



HHS Public Access

Author manuscript

Am J Psychiatry. Author manuscript; available in PMC 2022 August 09.

Published in final edited form as:

Am J Psychiatry. 2020 October 01; 177(10): 944–954. doi:10.1176/appi.ajp.2020.19111158.

An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression

Karmel W. Choi, PhD,

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston

Murray B. Stein, MD, MPH,

Departments of Psychiatry and Family Medicine and Public Health, University of California, San Diego, La Jolla

Kristen M. Nishimi, MPH,

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston

Tian Ge, PhD,

Jonathan R.I. Coleman, PhD,

Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London

Chia-Yen Chen, PhD,

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston

Biogen, Cambridge, Mass.

Andrew Ratanatharathorn, PhD Candidate,

Department of Epidemiology, Columbia University Mailman School of Public Health, New York

Amanda B. Zheutlin, PhD,

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston

Correspondence: Jordan W. Smoller, MD, ScD, Psychiatric & Neurodevelopmental Genetics Unit, 185 Cambridge Street Boston, MA, 02114, jsmoller@mgh.harvard.edu, Karmel Choi, PhD, Psychiatric & Neurodevelopmental Genetics Unit, 185 Cambridge Street, Boston MA, 02114, kwchoi@mgh.harvard.edu.

Disclosures:

Dr. Stein has in the past 3 years been a consultant for Actelion, Aptinyx, Dart Neuroscience, Healthcare Management Technologies, Janssen, Jazz Pharmaceuticals, Neurocrine Biosciences, Oxeia Biopharmaceuticals, and Pfizer. Dr. Smoller is an unpaid member of the Bipolar/Depression Research Community Advisory Panel of 23andMe.

Location of work:

Psychiatric & Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston MA

Erin C. Dunn, PhD,

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine,
Massachusetts General Hospital, Boston

23andMe Research Team,

Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Gerome Breen, PhD,

Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and
Neuroscience, King's College London

Karestan C. Koenen, PhD,

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine,
Massachusetts General Hospital, Boston

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston

Jordan W. Smoller, MD, ScD

Department of Psychiatry, Massachusetts General Hospital, Boston

Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine,
Massachusetts General Hospital, Boston

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston

Abstract

Objective: Efforts to prevent depression—the leading cause of disability worldwide—have focused on a limited number candidate factors. Using phenotypic and genomic data from over 100,000 UK Biobank participants, the authors aimed to systematically screen and validate a wide range of potential modifiable factors for depression.

Methods: Baseline data were extracted for 106 modifiable factors, including lifestyle (e.g., exercise, sleep, media, diet), social (e.g., support, engagement), and environmental (e.g., greenspace, pollution) variables. Incident depression was defined as minimal depressive symptoms at baseline and clinically significant depression at follow-up. At-risk individuals for incident depression were identified based on (i) polygenic risk scores, or (ii) reported traumatic life events. An exposure-wide association scan (ExWAS) was conducted to identify factors associated with incident depression in the full sample and among at-risk individuals. Two-sample Mendelian randomization (MR) was then used to validate potentially causal relationships between identified factors and depression.

Results: Numerous factors across social, sleep, media, dietary, and exercise-related domains were prospectively associated with depression, even among at-risk individuals. However, only a subset of factors was verified by MR, including confiding in others (OR=0.76 [0.67–0.86], $p=2.53E-05$), TV use (OR=1.09 [1.05–1.13], $p=6.81E-06$), and daytime napping (OR=1.34 [1.17–1.53], $p=1.82E-05$).

Conclusions: Using a two-stage approach, this study validates several actionable targets for preventing depression. It also demonstrates that not all factors associated with depression in observational research may translate into robust targets for prevention. A large-scale exposure-wide approach combined with genetically informed methods for causal inference may help prioritize strategies for multi-modal prevention in psychiatry.

Introduction

Depression is the leading cause of disability worldwide¹, but knowledge of actionable strategies that could mitigate depression risk remains relatively limited. A number of critical research gaps have remained. First, literature to date has focused on validating a limited set of hypothesized modifiable factors for depression, such as physical activity^{2,3} or social support⁴. Without broader investigation, additional factors may be overlooked or unknown. Investigating a wide range of factors could help confirm existing relationships and also identify potentially novel prevention targets. Systematically testing the relationship between many variables and a single outcome for hypothesis-free discovery is now common practice in other fields in the form of genome- or phenome-wide association studies and has led to new insights about underlying associations^{5,6}, but has not yet been applied to identifying modifiable factors for depression.

Second, few studies to our knowledge have appraised the relative influences of multiple modifiable factors within the same population. Some factors (e.g., specific nutrients or individual foods) that show statistically significant effects when studied alone may not prove robust or as clinically relevant when considered alongside other factors⁷. Understanding the relative importance of different modifiable factors that could be integrated into prevention packages has been limited to date by modest sample sizes for multiple testing and lack of comprehensive measurements in a single study. The availability of large cohort studies, such as the UK Biobank,⁸ now make comprehensive and well-powered inquiries possible.

Third, we do not know which modifiable factors may help prevent depression among individuals at elevated risk. Two of the best substantiated risk factors for depression—genetic vulnerability and early life adversity^{9,10}—are effectively unmodifiable in adults. What generally helps prevent depression in most people may not necessarily be most relevant for those with specific risk profiles¹¹, and vice versa. Depression is now recognized as a polygenic condition¹²—influenced by many variants across the genome with individually small effects¹³. As we increasingly are able to quantify polygenic risk for depression¹⁴ and possibly return this information to individuals in the future¹⁵, it becomes vital to expand knowledge of effective actionable measures for those identified at elevated risk. Similarly, life history factors such as traumatic events are known to increase risk for depression¹⁶. As we more systematically assess established sources of genetic and environmental risk in a precision medicine framework¹⁷, evidence of modifiable factors that benefit high-risk individuals could guide recommendations to offset pre-existing vulnerabilities for depression¹⁸.

Finally, modifiable factors may be associated with depression for non-causal reasons, including unaccounted third variables (i.e., confounding) and reverse causation (e.g.,

whereby depression risk influences behavioral patterns). To strengthen conclusions about which modifiable factors may be high-priority intervention targets, Mendelian randomization (MR) analyses can be used to further test relationships between identified factors and depression. MR is an alternative strategy for causal inference that uses genetic variants inherited at birth as statistical “instruments” to approximate a natural experiment in which individuals are assigned to varying average lifetime levels of an exposure (e.g., social affiliation) in relation to an outcome of interest (e.g., depression)¹⁹. This use of genetic data bypasses typical sources of confounding in observational data and allows for triangulation of findings²⁰. We previously leveraged MR to validate a protective relationship between objectively measured physical activity and depression risk³. Here, we extend the MR approach to evaluate a wide range of possible modifiable factors.

In this study, using phenotypic and genomic data from over 100,000 UK Biobank participants without active depressive symptoms at baseline, we used an exposure-wide association study (ExWAS) design to test the relationship between 106 modifiable factors and clinically significant depression at follow-up (Figure 1). Given the established role of genetics and traumatic life events on depression risk, we also aimed to identify factors that may influence depression even in the context of these risks. Finally, we used two-sample MR to further assess directional effects and potential causal relationships between identified factors and depression.

Methods

Sample and Procedures

Our initial sample consisted of 123,794 adults of white British ancestry who enrolled in the UK Biobank, had high-quality genomic data (for quality control, see Methods S1), and completed an online follow-up mental health survey approximately six to eight years after their initial enrollment (Figure 1). Data analytic procedures were approved by the institutional review board at Partners HealthCare and conducted as part of UK Biobank application #32568. Primary data processing and statistical analyses were conducted between October 2018 and August 2019.

Measures

Depression.—At baseline, participants endorsing depressed mood and/or anhedonia (for details, see Methods S2) for more than half the days in the past two weeks were considered to have elevated depressive symptoms²¹ and were excluded from this study (n=5,416; leaving n=118,378). At follow-up, symptoms of depression were measured using all nine PHQ-9²² items, summed to create an overall score ranging from 0 to 27. To derive predicted probabilities of depression to stratify at-risk groups, we created a binary variable for clinically significant incident depression based on a score cut-off of 10²³.

Modifiable factors.—We curated data on 106 potentially modifiable factors (Table S1a) as measured or derived at baseline. These factors included behavioral (e.g., exercise, sleep, media use, diet), social (e.g., activities, support), and environmental (e.g., greenspace, pollution) variables. We selected these variables by inspecting the UK Biobank data

showcase (<http://biobank.ctsu.ox.ac.uk/crystal/index.cgi>). After review by three authors (KWC, JWS, KN), we included any variables in a domain that were (a) unlikely a close comorbidity of mental health (e.g., substance use or cognitive functioning); (b) putatively modifiable at an individual and/or societal level (e.g., lifestyle or environmental factors); and (c) largely available for most participants and not just collected for a small subset (e.g., based on branching response options). Potentially correlated variables within a category (e.g., 16-hour and 24-hour noise pollution) were retained to assess the relative influences of all available variables. As negative controls, we also selected two non-modifiable variables hypothesized to be unrelated to depression, i.e., natural hair color and skin tanning ability. Data processing was performed on all variables (described in Methods S3 and Table S1a).

Traumatic life experiences.—In the online follow-up, participants reported on their history of traumatic life experiences—including childhood physical, sexual, and emotional abuse; partner-based physical, sexual, and emotional abuse; and other lifetime traumatic events including exposure to sexual assault, violent crime, life-threatening accident, and witnessing violent death (for details, see Methods S2).

Covariates.—Baseline variables were extracted for participant characteristics (i.e., sex, age, assessment center); sociodemographic factors (i.e., socioeconomic deprivation, employment status, household income, completion of higher education, urbanicity, household size); and physical health factors (i.e., BMI, and physical illness/disability) (for details and inclusion rationale, see Methods S2).

Polygenic risk scoring

Polygenic risk scores (PRS) were generated based on large-scale genome-wide association results for major depression¹². Specifically, we used summary statistics (discovery GWAS $n=431,394$) from the Psychiatric Genomics Consortium leaving out UK Biobank data to minimize sample overlap and including 23andMe data for improved statistical power. We retained SNPs with minor allele frequency > 0.01 and INFO quality score > 0.80 . To generate PRS, we applied PRS-CS²⁴—a Bayesian polygenic prediction method that places a continuous shrinkage (CS) prior on effect sizes for all HapMap3 SNPs and infers posterior SNP weights using GWAS summary statistics combined with an external LD reference panel, such as the 1000 Genomes Project European sample (for more details and comparison with conventional clumping and thresholding, see Methods S4). We set the global shrinkage parameter at 0.01 to reflect the likely polygenic architecture of major depression. Scores were calculated by summing the number of risk alleles at each SNP multiplied by the posterior SNP weight inferred using PRS-CS, with a total of 1,090,207 included SNPs. For the distribution of scores, see Methods S4. We then extracted residuals from a model in which PRS were regressed on the top 10 European ancestry principal components provided by the UKB for use as stratification-adjusted PRS in subsequent analyses.

Stratifying participants at risk for incident depression

Among individuals with available data on later depression and risk variables (i.e., polygenic risk and reported traumatic life events) ($n=113,589$; 4.3% with incident depression), we removed a holdout training sample of 1,000 participants consisting of an even split of

randomly selected cases and controls for incident depression (for rationale, see Methods S5). In this holdout training sample, we regressed incident depression against (a) polygenic risk, or (b) reported traumatic life events (for descriptive distributions, see Table S1b). Here, each traumatic life event was entered as a separate independent variable within a multivariable model to estimate relative weights of each event on depression risk, rather than assuming equal influences. We obtained regression coefficients for each set of risk variables from the training sample (Methods S5) and used these coefficients as weights to generate predicted probability scores for incident depression for individuals in the testing sample (n=112,589—based on (a) polygenic risk, or (b) reported traumatic life events). Selecting individuals with high predicted probability scores (> 90th percentile), we obtained three groups: (i) individuals in the full sample unselected for risk (*full*; maximum n=112,589), (ii) individuals at risk based on genetic factors (*PRS*; maximum n=11,258), and (iii) individuals at risk based on reported traumatic life events (*TLE*; maximum n=11,258). Only 1563 individuals belonged to both PRS and TLE groups (13.9% of each), suggesting modest overlap and potentially distinct influences on depression (for exploratory results in this reduced sample, see Tables S2j–l and Figures S4a–c).

Exposure-wide association scan

Using an ExWAS approach with logistic regression (Methods S6), we tested associations between each baseline modifiable factor and incident depression in each of these samples (Figure 1), with a conservative Bonferroni-corrected threshold for establishing top hits ($p=0.000157$, i.e., 0.05 divided by 106 tests across three main analytic samples). All associations were adjusted for sex, baseline age, and assessment center (Model 0). We further adjusted for sociodemographic factors described earlier (Model 1), and also added physical health factors (Model 2). All analytic samples were restricted to participants who had not withdrawn from the UK Biobank (as of February 2020) and had full covariate data (*full*: maximum n=100,517; *PRS*: maximum n=10,093; *TLE*: maximum n=10,154) to ensure differences in results between successively adjusted models reflected the addition of covariates, rather than varying sample size. We also descriptively assessed the overlap between significant factors in each at-risk sample versus the full sample and between at-risk samples.

Mendelian randomization (MR) analyses

We performed bidirectional two-sample MR analyses (Methods S7) between depression and modifiable factors identified in the fully adjusted ExWAS (Model 2) in the overall sample. For each factor, we accessed the GWAS Atlas database²⁵ (<https://atlas.ctglab.nl>) to obtain publicly available UK Biobank-based summary statistics. For depression, we retained the Psychiatric Genomics Consortium summary statistics used for polygenic scoring¹². As instruments for each factor, we extracted highly associated SNPs ($p<5\times 10E-7$; for rationale, see Methods S7a) that were clumped for independence at $r^2>0.001$. Using the *TwoSampleMR* package in R²⁶, we conducted MR analyses to estimate the effect of each modifiable factor on the risk of depression, and vice versa. For primary MR analyses, we combined per-SNP effects using inverse variance weighted (IVW) meta-analysis, where the resulting estimate represents the slope of a weighted regression of SNP-outcome effects on SNP-exposure effects in which the intercept is constrained to zero. We applied

MR-PRESSO²⁷ with additional tests (i.e., Cook's distance, studentized residuals, Q-value outliers) to detect statistical outliers reflecting potential bias²⁸, and removed these outliers to generate reported estimates. We relaxed the instrument p-value threshold for several traits ($p < 5 \times 10^{-6}$; i.e., vitamin B; walking frequency) lacking sufficient SNPs (> 3) following outlier removal. We then compared the pattern of IVW results to other established MR methods—the weighted median approach²⁹ and MR Egger regression³⁰—whose estimates rely on different assumptions and are relatively robust to horizontal pleiotropy, i.e., violation of MR assumption that genetic instruments act on the outcome only via their effects on the exposure. For significant results, we further assessed horizontal pleiotropy using leave-one-SNP-out analyses, modified Cochran's Q statistic, MR Egger intercept test³¹, and manual SNP lookups. Further details are noted in Methods S7. Reported estimates were converted to odds ratios where the outcome was binary, and interpreted using a conservative p-value threshold (0.05/number of factors with available summary statistics).

Results

Modifiable factors prospectively associated with depression status in the full sample

In the full sample (for descriptives, see Methods S2d and Table S1c), 49 factors spanning multiple (e.g., physical activity, media use, sleep, social, environmental, and dietary) domains were significantly associated with depression (Model 0) (Figure S1a, Figure 1c, and Table S2a). After adjusting for sociodemographic factors (Model 1), 39 factors were significantly associated with depression (Figure S1b, Figure S1d, and Table S2b). After further adjusting for physical health factors (Model 2), 29 factors remained significantly associated with depression (Figure 2, Figure S1e, and Table S2c). Of these, 18 factors were associated with reduced odds of depression and 11 were associated with increased odds of depression (all continuous factors were standardized to mean=0 and SD=1; for variable type/scaling of each factor, see Table S1a). The top ten included six protective factors: confiding in others (aOR=0.83, 95% CI [0.82–0.85], $p=9.66E-100$); sleep duration (aOR=0.83 [0.80–0.85], $p=5.37E-33$); engaging in exercises like swimming or cycling (aOR=0.70 [0.66–0.75], $p=2.91E-25$); walking pace (aOR=0.79 [0.74–0.84], $p=3.37E-15$); being part of gym/club (aOR=0.77 [0.72–0.83], $p=3.98E-12$); and cereal intake (aOR=0.89 [0.87–0.92], $p=9.57E-12$); and four risk factors: daytime napping (aOR=1.29 [1.22–1.37], $p=1.20E-19$); computer use time (aOR=1.10 [1.07–1.13], $p=9.36E-12$); TV watching time (aOR=1.12 [1.08–1.16], $p=6.07E-12$); and cell phone use (aOR=1.10 [1.07–1.13], $p=1.25E-11$).

Factors associated with depression among at-risk individuals based on polygenic risk

Among individuals at high predicted probability for depression based on PRS, 12 factors were identified to be significantly associated with depression (Model 0) (Figure S2a, Figure S2e, and Table S2d). These reduced to ten (Model 1; Figure S2b, Figure S2f, and Table S2e) and four top factors (Model 2; Figure S2c, Figure S2g, and Table S2f) following adjustment for sociodemographic and health factors. Notably, these factors had been identified in the full sample. Of these, two appeared protective: frequency of confiding in others (aOR=0.85 [0.81–0.89], $p=2.87E-13$) and sleep duration (aOR=0.81 [0.75–0.88], $p=4.07E-07$), while the other two appeared to increase risk: computer use time (aOR=1.17 [1.09–1.26], $p=1.19E-05$), and salt intake (aOR=1.21 [1.10–1.33], $p=1.31E-04$).

Factors associated with depression among at-risk individuals based on traumatic life events

Among individuals with high predicted probability for depression based on their reported traumatic life events, 18 factors were significantly associated with depression (Model 0) (Figure S3a, Figure S3e, and Table S2g). These reduced to 16 (Model 1; Figure S3b, Figure S3f, and Table S2h) and four top factors (Model 2; Figure S3c, Figure S3g, and Table S2i) following adjustment for sociodemographic and health factors. Again, these factors had been identified in the full sample. Of these, three appeared protective: frequency of confiding in others (aOR=0.85 [0.82–0.88], $p=2.00E-22$); engaging in exercises like swimming/cycling (aOR=0.66 [0.59–0.75], $p=2.31E-10$); sleep duration (aOR=0.83 [0.79–0.89], $p=3.93E-09$); while one factor appeared to increase risk: TV watching time (aOR=1.15 [1.08–1.23], $p=5.85E-06$). Two factors (confiding in others, sleep duration) had also been identified as top factors in the PRS group, and TV time showed a similar estimate in the PRS group (aOR=1.17 [1.08–1.27]) as well. The remaining top factor (i.e., computer use, salt intake, and exercises like swimming/cycling) showed overlapping confidence intervals between PRS and TLE groups, suggesting associations may be relatively comparable across genetic or environmental risk despite not meeting the defined threshold.

Follow-up Mendelian randomization (MR) analyses

We tested all modifiable factors identified in the adjusted full sample (Model 2) with available GWAS summary statistics. Bidirectional MR analyses between each factor and depression revealed a number of findings suggesting causal relationships (Figure 4 for IVW results; weighted median results shown in Figures S5a–b and all estimates in Table S3).

MR evidence supported a beneficial effect of confiding in others (OR=0.76 [0.67–0.86], $p=2.53E-05$; 10 SNPs, Figure S6a), with non-significant effects in the reverse direction. We also found MR evidence supporting a deleterious effect of TV use (OR=1.09 [1.05–1.13], $p=6.81E-06$; 145 SNPs, Figure S6b), with non-significant effects in the reverse direction. No evidence of effect heterogeneity or horizontal pleiotropy was observed for either factor (Methods S7c). Daytime napping showed bidirectional effects with depression, such that daytime napping was linked to higher odds of depression (OR=1.34 [1.17–1.53], $p=1.82E-05$; 91 SNPs, Figure S6c) but depression was also associated with increased daytime napping (beta=0.05 [0.03–0.06], $p=8.45E-11$; 43 SNPs), with no evidence of effect heterogeneity or horizontal pleiotropy in either direction. Surprisingly, MR evidence suggested that multivitamin use was also linked to increased odds of depression (OR=1.28 [1.11–1.47], $p=6.04E-04$; 6 SNPs, Figure S6d). Given the lower number of SNPs tested, this effect was notably attenuated when further relaxing the instrument SNP threshold to $p<5\times 10E-6$ (OR=1.07 [1.0–1.14], $p=0.0498$; 30 SNPs). Depression was also nominally associated with increased intake of multivitamins (OR=1.06 [1.003–1.13], $p=4.07E-2$; 44 SNPs). Other nominal results at the $p<0.05$ threshold (Figure 4 and Methods S7d) included potential effects of tea intake, family/friend visits, and exercises such as cycling/swimming (protective) and salt intake (risk-increasing) on depression risk—and in the reverse direction, potential effects of depression risk on social group attendance, driving time (reduced), and computer use (increased).

Discussion

Although depression is a major source of suffering and lost productivity globally, successful prevention remains challenging. Using phenotypic and genomic data from the UK Biobank, we used a novel two-stage approach to screen and validate a broad panel of modifiable factors as potential prevention targets. Consistent with the multifactorial nature of depression³², we first identified a range of factors across social, media, sleep, dietary, and physical activity-related domains that were associated with incident depression over the course of study participation, both in general and among at-risk individuals. In subsequent Mendelian randomization analyses, we identified factors with convergent support across both methods, and other with discrepant evidence that may require further validation before targeting in resource-intensive trials or policy.

Among factors with convergent support, confiding in others showed the strongest phenotypic associations—even among at-risk individuals—that were substantiated by robust MR results, validating the impact of trusted social connections as causally protective for depression. Visiting with family and friends was also supported by nominally significant MR results, pointing to frequent social interactions as an additional key facet of social engagement that may be protective. Findings align with literature on social connections and mental wellbeing⁴ and with our recent study in military personnel demonstrating that greater social cohesion was linked to reduced risk of incident depression despite high genetic or environmental risk³³. Emergence of social factors as most robust among many other modifiable targets suggests that efforts to counteract disconnection at the societal and individual level—whether by social activity prescriptions³⁴ or reducing stigma of seeking emotional support—should be central to an effective depression prevention agenda. Our two-stage analyses also validated TV use as a risk factor for depression³⁵. Future work is needed to determine whether this effect is due to screen time or media exposure per se, or whether TV time serves more generally as a proxy for sedentary behavior, which was not explicitly measured in the full sample but has been identified as a risk factor for depression³⁶. Regardless, findings suggest that assessing media use patterns in adults (e.g., by health providers) and providing psychoeducation around potential mood impacts of excess TV watching could represent another effective component of depression prevention. Finally, daytime napping emerged unexpectedly with bidirectional influences in the MR context; that is, a tendency for daytime napping in adults appeared to increase risk of depression but depression itself may be a cause of increased napping.

A substantial number of associated factors were not supported by current MR evidence, for several possible reasons. First, not all modifiable factors—even those prospectively related to depression—may be causal in their effects on depression risk and thus represent weaker targets for prevention. For example, bidirectional MR evidence suggested that factors such as increased computer use or vitamin B supplementation are more likely to be consequences of depression than causes, such that depressed individuals may tend to spend more time on the computer or take supplements. It may be useful to leverage these factors as early indicators of depression rather than direct modifiers of depression risk. Causality notwithstanding, the co-occurrence of depression risk with a range of health-relevant behaviors highlights a potential mechanism for physical morbidities (e.g.,

cardio-metabolic disease, premature mortality) often associated with this condition, which could inform preventative interventions to reduce health disparities individuals with/at risk of depression³⁷.

Second, the relationship between certain modifiable factors and depression may not be straightforward, requiring more nuanced study. For example, overall reported sleep duration, which was related to incident depression but not substantiated by MR, may have complex and non-linear effects³⁸ that could not be fully explored in this study but could be probed in future MR studies with more detailed sleep-related phenotypes³⁹. Geo-coded environmental exposures such as pollution or natural space also showed associations⁴⁰ that did not persist after adjusting for sociodemographic factors and were thus not tested in MR. It may be that such environmental exposures exert stronger influences earlier in development⁴¹ or depend on heterogeneous features (e.g., tree canopy versus grass coverage⁴²) under consideration.

Third, although we adjusted for sociodemographic and health factors, residual confounding could explain some observed associations. For example, various dietary factors associated with depression (e.g., cereal consumption, lamb intake; vitamin B supplementation) were not supported in MR and may instead reflect behavioral patterns such as daily routines, social rituals, or health concerns that affect mental health more broadly. Despite popular views of vitamin B as a mood-boosting supplement, our findings align with a current lack of randomized trial evidence supporting beneficial effects on depression⁴³. Among the more surprising findings, multivitamin use was not only associated with increased depression but also supported by MR evidence, though attenuated in sensitivity analyses. Given sparse evidence to date⁴⁴, this finding should be interpreted with caution unless supported by further data, though an intriguing (but non-significant) trend for multivitamin supplementation and higher odds of depression was recently observed in a multi-site randomized trial for depression prevention⁴⁵. We also found evidence suggesting reverse causation, whereby depressed individuals may tend to take multivitamins.

Fourth, the strength of current genetic instruments may have contributed to discrepancies between phenotypic and MR associations. Although physical activity variables showed some of the largest protective relationships with incident depression, their effects were not bolstered in MR. We previously observed that while influences of objectively measured physical activity (not included here) on depression were validated in MR³, self-report measures did not show these patterns. Objective measures—capturing a broad tendency for movement—demonstrate higher heritability⁴⁶ and may yield more powerful genetic instruments. Indeed, self-reported activity variables, as well as dietary factors, tended to have fewer genome-wide significant SNPs than other traits (e.g., media use). Nonetheless, nominal MR results suggested that liability for engaging in exercises like swimming/cycling (protective) and salt intake (risk) may affect depression risk, meriting further inquiry.

Our study should be evaluated in light of several limitations. First, while we considered a wide array of lifestyle and environmental factors, we were limited by available variables in the UK Biobank database. These did not include modifiable psychological factors (e.g., coping styles) that could also influence depression risk. Second, although the exposure-wide design is a major strength, some associations—potentially noteworthy if studied alone—may

Acknowledgements:

This research was conducted using the UK Biobank resource under an approved data request (#32568). This work involved the use of the Enterprise Research Infrastructure & Services (ERIS) at Partners HealthCare. K.W.C. was supported in part by a NIMH T32 Training Fellowship (T32MH017119). J.W.S is a Tepper Family MGH Research Scholar and supported in part by the Demarest Lloyd, Jr. Foundation. T.G. is supported in part by NIA grant K99AG054573. J.R.I.C and G.B. are funded partly by the UK National Institute of Health Research (NIHR), and partly by a grant from Cohen Veterans Bioscience. This paper represents independent research funded in part by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NIH, NHS, NIHR or the Department of Health and Social Care.

We would also like to thank the research participants and employees of 23andMe, Inc. for making this work possible. The following members of the 23andMe Research Team contributed to this study: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Xin Wang, Catherine H. Wilson.

References

1. World Health Organization. Depression. Fact sheets. March 2018. <https://www.who.int/news-room/fact-sheets/detail/depression>.
2. Schuch FB, Vancampfort D, Firth J, et al. Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. *American Journal of Psychiatry*. April 2018;appi.ajp.2018.1. doi:10.1176/appi.ajp.2018.17111194
3. Choi KW, Chen C-Y, Stein MB, et al. Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry*. January 2019. doi:10.1001/jamapsychiatry.2018.4175
4. Santini ZI, Koyanagi A, Tyrovolas S, Mason C, Haro JM. The association between social relationships and depression: A systematic review. *Journal of Affective Disorders*. 2015;175:53–65. doi:10.1016/j.jad.2014.12.049 [PubMed: 25594512]
5. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5–22. doi:10.1016/j.ajhg.2017.06.005 [PubMed: 28686856]
6. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annual Review of Genomics and Human Genetics*. 2016;17(1):353–373. doi:10.1146/annurev-genom-090314-024956
7. Ioannidis JPA. Neglecting Major Health Problems and Broadcasting Minor, Uncertain Issues in Lifestyle Science. *JAMA*. October 2019;1. doi:10.1001/jama.2019.17576
8. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779 [PubMed: 25826379]
9. Sullivan PF, Neale MC, Kendler KS. Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *American Journal of Psychiatry*. 2000;157(10):1552–1562. doi:10.1176/appi.ajp.157.10.1552 [PubMed: 11007705]
10. Köhler CA, Evangelou E, Stubbs B, et al. Mapping risk factors for depression across the lifespan: An umbrella review of evidence from meta-analyses and Mendelian randomization studies. *Journal of Psychiatric Research*. 2018;103:189–207. doi:10.1016/j.jpsychires.2018.05.020 [PubMed: 29886003]
11. Patel V, Goodman A. Researching protective and promotive factors in mental health. *Int J Epidemiol*. 2007;36(4):703–707. doi:10.1093/ije/dym147 [PubMed: 17646185]
12. Wray NR, eQTLGen, 23andMe, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*. 2018;50(5):668–681. doi:10.1038/s41588-018-0090-3 [PubMed: 29700475]

13. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 2014;55(10):1068–1087. doi:10.1111/jcpp.12295 [PubMed: 25132410]
14. McIntosh AM, Sullivan PF, Lewis CM. Uncovering the Genetic Architecture of Major Depression. *Neuron*. 2019;102(1):91–103. doi:10.1016/j.neuron.2019.03.022 [PubMed: 30946830]
15. Palk AC, Dalvie S, de Vries J, Martin AR, Stein DJ. Potential use of clinical polygenic risk scores in psychiatry - ethical implications and communicating high polygenic risk. *Philos Ethics Humanit Med*. 2019;14(1):4. doi:10.1186/s13010-019-0073-8 [PubMed: 30813945]
16. Tennant C Life Events, Stress and Depression: A Review of Recent Findings. *Australian & New Zealand Journal of Psychiatry*. 2002;36(2):173–182. doi:10.1046/j.1440-1614.2002.01007.x
17. Prendes-Alvarez S, Nemeroff CB. Personalized medicine: Prediction of disease vulnerability in mood disorders. *Neurosci Lett*. 2018;669:10–13. doi:10.1016/j.neulet.2016.09.049 [PubMed: 27746310]
18. Choi KW, Stein MB, Dunn EC, Koenen KC, Smoller JW. Genomics and psychological resilience: a research agenda. *Mol Psychiatry*. July 2019. doi:10.1038/s41380-019-0457-6
19. Byrne EM, Yang J, Wray NR. Inference in Psychiatry via 2-Sample Mendelian Randomization- From Association to Causal Pathway? *JAMA Psychiatry*. 2017;74(12):1191–1192. doi:10.1001/jamapsychiatry.2017.3162 [PubMed: 29094155]
20. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. *Nature*. 2018;553(7689):399–401. doi:10.1038/d41586-018-01023-3
21. Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med*. 2010;8(4):348–353. doi:10.1370/afm.1139 [PubMed: 20644190]
22. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613. doi:10.1046/j.15251497.2001.016009606.x [PubMed: 11556941]
23. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ*. 2012;184(3):E191–196. doi:10.1503/cmaj.110829 [PubMed: 22184363]
24. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;10(1). doi:10.1038/s41467-019-09718-5
25. Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*. 2019;51(9):1339–1348. doi:10.1038/s41588-019-0481-0 [PubMed: 31427789]
26. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7. doi:10.7554/eLife.34408
27. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693–698. doi:10.1038/s41588-018-0099-7 [PubMed: 29686387]
28. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;27(R2):R195–R208. doi:10.1093/hmg/ddy163 [PubMed: 29771313]
29. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology*. 2016;40(4):304–314. doi:10.1002/gepi.21965 [PubMed: 27061298]
30. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*. 2015;44(2):512–525. doi:10.1093/ije/dyv080 [PubMed: 26050253]
31. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology*. 2017;32(5):377–389. doi:10.1007/s10654-017-0255-x [PubMed: 28527048]

32. Sjöholm L, Lavebratt C, Forsell Y. A multifactorial developmental model for the etiology of Major Depression in a population-based sample. *Journal of Affective Disorders*. 2009;113(1–2):66–76. doi:10.1016/j.jad.2008.04.028 [PubMed: 18558436]
33. Choi KW, Chen C-Y, Ursano RJ, et al. Prospective study of polygenic risk, protective factors, and incident depression following combat deployment in US Army soldiers. *Psychological Medicine*. April 2019:1–9. doi:10.1017/S0033291719000527
34. Drinkwater C, Wildman J, Moffatt S. Social prescribing. *BMJ*. March 2019:11285. doi:10.1136/bmj.11285 [PubMed: 30923039]
35. de Wit L, van Straten A, Lamers F, Cuijpers P, Penninx B. Are sedentary television watching and computer use behaviors associated with anxiety and depressive disorders? *Psychiatry Research*. 2011;186(2–3):239–243. doi:10.1016/j.psychres.2010.07.003 [PubMed: 20692047]
36. Teychenne M, Ball K, Salmon J. Sedentary Behavior and Depression Among Adults: A Review. *International Journal of Behavioral Medicine*. 2010;17(4):246–254. doi:10.1007/s12529-010-9075-z [PubMed: 20174982]
37. Firth J, Siddiqi N, Koyanagi A, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *The Lancet Psychiatry*. 2019;6(8):675–712. doi:10.1016/S2215-0366(19)30132-4 [PubMed: 31324560]
38. Zhai L, Zhang H, Zhang D. Sleep duration and depression among adults: A meta-analysis of prospective studies. *Depression and Anxiety*. 2015;32(9):664–670. doi:10.1002/da.22386 [PubMed: 26047492]
39. Jones SE, van Hees VT, Mazzotti DR, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nature Communications*. 2019;10(1). doi:10.1038/s41467-019-09576-1
40. Sass V, Kravitz-Wirtz N, Karceski SM, Hajat A, Crowder K, Takeuchi D. The effects of air pollution on individual psychological distress. *Health & Place*. 2017;48:72–79. doi:10.1016/j.healthplace.2017.09.006 [PubMed: 28987650]
41. Engemann K, Pedersen CB, Arge L, Tsirogiannis C, Mortensen PB, Svenning J-C. Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. *Proceedings of the National Academy of Sciences*. 2019;116(11):5188–5193. doi:10.1073/pnas.1807504116
42. Astell-Burt T, Feng X. Association of Urban Green Space With Mental Health and General Health Among Adults in Australia. *JAMA Network Open*. 2019;2(7):e198209. doi:10.1001/jamanetworkopen.2019.8209 [PubMed: 31348510]
43. Firth J, Teasdale SB, Allott K, et al. The efficacy and safety of nutrient supplements in the treatment of mental disorders: a meta-review of meta-analyses of randomized controlled trials. *World Psychiatry*. 2019;18(3):308–324. doi:10.1002/wps.20672 [PubMed: 31496103]
44. Long S-J, Benton D. Effects of vitamin and mineral supplementation on stress, mild psychiatric symptoms, and mood in nonclinical samples: a meta-analysis. *Psychosom Med*. 2013;75(2):144–153. doi:10.1097/PSY.0b013e31827d5fbd [PubMed: 23362497]
45. Bot M, Brouwer IA, Roca M, et al. Effect of Multinutrient Supplementation and Food-Related Behavioral Activation Therapy on Prevention of Major Depressive Disorder Among Overweight or Obese Adults With Subsyndromal Depressive Symptoms: The MoodFOOD Randomized Clinical Trial. *JAMA*. 2019;321(9):858. doi:10.1001/jama.2019.0556 [PubMed: 30835307]
46. Klimentidis YC, Raichlen DA, Bea J, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including *CADM2* and *APOE*. *International Journal of Obesity*. June 2018. doi:10.1038/s41366-018-0120-3
47. Choi KW, Zheutlin AB, Karlson RA, et al. Physical activity offsets genetic risk for incident depression assessed via electronic health records in a biobank cohort study. *Depression and Anxiety*. 2020;37(2):106–114. doi:10.1002/da.22967 [PubMed: 31689000]
48. Adams MJ, Hill WD, Howard DM, et al. Factors associated with sharing e-mail information and mental health survey participation in large population cohorts. *Int J Epidemiol*. July 2019. doi:10.1093/ije/dyz134

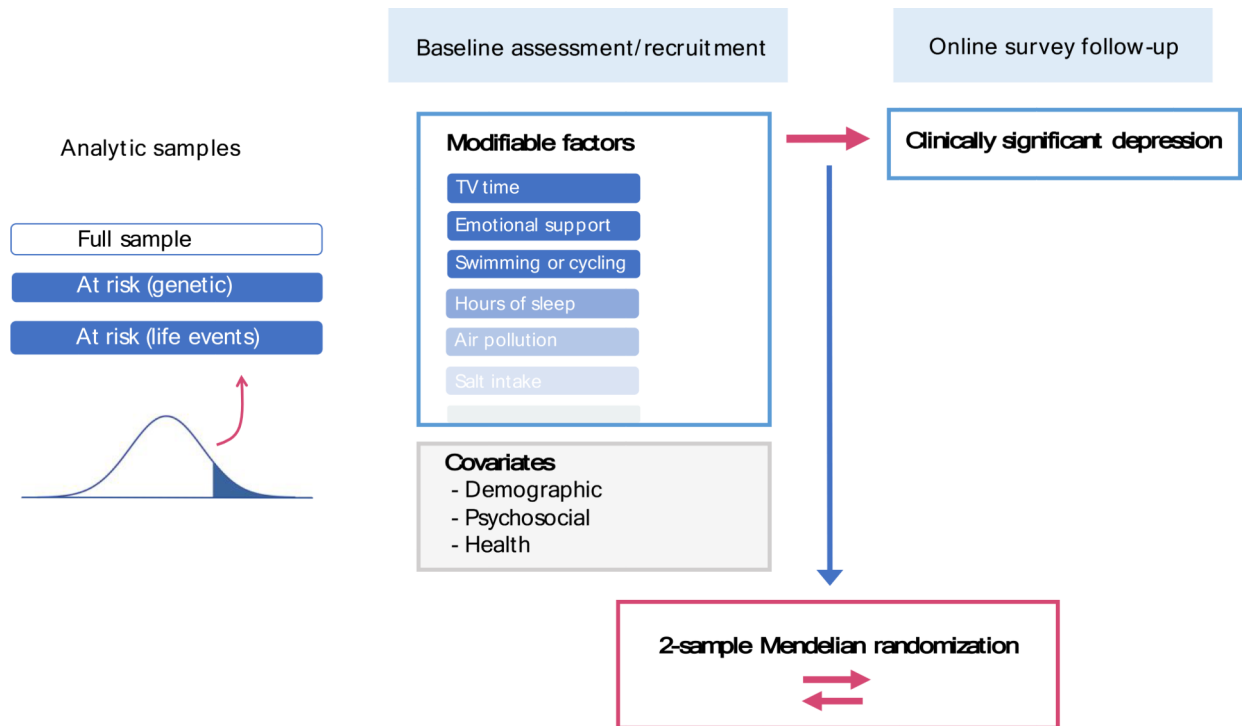


Figure 1. Overview of analytic design to test prospective associations between modifiable factors and subsequent depression.

Associations were tested in three analytic samples: (a) full sample; (b) at-risk individuals based on polygenic risk; and (c) at-risk individuals based on reported traumatic life events. To reduce bias in associations from contemporaneous reporting, modifiable factors were selected from those indexed to the baseline assessment, while subsequent depression was assessed at the follow-up survey approximately six to eight years later. Key distinctions with previous depression analyses in the UK Biobank are summarized in Supplementary Methods S0, emphasizing our targeted focus on modifiable factors for depression in a prospective design and among different risk groups.

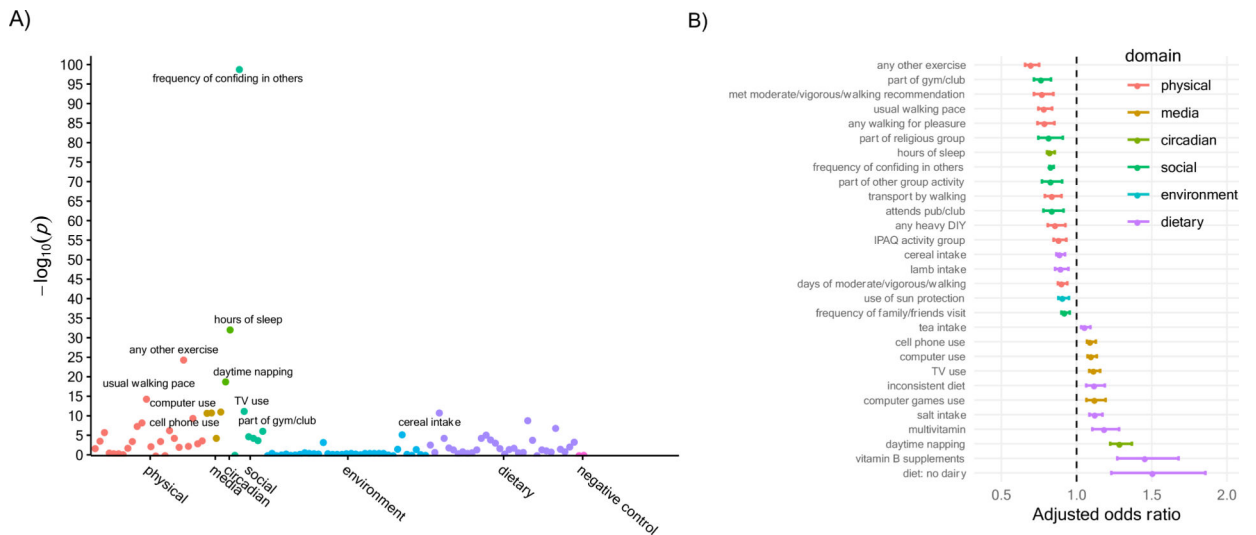


Figure 2. Association results between modifiable factors and clinically significant depression in the full sample, adjusted for sociodemographic and health factors (Model 2).

A) Association plot for modifiable factors in relation to incident depression, with x-axis organized by conceptual domains, y-axis showing statistical significance as $-\log_{10}$ of p-value, and red horizontal line showing the significance threshold corrected for multiple testing. **B)** Adjusted odds ratios for significant factors, in ascending order (i.e., from risk-reducing to risk-increasing). Full set of association results can be found in Tables S2a–c.

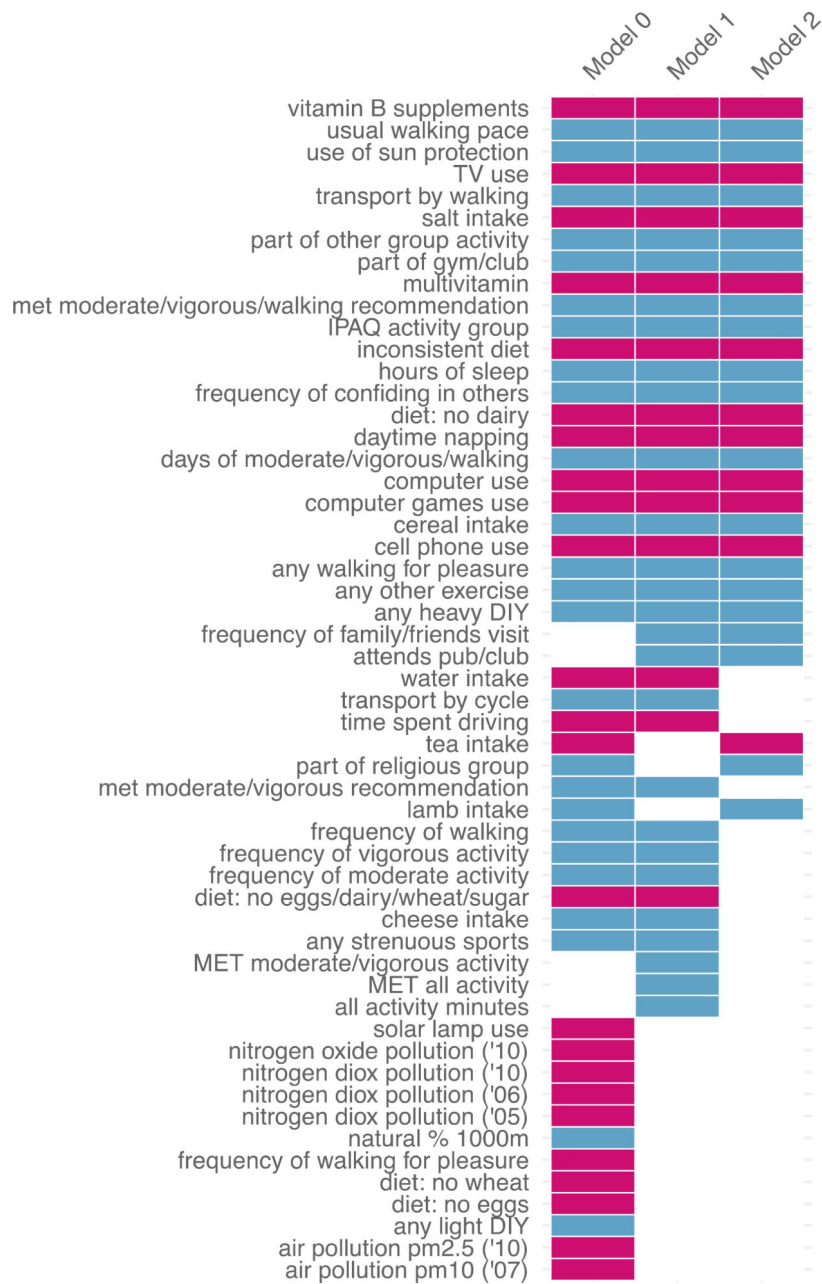


Figure 3. Consistency of top associated factors across levels of covariate adjustment.

Shown in order of consistency patterns across three, two, or one models, in descending alphabetical order within each pattern. Blue = reduced odds of depression; red = increased odds of depression. Results shown only for factors with significant associations in at least one model. Full set of association results can be found in Tables S2a–c.

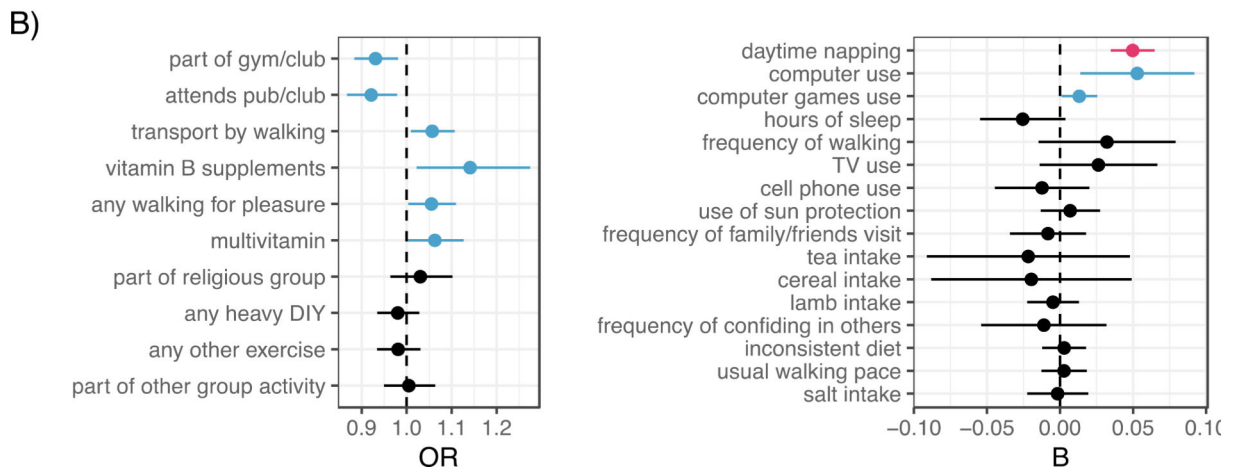
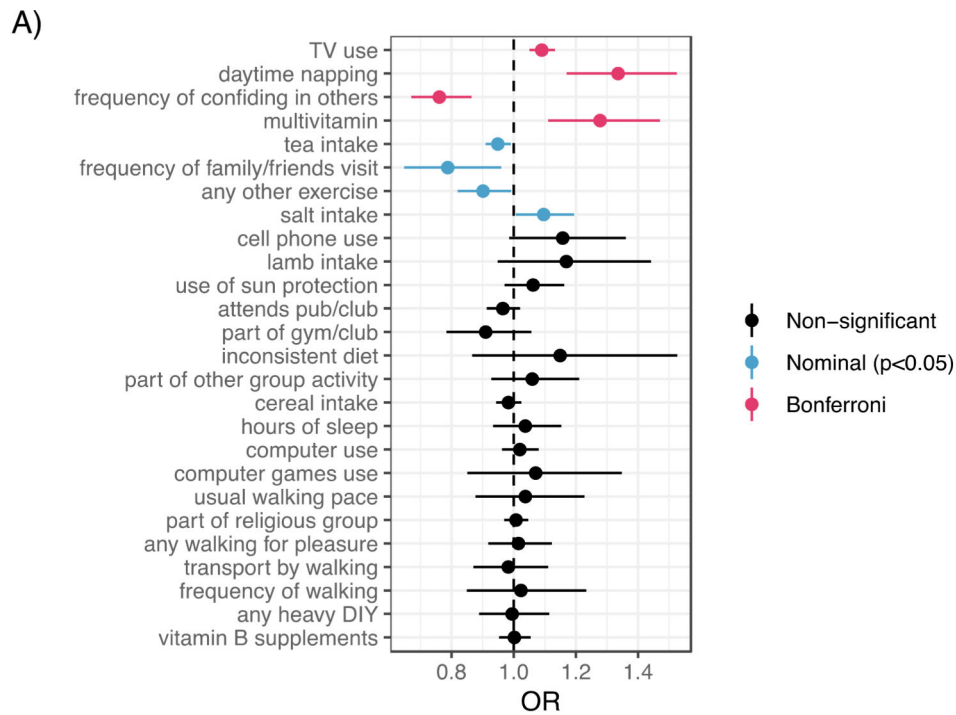


Figure 4. **A)** MR estimates of top modifiable factors → the risk of depression with outliers removed, based on the inverse-variance weighted method (for the weighted median method, see Figure S26). **B)** MR estimates of depression → top modifiable factors with outliers removed, based on the inverse-variance weighted method (for the weighted median method, see Figure S27). Odds ratio estimates shown on left for dichotomous factors as outcomes, and beta estimates shown on right for non-dichotomous factors as outcomes.