Journal of Biomedical Informatics 87 (2018) 88–95

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Association networks in a matched case-control design – Co-occurrence patterns of preexisting chronic medical conditions in patients with major depression versus their matched controls

Min-hyung Kim[a], Samprit Banerjee[b], Yize Zhao[b], Fei Wang[a], Yiye Zhang[a], Yongjun Zhu[c], Joseph DeFerio[a], Lauren Evans[b], Sang Min Park[d,e,*], Jyotishman Pathak[a,*]

[a] Division of Health Informatics, Department of Health Policy and Research, Weill Cornell Medical College of Cornell University, NY, USA
[b] Division of Biostatistics and Epidemiology, Department of Health Policy and Research, Weill Cornell Medical College of Cornell University, NY, USA
[c] Department of Library and Information Science, Sungkyungkwan University, Seoul, Republic of Korea
[d] Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[e] Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Matched pair analysis
Network analysis
Interaction analysis
Chronic medical conditions
Major depressive disorder

ABSTRACT

Objective: We present a method for comparing association networks in a matched case-control design, which provides a high-level comparison of co-occurrence patterns of features after adjusting for confounding factors. We demonstrate this approach by examining the differential distribution of chronic medical conditions in patients with major depressive disorder (MDD) compared to the distribution of these conditions in their matched controls.

Materials and methods: Newly diagnosed MDD patients were matched to controls based on their demographic characteristics, socioeconomic status, place of residence, and healthcare service utilization in the Korean National Health Insurance Service's National Sample Cohort. Differences in the networks of chronic medical conditions in newly diagnosed MDD cases treated with antidepressants, and their matched controls, were prioritized with a permutation test accounting for the false discovery rate. Sensitivity analyses for the associations between prioritized pairs of chronic medical conditions and new MDD diagnosis were performed with regression modeling.

Results: By comparing the association networks of chronic medical conditions in newly diagnosed depression patients and their matched controls, five pairs of such conditions were prioritized among 105 possible pairs after controlling the false discovery rate at 5%. In sensitivity analyses using regression modeling, four out of the five prioritized pairs were statistically significant for the interaction terms.

Conclusion: Association networks in a matched case-control design can provide a high-level comparison of comorbid features after adjusting for confounding factors, thereby supplementing traditional clinical study approaches. We demonstrate the differential co-occurrence pattern of chronic medical conditions in patients with MDD and prioritize the chronic conditions that have statistically significant interactions in regression models for depression.

1. Introduction

Depression is a highly prevalent disease that co-occurs with many chronic conditions over the course of an individual's lifetime [1–3]. Depression also has a large societal burden, including high utilization of healthcare services that are not directly related to the treatment of depression [4]. Depression is associated with many co-occurring medical conditions in a bi-directional manner [5–9] and may also affect the management of existing medical conditions [10–12]. Depression in the presence of other comorbid medical conditions is also more likely to be treatment-resistant [13–15].

Several statistical and machine learning methods have been employed to build risk prediction models for depression using a large number of predictors[16–19]. In our previous work, we quantified the

* Corresponding authors at: 101 Daehangno, Jongno-gu, Seoul 110-744, Republic of Korea (S.M. Park). 425 East 61st Street, Suite 301, New York, NY 10065, USA (J. Pathak).
E-mail addresses: fmpark1@snu.ac.kr (S.M. Park), jyp2001@med.cornell.edu (J. Pathak).

https://doi.org/10.1016/j.jbi.2018.09.016
Received 17 March 2018; Received in revised form 25 September 2018; Accepted 28 September 2018
Available online 06 October 2018
1532-0464/

impact of chronic conditions on the diagnosis of major depressive disorder (MDD) using conditional logistic regression and gradient boosting machine models using data from a regional electronic health record (EHR) system [18]. In a previous analysis, we employed a regularized logistic regression model with Elastic Net penalties on a randomized sample from a population-based longitudinal cohort. This demonstrated that inclusion of chronic medical conditions as predictors for MDD can improve the performance of the prediction model [19]. The improvement in predictive performance served as the motivation for the present study, where we explore the complex inter-relationships between chronic medical conditions in the context of depression in a comprehensive way using network analysis.

Network analysis is a method for understanding complex relationships among multiple entities [20,21]. In biomedicine, network analysis has been widely used in Genome-Wide Association Studies and Protein-Protein Interaction Network Studies to better understand the relationships between multiple genes and proteins, as well as to discover candidate genes and proteins that are related to the development of diseases, and to identify potential targets for treatment [20]. Network analysis has also been applied to the study of disease progression and patterns of disease co-occurrence using data from health insurance claims and EHRs [21–27]. However, to the best of our knowledge, network analysis has not yet been applied to a population-based clinical data with a matched case-control design.

In this study, we selected a random sample from a longitudinal cohort of Korean nationals provided by Korean National Health Insurance Service. Using this data, we present a method that compares the network of co-occurring chronic conditions in individuals with MDD to that observed in their matched controls. In addition, we prioritize chronic conditions that have a differential correlation in depressed patients compared to that in their matched controls, and show that some chronic conditions have statistically significant interactions in regression models for MDD.

2. Materials and methods

2.1. Study data

The data for this study was provided by the Korean National Health Insurance Service, a standardized twelve-year longitudinal cohort that provides public health researchers and policy makers with population-representative information regarding health status and healthcare services utilization [28]. A sample cohort of 1,025,340 individuals (2.2% of the total population of 46,605,433) was established in 2002 from a stratified random sampling with a proportional allocation from the Korean National Health Information Database [28–30]. The data include demographic characteristics, socioeconomic status, health insurance claims data (including inpatient, outpatient, and pharmacy claims), information derived from the death registry, and national health check-up data. The national health check-ups are provided to those who are aged 40 years or above to facilitate early detection of chronic medical conditions and risk factors. The diagnosis codes in the Korean National Health Information Database are based on the Korean Classification of Diseases, which is compatible with the International Classification of Diseases, Tenth Revision (ICD-10) by the World Health Organization (WHO) [28,31].

2.2. Study population

In this study, we had access to the data from January 1, 2002, to December 31, 2013. We used the first five years of data (2002–2006) to identify cohort inclusion and exclusion criteria, as well as predisposing features (or exposures). The cohort entry date was then assigned to January 1, 2007, and the follow-up observation period extended from this entry date through December 31, 2013 to identify newly diagnosed MDD cases.

As of the entry date, we included all individuals in the cohort who (1) were aged 40 years or above, (2) did not have any diagnosis of MDD (ICD-10 codes F32.x and F33.x) prior to the cohort entry date, and (3) did not have a diagnosis of either bipolar disorder (ICD-10 codes F31.x) or schizophrenia (ICD-10 codes F20.x).

Within the included cohort members, we defined the members of the nested case group as those who (1) received a new diagnosis of MDD on the cohort entry date or later, (2) the diagnosis of MDD was recorded at least twice, and (3) were prescribed antidepressants, but (4) did not have a diagnosis of either bipolar disorder or schizophrenia *before* the diagnosis of MDD. We chose this operational definition because a claims-based patient identification algorithm requiring at least two diagnoses of depression and an antidepressant prescription has shown to have high agreement with clinically diagnosed depression [32].

Within the included cohort members, we matched controls among the cohort members who were at risk of becoming cases (i.e., alive, but not yet diagnosed with MDD, bipolar disorder, or schizophrenic disorder) on the same day using the incidence density sampling scheme [33,34]. Specifically, for each time point when the nested case was defined, we sampled a control matched on demographic characteristics (i.e., age and sex), socioeconomic status (i.e., household income decile and financial dependency status), health insurance status (i.e., medical aid status), geographical location (i.e., metropolitan residence), and healthcare services utilization (i.e., number of hospitalizations per year, number of outpatient visits per year).

Individuals who met the inclusion criteria and entered the cohort on January 1, 2007, were followed up until December 31, 2013, and a total of 10,299 new MDD cases were identified and matched with a control. Of note, among the 10,299 cases, 163 patients were later diagnosed with bipolar disorder, and 41 patients were later diagnosed with schizophrenia by December 31, 2013.

2.3. Chronic medical conditions for analysis

A total of 15 chronic medical conditions were analyzed in this study: anemia, arthritis, asthma, cataract, chronic kidney disease, chronic obstructive pulmonary disease (COPD), diabetes mellitus, dementia, hyperlipidemia, heart failure, hypothyroidism, hypertension, ischemic heart disease, osteoporosis, and stroke. Chronic medical conditions were identified from the diagnosis codes recorded *prior* to the cohort entry date. The diagnosis codes for chronic medical conditions were classified using the Chronic Conditions Data Warehouse Condition Algorithms (rev. 11/2017) by Centers for Medicare & Medicaid Services (CMS) [35], which is based on the modified version of ICD-10-CM [36].

2.4. Analytic methods

To compare the overall association network of chronic medical conditions in the MDD cases versus controls, we employed the network-theoretic approach with the “guilt by rewiring” principle suggested by Hou et al. [37], in which disease-associated entities are assumed to involve dynamic changes, or rewiring, of the network architecture [37–41].

We first looked at pairwise partial correlations between chronic medical conditions separately in MDD cases and controls. Partial correlations between pairwise combinations of chronic medical conditions were calculated as follows:

$$\hat{\rho}_{X_p X_q \cdot \boldsymbol{Z}} = \frac{N\sum_{i=1}^{N} \hat{\epsilon}_{X_p,i}\hat{\epsilon}_{X_q,i} - \sum_{i=1}^{N} \hat{\epsilon}_{X_p,i} \sum_{i=1}^{N} \hat{\epsilon}_{X_q,i}}{\sqrt{N\sum_{i=1}^{N} \hat{\epsilon}_{X_p,i}^{\,2} - \left(\sum_{i=1}^{N} \hat{\epsilon}_{X_p,i}\right)^2}\sqrt{N\sum_{i=1}^{N} \hat{\epsilon}_{X_q,i}^{\,2} - \left(\sum_{i=1}^{N} \hat{\epsilon}_{X_q,i}\right)^2}},$$

where $\hat{\epsilon}_{X_p,i}$ denotes the i-th residual from regressing X_p on a set of covariates $\boldsymbol{Z}$, and $\hat{\epsilon}_{X_q,i}$ denotes the i-th residual from regressing X_q on a set of covariates $\boldsymbol{Z}$ (i.e., demographic characteristics, socioeconomic

status, geographical locations, and healthcare services utilization). Using the partial correlation coefficients as weights of attractive forces, we constructed association networks of chronic medical conditions separately in MDD cases and controls with the Fruchterman Reingold algorithm, which is a force-directed layout algorithm that iterates until the sum of the force vectors reach equilibrium [42].

To adjust for confounding variables and to account for our matched pair design, we revised the rewire metric as a 2-sided test under the null hypothesis that the Fisher-transformed difference between the partial correlations in cases and controls equals zero. The revised rewire metric can be then interpreted as a p-value for rejecting the null hypothesis. This p-value is calculated under the null permutation distribution by shuffling the case/control labels within each matched pair. The overall false discovery rate (FDR) is controlled for all pairwise tests with the Benjamini-Hochberg procedure [43].

Specifically, we revised and modified the existing rewire metric as follows:

$$P_{\text{rewire}_{X_pX_q}}(revised) = Pr\left(\left| \frac{\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{null}^{case}}) - \tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{null}^{control}})}{\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{observed}^{case}}) - \tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{observed}^{control}})} \right| > 1 \right),$$

where $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{observed}^{case}}$ and $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{observed}^{control}}$ denote partial correlations between chronic medical conditions X_p and X_q adjusting for a set of potential confounders $\boldsymbol{Z}$ (i.e., demographic characteristics, socioeconomic status, and geographical locations) observed in the study samples of cases and controls, respectively. The terms, $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{null}^{case}}$ and $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}_{null}^{control}}$, denote a sample from the permutation distribution of the partial correlations under the null hypothesis of no correlation, where the permutation distribution is generated from 2000 iterations of re-randomization of the labels for being either case or control within each matched pair. A detailed comparison between the original and our revised rewire metric is provided in **Appendix A**.

Sensitivity analysis for the rewire metric analysis was performed with regression modeling. Conditional logistic regression modeling was performed for each pair of chronic medical conditions, prioritized with the network-theoretic analysis (i.e. X_p and X_q) and the corresponding pairwise interaction term (i.e. $X_p \cdot X_q$) as independent variables and diagnosis of MDD as the dependent variable.

R statistical software (version 3.5.0 for Windows; R Foundation for Statistical Computing) with tidyverse (version 1.2.1) [44] was used for all analyses. The matching procedure was based on the libraries lsmeans (version 2.27-62) [45], epiR (version 0.9-96) [46], tableone (version 0.9.3) [47], and Matching (version 4.9-3) [48]. Construction of association network and visualization was based on the libraries igraph (version 1.2.1) [49], and latticeExtra (version 0.6-28) [50]. All software and code used in this work are archived in GitHub [51].

3. Results

Table 1 shows the baseline demographics, socioeconomic status, place of residence, and healthcare services utilization, which were used for matching MDD patients and controls. Standardized mean difference (SMD) between MDD patients and their matched controls was 0.045 or less for age, sex, household income, financial independence, medical aid beneficiary status, whether the individual resided in a metropolitan area, whether the individual had any hospitalization, and whether the individual had more than three outpatient visits a year.

Table 2 shows the distribution of chronic medical conditions in MDD patients and their matched controls. When comparing the distribution of chronic medical conditions in MDD patients to that observed in their matched controls, the strongest SMD of 0.126 was observed with arthritis, followed by hyperlipidemia (SMD 0.094), ischemic heart disease (SMD 0.085), chronic obstructive pulmonary disorder (SMD 0.082), heart failure (SMD 0.082), and osteoporosis (SMD 0.082). Fig. 1a and b show the association networks of chronic medical conditions in the matched controls (Fig. 1a) and MDD patients (Fig. 1b).

Using our approach for network analysis in a matched case-control design, five pairs of chronic medical conditions were prioritized among 105 possible pairs of chronic conditions after controlling the false discovery rate at 5%. Table 3 shows the partial correlations between chronic medical conditions and the Fisher-transformed differences of partial correlation, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}})$, for the five prioritized pairs of chronic medical conditions. The strongest increase in $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}})$ comparing the cases vs. controls was observed between ischemic heart disease and stroke (0.08 in the cases vs. 0.01 in the controls, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}})$=4.38), followed by asthma and heart failure (0.06 in

Table 1
Baseline demographic characteristics, socioeconomic status, and geographical locations of major depressive disorder (MDD) patients and their matched controls.

	MDD patients (N = 10,299)	Matched controls (N = 10,299)	SMD
Age, years	55.9 (SD 10.4)	55.9 (SD 10.4)	< 0.001
Male	3003 (29.2%)	3003 (29.2%)	< 0.001
Female	7296 (70.8%)	7296 (70.8%)	< 0.001
Household income[†]	6.2 (SD 3.1)	6.2 (SD 3.1)	0.008
Financially dependent individuals[†]	6522 (63.3%)	6298 (61.2%)	0.045
Medical aid beneficiary	336 (3.3%)	336 (3.3%)	< 0.001
Metropolitan area[‡]	5796 (56.3%)	5796 (56.3%)	< 0.001
Individuals with any hospitalization	760 (7.4%)	731 (7.1%)	0.011
Individuals with more than 3 outpatient visits a year	7636 (74.1%)	7622 (74.0%)	0.003

Abbreviations: MDD = major depressive disorder; SD = standard deviation; SMD = standardized mean difference.

[†] Household income data was given in decile categories with higher values reflecting higher income. Household income is based on the nationwide income distribution, which determines the monthly premium payment for the individuals. For financially dependent individuals, the household income reflects that of the supporting individuals (family members).

[‡] The metropolitan areas include the cities of Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, and Ulsan.

Table 2
Chronic medical conditions of major depressive disorder (MDD) patients and their matched controls.

	MDD patients (N = 10,299)	Matched controls (N = 10,299)	SMD
Anemia	227 (2.2%)	152 (1.5%)	0.054
Arthritis	3151 (30.6%)	2571 (25.0%)	0.126
Asthma	317 (3.1%)	260 (2.5%)	0.034
Cataract	662 (6.4%)	501 (4.9%)	0.068
Chronic kidney disease	139 (1.3%)	131 (1.3%)	0.007
COPD	2139 (20.8%)	1805 (17.5%)	0.082
Dementia	19 (0.2%)	9 (0.1%)	0.026
Diabetes Mellitus	537 (5.2%)	502 (4.9%)	0.016
Glaucoma	29 (0.3%)	20 (0.2%)	0.018
Heart failure	622 (6.0%)	436 (4.2%)	0.082
Hyperlipidemia	659 (6.4%)	441 (4.3%)	0.094
Hypertension	29 (0.3%)	20 (0.2%)	0.018
Hypothyroidism	272 (2.6%)	258 (2.5%)	0.009
Ischemic heart disease	747 (7.3%)	535 (5.2%)	0.085
Osteoporosis	622 (6.0%)	436 (4.2%)	0.082
Stroke	532 (5.2%)	386 (3.7%)	0.069

Abbreviations: MDD = major depressive disorder; SMD = standardized mean difference; SD = Standard Deviation; COPD = chronic obstructive pulmonary disorder.

Note: Chronic medical conditions are identified from the diagnosis codes recorded prior to the cohort entry date of January 1, 2007.

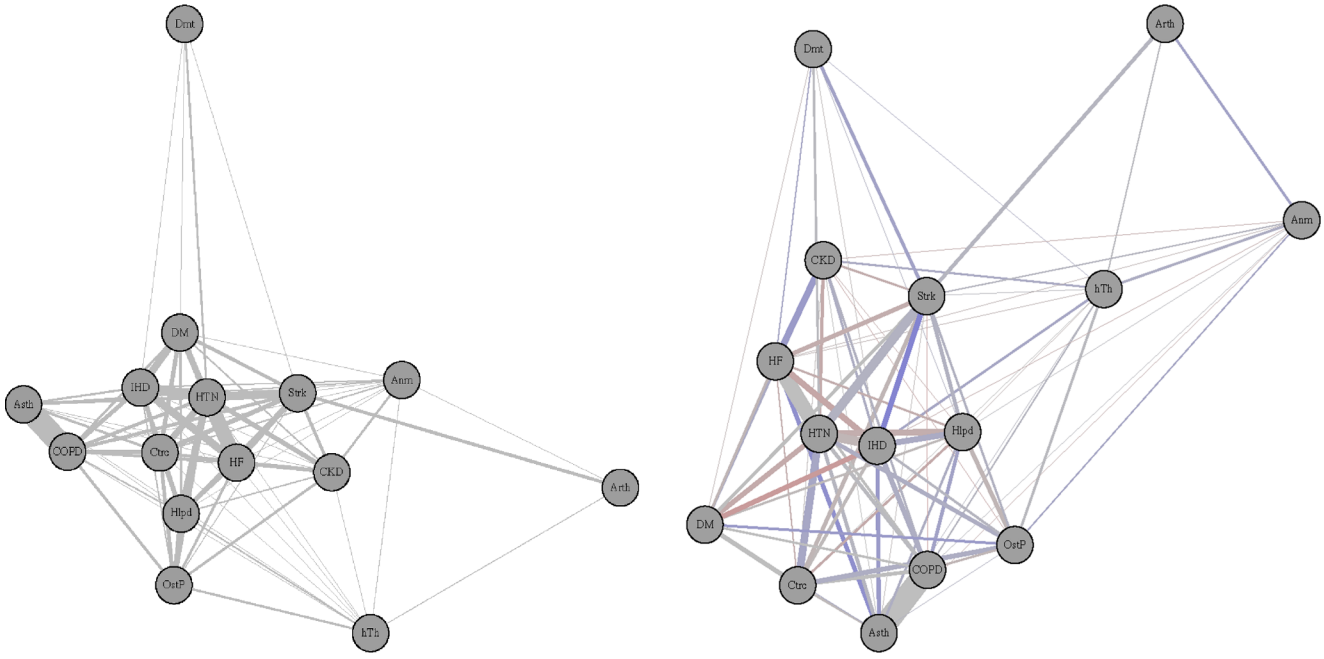

Fig. 1. Association network of chronic medical conditions in the matched controls (a) versus MDD patients (b). **Abbreviations:** Anm = anemia, Arth = arthritis, Asth = asthma, Ctrc = cataract, CKD = chronic kidney disease, COPD = chronic obstructive pulmonary disease, DM = Diabetes Mellitus, Dmt = dementia, Hlpd = hyperlipidemia, HF = heart failure, hTh = hypothyroidism, HTN = hypertension, IHD = ischemic heart disease, OstP = osteoporosis, Strk = stroke. **Note:** Edges between chronic medical conditions that showed increased partial correlation coefficients in the cases compared to the matched controls are colored blue; those that showed decreased partial correlation coefficient in the cases are colored red. The intensities of the blue and red colors are proportional to the difference in the magnitude of the partial correlation coefficient. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Partial correlations between chronic medical conditions and the Fisher-transformed differences, controlled for FDR < 0.05.

Chronic Condition Pair		Partial Correlation in Cases vs. Controls	Fisher-transformed difference†	FDR‡
Ischemic Heart Disease	Stroke	0.08 vs. 0.01	4.38	< 0.01
Asthma	Heart Failure	0.06 vs. 0.01	3.48	0.04
Diabetes	Osteoporosis	0.03 vs. 0.00	2.50	0.04
COPD	Stroke	0.01 vs. 0.04	−2.42	0.04
Heart Failure	Osteoporosis	−0.01 vs. 0.03	-2.58	0.04

Abbreviations: FDR = false discovery rate; COPD = chronic obstructive pulmonary disorder.

† The Fisher-transformed difference in partial correlation is calculated as $\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}\,observed}^{\;case}) - \tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}\,observed}^{\;control})$, where $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}\,observed}^{\;case}$ and $\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}\,observed}^{\;control}$ denote partial correlations between chronic medical conditions X_p and X_q adjusted for a set of covariates $\boldsymbol{Z}$ (i.e., demographic characteristics, socioeconomic status, place of residence, and healthcare services utilization) observed in the study samples of cases and controls respectively, and $\tanh^{-1}(\hat{\rho}) = \frac{1}{2} ln \frac{1+\hat{\rho}}{1-\hat{\rho}}$.

‡ False discovery rate (FDR) is controlled with the Benjamini-Hochberg procedure.

the cases vs. 0.01 in the controls, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}}) = 3.48$), and diabetes and osteoporosis (0.03 in the cases vs. 0.00 in the controls, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}}) = 2.50$). The strongest decrease in $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}})$comparing the cases vs. controls was observed between heart failure and osteoporosis (-0.01 in the cases vs. 0.03 in the controls, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}}) = -2.58$), followed by chronic obstructive pulmonary disorder and stroke (0.01 in the cases vs. 0.04 in the controls, $\Delta\tanh^{-1}(\hat{\rho}_{X_pX_q \cdot \boldsymbol{Z}}) = -2.42$).

In sensitivity analysis with regression modeling, the coefficient for the interaction term was greater than zero when the Fisher-transformed difference of partial correlations was increased in the cases vs. controls, and vice versa. A positive coefficient for the interaction term in the generalized linear (log-odds) scale indicates that the total effect of the co-occurrence is more than the additive effect of each pair, while a negative coefficient indicates that the effect is less than the additive effect of each pair. The interpretation of interactions in regression models is consistent with co-occurrence patterns in our network theoretic model, but four out of the five prioritized chronic medical condition pairs showed p-values for the interaction terms < 0.05, as shown in **Appendix B**. Specifically, the interaction terms between ischemic heart disease and stroke (coefficient 0.36, 95% CI (0.01, 0.72), p = 0.04), and between asthma and heart failure (coefficient 0.66, 95% CI (0.06, 1.25), p = 0.03) showed coefficients greater than zero with p-values < 0.05. The interaction terms between chronic obstructive pulmonary disorder and stroke (coefficient −0.34, 95% CI (−0.60, −0.07), p = 0.01), and between heart failure and osteoporosis (coefficient −0.73, 95% CI (−1.37, −0.09), p = 0.03) showed coefficients less than zero with p-values < 0.05. However, the interaction term between diabetes and osteoporosis showed a coefficient greater than zero with a p-value greater than 0.05 (coefficient 0.36, 95% CI (−0.03, 0.75), p = 0.07).

4. Discussion

In this study, we present a network analysis with a matched case-control design to explore all possible pairs of chronic medical conditions that differentially co-occur in newly diagnosed MDD patients compared to their matched controls, using a random sample from a nationally representative longitudinal database. This work contributes to the literature in two ways – first, we extend a method for network analysis with a rewire metric [37] to a matched design; and second, we provide a comprehensive graphical overview of chronic condition co-

occurrence in MDD patients and their matched controls, and statistically compare these networks.

Depression is a clinically and etiologically heterogeneous disorder [52–54], and synthesizing the evidence involving depression can be challenging because the illness is associated with various non-medical factors [55–58] that may confound results. In addition, depression in the presence of comorbid medical conditions is also more likely to be treatment-resistant depression [13–15]. Therefore, a better understanding of depression in the presence of comorbid medical conditions, after controlling for non-medical factors, contributes to the on-going efforts to improve the classification system that captures heterogeneous pathophysiology of depression [59].

The Korean National Health Information Database used in this study provides a unique opportunity to analyze population features on a large scale. The Korean National Health Insurance Service covers all Korean citizens, and the random sample is designed to be nationally representative [60]. Its universal healthcare system with income-based premiums allows people to seek healthcare with less financial burden. Furthermore, the system encourages check-ups for those aged 40 and above, thus minimizing the underdiagnosis of medical conditions. In addition, information regarding death is reliably captured in this nationwide database.

Our study design with matched pair sampling allows us to compare the association networks of chronic medical conditions in newly diagnosed MDD cases with their matched controls, adjusting for confounding factors [61]. A nested case-control design within a longitudinal cohort with an incidence density sampling scheme is a method for minimizing potential bias in observational studies [33,34]. In this study, we controlled for demographic characteristics, socioeconomic status, place of residence, and healthcare services utilization. We considered the individual conditions that were recorded before the initial diagnosis of MDD to identify the conditions with temporal precedence. By collecting individual features from matched controls in the same time frame, we may reduce potential bias that can result from differing observation periods, and regional and environmental factors that change over time, such as economic recession or natural disaster [62]. Temporal precedence is an essential component for causal inference [63]. However, it should be noted that the temporal precedence may not be well captured through routinely collected healthcare data, because patients with depression often have prodromal depression symptoms for many years before their diagnosis [52]. Furthermore, causal inference needs to be evaluated based on potential causal pathways, which requires hypothesis-specific adjustment methods for potential confounders [64,65]. Therefore, although we adjusted for certain confounders, it should be noted that the main advantage of our methodology is to provide a high-level overview of multi-dimensional data, rather than confirmation of causal associations.

The results of our analysis, consistent with many others involving depression, suggest that patients with newly diagnosed MDD tend to have more chronic medical conditions before the diagnosis of MDD, and that the co-occurrence pattern of chronic medical conditions is differential, when compared to their matched controls [1–3]. Table 2 shows that the newly diagnosed MDD patients are more likely to have chronic conditions compared to their matched controls, such as arthritis, hyperlipidemia, ischemic heart disease, chronic obstructive pulmonary disorder, heart failure, and osteoporosis, in a decreasing order of SMD. This study further considered 105 possible pairs of chronic medical conditions that differentially co-occur in newly diagnosed MDD patients compared to their matched controls. We discovered a total of five statistically significant pairs of chronic medical conditions after controlling for the study-wide false discovery rate at 5%, which could be prioritized for further investigation (Table 3). These prioritized relationships in our study involve three entities – two chronic medical conditions and MDD. The increase (decrease) in the correlation coefficient between two chronic medical conditions, when comparing the cases vs. controls, can be interpreted as the likelihood for co-occurrence rather than a solitary occurrence in the cases compared to the controls.

We further confirmed the findings from our network theoretic approach in sensitivity analyses with regression models on the matched pairs. In our study, four out of the five prioritized chronic medical condition pairs showed statistically significant interaction. The prioritized pairwise interactions of chronic medical conditions mean that the odds or probability for having MDD are differential for each combination of such medical conditions, and as a result, prediction models that include these variable interactions may have better predictive performance. An implication of our study for future research is that co-occurring conditions can have a non-linear effect on the risk of depression beyond the linear additive effects of each chronic condition [66–69].

Each of the chronic medical conditions prioritized in this work has been studied in the literature. First, our study prioritized a positive interaction between stroke and ischemic heart disease in response to depression (i.e., positive association between depression and the interaction term between stroke and ischemic heart disease), but a negative interaction between stroke and chronic obstructive pulmonary disorder in response to depression (i.e., negative association between depression and the interaction term between stroke and chronic obstructive pulmonary disorder) (Table 3, **Appendix B**). Post-stroke depression is an important clinical phenomenon, but despite various efforts, the prevention and treatment of post-stroke depression presents many challenges [9]. The association between ischemic heart disease and depression has also been widely noted, and the American Heart Association Prevention Committee published a clinical recommendation for depression screening in patients with coronary heart disease in 2008 [11]. Ischemic heart disease and stroke are elements of a larger category of cardiovascular disease, which share multiple risk factors, such as smoking [70,71], obesity [72,73], physical activity [74,75], hypertension [76,77], and hyperlipidemia [78,79]. Depression has also been associated with smoking [80,81], obesity [82,83], physical activity [84,85], hypertension [86,87], and hyperlipidemia [88,89] in the literature, but causal inference is challenging due to the complex nature of depression. The positive interaction between stroke and ischemic heart disease prioritized in this study may be related to the shared risk factors for these conditions, but further investigation is required to validate these findings. On the other hand, chronic obstructive pulmonary disorder (COPD) also has been associated with depression, but the co-occurrence is not suggestive of a single pathologic pathway, and may be suggestive of competing risks [90] or MDD patients' receiving less care for the medical conditions [91]. Therefore, further studies would be required to validate the negative interaction between stroke and COPD. Regarding heart failure, our study prioritized a positive interaction with asthma, but a negative interaction with osteoporosis, with respect to depression (Table 3, **Appendix B**). Heart failure [92], asthma [93], and osteoporosis [94] have been studied in the literature involving depression. In fact, heart failure and asthma patients may not only present as coexisting comorbidities [95], but also present with similar symptoms [96]. Therefore, the co-occurrence may imply both the complexity and the severity of the disease status, which may partly explain the positive interaction with respect to depression. However, the negative interaction between heart failure and osteoporosis is not explained similarly, although heart failure and osteoporosis have been associated with an increased disease burden [97]. This may be suggestive of competing risks, and the differential interactions may also depend on the specific causal pathway. Therefore, further studies are needed to validate these findings.

There are several limitations of the present analysis that should be noted. Diagnosis codes for depressive disorders, as well as other mental health diagnoses, tend to have a poor receiver operating characteristics [98–100]. In order to improve our ascertainment of cases, we employed an operational definition for MDD where we have included patients treated for MDD, who had at least two claims with a diagnosis of depression as well as a record of receiving a prescription for an antidepressant medication. On the other hand, those who received

antidepressants without diagnoses of MDD could be included in the matched controls. This will increase the similarity between the cases and the controls, and this may reduce the statistical power to detect the difference. However, the directionality of the bias is toward the null, and this may provide more conservative estimates thereby significantly reducing the risk of false discovery. In this study, 2.2% (230) of the controls received antidepressants without diagnoses of MDD. This proportion may be lower than what is expected in other healthcare settings [101] because Korea has a stringent policy against off-label use of medications.

The phenotyping method for identifying chronic medical conditions also has limitations. Because this work aims to demonstrate an application methodology of a network-theoretic principle to population data rather than to confirm specific associations among ontological condition concepts, we used the externally validated Chronic Conditions Data Warehouse Condition Algorithms. However, not all of the condition categories were used in this study. First, this study is limited to non-neoplastic chronic medical conditions because the diagnosis for the neoplastic disease is recorded with separate diagnosis-related group codes in the claims for cancer treatments, and ICD-10 alone may not distinguish the codes for screening tests. When a chronic condition category in the Chronic Conditions Data Warehouse Condition Algorithms is a subset of another condition category within the same algorithm, only the superset category was used for this work. For example, the category “Acute Myocardial Infarction” was not used because it is a subset of “Ischemic Heart Disease.” Similarly, the category “Alzheimer’s Disease” was not used because it is a subset of “Alzheimer’s Disease and Related Disorders or Senile Dementia.” Also, the category “Benign Prostatic Hyperplasia” was not used in this study because it is relevant only to male sex. The category “Atrial fibrillation” was not used because the algorithm does not capture most of the ICD-10 diagnosis codes used in Korea. For example, ICD-10 code I48.9 is used in Korea, but it is not included in the Chronic Conditions Data Warehouse Condition Algorithms (ICD-10 code I48.9 is “non-billable” code for Medicare). Finally, the category “Depression” was not used because we are using ICD-10 codes for defining “Major Depressive Disorder,” which is a subset of “Depression.” Of note, data-driven phenotyping may also be a promising method to incorporate the high dimensional diagnosis information [102], but requires an independent validation for the phenotyping method before analyzing the phenotypes derived from the phenotyping method. Because this study aimed to present an analytic method, this study used independently validated phenotyping algorithms that have been independently validated and published in the literature.

Another limitation is that although our study design allows us to prioritize the relationships among three entities – two chronic medical conditions and MDD – more complex relationships and higher-order interactions that involve more than three entities cannot be modeled using our current approach. On the other hand, our approach does explore multiple relationships among multiple entities, and hence the possibility of false discovery from multiple comparisons cannot be completely ruled out. To minimize such a possibility, we applied a revised rewire metric-based network-theoretic approach with permutation tests to estimate the precision and statistical significance of the findings and accounted for multiple comparisons. However, the extent of multiple comparisons would be increased in combinatorial scale if more entities were analyzed. In addition, the methodology used in this study is only comparing two association networks, one in the cases and another in the controls. More theoretical work would be necessary to apply this methodology to compare more than two association networks.

Although the above-mentioned limitations may not be completely avoidable in many observational studies, a comprehensive understanding of the complex relationships between multiple chronic medical conditions and depression that supplements traditional clinical studies is currently lacking. Our network-theoretic analysis with a matched case-control design allows us to compare the association networks of features after adjusting for confounding factors, and therefore, can supplement traditional study approaches. In addition to the advantage of providing a high-level overview of feature distribution, this may be a promising alternative to traditional meta-analysis when the homogeneity and consistency among the published literature are not guaranteed, and the baseline assumptions for conducting a meta-analysis are not met. When this methodology is applied to multiple health databases, more generalized evidence can be generated with greater confidence.

5. Conclusion

Association networks in a matched case-control design can provide a high-level comparison of comorbid features after adjusting for confounding factors, supplementing traditional clinical study approaches. With this method, the differential co-occurrence patterns of chronic medical conditions in the patients with MDD compared to those in the matched controls could be demonstrated, and the chronic conditions that have statistically significant interactions in the regression models could then be prioritized for further investigation.

Acknowledgments

The data used in this study were provided by the Korea National Health Insurance Service – National Sample Cohort (NHIS-NSC) 2002–2013. Authors thank Kyuwoong Kim and Jooyoung Chang at Seoul National University College of Medicine for assistance with data management and collaboration for this research.

Funding: This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education [2017R1D1A1B03033721]; United States National Institutes of Health (NIH) [R01MH105384]; [UL1TR000457-06]; [P50MH113838] and [T32MH019132].

Conflicts of interests

Authors have no conflicts of interests.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2018.09.016.

References

[1] R.C. Kessler, P. Berglund, O. Demler, R. Jin, K.R. Merikangas, E.E. Walters, Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication, Arch. Gen. Psychiatry 62 (2005) 593–602, https://doi.org/10.1001/archpsyc.62.6.593.
[2] R.H. Pietrzak, J. Kinley, T.O. Afifi, M.W. Enns, J. Fawcett, J. Sareen, Subsyndromal depression in the United States: prevalence, course, and risk for incident psychiatric outcomes, Psychol. Med. 43 (2013) 1401–1414, https://doi.org/10.1017/S0033291712002309.
[3] William W. Eaton, Medical and psychiatric comorbidity over the course of life, Am. Psychiatric Assoc. (2007).
[4] J.M. Donohue, H.A. Pincus, Reducing the societal burden of depression, Pharmacoeconomics 25 (2007) 7–24.
[5] R.M.A. Hirschfeld, The comorbidity of major depression and anxiety disorders: recognition and management in primary care, Prim. Care Companion J. Clin. Psychiatry 3 (2001) 244–254.
[6] A. Tylee, P. Gandhi, The importance of somatic symptoms in depression in primary care, Prim. Care Companion J. Clin. Psychiatry 7 (2005) 167–176.
[7] M.W.B. Zhang, R.C.M. Ho, M.W.L. Cheung, E. Fu, A. Mak, Prevalence of depressive symptoms in patients with chronic obstructive pulmonary disease: a systematic review, meta-analysis and meta-regression, Gen. Hosp. Psychiatry 33 (2011) 217–223, https://doi.org/10.1016/j.genhosppsych.2011.03.009.
[8] R.E. Jorge, R.G. Robinson, D. Moser, A. Tateno, B. Crespo-Facorro, S. Arndt, Major depression following traumatic brain injury, Arch. Gen. Psychiatry 61 (2004) 42–50, https://doi.org/10.1001/archpsyc.61.1.42.
[9] M.L. Hackett, K. Pickles, Part I: frequency of depression after stroke: an updated

systematic review and meta-analysis of observational studies, Int. J. Stroke Off. J. Int. Stroke Soc. 9 (2014) 1017–1025, https://doi.org/10.1111/ijs.12357.
[10] M.R. DiMatteo, H.S. Lepper, T.W. Croghan, Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence, Arch. Int. Med. 160 (2000) 2101–2107.
[11] J.H. Lichtman, J.T. Bigger, J.A. Blumenthal, N. Frasure-Smith, P.G. Kaufmann, F. Lespérance, D.B. Mark, D.S. Sheps, C.B. Taylor, E.S. Froelicher, American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, American Heart Association Council on Clinical Cardiology, American Heart Association Council on Epidemiology and Prevention, American Heart Association Interdisciplinary Council on Quality of Care and Outcomes Research, American Psychiatric Association, Depression and coronary heart disease: recommendations for screening, referral, and treatment: a science advisory from the American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, Council on Clinical Cardiology, Council on Epidemiology and Prevention, and Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Psychiatric Association, Circulation 118 (2008) 1768–1775, https://doi.org/10.1161/CIRCULATIONAHA.108.190769.
[12] American Diabetes Association, Standards of medical care in diabetes, Diabetes Care. 28 (Suppl 1) (2005) S4–S36.
[13] G.I. Keitner, C.E. Ryan, I.W. Miller, R. Kohn, N.B. Epstein, 12-month outcome of patients with major depression and comorbid psychiatric or medical illness (compound depression), Am. J. Psychiatry 148 (1991) 345–350, https://doi.org/10.1176/ajp.148.3.345.
[14] D.V. Iosifescu, A.A. Nierenberg, J.E. Alpert, M. Smith, S. Bitran, C. Dording, M. Fava, The impact of medical comorbidity on acute treatment in major depressive disorder, Am. J. Psychiatry. 160 (2003) 2122–2127, https://doi.org/10.1176/appi.ajp.160.12.2122.
[15] D.V. Iosifescu, N. Clementi-Craven, R. Fraguas, G.I. Papakostas, T. Petersen, J.E. Alpert, A.A. Nierenberg, M. Fava, Cardiovascular risk factors may moderate pharmacological treatment effects in major depressive disorder, Psychosom. Med. 67 (2005) 703–706, https://doi.org/10.1097/01.psy.0000170338.75346.d0.
[16] S.H. Huang, P. LePendu, S.V. Iyer, M. Tai-Seale, D. Carrell, N.H. Shah, Toward personalizing treatment for depression: predicting diagnosis and severity, J. Am. Med. Inform. Assoc. 21 (2014) 1069–1075.
[17] H. Jin, S. Wu, P. Di Capua, Development of a clinical forecasting model to predict comorbid depression among diabetes patients and an application in depression screening policy making, Prev. Chronic. Dis. 12 (2015) E142, https://doi.org/10.5888/pcd12.150047.
[18] E. Ryu, A.M. Chamberlain, R.S. Pendegraft, T.M. Petterson, W.V. Bobo, J. Pathak, Quantifying the impact of chronic conditions on a diagnosis of major depressive disorder in adults: a cohort study using linked electronic medical records, BMC Psychiatry 16 (2016) 114, https://doi.org/10.1186/s12888-016-0821-x.
[19] M. Kim, S. Banerjee, S.M. Park, J. Pathak, Improving risk prediction for depression via Elastic Net regression – Results from Korea National Health Insurance Services Data, AMIA, Annu. Symp. Proc. 2017 (2016) 1860–1869.
[20] Y. Moreau, L.-C. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery, Nat. Rev. Genet. 13 (2012) 523–536, https://doi.org/10.1038/nrg3253.
[21] J.C. Brunson, R.C. Laubenbacher, Applications of network analysis to routinely collected health care data: a systematic review, J. Am. Med. Inform. Assoc. JAMIA (2017), https://doi.org/10.1093/jamia/ocx052.
[22] C.A. Hidalgo, N. Blumm, A.-L. Barabási, N.A. Christakis, A dynamic network approach for the study of human phenotypes, PLoS Comput. Biol. 5 (2009) e1000353, , https://doi.org/10.1371/journal.pcbi.1000353.
[23] F.S. Roque, P.B. Jensen, H. Schmock, M. Dalgaard, M. Andreatta, T. Hansen, K. Søeby, S. Bredkjær, A. Juul, T. Werge, L.J. Jensen, S. Brunak, Using electronic patient records to discover disease correlations and stratify patient cohorts, PLoS Comput. Biol. 7 (2011) e1002141, , https://doi.org/10.1371/journal.pcbi.1002141.
[24] Y. Chen, R. Xu, Network analysis of human disease comorbidity patterns based on large-scale data mining, Bioinf. Res. Appl. (2014) 243–254.
[25] J.H. Kim, K.Y. Son, D.W. Shin, S.H. Kim, J.W. Yun, J.H. Shin, M.S. Kang, E.H. Chung, K.H. Yoo, J.M. Yun, Network analysis of human diseases using Korean nationwide claims data, J. Biomed. Inform. 61 (2016) 276–282, https://doi.org/10.1016/j.jbi.2016.05.002.
[26] V. Kannan, F. Swartz, N.A. Kiani, G. Silberberg, G. Tsipras, D. Gomez-Cabrero, K. Alexanderson, J. Tegnèr, Conditional disease development extracted from longitudinal health care cohort data using layered network construction, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep26170.
[27] E. Jeong, K. Ko, S. Oh, H.W. Han, Network-based analysis of diagnosis progression patterns using claims data, Sci. Rep. 7 (2017) 15561, https://doi.org/10.1038/s41598-017-15647-4.
[28] J. Lee, J.S. Lee, S.-H. Park, S.A. Shin, K. Kim, Cohort profile: the national health insurance service-national sample cohort (NHIS-NSC), South Korea, Int. J. Epidemiol. (2016) dyv319, https://doi.org/10.1093/ije/dyv319.
[29] L. Kim, J.-A. Kim, S. Kim, A guide for the utilization of health insurance review and assessment service national patient samples, Epidemiol. Health. (2014) e2014008, , https://doi.org/10.4178/epih/e2014008.
[30] W.G. Cochran, Sampling techniques-3, 1977.
[31] World Health Organization, ICD-10. http://apps.who.int/classifications/icd10 (accessed August 3, 2018).
[32] L. Townsend, J.T. Walkup, S. Crystal, M. Olfson, A systematic review of validated methods for identifying depression using administrative data, Pharmacoepidemiol. Drug Saf. 21 (Suppl 1) (2012) 163–173, https://doi.org/10.1002/pds.2310.
[33] M.J. Knol, J.P. Vandenbroucke, P. Scott, M. Egger, What do case-control studies estimate? Survey of methods and assumptions in published case-control research, Am. J. Epidemiol. 168 (2008) 1073–1081, https://doi.org/10.1093/aje/kwn217.
[34] S. Suissa, S. Dell'aniello, S. Vahey, C. Renoux, Time-window bias in case-control studies: statins and lung cancer, Epidemiol. Camb. Mass. 22 (2011) 228–231, https://doi.org/10.1097/EDE.0b013e3182093a0f.
[35] Y. Gorina, E.A. Kramarow, Identifying chronic conditions in Medicare claims data: evaluating the Chronic Condition Data Warehouse algorithm, Health Serv. Res. 46 (2011) 1610–1627, https://doi.org/10.1111/j.1475-6773.2011.01277.x.
[36] U.S. Department of Health & Human Services, ICD-10-CM – International Classification of Diseases, Tenth Revision, Clinical Modification, 2018. https://www.cdc.gov/nchs/icd/icd10cm.htm (accessed August 3, 2018).
[37] L. Hou, M. Chen, C.K. Zhang, J. Cho, H. Zhao, Guilt by rewiring: gene prioritization through network rewiring in genome wide association studies, Hum. Mol. Genet. 23 (2014) 2780–2790, https://doi.org/10.1093/hmg/ddt668.
[38] N.J. Hudson, A. Reverter, B.P. Dalrymple, A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation, PLoS Comput. Biol. 5 (2009) e1000382, , https://doi.org/10.1371/journal.pcbi.1000382.
[39] S. Bandyopadhyay, M. Mehta, D. Kuo, M.-K. Sung, R. Chuang, E.J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. Guénolé, H. van Attikum, K.M. Shokat, R.D. Kolodner, W.-K. Huh, R. Aebersold, M.-C. Keogh, N.J. Krogan, T. Ideker, Rewiring of genetic networks in response to DNA damage, Science 330 (2010) 1385–1389, https://doi.org/10.1126/science.1195618.
[40] P. Luo, L.-P. Tian, J. Ruan, F.-X. Wu, Identifying disease genes from PPI networks weighted by gene expression under different conditions, Bioinforma. Biomed. BIBM 2016 IEEE Int. Conf. On, IEEE (2016) 1259–1264.
[41] J.X. Hu, C.E. Thomas, S. Brunak, Network biology concepts in complex disease comorbidities, Nat. Rev. Genet. 17 (2016) 615–629, https://doi.org/10.1038/nrg.2016.87.
[42] T.M. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, Softw. Pract. Exp. 21 (1991) 1129–1164.
[43] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. (1995) 289–300.
[44] H. Wickham, The tidyverse, 2017.
[45] R.V. Lenth, Least-squares means: the R package lsmeans, J Stat Softw. 69 (2016) 1–33.
[46] M. Stevenson, M.M. Stevenson, I. BiasedUrn, Package 'epiR,' 2017.
[47] K. Yoshida, J. Bohn, Tableone: Create "Table 1" to Describe Baseline Characteristics, R Package Version 07. 3, 2015.
[48] J.S. Sekhon, Package 'Matching,' 2015, http://dirichlet.mat.puc.cl/web/packages/Matching/Matching.pdf.
[49] G. Csardi, T. Nepusz, The igraph software package for complex network research, Int. J. Complex Syst. 1695 (2006) 1–9.
[50] D. Sarkar, M.D. Sarkar, Package 'latticeExtra,' 2016.
[51] Min-Hyung Kim, Association Network Matched Case Control, 2018. https://github.com/mkim0710/AssociationNetworkMatchedCaseControl (accessed August 3, 2018).
[52] W.W. Eaton, J.C. Anthony, J. Gallo, G. Cai, A. Tien, A. Romanoski, C. Lyketsos, L.S. Chen, Natural history of diagnostic interview schedule/DSM-IV major depression. The baltimore epidemiologic catchment area follow-up, Arch. Gen. Psychiatry 54 (1997) 993–999.
[53] L. Chen, W.W. Eaton, J.J. Gallo, G. Nestadt, Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study, J. Affect. Disord. 59 (2000) 1–11.
[54] G. Hasler, Pathophysiology of depression: do we have any solid evidence of interest to clinicians?, World Psychiatry Off, J. World Psychiatr. Assoc. WPA 9 (2010) 155–161.
[55] F.M. Culbertson, Depression and gender. An international review, Am. Psychol. 52 (1997) 25–31.
[56] F.J. Zimmerman, W. Katon, Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship? Health Econ. 14 (2005) 1197–1215, https://doi.org/10.1002/hec.1011.
[57] M. Averina, O. Nilssen, T. Brenn, J. Brox, V.L. Arkhipovsky, A.G. Kalinin, Social and lifestyle determinants of depression, anxiety, sleeping disorders and self-evaluated quality of life in Russia–a population-based study in Arkhangelsk, Soc. Psychiatry Psychiatr. Epidemiol. 40 (2005) 511–518, https://doi.org/10.1007/s00127-005-0918-x.
[58] J.N. Rosenquist, J.H. Fowler, N.A. Christakis, Social network determinants of depression, Mol. Psychiatry 16 (2011) 273–281, https://doi.org/10.1038/mp.2010.13.
[59] T.R. Insel, The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry, Am. J. Psychiatry 171 (2014) 395–397, https://doi.org/10.1176/appi.ajp.2014.14020138.
[60] S.M. Park, K.Y. Son, J.-H. Park, B. Cho, Disparities in short-term and long-term all-cause mortality among Korean cancer patients with and without preexisting disabilities: a nationwide retrospective cohort study, Support. Care Cancer Off. J. Multinatl. Assoc. Support. Care Cancer. 20 (2012) 963–970, https://doi.org/10.1007/s00520-011-1168-y.
[61] D.B. Rubin, Matched Sampling for Causal Effects, first ed.,, Cambridge University Press, Cambridge; New York, 2006.
[62] S. Suissa, Immortal time bias in pharmaco-epidemiology, Am. J. Epidemiol. 167 (2008) 492–499, https://doi.org/10.1093/aje/kwm324.
[63] S. Kleinberg, G. Hripcsak, A review of causal inference for biomedical informatics,

J. Biomed. Inform. 44 (2011) 1102–1112, https://doi.org/10.1016/j.jbi.2011.07.001.
[64] J. Pearl, Causal inference in statistics: an overview, Stat. Surv. 3 (2009) 96–146.
[65] M.A. Hernán, J.M. Robins, Causal Inference, Chapman & Hall/CRC (2018).
[66] M.E. Charlson, P. Pompei, K.L. Ales, C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, J. Chronic Dis. 40 (1987) 373–383.
[67] R.A. Deyo, D.C. Cherkin, M.A. Ciol, Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases, J. Clin. Epidemiol. 45 (1992) 613–619.
[68] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity measures for use with administrative data, Med. Care 36 (1998) 8–27.
[69] C.N. Klabunde, A.L. Potosky, J.M. Legler, J.L. Warren, Development of a comorbidity index using physician claims data, J. Clin. Epidemiol. 53 (2000) 1258–1267.
[70] S.H. Jee, I. Suh, I.S. Kim, L.J. Appel, Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation Study, JAMA 282 (1999) 2149–2155.
[71] L.A. Bazzano, J. He, P. Muntner, S. Vupputuri, P.K. Whelton, Relationship between cigarette smoking and novel risk factors for cardiovascular disease in the United States, Ann. Int. Med. 138 (2003) 891–897.
[72] P.W.F. Wilson, S.R. Bozeman, T.M. Burton, D.C. Hoaglin, R. Ben-Joseph, C.L. Pashos, Prediction of first events of coronary heart disease and stroke with consideration of adiposity, Circulation 118 (2008) 124–130, https://doi.org/10.1161/CIRCULATIONAHA.108.772962.
[73] M.D. Jensen, D.H. Ryan, C.M. Apovian, J.D. Ard, A.G. Comuzzie, K.A. Donato, F.B. Hu, V.S. Hubbard, J.M. Jakicic, R.F. Kushner, C.M. Loria, B.E. Millen, C.A. Nonas, F.X. Pi-Sunyer, J. Stevens, V.J. Stevens, T.A. Wadden, B.M. Wolfe, S.Z. YanovskiAmerican College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society, AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society, J. Am. Coll. Cardiol. 63 (2014) (2013) 2985–3023, https://doi.org/10.1016/j.jacc.2013.11.004.
[74] H. Noda, H. Iso, H. Toyoshima, C. Date, A. Yamamoto, S. Kikuchi, A. Koizumi, T. Kondo, Y. Watanabe, Y. Wada, Y. Inaba, A. Tamakoshi, JACC Study Group, Walking and sports participation and mortality from coronary heart disease and stroke, J. Am. Coll. Cardiol. 46 (2005) 1761–1767, https://doi.org/10.1016/j.jacc.2005.07.038.
[75] L. Soares-Miranda, D.S. Siscovick, B.M. Psaty, W.T. Longstreth, D. Mozaffarian, Physical activity and risk of coronary heart disease and stroke in older adults: the cardiovascular health study, Circulation 133 (2016) 147–155, https://doi.org/10.1161/CIRCULATIONAHA.115.018323.
[76] K. Miura, M.L. Daviglus, A.R. Dyer, K. Liu, D.B. Garside, J. Stamler, P. Greenland, Relationship of blood pressure to 25-year mortality due to coronary heart disease, cardiovascular diseases, and all causes in young adult men: the Chicago Heart Association Detection Project in Industry, Arch. Int. Med. 161 (2001) 1501–1508.
[77] S. Lewington, R. Clarke, N. Qizilbash, R. Peto, R. CollinsProspective Studies Collaboration, Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies, Lancet Lond. Engl. 360 (2002) 1903–1913.
[78] Cholesterol Treatment Trialists' (CTT) Collaborators, B. Mihaylova, J. Emberson, L. Blackwell, A. Keech, J. Simes, E.H. Barnes, M. Voysey, A. Gray, R. Collins, C. Baigent, The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials, Lancet Lond. Engl. 380 (2012) 581–590, https://doi.org/10.1016/S0140-6736(12)60367-5.
[79] R. Chou, T. Dana, I. Blazina, M. Daeges, T.L. Jeanne, Statins for prevention of cardiovascular disease in adults: evidence report and systematic review for the US preventive services task force, JAMA 316 (2016) 2008–2024, https://doi.org/10.1001/jama.2015.15629.
[80] K.S. Kendler, M.C. Neale, C.J. MacLean, A.C. Heath, L.J. Eaves, R.C. Kessler, Smoking and major depression. A causal analysis, Arch. Gen. Psychiatry 50 (1993) 36–43.
[81] J.D. Kassel, L.R. Stroud, C.A. Paronis, Smoking, stress, and negative affect: correlation, causation, and context across stages of smoking, Psychol. Bull. 129 (2003) 270–304.
[82] M.S. Faith, P.E. Matz, M.A. Jorge, Obesity-depression associations in the population, J. Psychosom. Res. 53 (2002) 935–942.
[83] F.S. Luppino, L.M. de Wit, P.F. Bouvy, T. Stijnen, P. Cuijpers, B.W.J.H. Penninx, F.G. Zitman, Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies, Arch. Gen. Psychiatry 67 (2010) 220–229, https://doi.org/10.1001/archgenpsychiatry.2010.2.
[84] R.F. Zoeller Jr, Physical activity: depression, anxiety, physical activity, and cardiovascular disease: What's the connection? Am. J. Lifestyle Med. 1 (2007) 175–180.
[85] A. Ströhle, Physical activity, exercise, depression and anxiety disorders, J. Neural Transm. Vienna Austria 1996 (116) (2009) 777–784, https://doi.org/10.1007/s00702-008-0092-x.
[86] C.M.M. Licht, E.J.C. de Geus, A. Seldenrijk, H.P.J. van Hout, F.G. Zitman, R. van Dyck, B.W.J.H. Penninx, Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension, Hypertens. Dallas Tex 1979 (53) (2009) 631–638, https://doi.org/10.1161/HYPERTENSIONAHA.108.126698.
[87] Z. Li, Y. Li, L. Chen, P. Chen, Y. Hu, Prevalence of depression in patients with hypertension: a systematic review and meta-analysis, Medicine (Baltimore) 94 (2015) e1317, , https://doi.org/10.1097/MD.0000000000001317.
[88] C.-S. Chuang, T.-Y. Yang, C.-H. Muo, H.-L. Su, F.-C. Sung, C.-H. Kao, Hyperlipidemia, statin use and the risk of developing depression: a nationwide retrospective cohort study, Gen. Hosp. Psychiatry 36 (2014) 497–501, https://doi.org/10.1016/j.genhosppsych.2014.05.008.
[89] J. Oh, T.-S. Kim, Serum lipid levels in depression and suicidality: the Korea National Health and Nutrition Examination Survey (KNHANES) 2014, J. Affect. Disord. 213 (2017) 51–58, https://doi.org/10.1016/j.jad.2017.02.002.
[90] R.J. Norwood, A review of etiologies of depression in COPD, Int. J. Chron. Obstruct. Pulmon. Dis. 2 (2007) 485–491.
[91] I. Briskman, G. Bar, M. Boaz, M. Shargorodsky, Impact of co-morbid mental illness on the diagnosis and management of patients hospitalized for medical conditions in a general hospital, Int. J. Psychiatry Med. 43 (2012) 339–348, https://doi.org/10.2190/PM.43.4.d.
[92] T. Rutledge, V.A. Reis, S.E. Linke, B.H. Greenberg, P.J. Mills, Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes, J. Am. Coll. Cardiol. 48 (2006) 1527–1537, https://doi.org/10.1016/j.jacc.2006.06.055.
[93] M. Opolski, I. Wilson, Asthma and depression: a pragmatic review of the literature and recommendations for future research, Clin. Pract. Epidemiol. Ment. Health CP EMH 1 (2005) 18, https://doi.org/10.1186/1745-0179-1-18.
[94] G. Cizza, P. Ravn, G.P. Chrousos, P.W. Gold, Depression: a major, unrecognized risk factor for osteoporosis? Trends Endocrinol. Metab. TEM 12 (2001) 198–203.
[95] M.R. Patel, M.R. Janevic, S.G. Heeringa, A.P. Baptist, N.M. Clark, An examination of adverse asthma outcomes in US Adults with multiple morbidities, Ann. Am. Thorac. Soc. 10 (2013) 426–431, https://doi.org/10.1513/AnnalsATS.201302-032OC.
[96] P.A. McCullough, J.E. Hollander, R.M. Nowak, A.B. Storrow, P. Duc, T. Omland, J. McCord, H.C. Herrmann, P.G. Steg, A. Westheim, C.W. Knudsen, W.T. Abraham, S. Lamba, A.H.B. Wu, A. Perez, P. Clopton, P. Krishnaswamy, R. Kazanegra, A.S. Maisel, BNP Multinational Study Investigators, Uncovering heart failure in patients with a history of pulmonary disease: rationale for the early use of B-type natriuretic peptide in the emergency department, Acad. Emerg. Med. Off. J. Soc. Acad. Emerg. Med. 10 (2003) 198–204.
[97] S. Abou-Raya, A. Abou-Raya, Osteoporosis and congestive heart failure (CHF) in the elderly patient: double disease burden, Arch. Gerontol. Geriatr. 49 (2009) 250–254, https://doi.org/10.1016/j.archger.2008.09.004.
[98] R.H. Perlis, D.V. Iosifescu, V.M. Castro, S.N. Murphy, V.S. Gainer, J. Minnier, T. Cai, S. Goryachev, Q. Zeng, P.J. Gallagher, M. Fava, J.B. Weilburg, S.E. Churchill, I.S. Kohane, J.W. Smoller, Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model, Psychol. Med. 42 (2012) 41–50, https://doi.org/10.1017/S0033291711000997.
[99] D.A. Regier, W.E. Narrow, D.E. Clarke, H.C. Kraemer, S.J. Kuramoto, E.A. Kuhl, D.J. Kupfer, DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses, Am. J. Psychiatry 170 (2013) 59–70, https://doi.org/10.1176/appi.ajp.2012.12070999.
[100] K.A.S. Davis, C.L.M. Sudlow, M. Hotopf, Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses, BMC Psychiatry 16 (2016) 263, https://doi.org/10.1186/s12888-016-0963-x.
[101] J. Wong, A. Motulsky, M. Abrahamowicz, T. Eguale, D.L. Buckeridge, R. Tamblyn, Off-label indications for antidepressants in primary care: descriptive study of prescriptions from an indication based electronic prescribing system, BMJ 356 (2017) j603.
[102] T.H. McCoy, V.M. Castro, L. Snapper, K. Hart, J.L. Januzzi, J.C. Huffman, R.H. Perlis, Polygenic loading for major depression is associated with specific medical comorbidity, Transl. Psychiatry 7 (2017) e1238, , https://doi.org/10.1038/tp.2017.201.