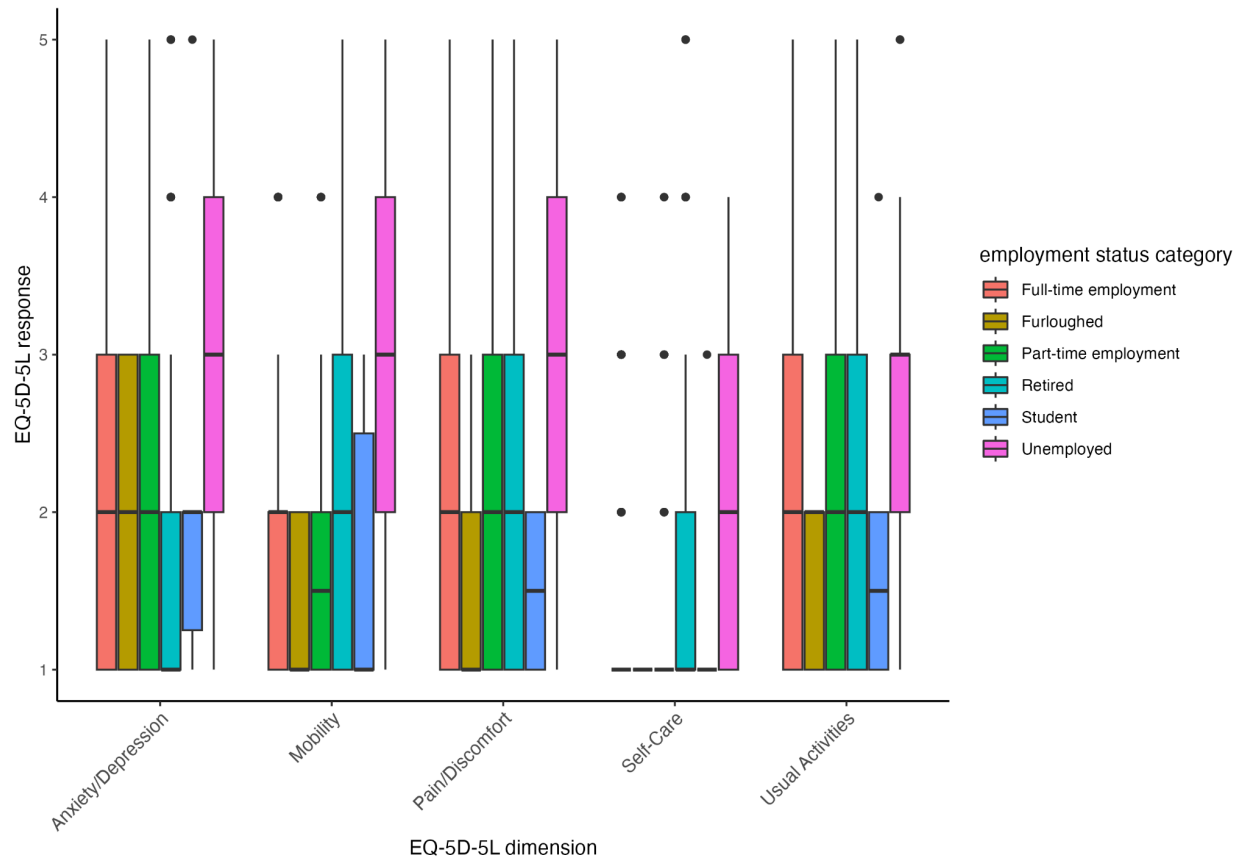


**Supplementary Figure S2. Assumed DAGs for the Norway, UK, and Russia cohorts with Sex (female vs male) as exposure. DAG for Sex in Norway (A), DAG for Sex in the UK (B), DAG for Sex in Russia (C). Note: htn = hypertension, in “Chronic Cardiac Disease (not htn).”**

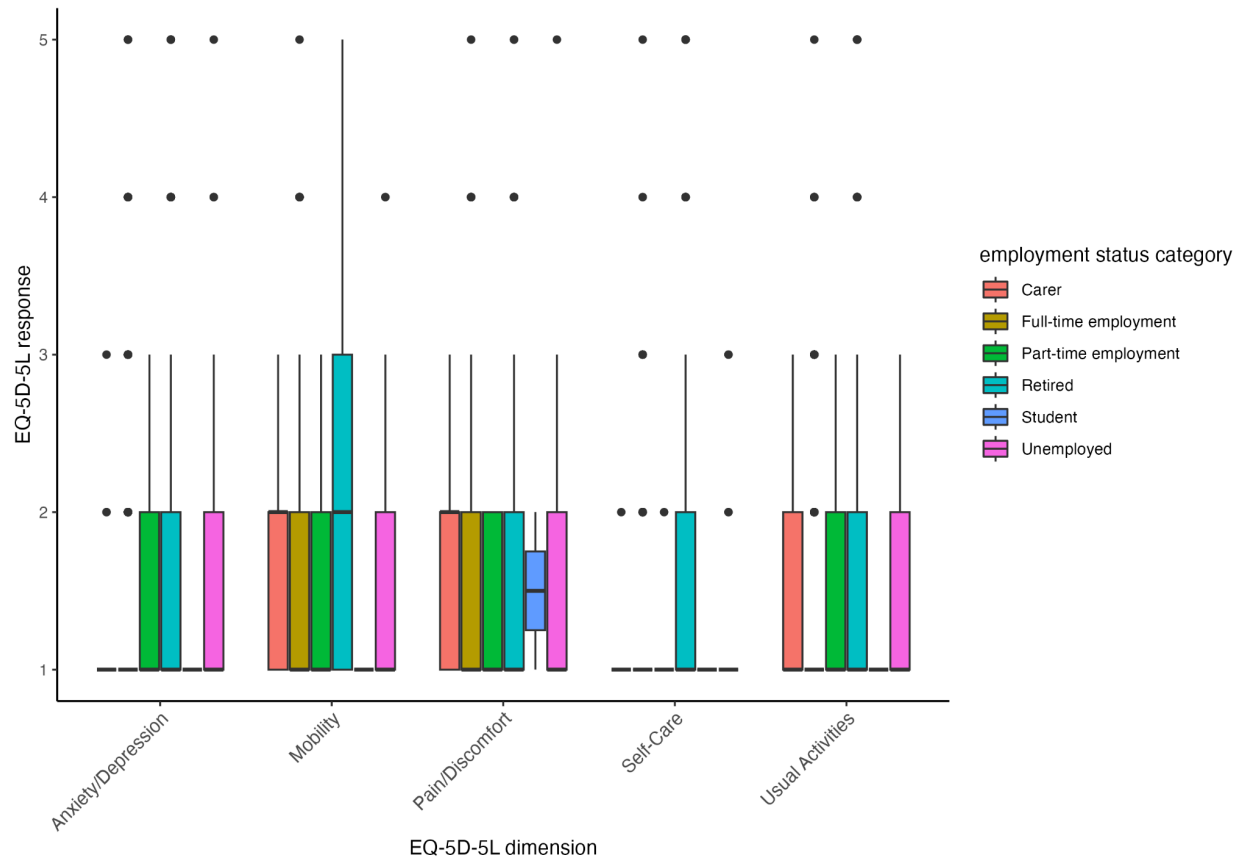
(A)



(B)



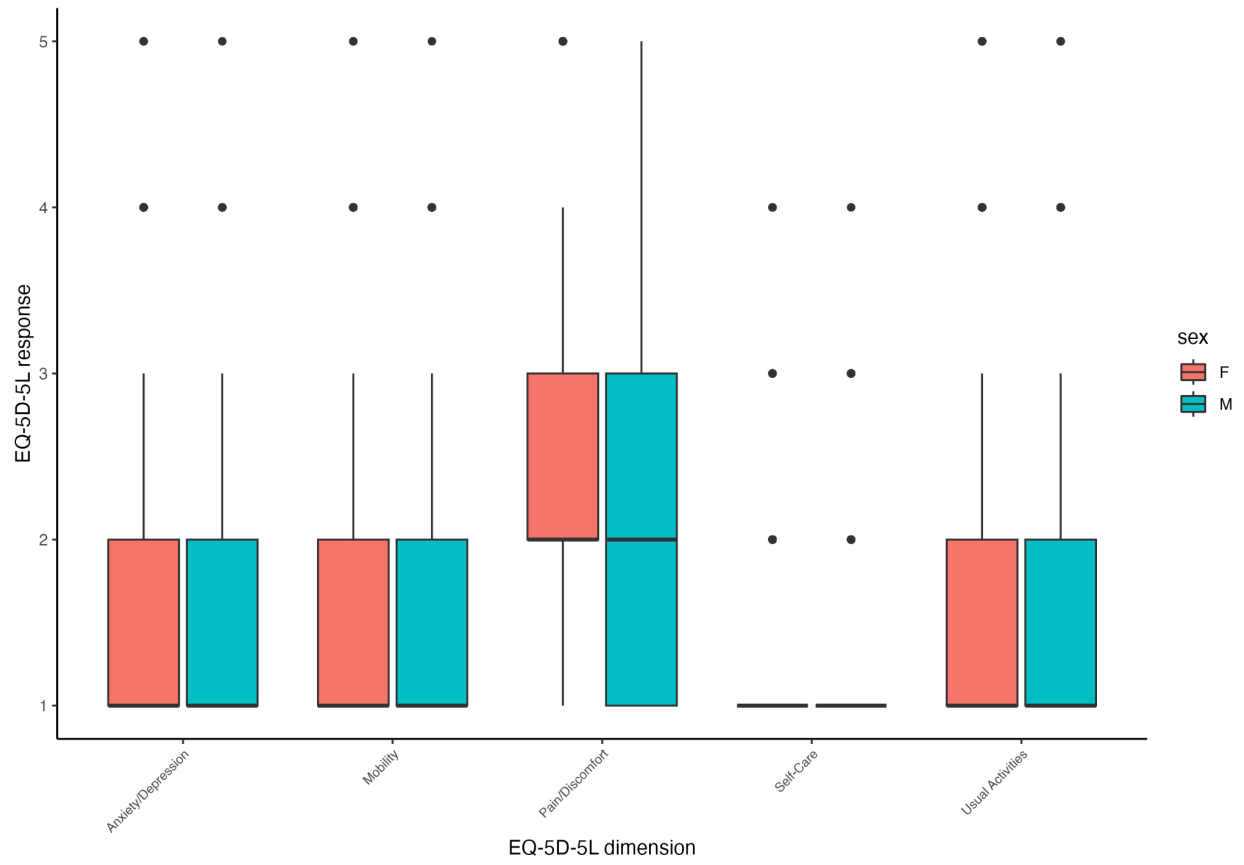
(C)



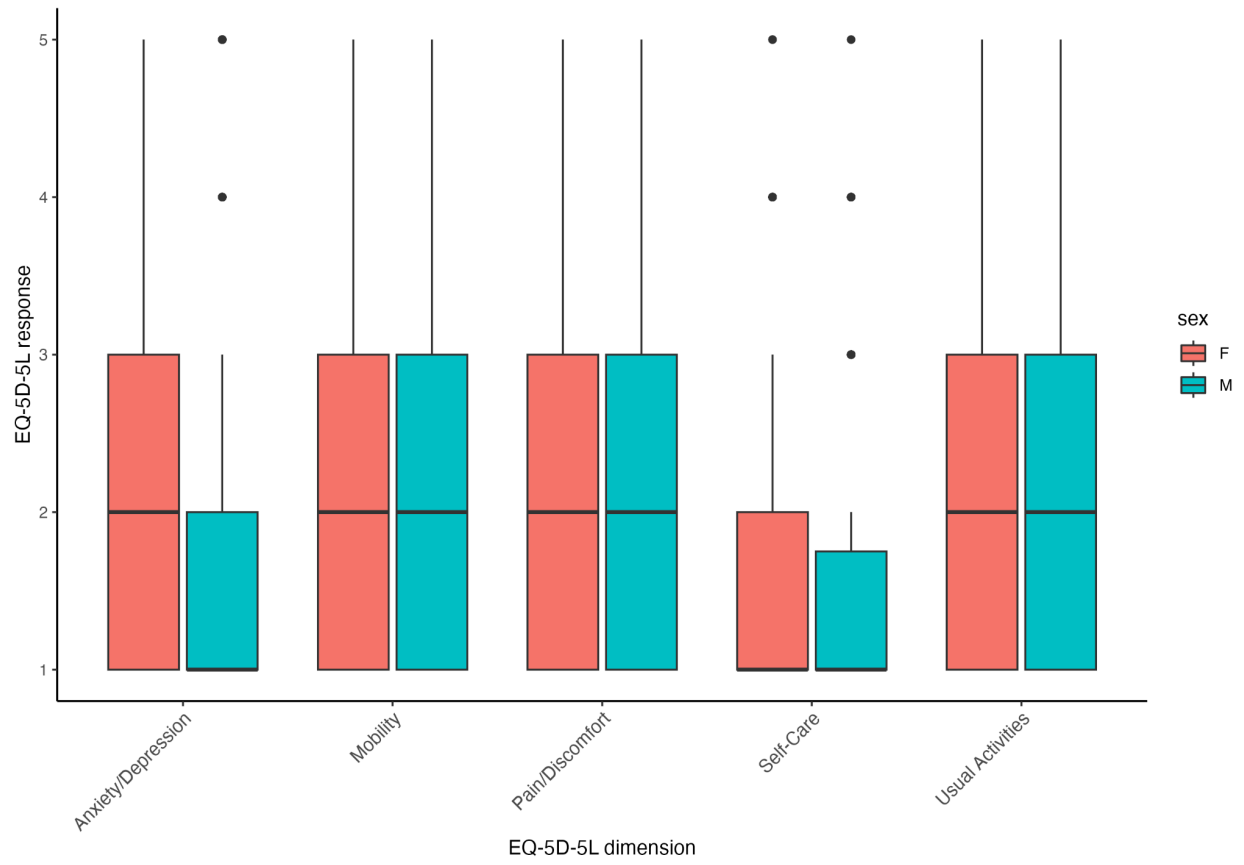
**Supplementary Figure S3. Boxplot comparisons of EQ-5D-5L responses (score of 1-5) by dimension and quintile of educational attainment in Norway (A) by dimension and employment status category in the UK (B), and by dimension and employment status category in Russia (C). Bars denote median scores. Points are drawn for scores which lie 1.5\*the interquartile range units above the upper quartile.**

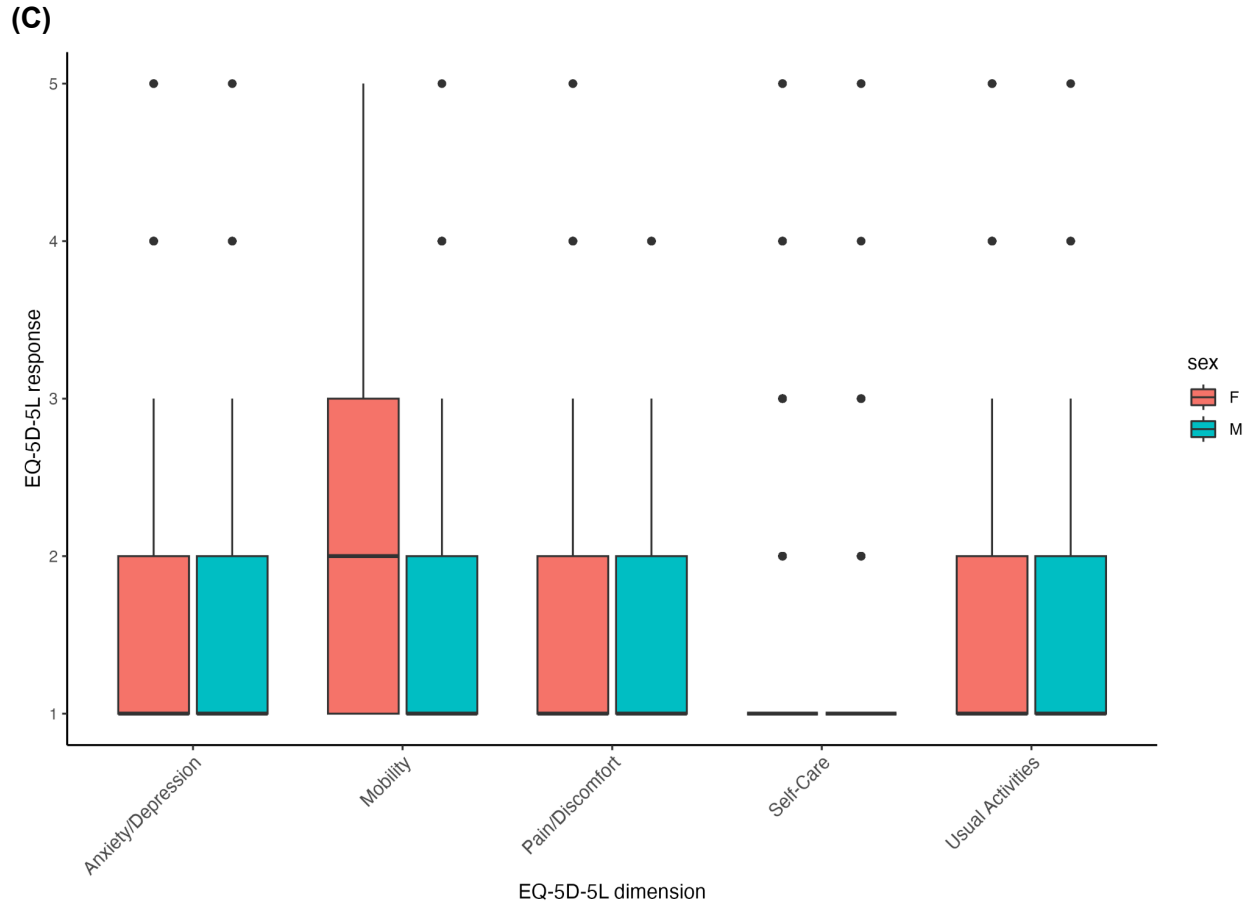


(A)



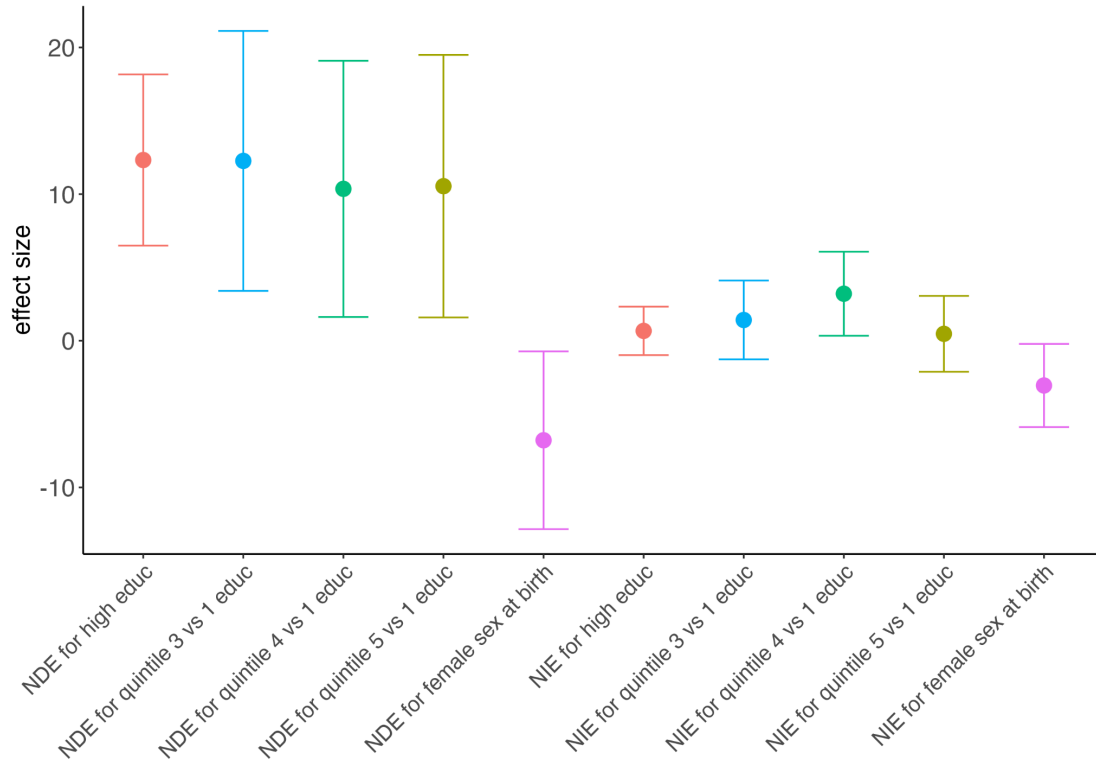
(B)



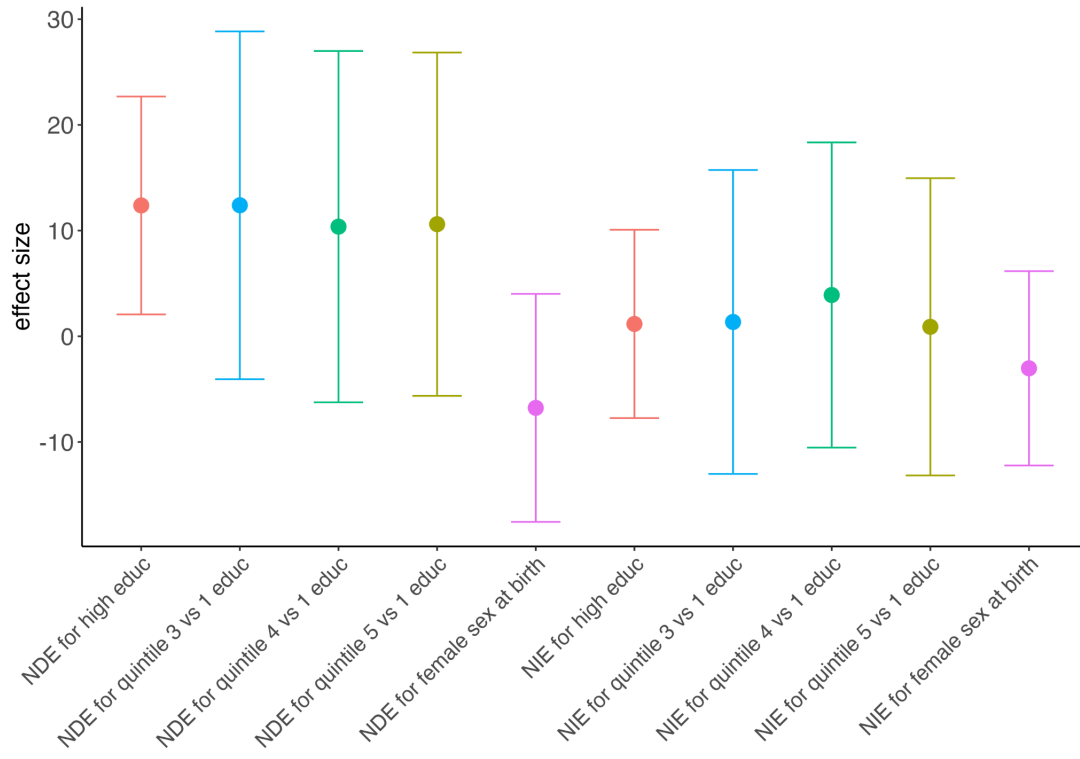


**Supplementary Figure S4. Boxplot comparisons of EQ-5D-5L responses by dimension and sex (Female (F), Male (M), and Unknown (U)) in Norway (A), the UK (B), and Russia (C). Points are drawn for scores which lie 1.5\*the interquartile range units above the upper quartile.**

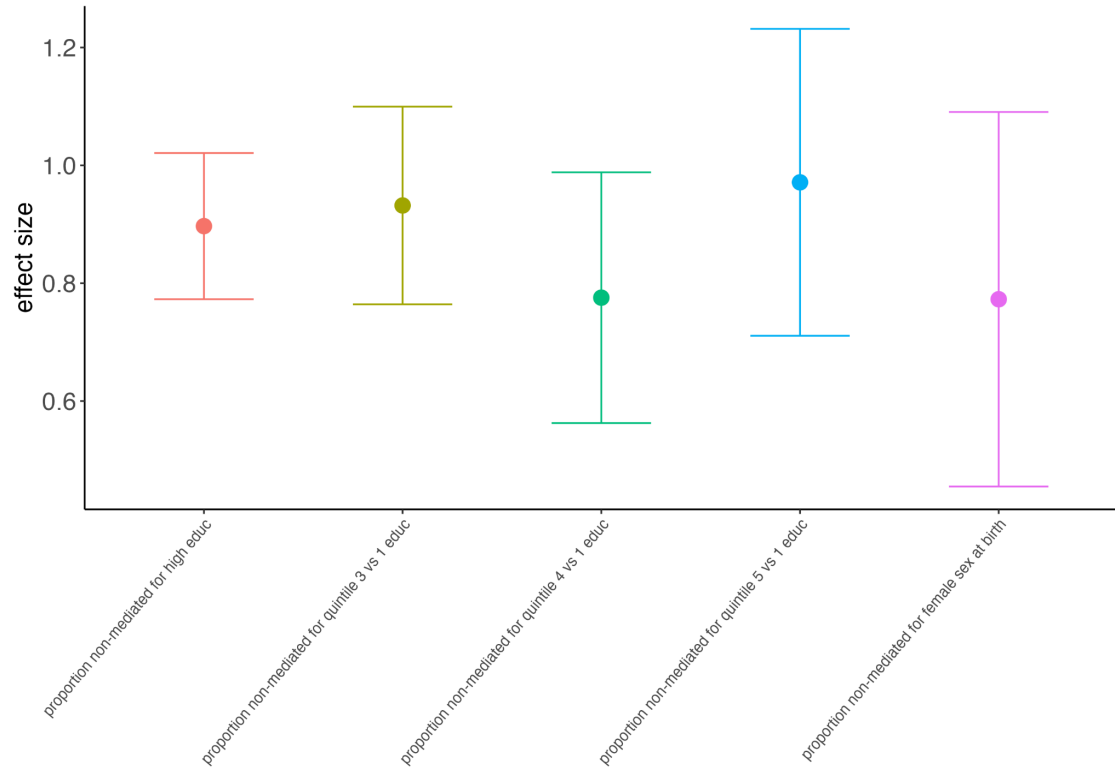
(A)



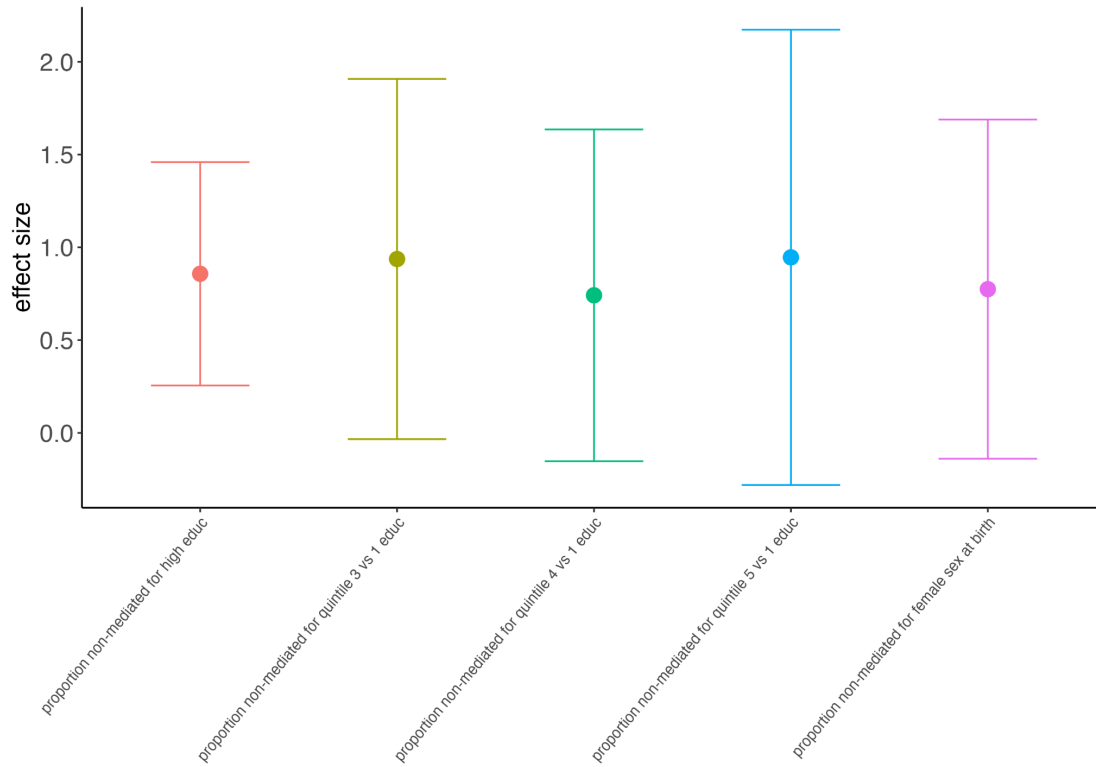
(B)



(C)

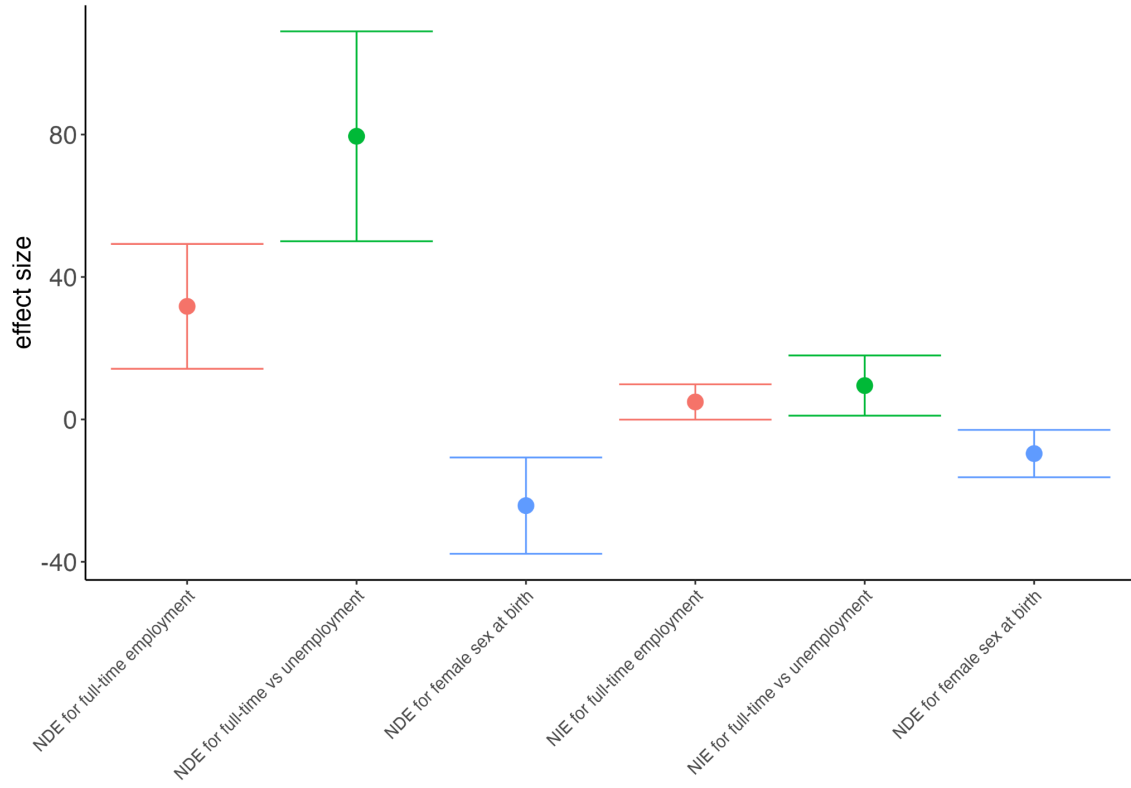


(D)

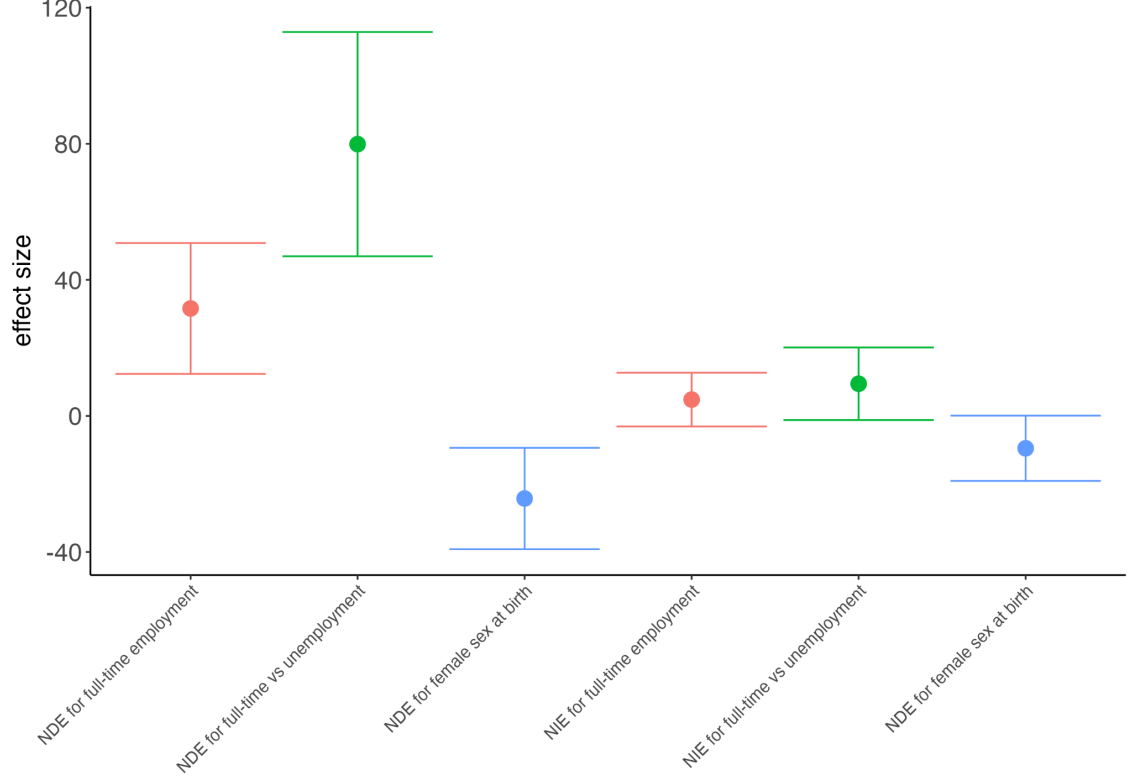


**Supplementary Figure S5. All output for Norway mediation analyses. Estimated NDEs and NIEs (point estimate and 95% confidence interval) in the main analysis (A) and in the sensitivity analysis (B). Colors denote the contrast of interest (pink: high vs low educational attainment; blue: quintile 3 vs 1 of educational attainment; green: quintile 4 vs 1 of educational attainment; khaki: quintile 5 vs 1 of educational attainment; purple: female vs male sex). Estimated proportions non-mediated (point estimate and 95% confidence interval) in the main analysis (C) and in the sensitivity analysis (D).**

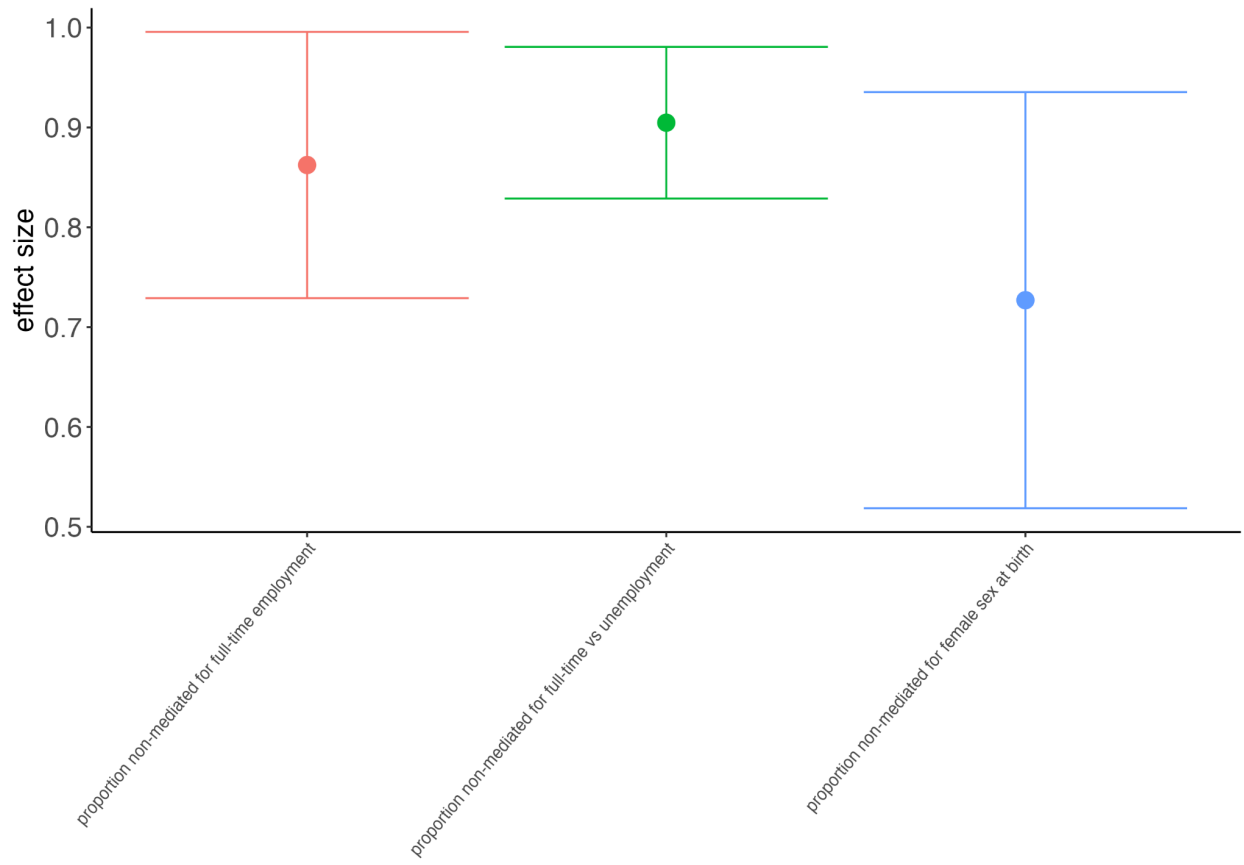
(A)



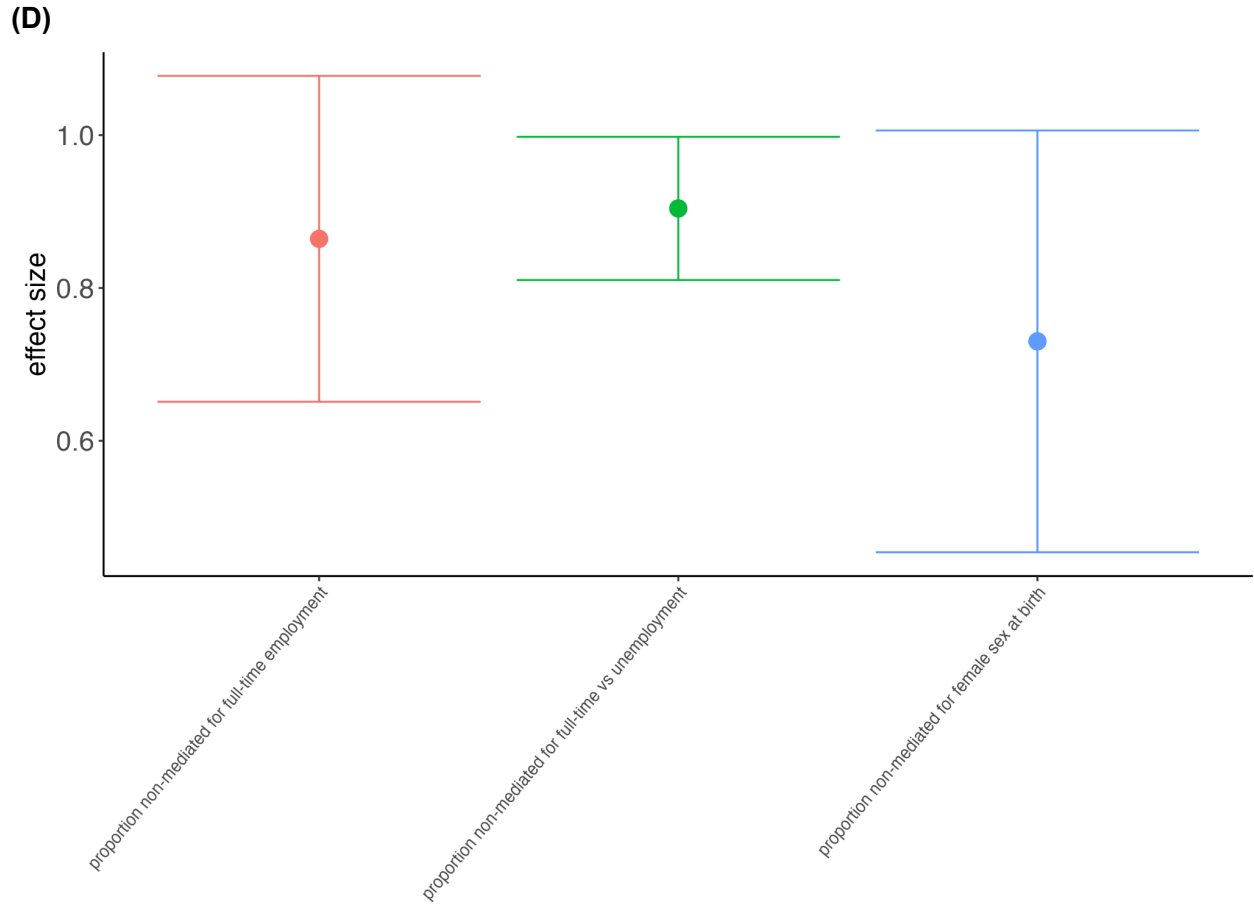
(B)



(c)

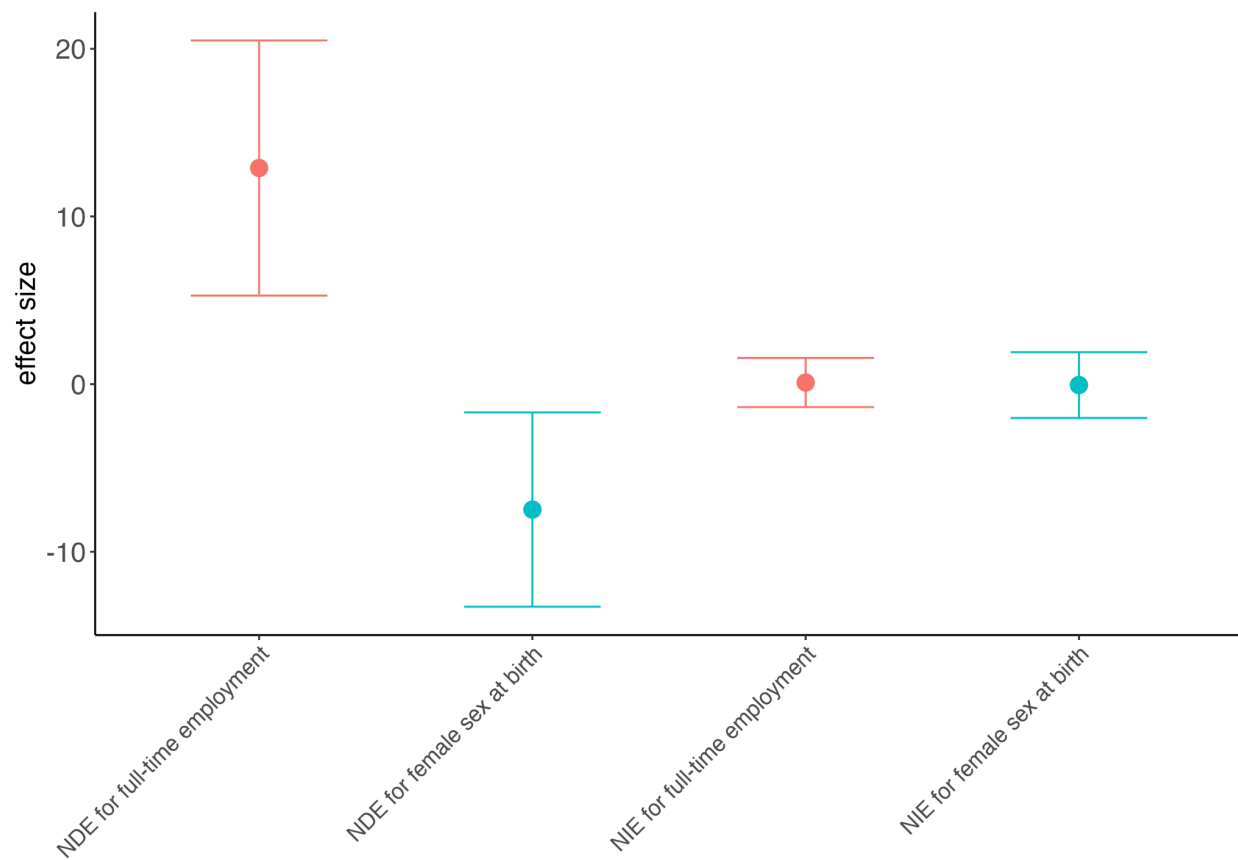




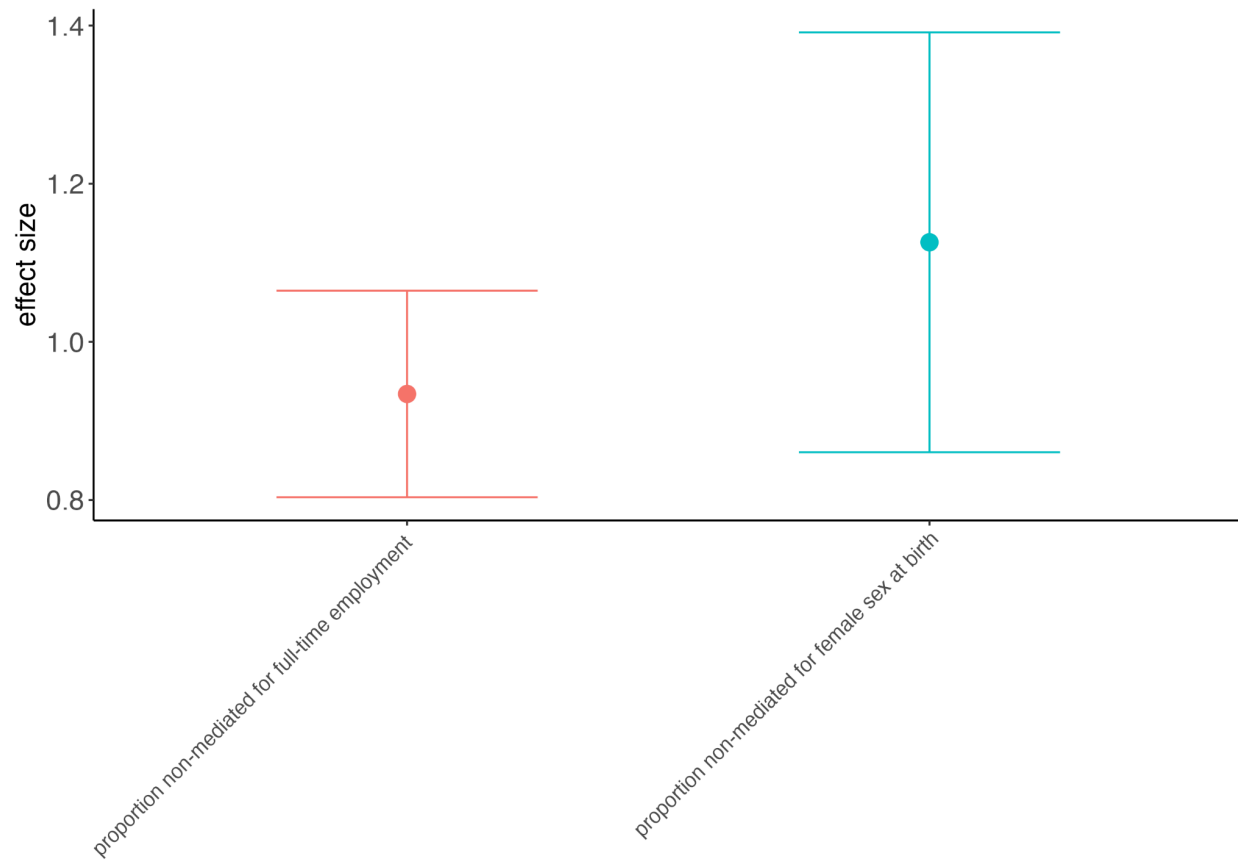


**Supplementary Figure S6. All output for UK mediation analyses. Estimated NDEs and NIEs (point estimate and 95% confidence interval) in the main analysis (A) and in the sensitivity analysis (B). Colors denote the contrast of interest (pink: full-time employment vs all other employment status categories; green: full-time employment vs unemployment; blue: female vs male sex). Estimated proportions non-mediated (point estimate and 95% confidence interval) in the main analysis (C) and in the sensitivity analysis (D).**

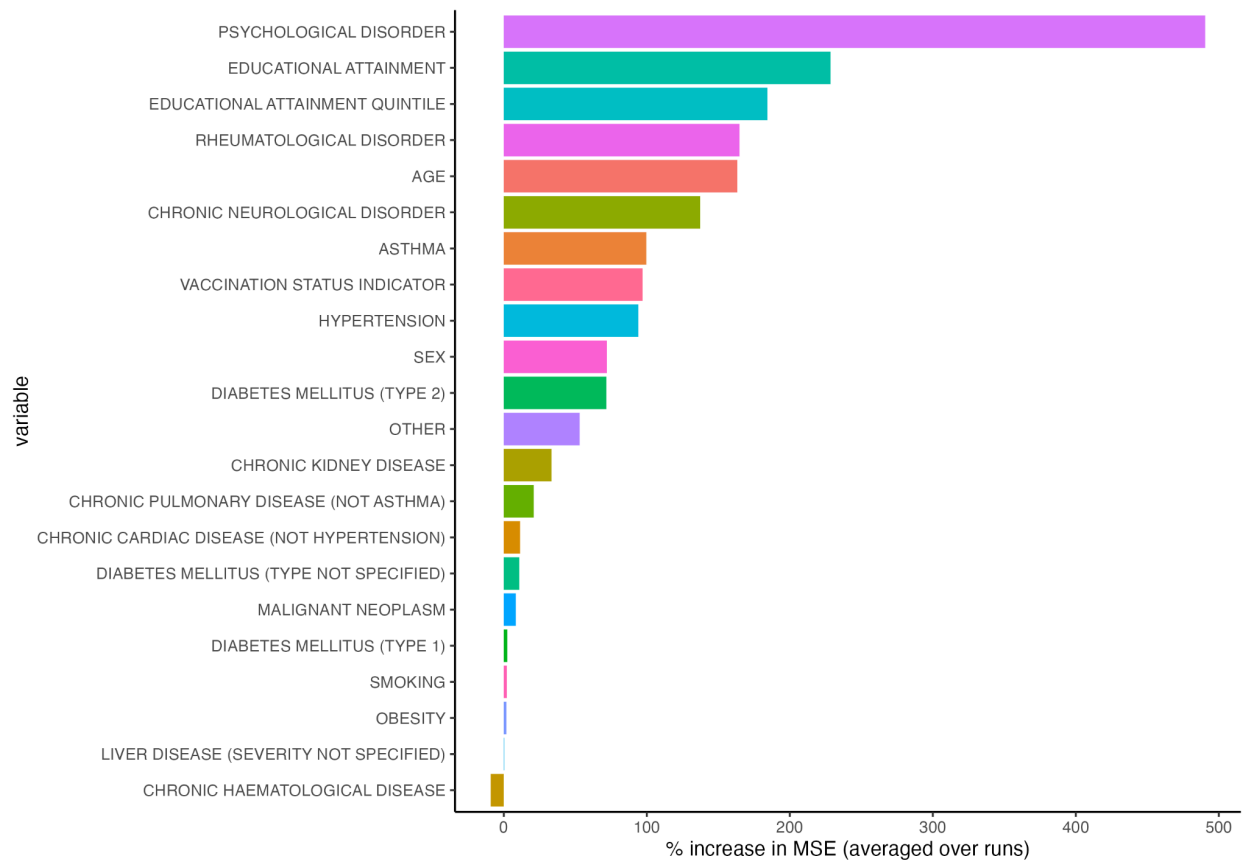
(A)



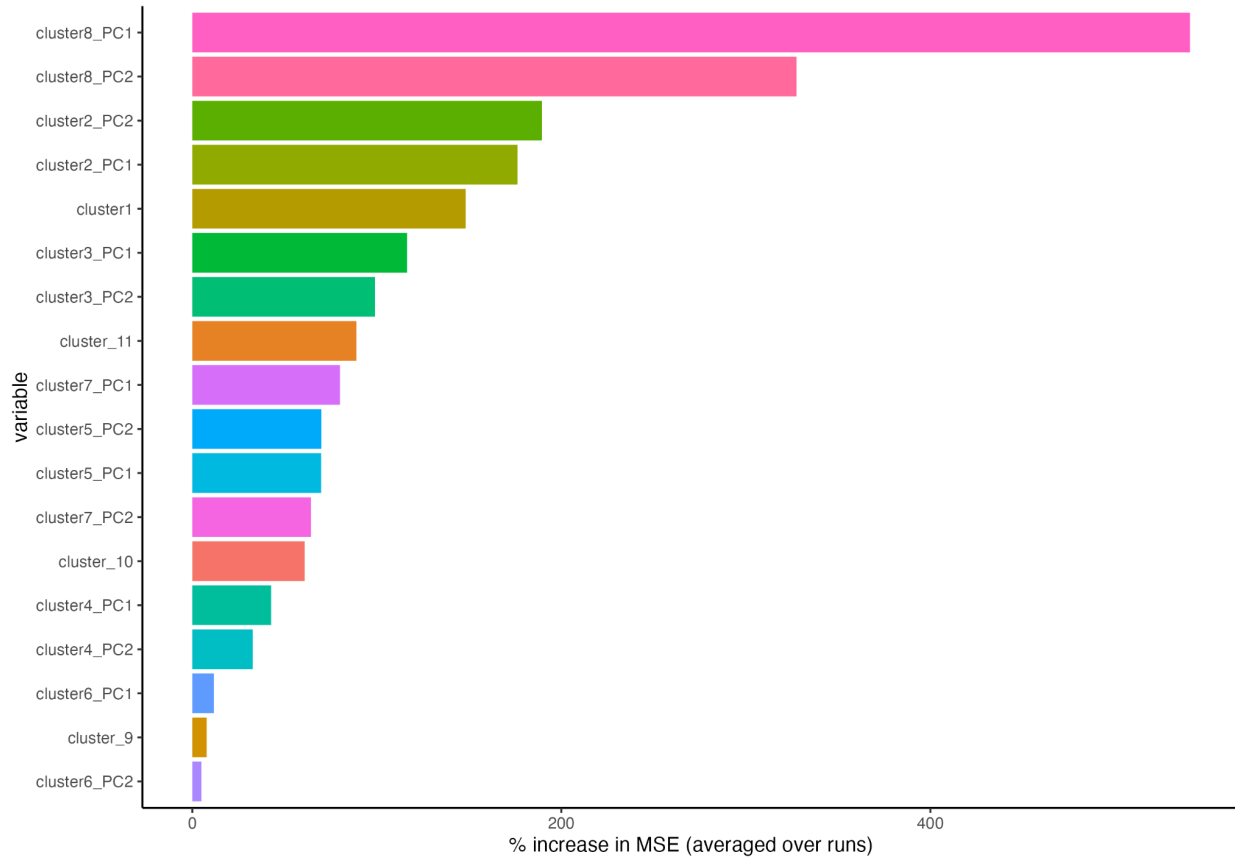
(B)



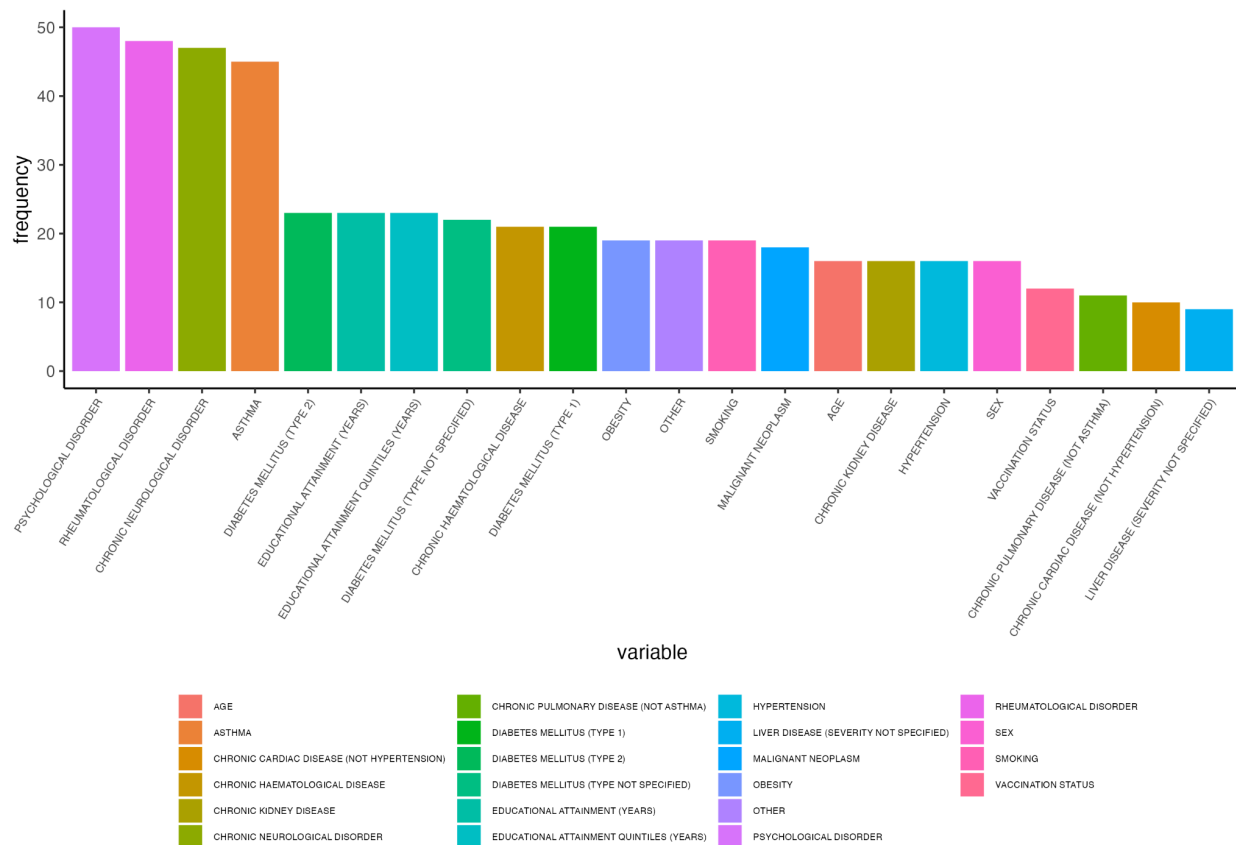
**Supplementary Figure S7. All output for the Russia mediation analyses. Estimated NDEs and NIEs (point estimate and 95% confidence interval) (A). Colors denote the contrast of interest (pink: full-time employment vs all other employment status categories; cyan: female vs male sex). Estimate proportions non-mediated (B).**



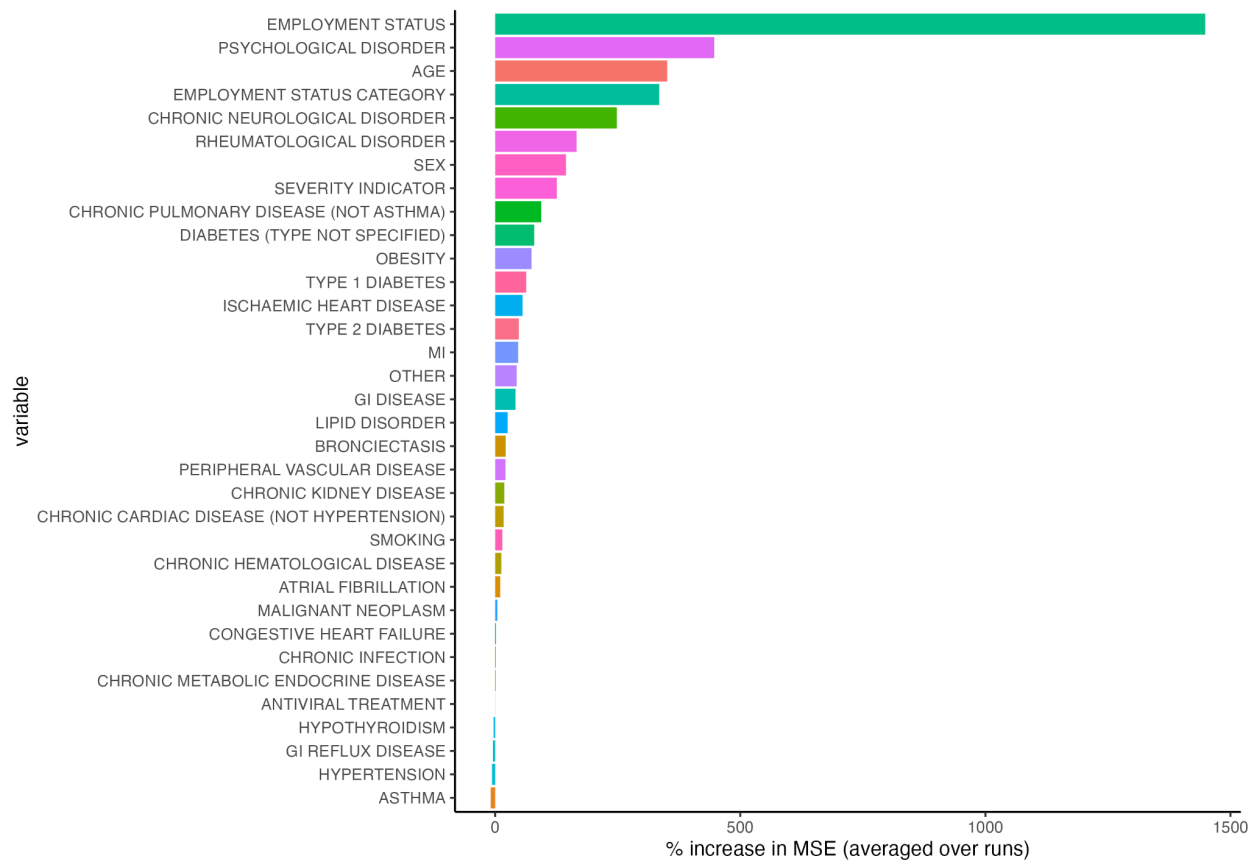
**Supplementary Figure S8a. Estimated variable importance measures, i.e. % increase in mean squared error or MSE, from individual random forest implementation (RF #1) for Norway (sex-based population correction sensitivity analysis).**



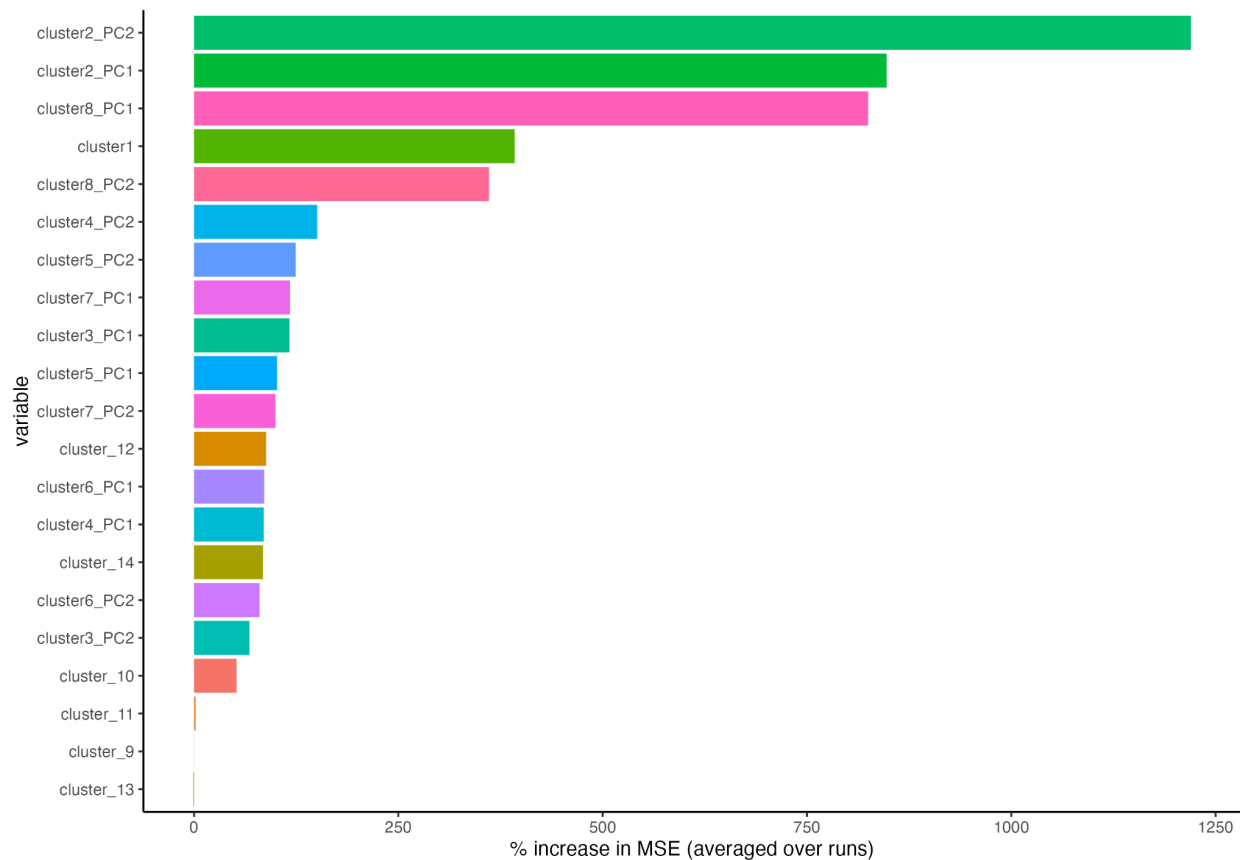
**Supplementary Figure 8b. Estimated variable importance measures, i.e. % increase in mean squared error or MSE, from pre-grouped random forest implementation (RF #2) for Norway. Rows indicate cluster names (a full list of variables belonging to each cluster can be found in Supplementary Table S3) and corresponding principal components, if the cluster consists of multiple variables. PC1 denotes principal component 1 and PC2 denotes principal component 2 (sex-based population correction sensitivity analysis).**



**Supplementary Figure S8c. Number of times (frequency) each variable appears in clusters selected for each CoV-VSURF run (RF #3) for Norway (sex-based population correction sensitivity analysis).**

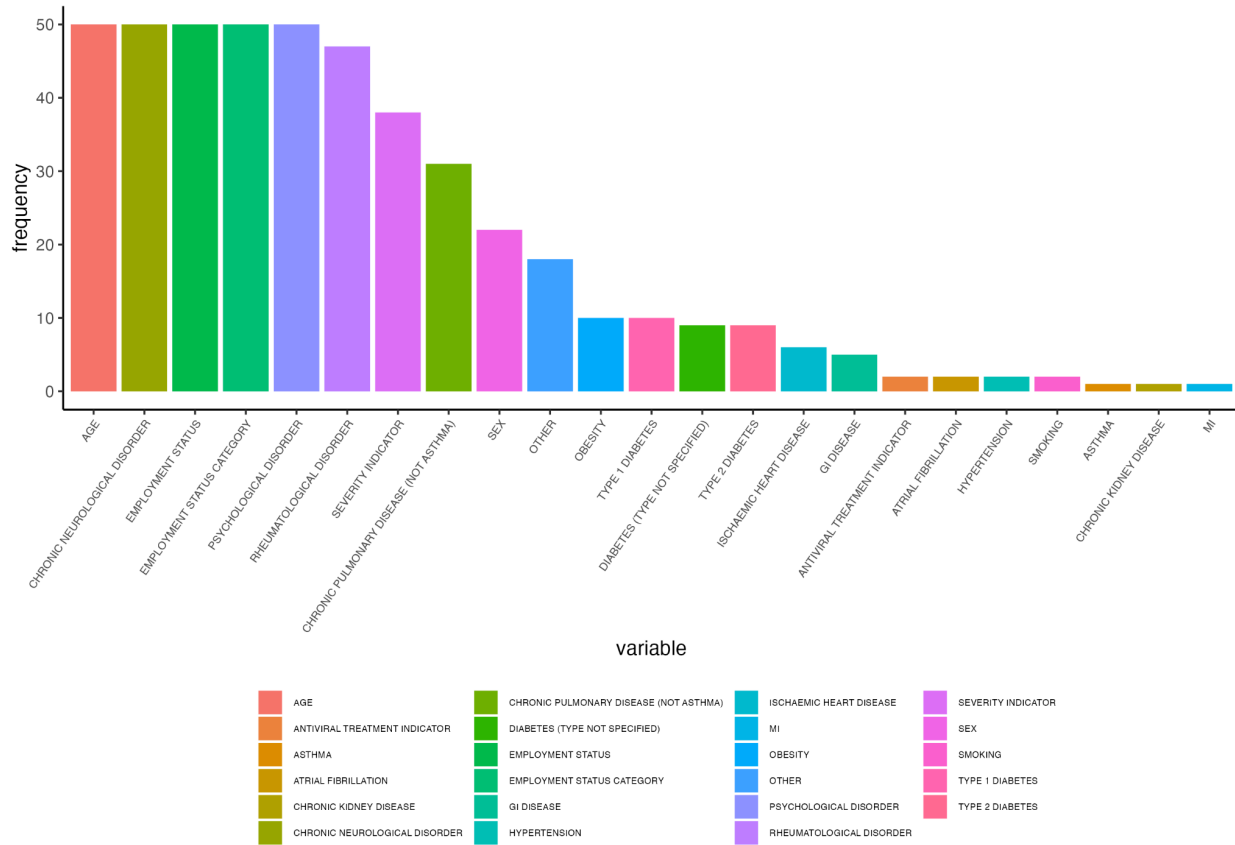


**Supplementary Figure S9a. Estimated variable importance measures, i.e. % increase in mean squared error or MSE, from individual random forest implementation (RF #1) for the UK (sex-based population correction sensitivity analysis).**



**Supplementary Figure S9b. Estimated variable importance measures, i.e. % increase in mean squared error or MSE, from pre-grouped random forest implementation (RF #2) for the UK. Rows indicate cluster names (a full list of variables belonging to each cluster can be found in Supplementary Table S3) and corresponding principal components, if the cluster consists of multiple variables. PC1 denotes principal component 1 and PC2 denotes principal component 2 (sex-based population correction sensitivity analysis).**





**Supplementary Figure S9c. Number of times (frequency) each variable appears in clusters selected for each CoV-VSURF run (RF #3) for the UK (sex-based population correction sensitivity analysis).**

## References

1. Reyes, L. F. *et al.* Clinical characteristics, risk factors and outcomes in patients with severe COVID-19 registered in the International Severe Acute Respiratory and Emerging Infection Consortium WHO clinical characterisation protocol: a prospective, multinational, multicentre, observational study. *ERJ Open Res* **8**, 00552–02021 (2022).
2. Bramante, C. T. *et al.* Outpatient treatment of COVID-19 and incidence of post-COVID-19 condition over 10 months (COVID-OUT): a multicentre, randomised, quadruple-blind, parallel-group, phase 3 trial. *The Lancet Infectious Diseases* **23**, 1119–1129 (2023).
3. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. DOI 10.18637/jss.v045.i03.
4. Pasek J (2018). anesrake: ANES Raking Implementation. R package version 0.80, <https://CRAN.R-project.org/package=anesrake>.
5. Matthew DeBell & Jon A. Krosnick (2009). *Computing Weights for American National Election Study Survey Data*. ANES Technical Report series, no. nes012427. <https://electionstudies.org/wp-content/uploads/2018/04/nes012427.pdf>.
6. ReStore National Centre for Research Methods. 5. Adjusting for non-response by weighting. <https://www.restore.ac.uk/PEAS/nonresponse.php>.
7. Pew Research Center. 1. How different weighting methods work. <https://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/>
8. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
9. A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* 2(3), 18–22.
10. Genuer R, Chavent M, Saracco J (2016). CoVVSURF: Combines clustering of variables with feature selection using random forests. R package version 0.1.
11. François Husson. Principal Component Analysis (PCA). [https://husson.github.io/MOOC\\_GB/PCA\\_course\\_slides.pdf](https://husson.github.io/MOOC_GB/PCA_course_slides.pdf)
12. Sebastien Le, Julie Josse, Francois Husson (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), 1-18. 10.18637/jss.v025.i01
13. van der Laan, M. J. & Rose, S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. (Springer Science+Business Media, 2011).
14. Hejazi, N. Causal Mediation Analysis. (2023). <https://tlverse.org/tlverse-workshops/causal-mediation-analysis.html>
15. Hastie, T. & Tibshirani, R. Generalized Additive Models. *Statistical Science* **1**, 297–318 (1986).
16. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29**, (2001).
17. Shi Y, Ke G, Soukhavong D, Lamb J, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T, Titov N (2023). lightgbm: Light Gradient Boosting Machine. R. package version 3.3.5, <https://CRAN.R-project.org/package=lightgbm>.
18. Salih, I. & Smith, S. H. Feedback neural network for pattern recognition. in (eds. Nasrabadi, N. M. & Katsaggelos, A. K.) 194–201 (San Jose, CA, 1999). doi:[10.1117/12.341120](https://doi.org/10.1117/12.341120).
19. Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
20. van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super Learner. *sagmb* **6**, (2007).

21. Coyle J, Hejazi N, Malenica I, Phillips R, Sofrygin O (2024). sl3: Pipelines for Machine Learning and Super Learning. doi:10.5281/zenodo.1342293. <https://doi.org/10.5281/zenodo.1342293>, R package version 1.4.4, <https://github.com/tlverse/sl3>.
22. Mullen KM, van Stokkum IHM (2012). nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS). R package version 1.4, <https://CRAN.R-project.org/package=nnls>.
23. Hejazi, N., Rudolph, K. & Díaz, I. medoutcon: Nonparametric efficient causal mediation analysis with machine learning in R. *JOSS* **7**, 3979 (2022).
24. Johannes Textor, Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, George T.H. Ellison (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* **45**(6):1887-1894.
25. Barrett M (2023). ggdag: Analyze and Create Elegant Directed Acyclic Graphs. R package version 0.2.10, <https://CRAN.R-project.org/package=ggdag>.
26. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
27. Sun, S., Chuang, L.-H., Sahlén, K.-G., Lindholm, L. & Norström, F. Estimating a social value set for EQ-5D-5L in Sweden. *Health Qual Life Outcomes* **20**, 167 (2022).
28. Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B. & Van Hout, B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics* **27**, 7–22 (2018).
29. Golicki, D., Jakubczyk, M., Graczyk, K. & Niewada, M. Valuation of EQ-5D-5L Health States in Poland: the First EQ-VT-Based Study in Central and Eastern Europe. *PharmacoEconomics* **37**, 1165–1176 (2019).