BMJ Open

A simulation study to demonstrate the biases in the diagnoses of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037022
Article Type:	Original research
Date Submitted by the Author:	26-Feb-2020
Complete List of Authors:	Chao, Yi-Sheng; Independent researcher, Independent researcher Lin, Kuan-Fu; National Taiwan University Hospital Yun-Lin Branch, Psychiatry Wu, Chao-Jung; UQAM, Département d'informatique Wu, Hsing-Chien; Taipei Hospital, Internal Medicine Hsu, Hui-Ting; Changhua Christian Healthcare System, Pathology Tsao, Lien-Cheng; Changhua Christian Healthcare System, Surgery Cheng, Yen-Po; Changhua Christian Healthcare System, Surgery Lai, Yi-Chun; National Yang Ming University Hospital, Chest Medicine Chen, Wei-Chih; Taipei Veterans General Hospital, Chest Medicine; National Yang-Ming University, Institute of Emergency and Critical Care Medicine
Keywords:	MENTAL HEALTH, Depression & mood disorders < PSYCHIATRY, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A simulation study to demonstrate the biases in the diagnoses of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Yi-Sheng Chao[1*], Kuan-Fu Lin,[2] Chao-Jung Wu[3], Hsing-Chien Wu[4], Hui-Ting Hsu[5], Lien-Cheng Tsao[5], Yen-Po Cheng[5], Yi-Chun Lai[6], Wei-Chih Chen[7,8]

[1]Independent researcher, Montréal, H2X 0A8 Canada, [2]National Taiwan University Hospital Yun-Lin Branch, Yunlin County, 640 Taiwan, [3]Département d'informatique Université du Québec à Montréal, Montréal H3B 1B4 Canada, [4]Taipei Hospital Ministry of Health and Welfare New Taipei city, 242 Taiwan, [5]Changhua Christian Hospital, Changhua County 526, Taiwan, [6]National Yang-Ming University Hospital, Yilan 260 Taiwan, [7]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 112, Taiwan, [8]Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei 112, Taiwan

[*]chaoyisheng@post.harvard.edu

Keywords: Frailty; bias; forward-stepwise regression; the Health and Retirement Study; index mining

Abstract

Objectives

Composite diagnostic criteria are likely to introduce biases to the diagnoses that subsequently have poor relationships with input symptoms. This study aims to understand the magnitudes of biases introduced to the diagnoses of three mental illnesses with large disease burdens (major depressive episodes, dysthymic disorder, and manic episodes) and the relationships between the diagnoses and the input symptoms.

Settings

Psychiatric care in general

Participants

Without real-world data available to the public, 100,000 subjects were simulated and the input symptoms were assigned based on the assumed prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations between symptoms (0, 0.1, 0.4, 0.7, and 0.9). The input symptoms were extracted from the diagnostic criteria of three mental illness. The diagnostic criteria were transformed to mathematical equations to convert the input symptoms to diagnoses.

Primary and secondary outcomes

Biases due to data censoring or categorization introduced to the intermediate variables and the three diagnoses were measured. The relationships between the input symptoms and diagnoses were interpreted using forward stepwise linear regressions.

Results

The prevalence rates of the diagnoses were lower than those of the input symptoms and proportional to the assumed prevalence rates and the correlations between the input symptoms. Certain input or bias variables consistently explained the diagnoses better than the others. Except for zero assumed correlations and 0.7 prevalence rates of the input symptoms for the diagnosis of dysthymic disorder, the input variables could not fully explain the diagnoses.

Conclusions

There are biases introduced to the diagnoses of three mental illnesses, major depressive episodes, dysthymic disorder, and manic episodes. The design of the diagnostic criteria determines the prevalence of the diagnoses, the relationships between the input symptoms and the diagnoses, and the biases introduced. The importance of the input variables has been largely distorted by the diagnostic criteria.

Trial registration

Not applicable

Strength and limitation

1. The prevalence of three mental illnesses were determined by the prevalence of the input symptoms and modified by the diagnostic criteria and correlations between the input variables in simulated populations.
2. Biases due to data censoring or categorization were introduced to the intermediate variables and the three diagnoses of mental illnesses in simulated populations.

3. The diagnostic criteria modified the importance of the input variables and certain input or bias variables were given more weights than expected in simulated populations.
4. The design of diagnostic criteria influenced the prevalence. Even with the same input variable prevalence, dysthymic disorder was the most prevalent and major depressive episodes were the lest prevalent in simulated populations.
5. This study is based on simulated data and needs to be verified with real-world data.

Background

The diagnoses of several mental illnesses in patients are made often based on a variety of criteria. These criteria often involve symptoms complained by the patients.[1, 2] For example, the diagnosis of major depressive disorder defined in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) requires at least one major depressive episodes.[1, 2] For each major depressive episode, the major criteria are “depressive mood and/or loss of interest or pleasure in life activities for at least 2 weeks”.[1, 2] In addition to qualify the major criteria, the patients need to report at least five of the nine symptoms that “cause clinically significant impairment in social, work, or other important areas of functioning almost every day”, including insomnia or hypersomnia and fatigue or loss of interest.[1, 2] In other words, patients need to match both the major and minor criteria before being diagnosed with a major depressive episode.

Historically this symptom-based diagnostic approach developed by Feighner et al. has been widely accepted.[3, 4] Since then, mental illnesses can be diagnosed through different sets of criteria. This approach is important because clinicians become capable of screening important symptoms before diagnosing and treating patients accordingly. In fact, these criteria can also be seen as composite measures that use multiple measures to capture disorders that may not be quantified with single variables.[5, 6] Recent studies on composite measures have found that composite measures are problematic because biases can be introduced while aggregating information from input variables.[6] The biases emerge while the sums of input variables are censored or while input variables are transformed inadequately.[6, 7] These biases have been proven vital to the diagnosis of frailty syndrome, a condition that often occurs in the elderly and is significant for several health outcomes.[6] For the diagnosis of frailty syndrome using the Biological Syndrome Model,[8] biases alone can explain more than 71% of the variances of the frailty diagnosis.[6]

Designed as composite measures, it is possible that the diagnostic criteria of mental illnesses also introduce biases to diagnoses so that the diagnoses could not be fully explained by the input symptoms listed in the criteria. This study aims to first understand the relationships between mental symptoms and diagnoses and then quantify the potential role of the biases regarding the diagnoses by simulating populations with different prevalence rates and between-variable correlations of mental symptoms.

Methods

Assumptions and simulation parameters

Simulated populations with mental symptoms of different prevalence rates and between-variable correlations were created to interpret the diagnoses and understand the potential magnitudes of biases that could be introduced via data processing (reproducible using data sets in the S 1 and S 2). Three diagnoses of mental illnesses were chosen for the leading associated disease burdens:[2] major depressive episodes for the diagnosis of major depressive disorder, dysthymic disorder, and manic episodes for the diagnosis of bipolar disorder.[1]

There were assumptions made to simulate the populations (Table 1). First, for each simulation the prevalence rates of the input symptoms were assumed to be similar for the

three diagnoses evaluated in this study. Second, the input symptoms for the diagnoses of major depressive episodes and dysthymic disorder correlated with the same correlation coefficients and those for the diagnosis of manic episodes correlated to one another.[9] Third, the input symptoms for the diagnosis of manic episodes were created independently of those for the diagnosis of the other two mental illnesses. The assumptions of the prevalence rates and between-variable correlations were made because there was no acceptable-quality data on the symptoms of mental illnesses published. There were studies on the prevalence of mental illnesses,[10, 11] but the information on the prevalence of mental symptoms was very limited. There were variables about depression or anxiety collected in national surveys, such as the items collected through the Center for Epidemiologic Studies Depression Scale.[6, 12-18] However, these variables were not the symptoms used in the DSM-IV-TR. Lastly, we assumed that the diagnoses were made accurately based on the presence of the input symptoms and the diagnostic criteria in the DSM-IV-TR. However, this did not hold in the real world.[19] For simplicity and practical reasons, we assumed perfect diagnostic quality by physicians and accurate reporting of the input symptoms by patients in the simulated populations.

Diagnostic criteria as mathematical functions

The input symptoms were extracted from the major and minor criteria of the diagnoses and listed in Table 2 to Table 4. The input symptoms, major and minor criteria, and the diagnoses were labelled with new variable names. All input symptoms, items or domains in the major or minor criteria, and the diagnoses were binomial variables, presenting zero and one for the absence and presence of the symptoms, criteria, and the diagnoses respectively. For example, “insomnia” and “hypersomnia” were extracted from one of the minor criteria for the diagnosis of major depressive episodes. “More talkative than usual” and “pressure to keep talking” were extracted from one of the minor criteria for the diagnosis of manic episodes.

Mathematical functions were generated based on the criteria to convert input symptoms into diagnoses. For example, one of the minor criteria of dysthymic disorder was “poor appetite or overeating”. This required two input symptoms and one bias variable to generate the criterion.[6] “Poor appetite or overeating” equaling the sum of two input variables, “poor appetite” and “overeating”, and a bias variable to achieve censoring of the sum of both variables.[6] The sum of two binomial variables could exceed one and the bias variable had values of -1 for certain subjects to obtain values less than or equal to one in all subjects.[6] In addition to adding variables together to derive an intermediate variable or a diagnosis, multiplication, categorization, and other more complicated methods were used in the diagnostic criteria to generate diagnosis variables and domain variables in the major or minor criteria. For example, the diagnosis of dysthymic disorder required the confirmation of both the major criteria, “depressed mood most of the day for more days than not, for at least 2 years” and the minor criteria, “the presence of two or more of the following symptoms”, at the same time. This was the same as multiplying two binomial variables to obtain the diagnosis of dysthymic disorder. Other equations to generate the intermediate variables and the diagnoses were listed and explained in Table 2 to Table 4.

Generation of bias variables

Bias variables were generated while binomial input symptoms were summed or multiplied to obtain binomial intermediate or diagnosis variables.[6] Therefore the number of bias variables

depended on the complexity of how the diagnoses were made. For example, six of the nine items or domains in the minor criteria for the diagnosis of major depressive episodes were the censored sums of the input symptoms and six bias variables were derived along with the intermediate variables that represented the items in the minor criteria. The other bias variables were described in Table 2 to Table 4.

Simulation parameters and simulated populations

We simulated populations of 100,000 subjects. There were five prevalence rates to simulate the input symptoms for the diagnosis of major depressive episodes, dysthymic disorder, and manic episodes: 0.05, 0.1, 0.3, 0.5, and 0.7. The correlations between the input symptoms were hypothesized to be 0, 0.1, 0.4, 0.7, and 0.9. There were 25 combinations of the assumed prevalence rates and between-variable correlations. The presence of the input symptoms were randomly assigned to the subjects after specifying the prevalence rates and between-variable correlations between the input symptoms.[20, 21] The intermediate and diagnosis variables were derived according to the equations in Table 2 to Table 4. For each combination of prevalence rates and between-variable correlations, the populations were simulated for 100 times to obtain the mean values and 95% confidence intervals (CIs) of derived prevalence rates, as well as the adjusted R squared and p values to approximate the diagnosis variables.

Diagnosis approximation

Due to the existence of the biases, the input symptoms were not likely to fully explain the diagnoses.[6] Therefore, the diagnoses were approximated by the input, bias, and intermediate variables individually or collectively.[6, 12, 14, 16] The approximation was conducted using forward-stepwise linear regressions.[6, 12, 14, 16, 22] The interpretability of the diagnoses by the input symptoms and bias variables was assessed via adjusted R square: zero suggesting that the input symptoms unrelated to the diagnosis; one suggesting that the input symptoms perfectly explained the diagnosis.[14, 15, 23-26]

All statistical analyses were conducted within R environment (v3.4.1)[27] and RStudio (v1.0.153).[28] P values less than 0.05 were considered statistical significant, two-tailed.

Results

The derived prevalence rates of the input symptoms of the three mental illnesses matched the assumed rates in Figure 1. The derived correlations between the input symptoms were close to assumed levels in S 3. The simulations were successful and accurate based on the assumed prevalence rates and correlations.

Prevalence of intermediate variables

The items in the major and minor criteria were the intermediate variables necessary to create the diagnoses. The methods to generate the intermediate variables were as important to the prevalence rates of the intermediate variables as the prevalence rates and correlations of the input symptoms in Figure 2. The intermediate variable, significant unintentional weight loss or gain, was created by summing and censoring two binomial variables with values of zero and one (significant unintentional weight loss; significant unintentional weight gain). The prevalence rates of the

intermediate variables were larger than those of the two input symptoms regardless of the assumed prevalence rates or between-variable correlations of the input symptoms.

In contrast, the diagnosis of dysthymic disorder was a multiplication product of two binomial variables, the major and minor criteria, and the prevalence rates of dysthymic disorder were lower than those of the major and minor criteria under all combinations of assumed correlations and prevalence rates in Figure 3.

Prevalence of mental illnesses

The prevalence rates of three diagnoses were plotted against the assumed prevalence rates and correlations of the input symptoms in Figure 3 to Figure 5 and listed in Table 5. None of the three diagnoses had prevalence rates exceeding those of the input symptoms. In general, higher prevalence rates or between-variable correlations of the input symptoms were associated with higher prevalence rates in the three diagnoses, except for manic episodes that had higher prevalence rates (0.692) assuming zero correlations and 0.7 prevalence rates than the prevalence rate (0.679) assuming 0.1 correlations and 0.7 prevalence rates of the input symptoms. When compared across Figure 3 to Figure 5, given the same assumed prevalence rates and between-variable correlations of the input symptoms, the diagnostic criteria of dysthymic disorder consistently generated diagnoses of the highest prevalence rates and the criteria of major depressive episodes created diagnoses of the least prevalence rates (see Table 5 for details).

Associations between the diagnoses and individual input symptoms and bias variables

The diagnoses were interpreted by the input symptoms (including intermediate variables) and the bias variables individually first. Take dysthymic disorder for example, the diagnosis was interpreted with the input symptoms, the bias variables, and both in Figure 6. For each simulation, the diagnosis of dysthymic disorder was approximated with an increasing number of the input symptoms, the bias variables, or both. After selecting the variables that best approximated the diagnosis based on adjusted R-squared, the input symptoms could explain a proportion of 0.955 of the diagnosis variance and the bias variables could explain at most a proportion of 0.405 of the diagnosis variance in Figure 6. With all variables used in the regression, the diagnosis could be perfectly explained by the input symptoms and bias variables (adjusted R-squared = 1). By repeating the same procedures to the diagnoses, the individual input symptoms and the bias variables that individually best explained the diagnoses were listed in **Error! Reference source not found.** and Table 7 respectively.

For the diagnosis of major depressive episodes, the first and second items in the major criteria (variable names: mde_ma1 for or mde_ma2 in Table 2) individually explained the diagnosis the best depending on the assumed prevalence rates and correlations in **Error! Reference source not found.**. For the diagnosis of dysthymic disorder, the major criteria (dys_ma in Table 3) consistently and individually explained the diagnosis the best. For the diagnosis of manic episodes, the third item of the major criteria (man_ma3 in Table 4) individually explained the diagnosis the best in all combinations of assumed prevalence rates and correlations. However, the proportions of diagnosis variances best explained by individual input symptoms varied in a large range between 0.001 to 0.974 depending on the assumed prevalence rates and between-variable correlations. Based on the adjusted R-squared for individual input symptoms, certain input variables were more important than other symptoms due to high correlation with the diagnoses, such as the major

criteria for the diagnosis of dysthymic disorder. The prevalence rates and between-variable correlations were important to determine the relationships between input symptoms and diagnoses.

Similarly, there were bias variables that consistently explained the diagnoses the best in Table 7. For the diagnosis of major depressive episodes, the bias due to categorization of the numbers of confirmed input symptoms up to three or four (mde_bias1 or mde_bias2 respectively in Table 2) were the leading bias variable. The diagnosis of major depressive episodes not explained by the input symptoms or information censoring (mde_bias in Table 2) were the leading bias variable in two combinations of the assumed prevalence rates and correlations. For the diagnosis of dysthymic disorder, the residual of the diagnosis not explained by the major and minor criteria (dys_bias in Table 3) and the bias due to categorization of the confirmed input symptoms in the minor criteria (dys_mi_bias) were the leading bias variables. For the diagnosis of manic episodes, the bias due to categorization of the number of confirmed input symptoms in the minor criteria up to three (man_bias1 in Table 4) was the leading bias variables, except for two combinations of the assumed prevalence rates and correlations, in which the bias due to categorization of the confirmed input symptoms in the minor criteria up to four (man_bias2 in Table 4) best explained the diagnosis. However, the proportions of diagnosis variances explained by individual bias variables varied in a wide range from zero to 0.87. Depending on the assumed prevalence rates and between-variable correlations of the input symptoms, certain bias variables were more important than other bias variables and even some input variables. The assumed prevalence rates and between-variable correlations were important factors for the relationships between the bias variables and the diagnoses.

In general, the proportions of the diagnosis variances could be explained by either individual input symptoms or single bias variables were low when the prevalence rates and between-variable correlations of the input symptoms were assumed low. With higher assumed prevalence rates or correlations, the proportions of the diagnoses explained by the single input symptoms or bias variables were higher. Across three diagnoses, the diagnosis of dysthymic disorder could be better explained by its own single input variables (higher adjusted R-squared) and the diagnosis of major depressive episodes was associated with the least adjusted R-squared. The bias variables of the diagnosis of manic episodes could explain the diagnosis individually better than the bias variables of the other two diagnoses.

Approximation of the diagnoses with input symptoms

When the diagnoses were approximated with all input symptoms of their own in Table 8, there were always some diagnosis variances that could not be explained by the input symptoms. In other words, the input symptoms could not fully explain the diagnoses, except for the diagnosis of dysthymic disorder that could be fully explained by the input symptoms (adjusted R-squared = 1) assuming zero between-variable correlations and 0.7 prevalence rates for the input symptoms. In Table 8, the proportions of diagnosis variances explained by input symptoms increased with higher assumed prevalence rates or between-variable correlations of the input symptoms in general. The input symptoms of dysthymic disorder explained the diagnosis better than those of the other two diagnoses under all combinations of assumed prevalence rates and between-variable correlations. The diagnosis of major depressive episodes was the worst approximated with its own input symptoms in terms of adjusted R-squared. However, the proportions of diagnosis variances explained by own

input symptoms varied in a wide range between 0.003 to 1.0. The assumed prevalence rates and between-variable correlations of the input symptoms and the design of the diagnostic criteria were all important for the relationships between input symptoms and diagnoses.

Approximating the diagnoses with bias variables

The diagnoses were approximated with the bias variables of their own. The bias variables always explained some of the diagnosis variances, except for the diagnosis of dysthymic disorder assuming zero between-variable correlations and 0.7 prevalence rates for the input symptoms (adjusted R-squared = 0). With increasing assumed between-variable correlations for the input symptoms, the adjusted R-squared increased. However, given the same assumed between-variable correlations, the proportions of diagnosis variances explained by the bias variables might increase or decrease with the assumed prevalence rates. Compared to the adjusted R-squared in Table 8, the proportions of the diagnosis variances explained by the bias variables were always smaller than those explained by the input symptoms in Table 9. However, the proportions of diagnosis variances explained by bias variables also varied in a wide range from zero to 0.89. The assumed prevalence rates and between-variable correlations of input symptoms and the design of the diagnostic criteria were important for the relationship between the bias variables and the diagnoses. Only when the input symptoms for the diagnosis of dysthymic disorder were randomly and independently prevalent to 70% of the simulated populations, the bias variables became irrelevant to the diagnosis.

Discussion

This study is the first attempt to understand the relationships between the input symptoms and the diagnoses of three mental illnesses: major depressive episodes (at least one episode required for the diagnosis of major depressive disorder), dysthymic disorder, and manic episodes. The diagnostic criteria of three mental illnesses have been reviewed and rewrote as mathematical functions. Simulated populations, 100,000 for each of 100 simulations, with input symptoms of the three diagnoses were created. For simplicity and practicality reasons, the presence of the input symptoms was randomly assigned and the input symptoms were assumed to have uniform prevalence rates and between-variable correlations. There were 25 combinations of assumed prevalence rates and between-variable correlations simulated.

Mathematically, the diagnostic criteria are functions and composite measures to transform the information from the input variables to diagnoses. There are bias variables created in the process of information transformation.[6] There are three major mechanisms of introducing biases, censoring, data categorization[7] and multiplication of input symptoms with values of zero and one presenting the absence and presence of the symptoms.[6] These mechanisms introduce information or biases that cannot be fully explained by the input symptoms.[6] The biases introduced can sometimes explain more than half of the variances of the diagnoses depending on the prevalence rates and between-variable correlations of the input symptoms (e.g. assuming input symptoms with 0.7 or 0.9 prevalence rates for the three diagnoses). The findings show that the design of the diagnostic criteria important for bias introduction and significant for the prevalence of the diagnoses in populations, the relationships between the input symptoms and the diagnoses, and the relationships between the bias variables and the diagnoses.

The impact of the diagnostic criteria

With the same assumptions in the prevalence rates and between-variable correlations of the input symptoms, the design of the diagnostic criteria of three mental illnesses can be compared to each other. The design of diagnostic criteria transform input symptoms to various diagnosis prevalence rates with implicit upper limits (i.e. no more prevalent than the input symptoms), unacknowledged differential weights on the input symptoms (i.e. certain input symptoms explaining the diagnoses better), and the introduction of biases (i.e. due to censoring, data categorization, or multiplication).

We were the first to notice that the prevalence rates of the three diagnoses were lower than those of the input symptoms, if randomly distributed with uniform prevalence rates and correlations. Given similar assumed input symptom prevalence and correlations, dysthymic disorder is the most prevalent and major depressive episodes are the least. The diagnosis of dysthymic disorder can be better explained by own input symptoms individually or collectively. The diagnosis of major depressive episodes is the worst explained by own input symptoms individually or collectively. As expected, the diagnosis of the three mental illness are similar to composite measures or indices and are subject to the biases introduced by data processing given all combinations of the assumed prevalence rates and between-variable correlations of the input symptoms.[6] There is only one exception: dysthymic disorder with the input symptoms that are randomly and independently present in 70% of the population. This is because the diagnosis of dysthymic disorder is a multiplication product of the major and minor criteria. Without correlations, everyone in the population is certain to qualify for the minor criteria (probability of 100% because of having at least two out of six item in the minor criteria: mathematically $[C(2,6) + C(3,6) + C(4,6) + C(5,6) + C(6,6)] \times (0.7)^6) = 37 \times 0.117 = 4.35 > 100\%$). When 70% of the population are also randomly assigned with the major criteria, the diagnosis of dysthymic disorder can be fully explained by the major criteria alone. In fact, without correlations between input symptoms it only requires each of the six items in the minor criteria to be randomly assigned to 54.8% $[(1/37)^{(1/6)}]$ of the population for everyone to qualify for the minor criteria and the diagnosis can be fully explained by the minor and major criteria.

Distortion of the input symptoms

The importance of the input symptoms has been distorted due to the functions to generate the diagnoses. This has been proven in the diagnosis of frailty.[6] In other words, based on the functions to generate the diagnoses, the input symptoms are differentially weighted without the weights being explicitly acknowledged. The most prominent is the diagnosis of dysthymic disorder, more than 90% of whose variance can be explained by its major criteria assuming 0.7 or 0.9 between-variable correlations for the input symptoms in Table 6. Another example is that the third item of the major criteria for the diagnosis of manic episodes, “irritable mood”, individually predicts the diagnosis better than any other input symptoms. Assuming 0.9 correlations between input symptoms, this input symptom has been put more weight than others and can explain more than 91.8% of the diagnosis variance. Based on the texts in the DSM-IV-TR, we don’t think this symptom should be emphasized to this degree and consider the diagnostic criteria are imposing implicit and unequal weights to the input symptoms, as well as introducing biases.

Future directions

We think it important to rethink the role and importance of the diagnostic system. Current approaches are embedded with implicit assumptions of the prevalence rates of the diagnoses (no higher than input symptoms), unacknowledged weights to input symptoms (certain input symptoms explaining the diagnoses much better), and biases that could not be explained by the input symptoms. The diagnosis of dysthymic disorder is probably accidentally "designed" to be more prevalent than that of major depressive episodes or manic episodes based on the diagnostic criteria assuming input symptoms with the same prevalence rates. In the real world, there are other important issues related to the diagnostic criteria. For example, diagnosis is not closely linked to treatment,[19, 29] diagnosis is not well made particularly by non-psychiatrists,[30] and there are two diagnostic systems (DSM and International Classification of Disease) that require efforts to harmonize.[31] Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved in a way that they can be easier to understand and use without introducing biases and can be closely linked to clinical decisions. We are developing methods to better detect symptom-based conditions and proposing methods to search for neglected mental symptoms.

Limitations

The strength of this study is the use of simple assumptions in simulated populations that enables the comparison of the diagnostic criteria of three mental illnesses. However, the assumptions in the prevalence rates and between-variable correlations for the input symptoms might not be realistic. Some of the assumptions are unlikely to hold in the real world. However, this is the only option for us due to the lack of real-world data on the prevalence of the input symptoms. In addition, the translation from symptoms to diagnoses was assumed to be perfect based on the diagnostic criteria.

Conclusion

To the best of our knowledge, there is no study on the relationships between the input symptoms and the diagnoses. The input symptoms were extracted from the diagnostic criteria and the diagnostic criteria were transformed to mathematical equations. Without mental illness data available to the public, 100,000 subjects were simulated with different assumptions on the prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations (0, 0.1, 0.4, 0.7, and 0.9) of the input symptoms. We found that biases were introduced to the diagnoses of three mental illnesses, major depressive episodes, dysthymic disorder, and manic episodes. The prevalence rates of the diagnoses were proportional to the assumed prevalence rates and between-variable correlations of the input symptoms. Certain input symptoms were more important than the others to explain the diagnoses. However, the input symptoms could not fully explain the diagnoses, except when the input symptoms independent to each other with 0.7 prevalence rates were used for the diagnosis of dysthymic disorder.

Declarations

Patient and Public Involvement

This is a simulation study that does not require the input from patients or the public.

Acknowledgments

Not applicable.

Ethical Statement

Not applicable.

Funding Statement

The authors received no specific funding for this work.

Consent to participate

Not applicable

Consent for publication

Not applicable

Data Availability

No real-world data used. All analysis based on simulations reproducible with the files in the Supplemental materials.

Competing Interests

YSC is currently employed by the Canadian Agency for Drugs and Technologies in Health. The other authors have declared that no competing interests exist.

Authors' Contributions

YSC conceptualized and designed this study, managed and analyzed data and drafted the manuscript. KFL assisted in the interpretation of the diagnostic criteria. CJW assisted in data management and computation. HCW, HTH, LCT, YPC, YCL, and WCC participated in the design of this study. All authors reviewed and approved the manuscript.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR®). Washington, DC: American Psychiatric Association Publishing; 2010.
2. Center for Substance Abuse Treatment. Managing Depressive Symptoms in Substance Abuse Clients During Early Recovery. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2008.
3. Feighner JP, Robins E, Guze SB, Woodruff RA, Jr., Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry. 1972;26(1):57-63. Epub 1972/01/01.
4. Kendler KS, Munoz RA, Murphy G. The development of the Feighner criteria: a historical perspective. The American journal of psychiatry. 2010;167(2):134-42. Epub 2009/12/17. doi: 10.1176/appi.ajp.2009.09081155.
5. Chao Y-S, Wu C-J. PP46 When Composite Measures Or Indices Fail: Data Processing Lessons. International Journal of Technology Assessment in Health Care. 2019;34(S1):83-. Epub 01/03. doi: 10.1017/S0266462318002088.
6. Chao Y-S, Wu H-C, Wu C-J, Chen W-C. Index or illusion: The case of frailty indices in the Health and Retirement Study. PLOS ONE. 2018;13(7):e0197859. doi: 10.1371/journal.pone.0197859.
7. Barnwell-Menard JL, Li Q, Cohen AA. Effects of categorization method, regression type, and variable distribution on the inflation of Type-I error rate when categorizing a confounding variable. Stat Med. 2015;34(6):936-49. Epub 2014/12/17. doi: 10.1002/sim.6387.
8. Cigolle CT, Ofstedal MB, Tian Z, Blaum CS. Comparing models of frailty: the Health and Retirement Study. J Am Geriatr Soc. 2009;57(5):830-9. Epub 2009/05/21. doi: 10.1111/j.1532-5415.2009.02225.x.
9. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behaviour research and therapy. 1997;35(1):79-89.
10. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. Scientific reports. 2018;8(1):2861-. doi: 10.1038/s41598-018-21243-x. PubMed PMID: 29434331.
11. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. PLOS ONE. 2013;8(11):e75362. doi: 10.1371/journal.pone.0075362.
12. Chao Y-S, Wu C-J. PD26 Principal Component Approximation: Canadian Health Measures Survey. International Journal of Technology Assessment in Health Care. 2019;34(S1):138-9. Epub 01/03. doi: 10.1017/S026646231800301X.
13. Chao Y-S, Wu C-J, Wu H-C, Chen W-C. Trend analysis for national surveys: Application to all variables from the Canadian Health Measures Survey cycle 1 to 4. PLOS ONE. 2018;13(8):e0200127. doi: 10.1371/journal.pone.0200127.
14. Chao Y-S, Wu H-C, Wu C-J, Chen W-C. Principal Component Approximation and Interpretation in Health Survey and Biobank Data. Frontiers in Digital Humanities. 2018;5(11). doi: 10.3389/fdigh.2018.00011.
15. Chao YS, Wu HC, Wu CJ, Chen WC. Stages of Biological Development across Age: An Analysis of Canadian Health Measure Survey 2007-2011. Front Public Health. 2018;5(2296-2565 (Print)). doi: 10.3389/fpubh.2017.00355. eCollection 2017.
16. Chao Y-S, Wu C-J. PD25 Principal Component Approximation: Medical Expenditure Panel Survey. International Journal of Technology Assessment in Health Care. 2019;34(S1):138-. Epub 01/03. doi: 10.1017/S0266462318003008.
17. Chao Y-S, Wu C-J, Chen T-S. Risk adjustment and observation time: comparison between cross-sectional and 2-year panel data from the Medical Expenditure Panel Survey (MEPS). Health Information Science and Systems. 2014;2:5. doi: 10.1186/2047-2501-2-5. PubMed PMID: PMC4340859.

18. Chao YS, Wu HT, Scutari M, Chen TS, Wu CJ, Durand M, et al. A network perspective on patient experiences and health status: the Medical Expenditure Panel Survey 2004 to 2011. BMC health services research. 2017;17(1472-6963 (Electronic)). doi: 10.1186/s12913-017-2496-5.
19. Bonnin JE. Treating without diagnosis: psychoanalysis in medical settings in Argentina. 2015.
20. Leisch F, Weingessel A, Hornik K. On the generation of correlated artificial binary data. 1998.
21. Leisch F, Weingessel A, Hornik K. bindata: Generation of Artificial Binary Data, 2012. URL http://cran/ R-project org/package= bindata R package version 09-19.
22. Lumley T, Lumley MT. Package 'leaps'. Regression Subset Selection Thomas Lumley Based on Fortran Code by Alan Miller Available online: http://cran/ R-project org/package= leaps (accessed on 18 March 2018). 2013.
23. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition: Springer New York; 2009.
24. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R: Springer New York; 2013.
25. Chao Y-S, Wu C-J. Principal component-based weighted indices and a framework to evaluate indices: Results from the Medical Expenditure Panel Survey 1996 to 2011. PLoS ONE. 2017;12(9):e0183997. doi: 10.1371/journal.pone.0183997. PubMed PMID: PMC5590867.
26. Chao YS, Wu HT, Wu CJ. Feasibility of Classifying Life Stages and Searching for the Determinants: Results from the Medical Expenditure Panel Survey 1996-2011. Front Public Health. 2017;5:247(2296-2565 (Print)). doi: 10.3389/fpubh.2017.00247. eCollection 2017.
27. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
28. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2016.
29. Demyttenaere K, Bonnewyn A, Bruffaerts R, De Girolamo G, Gasquet I, Kovess V, et al. Clinical factors influencing the prescription of antidepressants and benzodiazepines:: Results from the European study of the epidemiology of mental disorders (ESEMeD). Journal of affective disorders. 2008;110(1-2):84-93.
30. Margolis RL. Nonpsychiatrist house staff frequently misdiagnose psychiatric disorders in general hospital inpatients. Psychosomatics. 1994;35(5):485-91.
31. First MB. Harmonisation of ICD–11 and DSM–V: opportunities and challenges. The British Journal of Psychiatry. 2009;195(5):382-90.

Table 1. The assumptions and parameters in the simulations

	Assumptions	
1	Equal prevalence rates for the input symptoms of the same diagnosis; presence of input symptoms assigned randomly	
2	Same correlations between the input symptoms of the diagnoses of major depressive episodes and dysthymic disorder; same correlations between the input symptoms of manic episodes	
3	The input symptoms of manic episodes created independent of those of major depressive episodes and dysthymic disorder	
4	Diagnoses made accurately based on the diagnostic criteria and symptoms reported accurately by patients	
	Parameters of input symptoms of the same diagnosis for each simulation	
1	Population sizes	10,000
2	Prevalence rates (uniform for all input symptoms in a simulation)	0.05, 0.1, 0.3, 0.5, and 0.7
3	Correlations (uniform between all input symptoms of the same diagnosis in a simulation)	0, 0.1, 0.4, 0.7, and 0.9
4	Number of simulations for each combination of the assumed prevalence rates and between-variable correlations of the input symptoms	100

519 **Table 2. The input symptoms, intermediate variables, and bias variables for the diagnosis of major depressive episodes.**

Classification of symptoms	Criterion variable	Domains in the major or minor criteria	Domain variables	Symptoms	Symptom variables	Equations to derive diagnosis or domain variables	Approximation by linear regression	Mechanisms related to introducing biases
Major depressive episode (variable = mde)						mde = mde_ma1 x mde_ma2 x (mde_mi3 + mde_mi4 + mde_mi5 + mde_mi6 + mde_mi7 + mde_mi8 + mde_mi9 + mde_bias1) + (1- mde_ma1 x mde_ma2) x (me_ma1 x mde_ma2) x (mde_mi3 + mde_mi4 + mde_mi5 + mde_mi6 + mde_mi7 + mde_mi8 + mde_mi9 + mde_bias2)	mde = intercept + coef1 x mde_ma1 + coef2 x mde_ma2 + coef3 x mde_mi3 + coef4 x mde_mi4 + coef5 x mde_mi5 + coef6 x mde_mi6 + coef7 x mde_mi7 + coef8 x mde_mi8 + coef9 x mde_mi9 + coef10 x mde_bias	1) Multiplication to create the situations when one or two symptoms in the major criteria confirmed and the bias (mde_bias) calculated by extracting the information of the diagnosis not explained by the input symptoms and two bias variables generated by censoring (mde_bias1 and mde_bias2) 2) Categorizing of the sum of the input symptoms in the minor criteria at the threshold of three or four (mde_bias1 and mde_bias2)
Major criteria, essential for diagnosis								
		Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.						
		Depressed mood for more than two weeks.	mde_ma1					
		Loss of interest or pleasure in daily activities for more than two weeks.	mde_ma2					
Minor criteria (at least 5 of the symptoms including the two in major criteria)	mde_mi							
		Significant unintentional weight loss or gain	mde_mi3			mde_mi3 = mde_mi3_1 + mde_mi3_2 + mde_mi3_bias		Censoring of the sum of multiple input variables
				Significant unintentional weight gain	mde_mi3_1			
				Significant unintentional weight loss	mde_mi3_2			
				Information of the domain not explained by the input variables	mde_mi3_bias			
		Insomnia or sleeping too much$	mde_mi4			mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias		Censoring of the sum of multiple input variables
				Insomnia	mde_mi4_1			
				Sleeping too much	mde_mi4_2			
				Information of the domain not explained by the input variables	mde_mi4_bias			
		Agitation or psychomotor retardation noticed by others	mde_mi5			mde_mi5 = mde_mi5_1 + mde_mi5_2 + mde_mi5_bias		Censoring of the sum of multiple input variables
				Agitation	mde_mi5_1			

				Psychomotor retardation noticed by others	mde_mi5_2		
				Information of the domain not explained by the input variables	mde_mi5_bias		
		Fatigue or loss of energy	mde_mi6			mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias	Censoring of the sum of multiple input variables
				Fatigue	mde_mi6_1		
				Loss of energy	mde_mi6_2		
				Information of the domain not explained by the input variables	mde_mi6_bias		
		Feelings of worthlessness or excessive guilt	mde_mi7			mde_mi7 = mde_mi7_1 + mde_mi7_2 + mde_mi7_bias	Censoring of the sum of multiple input variables
				Feelings of worthlessness	mde_mi7_1		
				Feelings of excessive guilt	mde_mi7_2		
				Information of the domain not explained by the input variables	mde_mi7_bias		
		Diminished ability to think or concentrate, or indecisiveness+	mde_mi8			mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias	Censoring of the sum of multiple input variables
				Diminished ability to think or concentrate	mde_mi8_1		
				Indecisiveness	mde_mi8_2		
				Information of the domain not explained by the input variables	mde_mi8_bias		
		Recurrent thoughts of death	mde_mi9				
Information due to categorization (choosing three domains in minor criteria)	mde_bias1						Bias introduced to categorize the sum of the number of confirmed symptoms in the minor criteria
Information due to categorization (choosing four domains in minor criteria)	mde_bias2						Bias introduced to categorize the sum of the number of confirmed symptoms in the minor criteria
Information of diagnosis not explained by the domains	mde_bias						Information of the diagnosis not explained by the input variables and two bias variables generated due to data categorization

523 Table 3. The input symptoms, intermediate variables, and bias variables for the diagnosis of dysthymic disorder.

Classification of symptoms	Criterion variable	Major or minor criteria (domains)	Intermediate variables	Symptoms	Symptom variables	Equations to generate diagnosis or domain variables	Approximation	Mechanisms related to introducing biases
Dysthymia (variable = dys)						dys = dys_ma x dys_mi	dys = intercept + coef1 x dys_ma + coef2 x dys_mi + coef3 x dys_bias	Multiplication to create the situations where both the major and minor criteria met (union of two binomial variables, mde_ma x mde_mi) and the bias variable (dys_bias) equivalent to the residual of the diagnosis not explained by the input symptoms and the bias variables due to censoring and categorization
Major criteria, essential for diagnosis								
		Depressed mood most of the day for more days than not, for at least 2 years	dys_ma					
Minor criteria (at least 2 items)			dys_mi			dys_mi = dys_mi1 + dys_mi2 + dys_mi3 + dys_mi4 + dys_mi5 + dys_mi6 + dys_mi_bias		Categorizing of the sum of multiple input variables
		Poor appetite or overeating	dys_mi1			dys_mi1 = dys_mi1_1 + dys_mi1_2 + dys_mi1_bias		Censoring of the sum of multiple input variables
				Poor appetite	dys_mi1_1			
				Overeating	dys_mi1_2			
				Information of the domain not explained by the input variables	dys_mi1_bias			
		Insomnia or sleeping too much*	dys_mi2/mde_mi4			dys_mi2 = mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias		Censoring of the sum of multiple input variables
				Insomnia	mde_mi4_1			
				Sleeping too much	mde_mi4_2			
				Information of the domain not explained by the input variables	mde_mi4_bias			
		Low energy or fatigue*	dys_mi3/mde_mi6			dys_mi3 = mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias		Censoring of the sum of multiple input variables
				Fatigue	mde_mi6_1			
				Loss of energy (low energy)	mde_mi6_2			
				Information of the domain not explained by the input variables	mde_mi6_bias			
		Low self-esteem	dys_mi4					
		Poor concentration or difficulty making decisions*	dys_mi5/mde_mi8			dys_mi5 = mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias		Censoring of the sum of multiple input variables
				Diminished ability to think or concentrate (Poor concentration)	mde_mi8_1			
				difficulty making decisions (indecisiveness)	mde_mi8_2			
				Information of the domain not explained by the input variables	mde_mi8_bias			
		Feelings of hopelessness	dys_mi6					

		Information of minor criteria not explained by input variables	dys_mi_bias	Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis not explained by major or minor criteria	dys_bias			Information of the diagnosis not explained by the input symptoms and the bias variables generated due to data categorization (dys_mi_bias)

*The same input symptoms used for the diagnosis of major depressive episodes.

527 Table 4. The input symptoms, intermediate variables, and bias variables for the diagnosis of manic episodes.

Classification of symptoms	Criterion variable	Major or minor criteria (domains)	Domain variables	Symptoms	Symptom variables	Equations	Approximation	Mechanisms related to introducing biases
Manic episode (variable = manic)						manic = (1- man_ma1 x man_ma2) x (man_ma1 + man_ma2) x man_ma3 x (man_mi1 + man_mi2 + man_mi3 + man_mi4 + man_mi5 + man_mi6 + man_mi7 + man_bias1) + [1 - (1 - man_ma1 x man_ma2)(man_ma1 + man_ma2)] x man_ma3 x (man_mi1 + man_mi2 + man_mi3 + man_mi4 + man_mi5 + man_mi6 + man_mi7 + man_bias2)	manic = intercept + coef1 x man_ma1 + coef2 x man_ma2 + coef3 x man_ma3 + coef4 x man_mi1 + coef5 x man_mi2 + coef6 x man_mi3 + coef7 x man_mi4 + coef8 x man_mi5 + coef9 x man_mi6 + coef10 x man_mi7 + coef11 x man_bias	1) Multiplication to create the situations where one of the symptom in the major criteria met (union of three binomial variables, such as man_ma1 + man_ma2 and man_ma1 x man_ma2), \n 2) multiplication for the condition of presenting irritable mood (... x man_ma3), and 3) the bias variable (man_bias) equivalent to the residual of the diagnosis not explained by the input symptoms and the bias variables due to censoring; 4) the bias variables introduced by categorizing the number of input symptoms confirmed in the minor criteria (man_bias1 and man_bias2)
Major criteria, essential for the diagnosis of a manic episode (more than one bipolar episode required to diagnose bipolar disorder)								
		A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary)						
				Elevated mood, lasting at least 1 week	man_ma1			
				Expansive mood, lasting at least 1 week	man_ma2			
				Irritable mood, lasting at least 1 week	man_ma3			
Minor criteria (3 or more of the following symptoms have persisted; 4 if the mood is only irritable)								
		Increased self-esteem or grandiosity	man_mi1			man_mi1 = man_mi1_1 + man_mi1_2 + man_mi1_bias		Censoring of the sum of multiple input variables
				Increased self-esteem	man_mi1_1			
				Grandiosity	man_mi1_2			
				Information of the domain not explained by the input variables	man_mi1_bias			

		Decreased need for sleep (e.g., feels rested after only 3 hours of sleep)	man_mi2			
		More talkative than usual or pressure to keep talking	man_mi3		man_mi3 = man_mi3_1 + man_mi3_2 + man_mi3_bias	Censoring of the sum of multiple input variables
				More talkative than usual	man_mi3_1	
				Pressure to keep talking	man_mi3_2	
				Information of the domain not explained by the input variables	man_mi3_bias	
		Flight of ideas or subjective experience that thoughts are racing	man_mi4		man_mi4 = man_mi4_1 + man_mi4_2 + man_mi4_bias	Censoring of the sum of multiple input variables
				Flight of ideas	man_mi4_1	
				Subjective experience that thoughts are racing	man_mi4_2	
				Information of the domain not explained by the input variables	man_mi4_bias	
		Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)	man_mi5			
		Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation	man_mi6		man_mi6 = man_mi6_1 + man_mi6_2 + man_mi6_bias	Censoring of the sum of multiple input variables
				Increase in goal-directed activity	man_mi6_1	
				Psychomotor agitation	man_mi6_2	
				Information of the domain not explained by the input variables	man_mi6_bias	
		Excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)"	man_mi7			
Information of diagnosis due to categorization (choosing at least three symptoms)	man_bias1					Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis due to categorization (choosing at least four symptoms)	man_bias2					Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis not explained by symptoms	man_bias					Information of the diagnosis not explained by the input symptoms and the bias variables generated due to data categorization, man_bias1 and man_bias2

Table 5. The derived prevalence rates of the diagnoses of major depressive episodes, dysthymic disorder, and manic episodes based on the assumed prevalence rates and between-variable correlations of the input symptoms

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0 (95% CI = 0 to 0)	0.004 (95% CI = 0.004 to 0.004)	0 (95% CI = 0 to 0)
0	0.1	0.001 (95% CI = 0.001 to 0.001)	0.025 (95% CI = 0.025 to 0.025)	0.002 (95% CI = 0.002 to 0.002)
0	0.3	0.067 (95% CI = 0.067 to 0.067)	0.249 (95% CI = 0.249 to 0.249)	0.136 (95% CI = 0.135 to 0.136)
0	0.5	0.245 (95% CI = 0.244 to 0.245)	0.493 (95% CI = 0.493 to 0.493)	0.436 (95% CI = 0.436 to 0.436)
0	0.7	0.49 (95% CI = 0.49 to 0.49)	0.7 (95% CI = 0.7 to 0.7)	0.692 (95% CI = 0.692 to 0.693)
0.1	0.05	0.004 (95% CI = 0.004 to 0.004)	0.018 (95% CI = 0.018 to 0.018)	0.007 (95% CI = 0.007 to 0.007)
0.1	0.1	0.011 (95% CI = 0.011 to 0.011)	0.049 (95% CI = 0.049 to 0.049)	0.022 (95% CI = 0.021 to 0.022)
0.1	0.3	0.094 (95% CI = 0.094 to 0.094)	0.25 (95% CI = 0.25 to 0.25)	0.172 (95% CI = 0.171 to 0.172)
0.1	0.5	0.267 (95% CI = 0.267 to 0.268)	0.482 (95% CI = 0.482 to 0.482)	0.425 (95% CI = 0.425 to 0.425)
0.1	0.7	0.51 (95% CI = 0.509 to 0.51)	0.697 (95% CI = 0.697 to 0.697)	0.679 (95% CI = 0.679 to 0.679)
0.4	0.05	0.019 (95% CI = 0.019 to 0.019)	0.037 (95% CI = 0.037 to 0.037)	0.029 (95% CI = 0.029 to 0.029)
0.4	0.1	0.042 (95% CI = 0.042 to 0.042)	0.078 (95% CI = 0.078 to 0.078)	0.062 (95% CI = 0.062 to 0.062)
0.4	0.3	0.166 (95% CI = 0.166 to 0.167)	0.267 (95% CI = 0.267 to 0.267)	0.231 (95% CI = 0.231 to 0.231)
0.4	0.5	0.344 (95% CI = 0.344 to 0.344)	0.476 (95% CI = 0.476 to 0.476)	0.44 (95% CI = 0.44 to 0.441)
0.4	0.7	0.57 (95% CI = 0.57 to 0.57)	0.689 (95% CI = 0.688 to 0.689)	0.666 (95% CI = 0.666 to 0.666)
0.7	0.05	0.035 (95% CI = 0.035 to 0.035)	0.046 (95% CI = 0.046 to 0.046)	0.042 (95% CI = 0.042 to 0.042)
0.7	0.1	0.071 (95% CI = 0.071 to 0.071)	0.092 (95% CI = 0.092 to 0.092)	0.085 (95% CI = 0.085 to 0.085)
0.7	0.3	0.233 (95% CI = 0.233 to 0.234)	0.285 (95% CI = 0.285 to 0.285)	0.27 (95% CI = 0.27 to 0.27)
0.7	0.5	0.422 (95% CI = 0.421 to 0.422)	0.486 (95% CI = 0.485 to 0.486)	0.469 (95% CI = 0.468 to 0.469)
0.7	0.7	0.635 (95% CI = 0.635 to 0.635)	0.69 (95% CI = 0.69 to 0.691)	0.678 (95% CI = 0.677 to 0.678)
0.9	0.05	0.042 (95% CI = 0.042 to 0.042)	0.048 (95% CI = 0.048 to 0.048)	0.046 (95% CI = 0.046 to 0.046)
0.9	0.1	0.085 (95% CI = 0.085 to 0.085)	0.096 (95% CI = 0.096 to 0.097)	0.093 (95% CI = 0.093 to 0.093)
0.9	0.3	0.268 (95% CI = 0.268 to 0.268)	0.293 (95% CI = 0.293 to 0.293)	0.286 (95% CI = 0.286 to 0.287)
0.9	0.5	0.463 (95% CI = 0.463 to 0.463)	0.493 (95% CI = 0.492 to 0.493)	0.485 (95% CI = 0.485 to 0.486)
0.9	0.7	0.669 (95% CI = 0.669 to 0.669)	0.695 (95% CI = 0.694 to 0.695)	0.688 (95% CI = 0.688 to 0.688)

Table 6. The individual input symptoms that best explained the diagnoses: major depressive episodes, dysthymic disorder, and manic episodes

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	mde_ma1	dys_ma	man_ma3
0	0.05	0.001 (95% CI = 0.001 to 0.001)	0.076 (95% CI = 0.075 to 0.077)	0.002 (95% CI = 0.002 to 0.002)
0	0.1	mde_ma1	dys_ma	man_ma3
0	0.1	0.01 (95% CI = 0.01 to 0.01)	0.228 (95% CI = 0.227 to 0.229)	0.021 (95% CI = 0.02 to 0.021)
0	0.3	mde_ma1	dys_ma	man_ma3
0	0.3	0.167 (95% CI = 0.167 to 0.167)	0.774 (95% CI = 0.773 to 0.774)	0.366 (95% CI = 0.366 to 0.367)
0	0.5	mde_ma2	dys_ma	man_ma3
0	0.5	0.324 (95% CI = 0.324 to 0.325)	0.971 (95% CI = 0.971 to 0.971)	0.773 (95% CI = 0.772 to 0.773)
0	0.7	mde_ma2	dys_ma	man_ma3
0	0.7	0.412 (95% CI = 0.412 to 0.412)	0.999 (95% CI = 0.999 to 0.999)	0.964 (95% CI = 0.964 to 0.964)
0.1	0.05	mde_ma2	dys_ma	man_ma3
0.1	0.05	0.07 (95% CI = 0.07 to 0.071)	0.353 (95% CI = 0.352 to 0.355)	0.136 (95% CI = 0.135 to 0.137)
0.1	0.1	mde_ma1	dys_ma	man_ma3
0.1	0.1	0.101 (95% CI = 0.1 to 0.101)	0.462 (95% CI = 0.461 to 0.463)	0.199 (95% CI = 0.198 to 0.199)
0.1	0.3	mde_ma2	dys_ma	man_ma3
0.1	0.3	0.242 (95% CI = 0.242 to 0.243)	0.777 (95% CI = 0.777 to 0.778)	0.483 (95% CI = 0.483 to 0.484)
0.1	0.5	mde_ma2	dys_ma	man_ma3
0.1	0.5	0.365 (95% CI = 0.365 to 0.366)	0.932 (95% CI = 0.931 to 0.932)	0.74 (95% CI = 0.74 to 0.741)
0.1	0.7	mde_ma2	dys_ma	man_ma3
0.1	0.7	0.445 (95% CI = 0.445 to 0.446)	0.986 (95% CI = 0.986 to 0.986)	0.906 (95% CI = 0.906 to 0.907)
0.4	0.05	mde_ma1	dys_ma	man_ma3
0.4	0.05	0.375 (95% CI = 0.373 to 0.376)	0.731 (95% CI = 0.729 to 0.732)	0.561 (95% CI = 0.559 to 0.562)
0.4	0.1	mde_ma1	dys_ma	man_ma3
0.4	0.1	0.395 (95% CI = 0.394 to 0.396)	0.763 (95% CI = 0.762 to 0.764)	0.595 (95% CI = 0.594 to 0.596)
0.4	0.3	mde_ma1	dys_ma	man_ma3
0.4	0.3	0.465 (95% CI = 0.465 to 0.466)	0.851 (95% CI = 0.85 to 0.851)	0.701 (95% CI = 0.701 to 0.702)
0.4	0.5	mde_ma2	dys_ma	man_ma3
0.4	0.5	0.525 (95% CI = 0.524 to 0.525)	0.908 (95% CI = 0.908 to 0.908)	0.787 (95% CI = 0.786 to 0.787)
0.4	0.7	mde_ma2	dys_ma	man_ma3
0.4	0.7	0.568 (95% CI = 0.568 to 0.569)	0.946 (95% CI = 0.946 to 0.947)	0.855 (95% CI = 0.854 to 0.855)
0.7	0.05	mde_ma2	dys_ma	man_ma3
0.7	0.05	0.688 (95% CI = 0.687 to 0.69)	0.909 (95% CI = 0.908 to 0.909)	0.831 (95% CI = 0.83 to 0.832)
0.7	0.1	mde_ma1	dys_ma	man_ma3
0.7	0.1	0.688 (95% CI = 0.687 to 0.689)	0.912 (95% CI = 0.911 to 0.913)	0.836 (95% CI = 0.835 to 0.836)
0.7	0.3	mde_ma2	dys_ma	man_ma3
0.7	0.3	0.71 (95% CI = 0.709 to 0.711)	0.93 (95% CI = 0.93 to 0.93)	0.862 (95% CI = 0.861 to 0.862)
0.7	0.5	mde_ma2	dys_ma	man_ma3
0.7	0.5	0.729 (95% CI = 0.728 to 0.729)	0.944 (95% CI = 0.943 to 0.944)	0.882 (95% CI = 0.882 to 0.883)
0.7	0.7	mde_ma1	dys_ma	man_ma3
0.7	0.7	0.745 (95% CI = 0.744 to 0.745)	0.954 (95% CI = 0.954 to 0.955)	0.9 (95% CI = 0.9 to 0.9)
0.9	0.05	mde_ma1	dys_ma	man_ma3
0.9	0.05	0.828 (95% CI = 0.827 to 0.829)	0.958 (95% CI = 0.957 to 0.958)	0.918 (95% CI = 0.917 to 0.919)
0.9	0.1	mde_ma2	dys_ma	man_ma3
0.9	0.1	0.838 (95% CI = 0.838 to 0.839)	0.961 (95% CI = 0.961 to 0.961)	0.925 (95% CI = 0.924 to 0.925)
0.9	0.3	mde_ma2	dys_ma	man_ma3
0.9	0.3	0.856 (95% CI = 0.856 to 0.857)	0.969 (95% CI = 0.968 to 0.969)	0.937 (95% CI = 0.936 to 0.937)
0.9	0.5	mde_ma2	dys_ma	man_ma3
0.9	0.5	0.862 (95% CI = 0.862 to 0.863)	0.972 (95% CI = 0.972 to 0.972)	0.942 (95% CI = 0.942 to 0.943)
0.9	0.7	mde_ma2	dys_ma	man_ma3
0.9	0.7	0.865 (95% CI = 0.865 to 0.866)	0.974 (95% CI = 0.974 to 0.974)	0.946 (95% CI = 0.946 to 0.946)

Table 7. The individual bias variables that best explained the diagnoses: major depressive episodes, dysthymic disorder, and manic episodes

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	mde_bias2	dys_bias	man_bias2
0	0.05	0 (95% CI = 0 to 0)	0.028 (95% CI = 0.028 to 0.028)	0.001 (95% CI = 0.001 to 0.001)
0	0.1	mde_bias2	dys_bias	man_bias2
0	0.1	0.004 (95% CI = 0.004 to 0.004)	0.053 (95% CI = 0.053 to 0.054)	0.011 (95% CI = 0.011 to 0.011)
0	0.3	mde_bias2	dys_bias	man_bias1
0	0.3	0.015 (95% CI = 0.015 to 0.015)	0.045 (95% CI = 0.045 to 0.045)	0.089 (95% CI = 0.089 to 0.09)
0	0.5	mde_bias	dys_bias	man_bias1
0	0.5	0.013 (95% CI = 0.013 to 0.014)	0.007 (95% CI = 0.007 to 0.007)	0.035 (95% CI = 0.034 to 0.035)
0	0.7	mde_bias	dys_bias	man_bias1
0	0.7	0.01 (95% CI = 0.01 to 0.01)	0 (95% CI = 0 to 0)	0.002 (95% CI = 0.002 to 0.002)
0.1	0.05	mde_bias2	dys_bias	man_bias1
0.1	0.05	0.037 (95% CI = 0.037 to 0.037)	0.113 (95% CI = 0.113 to 0.114)	0.083 (95% CI = 0.083 to 0.084)
0.1	0.1	mde_bias2	dys_bias	man_bias1
0.1	0.1	0.047 (95% CI = 0.047 to 0.048)	0.122 (95% CI = 0.121 to 0.122)	0.116 (95% CI = 0.115 to 0.116)
0.1	0.3	mde_bias2	dys_mi_bias	man_bias1
0.1	0.3	0.077 (95% CI = 0.077 to 0.077)	0.105 (95% CI = 0.105 to 0.106)	0.198 (95% CI = 0.197 to 0.198)
0.1	0.5	mde_bias2	dys_mi_bias	man_bias1
0.1	0.5	0.079 (95% CI = 0.079 to 0.08)	0.073 (95% CI = 0.073 to 0.073)	0.166 (95% CI = 0.166 to 0.167)
0.1	0.7	mde_bias2	dys_mi_bias	man_bias1
0.1	0.7	0.065 (95% CI = 0.065 to 0.065)	0.047 (95% CI = 0.046 to 0.047)	0.094 (95% CI = 0.093 to 0.094)
0.4	0.05	mde_bias1	dys_mi_bias	man_bias1
0.4	0.05	0.294 (95% CI = 0.293 to 0.295)	0.415 (95% CI = 0.413 to 0.416)	0.432 (95% CI = 0.431 to 0.433)
0.4	0.1	mde_bias1	dys_mi_bias	man_bias1
0.4	0.1	0.304 (95% CI = 0.303 to 0.304)	0.419 (95% CI = 0.418 to 0.42)	0.445 (95% CI = 0.444 to 0.445)
0.4	0.3	mde_bias1	dys_mi_bias	man_bias1
0.4	0.3	0.335 (95% CI = 0.334 to 0.335)	0.411 (95% CI = 0.411 to 0.412)	0.473 (95% CI = 0.472 to 0.473)
0.4	0.5	mde_bias1	dys_mi_bias	man_bias1
0.4	0.5	0.354 (95% CI = 0.354 to 0.355)	0.395 (95% CI = 0.395 to 0.396)	0.475 (95% CI = 0.474 to 0.475)
0.4	0.7	mde_bias1	dys_mi_bias	man_bias1
0.4	0.7	0.356 (95% CI = 0.355 to 0.356)	0.367 (95% CI = 0.366 to 0.367)	0.451 (95% CI = 0.45 to 0.451)
0.7	0.05	mde_bias1	dys_mi_bias	man_bias1
0.7	0.05	0.616 (95% CI = 0.615 to 0.617)	0.705 (95% CI = 0.704 to 0.706)	0.723 (95% CI = 0.722 to 0.724)
0.7	0.1	mde_bias1	dys_mi_bias	man_bias1
0.7	0.1	0.611 (95% CI = 0.611 to 0.612)	0.699 (95% CI = 0.698 to 0.699)	0.72 (95% CI = 0.72 to 0.721)
0.7	0.3	mde_bias1	dys_mi_bias	man_bias1
0.7	0.3	0.623 (95% CI = 0.623 to 0.624)	0.699 (95% CI = 0.699 to 0.7)	0.728 (95% CI = 0.728 to 0.729)
0.7	0.5	mde_bias1	dys_mi_bias	man_bias1
0.7	0.5	0.632 (95% CI = 0.632 to 0.633)	0.696 (95% CI = 0.696 to 0.697)	0.731 (95% CI = 0.731 to 0.732)
0.7	0.7	mde_bias1	dys_mi_bias	man_bias1
0.7	0.7	0.639 (95% CI = 0.638 to 0.639)	0.693 (95% CI = 0.692 to 0.693)	0.732 (95% CI = 0.731 to 0.732)
0.9	0.05	mde_bias1	dys_mi_bias	man_bias1
0.9	0.05	0.777 (95% CI = 0.776 to 0.778)	0.835 (95% CI = 0.834 to 0.835)	0.847 (95% CI = 0.847 to 0.848)
0.9	0.1	mde_bias1	dys_mi_bias	man_bias1
0.9	0.1	0.788 (95% CI = 0.788 to 0.789)	0.842 (95% CI = 0.841 to 0.843)	0.855 (95% CI = 0.854 to 0.855)
0.9	0.3	mde_bias1	dys_mi_bias	man_bias1
0.9	0.3	0.807 (95% CI = 0.806 to 0.807)	0.854 (95% CI = 0.853 to 0.854)	0.867 (95% CI = 0.867 to 0.868)
0.9	0.5	mde_bias1	dys_mi_bias	man_bias1
0.9	0.5	0.811 (95% CI = 0.811 to 0.811)	0.855 (95% CI = 0.855 to 0.856)	0.87 (95% CI = 0.87 to 0.871)
0.9	0.7	mde_bias1	dys_mi_bias	man_bias1
0.9	0.7	0.812 (95% CI = 0.811 to 0.812)	0.853 (95% CI = 0.853 to 0.853)	0.869 (95% CI = 0.869 to 0.87)

Table 8. Approximating the diagnoses using input symptoms and derived adjusted R-squared

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0.003 (95% CI = 0.002 to 0.003)	0.122 (95% CI = 0.121 to 0.123)	0.004 (95% CI = 0.004 to 0.005)
0	0.1	0.024 (95% CI = 0.023 to 0.024)	0.305 (95% CI = 0.304 to 0.306)	0.039 (95% CI = 0.038 to 0.039)
0	0.3	0.348 (95% CI = 0.348 to 0.349)	0.842 (95% CI = 0.841 to 0.842)	0.483 (95% CI = 0.482 to 0.483)
0	0.5	0.649 (95% CI = 0.649 to 0.649)	0.986 (95% CI = 0.986 to 0.986)	0.817 (95% CI = 0.817 to 0.817)
0	0.7	0.823 (95% CI = 0.823 to 0.823)	1 (95% CI = 1 to 1)	0.967 (95% CI = 0.967 to 0.967)
0.1	0.05	0.143 (95% CI = 0.141 to 0.144)	0.435 (95% CI = 0.433 to 0.436)	0.212 (95% CI = 0.211 to 0.213)
0.1	0.1	0.198 (95% CI = 0.197 to 0.199)	0.539 (95% CI = 0.538 to 0.54)	0.29 (95% CI = 0.289 to 0.291)
0.1	0.3	0.45 (95% CI = 0.45 to 0.451)	0.826 (95% CI = 0.826 to 0.827)	0.588 (95% CI = 0.588 to 0.589)
0.1	0.5	0.663 (95% CI = 0.663 to 0.664)	0.952 (95% CI = 0.952 to 0.952)	0.799 (95% CI = 0.799 to 0.799)
0.1	0.7	0.809 (95% CI = 0.809 to 0.809)	0.991 (95% CI = 0.991 to 0.991)	0.922 (95% CI = 0.922 to 0.922)
0.4	0.05	0.587 (95% CI = 0.585 to 0.588)	0.782 (95% CI = 0.781 to 0.783)	0.675 (95% CI = 0.674 to 0.676)
0.4	0.1	0.607 (95% CI = 0.606 to 0.608)	0.807 (95% CI = 0.807 to 0.808)	0.698 (95% CI = 0.697 to 0.698)
0.4	0.3	0.688 (95% CI = 0.688 to 0.689)	0.878 (95% CI = 0.877 to 0.878)	0.775 (95% CI = 0.774 to 0.775)
0.4	0.5	0.761 (95% CI = 0.761 to 0.762)	0.925 (95% CI = 0.924 to 0.925)	0.838 (95% CI = 0.838 to 0.838)
0.4	0.7	0.821 (95% CI = 0.821 to 0.822)	0.956 (95% CI = 0.956 to 0.956)	0.887 (95% CI = 0.887 to 0.888)
0.7	0.05	0.813 (95% CI = 0.812 to 0.814)	0.925 (95% CI = 0.925 to 0.926)	0.877 (95% CI = 0.877 to 0.878)
0.7	0.1	0.826 (95% CI = 0.826 to 0.827)	0.928 (95% CI = 0.927 to 0.928)	0.881 (95% CI = 0.881 to 0.882)
0.7	0.3	0.86 (95% CI = 0.86 to 0.86)	0.942 (95% CI = 0.942 to 0.942)	0.9 (95% CI = 0.9 to 0.9)
0.7	0.5	0.88 (95% CI = 0.88 to 0.88)	0.953 (95% CI = 0.953 to 0.953)	0.913 (95% CI = 0.913 to 0.913)
0.7	0.7	0.895 (95% CI = 0.895 to 0.895)	0.962 (95% CI = 0.962 to 0.962)	0.925 (95% CI = 0.925 to 0.925)
0.9	0.05	0.903 (95% CI = 0.903 to 0.904)	0.965 (95% CI = 0.965 to 0.966)	0.941 (95% CI = 0.94 to 0.941)
0.9	0.1	0.91 (95% CI = 0.91 to 0.911)	0.968 (95% CI = 0.968 to 0.968)	0.945 (95% CI = 0.945 to 0.945)
0.9	0.3	0.923 (95% CI = 0.923 to 0.923)	0.974 (95% CI = 0.974 to 0.974)	0.954 (95% CI = 0.953 to 0.954)
0.9	0.5	0.928 (95% CI = 0.928 to 0.928)	0.976 (95% CI = 0.976 to 0.977)	0.958 (95% CI = 0.957 to 0.958)
0.9	0.7	0.932 (95% CI = 0.932 to 0.932)	0.978 (95% CI = 0.978 to 0.978)	0.96 (95% CI = 0.96 to 0.96)

557 Table 9. Approximating the diagnoses using bias variables and derived R-squared

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0.003 (95% CI = 0.002 to 0.003)	0.029 (95% CI = 0.029 to 0.03)	0.004 (95% CI = 0.004 to 0.004)
0	0.1	0.013 (95% CI = 0.012 to 0.013)	0.056 (95% CI = 0.056 to 0.056)	0.017 (95% CI = 0.017 to 0.017)
0	0.3	0.083 (95% CI = 0.083 to 0.083)	0.047 (95% CI = 0.047 to 0.047)	0.098 (95% CI = 0.098 to 0.099)
0	0.5	0.111 (95% CI = 0.111 to 0.112)	0.007 (95% CI = 0.007 to 0.007)	0.039 (95% CI = 0.038 to 0.039)
0	0.7	0.095 (95% CI = 0.095 to 0.095)	0 (95% CI = 0 to 0)	0.012 (95% CI = 0.012 to 0.013)
0.1	0.05	0.083 (95% CI = 0.082 to 0.084)	0.145 (95% CI = 0.144 to 0.146)	0.126 (95% CI = 0.125 to 0.127)
0.1	0.1	0.096 (95% CI = 0.095 to 0.097)	0.156 (95% CI = 0.155 to 0.156)	0.154 (95% CI = 0.153 to 0.154)
0.1	0.3	0.145 (95% CI = 0.144 to 0.145)	0.139 (95% CI = 0.138 to 0.139)	0.216 (95% CI = 0.216 to 0.216)
0.1	0.5	0.172 (95% CI = 0.172 to 0.173)	0.097 (95% CI = 0.097 to 0.097)	0.182 (95% CI = 0.181 to 0.182)
0.1	0.7	0.175 (95% CI = 0.175 to 0.175)	0.065 (95% CI = 0.064 to 0.065)	0.115 (95% CI = 0.115 to 0.116)
0.4	0.05	0.421 (95% CI = 0.419 to 0.423)	0.455 (95% CI = 0.453 to 0.456)	0.505 (95% CI = 0.504 to 0.506)
0.4	0.1	0.422 (95% CI = 0.421 to 0.423)	0.454 (95% CI = 0.453 to 0.455)	0.507 (95% CI = 0.506 to 0.508)
0.4	0.3	0.435 (95% CI = 0.434 to 0.435)	0.442 (95% CI = 0.442 to 0.443)	0.512 (95% CI = 0.512 to 0.513)
0.4	0.5	0.452 (95% CI = 0.452 to 0.453)	0.427 (95% CI = 0.427 to 0.427)	0.506 (95% CI = 0.505 to 0.506)
0.4	0.7	0.46 (95% CI = 0.459 to 0.46)	0.403 (95% CI = 0.402 to 0.403)	0.481 (95% CI = 0.481 to 0.482)
0.7	0.05	0.728 (95% CI = 0.727 to 0.729)	0.729 (95% CI = 0.728 to 0.731)	0.764 (95% CI = 0.763 to 0.765)
0.7	0.1	0.722 (95% CI = 0.721 to 0.723)	0.723 (95% CI = 0.722 to 0.724)	0.76 (95% CI = 0.759 to 0.761)
0.7	0.3	0.726 (95% CI = 0.726 to 0.727)	0.722 (95% CI = 0.722 to 0.723)	0.761 (95% CI = 0.761 to 0.762)
0.7	0.5	0.732 (95% CI = 0.731 to 0.732)	0.72 (95% CI = 0.719 to 0.72)	0.76 (95% CI = 0.76 to 0.761)
0.7	0.7	0.737 (95% CI = 0.736 to 0.737)	0.717 (95% CI = 0.716 to 0.717)	0.758 (95% CI = 0.758 to 0.759)
0.9	0.05	0.852 (95% CI = 0.851 to 0.853)	0.85 (95% CI = 0.849 to 0.851)	0.871 (95% CI = 0.871 to 0.872)
0.9	0.1	0.86 (95% CI = 0.859 to 0.861)	0.857 (95% CI = 0.856 to 0.857)	0.876 (95% CI = 0.876 to 0.877)
0.9	0.3	0.872 (95% CI = 0.871 to 0.872)	0.867 (95% CI = 0.867 to 0.868)	0.886 (95% CI = 0.886 to 0.886)
0.9	0.5	0.874 (95% CI = 0.874 to 0.875)	0.869 (95% CI = 0.868 to 0.869)	0.888 (95% CI = 0.887 to 0.888)
0.9	0.7	0.874 (95% CI = 0.874 to 0.875)	0.867 (95% CI = 0.866 to 0.867)	0.886 (95% CI = 0.886 to 0.886)

Figure 1. The assumed and derived prevalence rates for major depressive episodes, dysthymic disorder, and manic episodes.

Note: each of the combinations of assumed prevalence rates and between-variable correlations of the input symptoms for major depressive episodes, dysthymic disorder, and manic episodes was represented by one circle, but the circles overlapped in the graph.

Figure 2. The prevalence rates of an intermediate variable for the diagnosis of major depressive episodes.

Note: The intermediate variable is “significant unintentional weight loss or gain” and the input symptoms are “significant unintentional weight loss” and “significant unintentional weight gain”. The black line represents the situation where the prevalence rates of the input symptoms are the same as that of the intermediate variable. Lines above the black lines have prevalence rates larger than those of the input symptoms.

Figure 3. The prevalence rates of dysthymic disorder.

Note: Dysthymic disorder is diagnosed when both the major (depressed mood most of the day for more days than not, for at least 2 years) and minor criteria (at least two of the six items) are confirmed. The black line represents the situation where the prevalence rates of the input symptoms are the same as those of the intermediate variable. Lines below the black lines have prevalence rates lower than those of the input symptoms.

Figure 4. The prevalence rates of major depressive episodes.

Note: Major depressive episodes are diagnosed when both major and minor criteria are confirmed. The black line represents the situation where the prevalence rates of the input symptoms are the same as that of the intermediate variable. Lines below the black lines have prevalence rates lower than those of the input symptoms.

Figure 5. The prevalence rates of manic episodes

Note: Manic episodes are diagnosed when the symptoms present as described in the diagnostic manual. The black line represents the situation where the prevalence rates of the input symptoms are the same as those of the input symptoms. Lines below the black lines have prevalence rates lower than those of the input symptoms.

Figure 6. The approximation of the diagnosis of dysthymic disorder by the input symptoms, the bias variables, and both, measured by R-squared

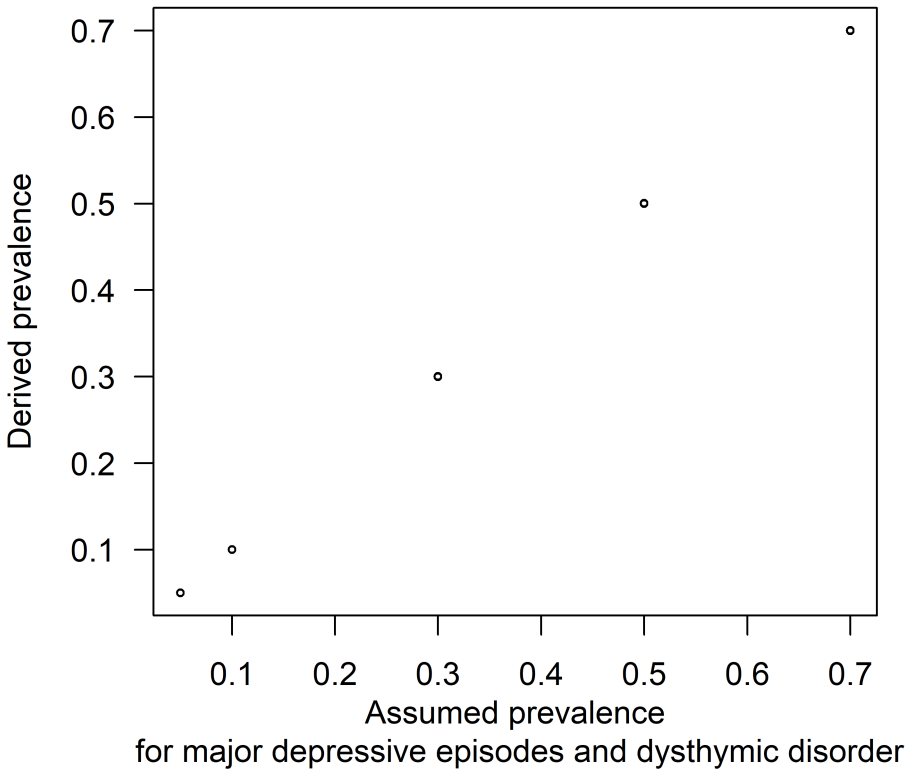
Note: the diagnosis of dysthymic disorder is approximated by the input symptoms, the bias variables, and both using forward-stepwise regression. The selection of the variables was determined by adjusted R-squared. See Table 4 for the details in the input symptoms and the bias variables. The assumed correlations between the input symptoms are 0.4 and the assumed prevalence rates of the input symptoms are 0.7.

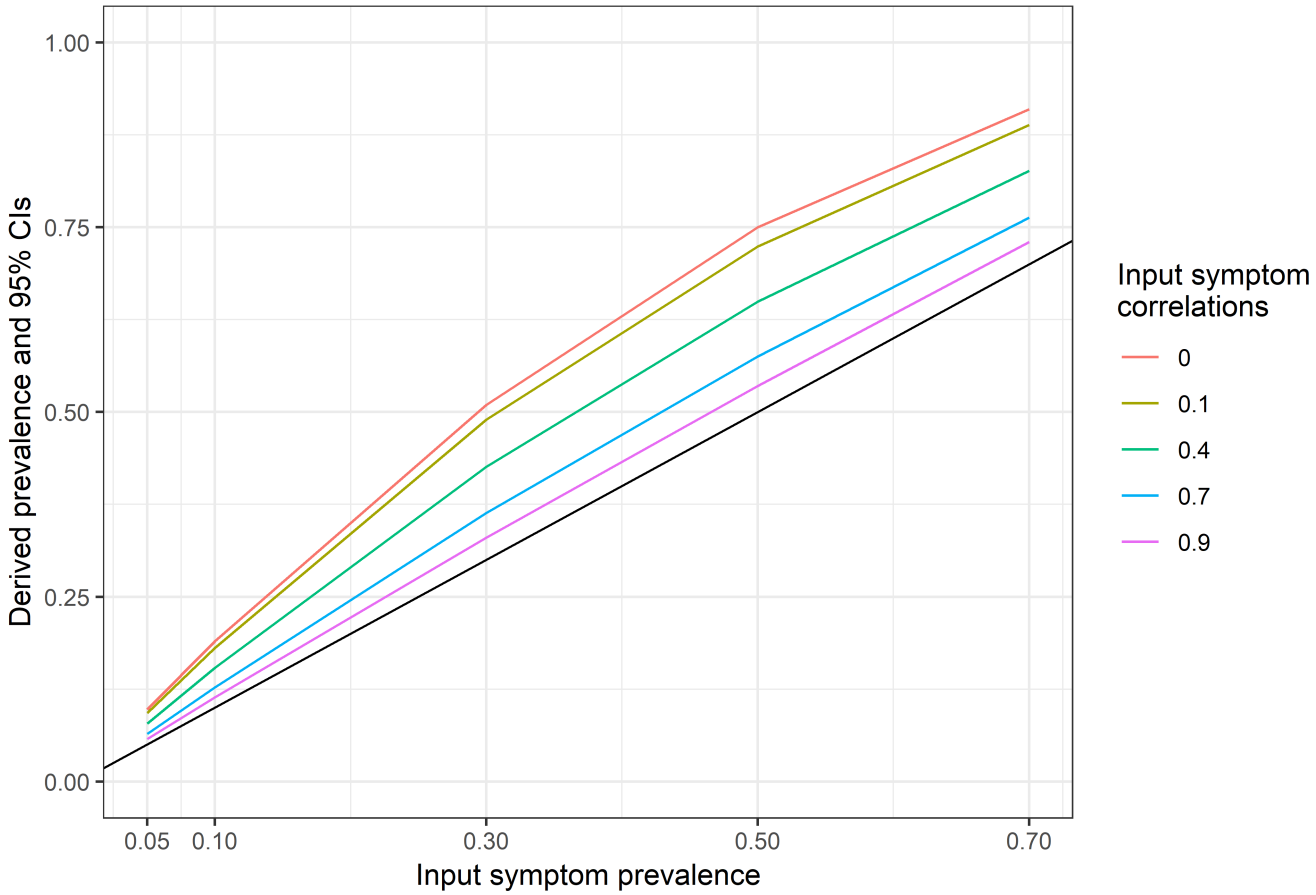
Supplemental materials

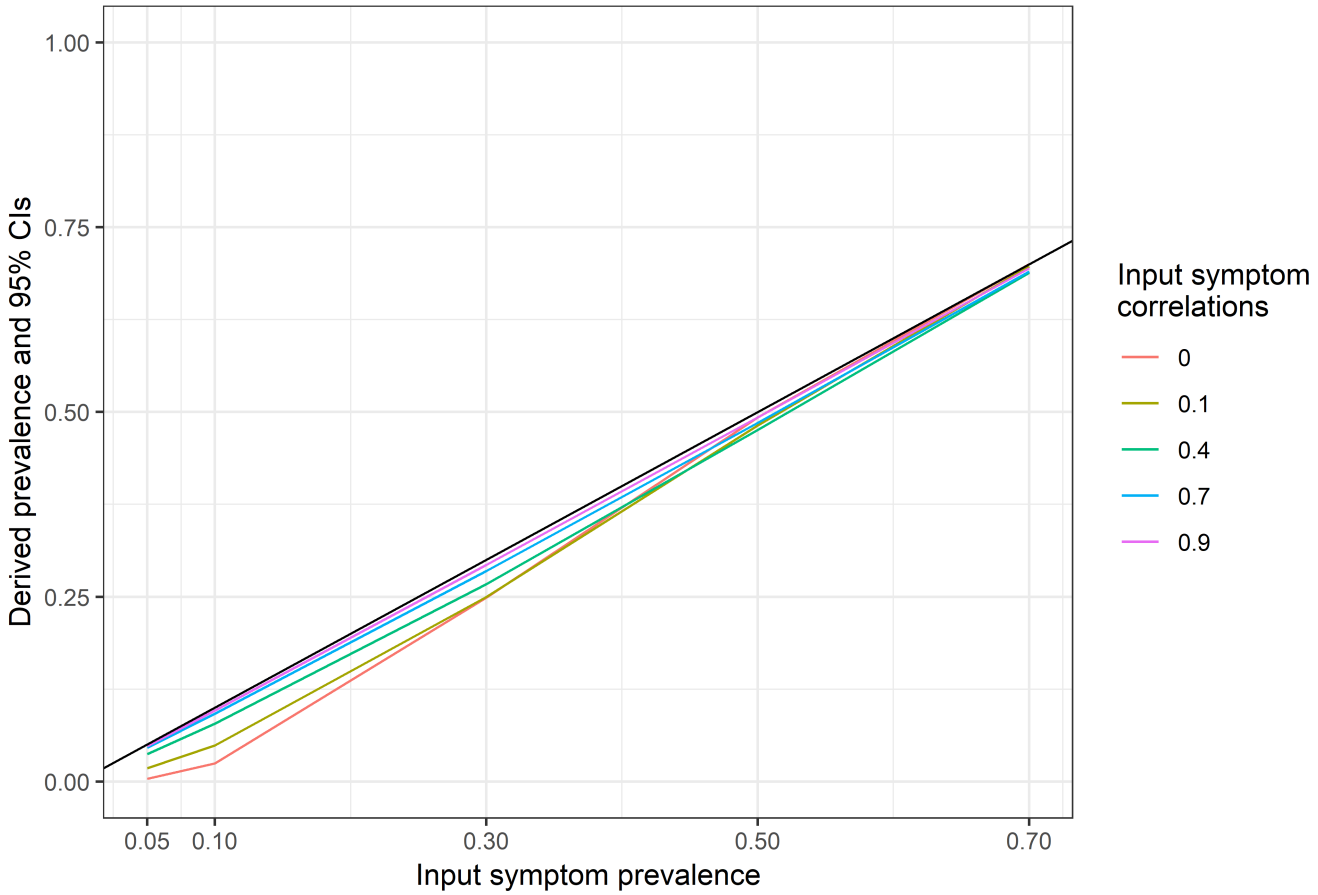
S 1. Characteristics of the input symptoms for simulations

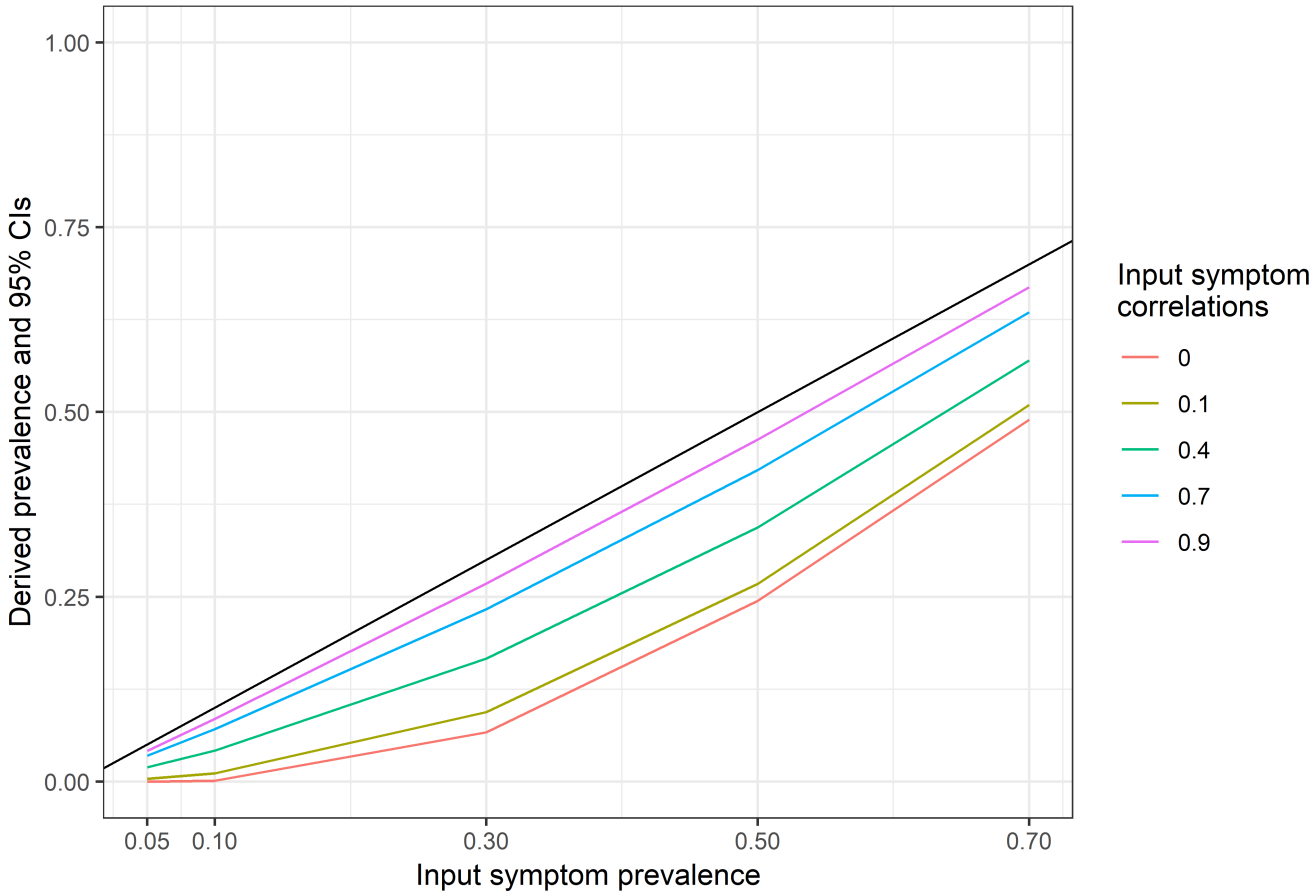
S 2. R codes to be used with S1 to simulate populations

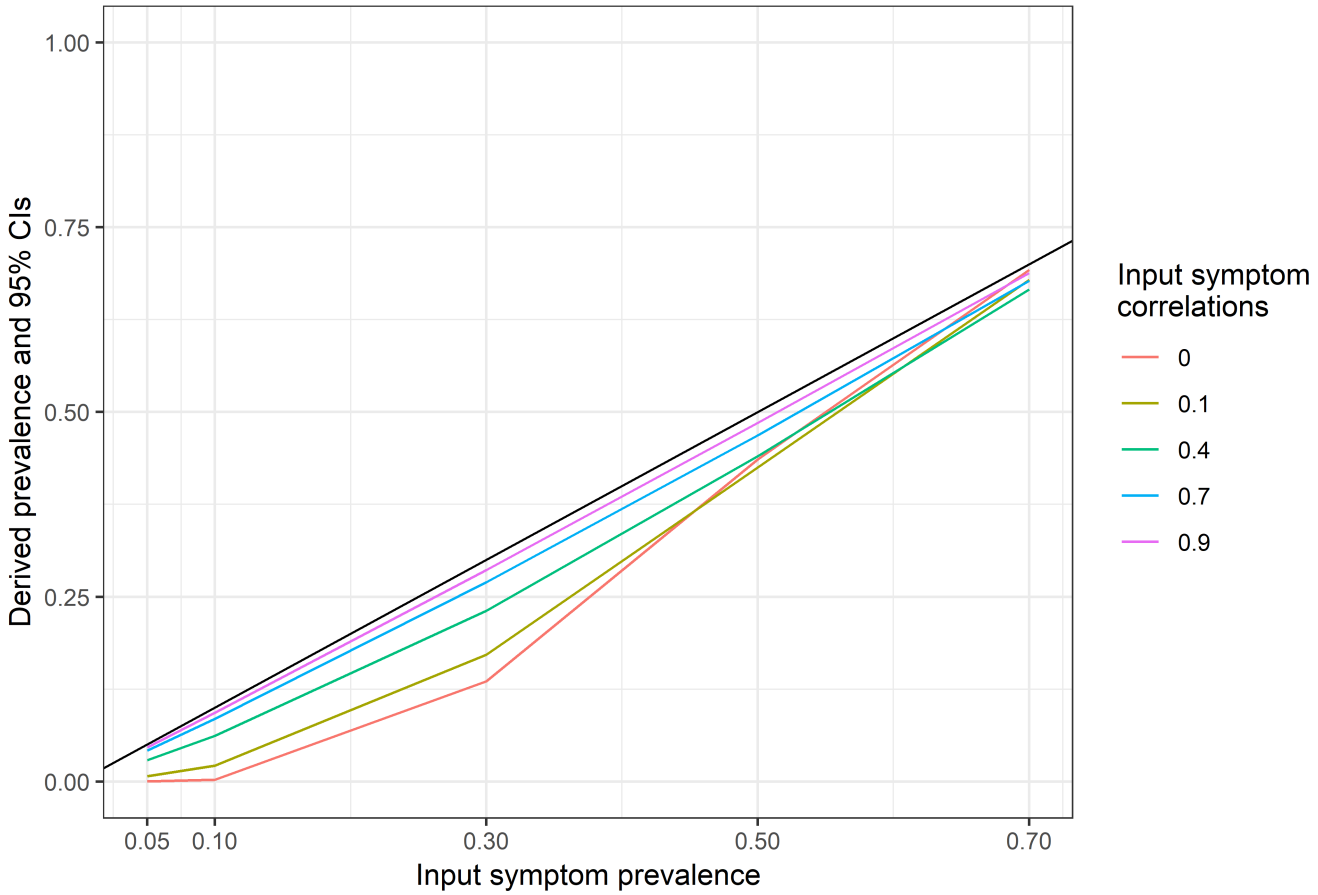
S 3. Correlations between the symptoms

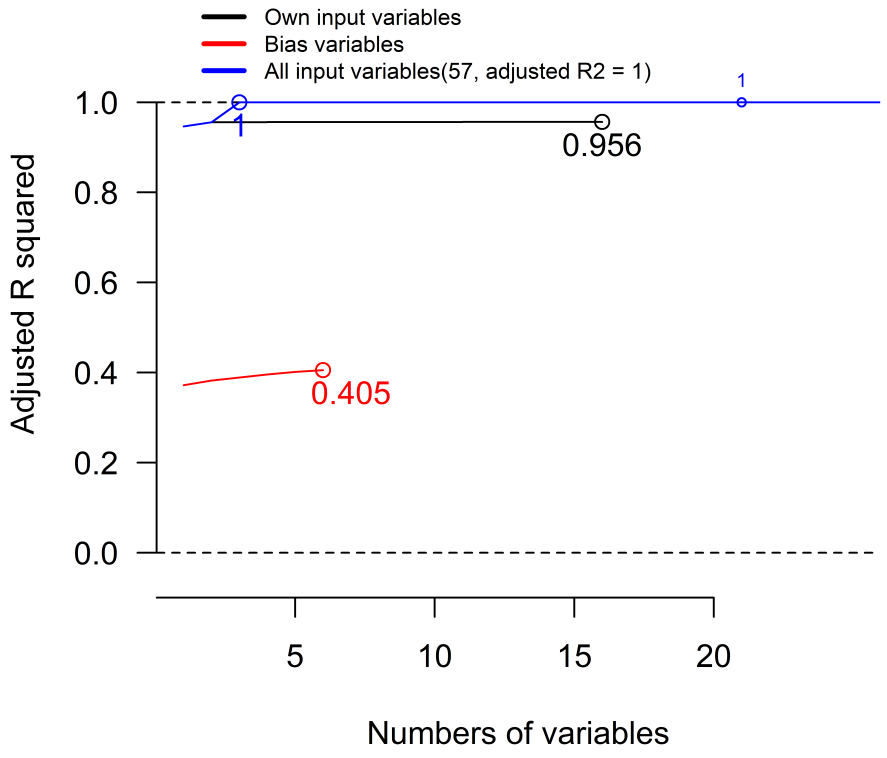

Derived prevalence
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Assumed prevalence
for major depressive episodes and dysthymic disorder


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


Own input variables
Bias variables
All input variables(57, adjusted R2 = 1)
Adjusted R squared
1.0
0.8
0.6
0.4
0.2
0.0
1
1
0.956
0.405
5
10
15
20
Numbers of variables

A simulation study to demonstrate the biases in the diagnoses of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Yi-Sheng Chao[1*], Kuan-Fu Lin,[2] Chao-Jung Wu[3], Hsing-Chien Wu[4], Hui-Ting Hsu[5], Lien-Cheng Tsao[5], Yen-Po Cheng[5], Yi-Chun Lai[6], Wei-Chih Chen[7,8]

[1]Independent researcher, Montréal, H2X 0A8 Canada, [2]National Taiwan University Hospital Yun-Lin Branch, Yunlin County, 640 Taiwan, [3]Département d'informatique Université du Québec à Montréal, Montréal H3B 1B4 Canada, [4]Taipei Hospital Ministry of Health and Welfare New Taipei city, 242 Taiwan, [5]Changhua Christian Hospital, Changhua County 526, Taiwan, [6]National Yang-Ming University Hospital, Yilan 260 Taiwan, [7]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 112, Taiwan, [8]Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei 112, Taiwan
[*]chaoyisheng@post.harvard.edu

Keywords: Frailty; bias; forward-stepwise regression; the Health and Retirement Study; index mining

Mental illness illness variable	Classification	Criterion var	Major or min	Domain varia	Symptoms	Symptom var	Equations	Approximation	Assumed pre	Derived prev	definition	variable	R	formula	code	outcome	suboutcome
Major Depre mde							mde = mde_nmde + intercept + coef1 x mde_ma1 + coef2 x mde_ma2 + coef3 x				Major Depre mde			mde = mde_nsimSmde + simde			
	Major criteria, essential for diagnosis										Major criteria						
		Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.									Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.						
		Depressed m mde_ma1						+ rbinom(n=	0.4)		Depressed m mde_ma1			mde_ma1 = nsimSmde_mamde			
		Loss of interes mde_ma2						+ rbinom(n=	0.4)		Loss of interes mde_ma2			mde_ma2 = nsimSmde_mamde			
	Minor criteria mde_mi										Minor criteria (at least 5 of the symptoms including the two in major criteria)						
		Significant un mde_mi3					mde_mi3 = mde_mi3_1 + mde_mi3_2 + mde_mi3_bias				Significant un mde_mi3			mde_mi3 = 1*(mde_mi3_mde		mde_mi3	
					Significant un mde_mi3_1			+ rbinom(n=	0.4)		Significant un mde_mi3_1			mde_mi3_1 =simSmde_mimde		mde_mi3	
					Significant un mde_mi3_2			+ rbinom(n=	0.4)		Significant un mde_mi3_2			mde_mi3_2 =simSmde_mimde		mde_mi3	
														simSmde_mi3 = 1*((simSmde_mi3_1 + simSmde_mi3_2) > 0)			
		Information mde_mi3_bias									Information mde_mi3_bias			mde_mi3_biassimSmde_mimde		mde_mi3	
					Insomnia or mde_mi4		mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias				Insomnia or mde_mi4			mde_mi4 = 1*(mde_mi4_mde		mde_mi4	
					Insomnia	mde_mi4_1		+ rbinom(n=	0.4)		Insomnia	mde_mi4_1		mde_mi4_1 =simSmde_mimde		mde_mi4	
					Sleeping too	mde_mi4_2		+ rbinom(n=	0.4)		Sleeping too	mde_mi4_2		mde_mi4_2 =simSmde_mimde		mde_mi4	
														simSmde_mi4 = 1*((simSmde_mi4_1 + simSmde_mi4_2) > 0)			
		Information mde_mi4_bias									Information mde_mi4_bias			mde_mi4_biassimSmde_mimde		mde_mi4	
					Agitation or p mde_mi5		mde_mi5 = mde_mi5_1 + mde_mi5_2 + mde_mi5_bias				Agitation or p mde_mi5			mde_mi5 = 1*(mde_mi5_mde		mde_mi5	
					Agitation	mde_mi5_1		+ rbinom(n=	0.4)		Agitation	mde_mi5_1		mde_mi5_1 =simSmde_mimde		mde_mi5	
					Psychomotor mde_mi5_2			+ rbinom(n=	0.4)		Psychomotor mde_mi5_2			mde_mi5_2 =simSmde_mimde		mde_mi5	
														simSmde_mi5 = 1*((simSmde_mi5_1 + simSmde_mi5_2) > 0)			
		Information mde_mi5_bias									Information mde_mi5_bias			mde_mi5_biassimSmde_mimde		mde_mi5	
					Fatigue or lo mde_mi6		mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias				Fatigue or lo mde_mi6			mde_mi6 = 1*(mde_mi6_mde		mde_mi6	
					Fatigue	mde_mi6_1		+ rbinom(n=	0.4)		Fatigue	mde_mi6_1		mde_mi6_1 =simSmde_mimde		mde_mi6	
					Loss of energ mde_mi6_2			+ rbinom(n=	0.4)		Loss of energ mde_mi6_2			mde_mi6_2 =simSmde_mimde		mde_mi6	
														simSmde_mi6 = 1*((simSmde_mi6_1 + simSmde_mi6_2) > 0)			
		Information mde_mi6_bias									Information mde_mi6_bias			mde_mi6_biassimSmde_mimde		mde_mi6	
					Feelings of w mde_mi7		mde_mi7 = mde_mi7_1 + mde_mi7_2 + mde_mi7_bias				Feelings of w mde_mi7			mde_mi7 = 1*(mde_mi7_mde		mde_mi7	
					Feelings of w	mde_mi7_1		+ rbinom(n=	0.4)		Feelings of w	mde_mi7_1		mde_mi7_1 =simSmde_mimde		mde_mi7	
					Feelings of w	mde_mi7_2		+ rbinom(n=	0.4)		Feelings of w	mde_mi7_2		mde_mi7_2 =simSmde_mimde		mde_mi7	
														simSmde_mi7 = 1*((simSmde_mi7_1 + simSmde_mi7_2) > 0)			
		Information mde_mi7_bias									Information mde_mi7_bias			mde_mi7_biassimSmde_mimde		mde_mi7	
					Diminished a mde_mi8		mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias				Diminished a mde_mi8			mde_mi8 = 1*(mde_mi8_mde		mde_mi8	
					Diminished a	mde_mi8_1		+ rbinom(n=	0.4)		Diminished a	mde_mi8_1		mde_mi8_1 =simSmde_mimde		mde_mi8	
					Indecisiveness	mde_mi8_2		+ rbinom(n=	0.4)		Indecisiveness	mde_mi8_2		mde_mi8_2 =simSmde_mimde		mde_mi8	
														simSmde_mi8 = 1*((simSmde_mi8_1 + simSmde_mi8_2) > 0)			
		Information mde_mi8_bias									Information mde_mi8_bias			mde_mi8_biassimSmde_mimde		mde_mi8	
					Recurrent thm mde_mi9			+ rbinom(n=	0.4)		Recurrent thm mde_mi9			mde_mi9 = rbinSmde_mimde			
		Information of the minor criteria not expained by the major or minir criteria									Information of the minor criteria not expained by the major or minir criteria						
														simSmde = simSmde_ma1 * simSmde_ma2 * (simSmde_mi3 + simSmde_mi4 + simSmde_mi5 + simSmde_mi6 + simSmde_mi7 + simSmde_mi8 + simSmde_mi9 + simSmde_bias1) + (1- simSmde_ma1 * simSmde_ma2) * (mde_ma1 * simSmde_ma2) * (simSmde_mi3 + simSmde_mi4 + simSmde_mi5 + simSmde_mi6 + simSmde_mi7 + simSmde_mi8 + simSmde_mi9 + simSmde_bias2)			
	Information omde_bias1										Information omde_bias1			mde_bias1 = simSmde_biamde			
	Information omde_bias2										Information omde_bias2			mde_bias2 = simSmde_biamde			
	Information omde_bias										Information omde_bias			mde_bias = nsimSmde_biamde			
Dysthymic Di dys							dys = dys_madys + intercept + coef1 x dys_ma + coef2 x dys_mi + coef3 x dys_b				Dysthymic Di dys			dys = dys_maxsimSdys + sindys			
	Major criteria, essential for diagnosis										Major criteria						
		dys_ma	Depressed mood most of the day for more days than not, for at least 2 years					+ rbinom(n=	0.4)		Depressed m dys_ma			dys_ma = rbinsimSdys_ma dys			
	Minor criteria dys_mi						dys_mi = dys_mi1 + dys_mi2 + dys_mi3 + dys_mi4 + dys_mi5 + dys_mi6 + dys_mi7 + dys_mi8 + dys_b				Minor criteria dys_mi			dys_mi = 1*(simSdys_mi1 +dys			
		Poor appetite dys_mi1					dys_mi1 = dys_mi1_1 + dys_mi1_2 + dys_mi1_bias				Poor appetite dys_mi1			dys_mi1 = 1*((dys_mi1_1 +dys		dys_mi1	
					Poor appetite dys_mi1_1			+ rbinom(n=	0.4)		Poor appetite dys_mi1_1			dys_mi1_1 = simSdys_mi1dys		dys_mi1	
					Overeating	dys_mi1_2		+ rbinom(n=	0.4)		Overeating	dys_mi1_2		dys_mi1_2 = simSdys_mi1dys		dys_mi1	
														simSdys_mi1 = 1*((simSdys_mi1_1 + simSdys_mi1_2) > 0)			
		Information ody_mi1_bias									Information ody_mi1_bias			dys_mi1_biassimSdys_mi1dys		dys_mi1	
		Insomnia or sdys_mi2					dys_mi2 = mde_mi4 + mde_mi4_1 + mde_mi4_2 + mde_mi4_bias				Insomnia or smde_mi4			mde_mi4 = 1*(mde_mi4_dys			
					Insomnia	mde_mi4_1		+ rbinom(n=	0.4)		Insomnia	mde_mi4_1		mde_mi4_1 =simSmde_midys		mde_mi4	
					Sleeping too	mde_mi4_2		+ rbinom(n=	0.4)		Sleeping too	mde_mi4_2		mde_mi4_2 =simSmde_midys		mde_mi4	
														simSmde_mi4 = 1*((simSmde_mi4_1 + simSmde_mi4_2) > 0)			
		Information mde_mi4_bias									Information mde_mi4_bias			mde_mi4_biassimSmde_midys			
		Low energy ody					dys_mi3 = mde_mi6 + mde_mi6_1 + mde_mi6_2 + mde_mi6_bias				Low energy cmde_mi6			mde_mi6 = 1*(mde_mi6_dys			
					Fatigue	mde_mi6_1		+ rbinom(n=	0.4)		Fatigue	mde_mi6_1		mde_mi6_1 =simSmde_midys			
					Loss of energ	mde_mi6_2		+ rbinom(n=	0.4)		Loss of energ	mde_mi6_2		mde_mi6_2 =simSmde_midys			
														simSmde_mi6 = 1*((simSmde_mi6_1 + simSmde_mi6_2) > 0)			
		Information mde_mi6_bias									Information mde_mi6_bias			mde_mi6_biasSmde_midys			
		Low self-esteem dys_mi4						+ rbinom(n=	0.4)		Low self-esteem dys_mi4			dys_mi4 = rbinSdys_mi4dys			
		Poor concentration dys_mi5					dys_mi5 = mde_mi8 + mde_mi8_1 + mde_mi8_2 + mde_mi8_bias				Poor concentration mi8			mde_mi8 = 1*(mde_mi8_dys			
					Diminished a	mde_mi8_1		+ rbinom(n=	0.4)		Diminished a	mde_mi8_1		mde_mi8_1 =simSmde_midys			
					difficulty ma	mde_mi8_2		+ rbinom(n=	0.4)		difficulty ma	mde_mi8_2		mde_mi8_2 =simSmde_midys			
														simSmde_mi8 = 1*((simSmde_mi8_1 + simSmde_mi8_2) > 0)			
		Information mde_mi8_bias									Information mde_mi8_bias			mde_mi8_biasSmde_midys			
		Feelings of h dys_mi6						+ rbinom(n=	0.4)		Feelings of h dys_mi6			dys_mi6 = rbinSdys_mi6dys			
		Information ody_mi_bias									Information ody_mi_bias			dys_mi_bias simSdys_mi_dys			
														simSdys = simSdys_ma * simSdys_mi			
	Information ody_bias										Information dys_bias			dys_bias = dysimSdys_biasdys			
Manic episod manic							manic = (1- mmanic + intercept + coef1 x man_ma1 + coef2 x man_ma2 + coef3				Manic episod manic			manic = (1- mnimSmanic + dys			
	Major criteria, essential for the diagnosis of a manic episode (more than one bipolar episode required to diagnose bipolar disorder)										Major criteria						
		A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hosp									A distinct per						
					Elevated ma man_ma1			+ rbinom(n=	0.1)		Elevated ma man_ma1			man_ma1 = nsimSman_mamanic			
					Expansive m man_ma2			+ rbinom(n=	0.1)		Expansive m man_ma2			man_ma2 = nsimSman_mamanic			
					Irritable mo man_ma3			+ rbinom(n=	0.1)		Irritable mo man_ma3			man_ma3 = nsimSman_mamanic			
	Minor criteria (3 or more of the following symptoms have persisted; 4 if the mood is only irritable)										Minor criteria						
		Increased self man_mi1					man_mi1 = man_mi1_1 + man_mi1_2 + man_mi1_bias				Increased self man_mi1			man_mi1 = 1*(man_mi1_manic		man_mi1	
					Increased self man_mi1_1			+ rbinom(n=	0.1)		Increased self man_mi1_1			man_mi1_1 =simSman_mimanic		man_mi1	
					Grandiosity	man_mi1_2		+ rbinom(n=	0.1)		Grandiosity	man_mi1_2		man_mi1_2 =simSman_mimanic		man_mi1	
														simSman_mi1 = 1*((simSman_mi1_1 + simSman_mi1_2) >0)			
		Information cman_mi1_bias									Information cman_mi1_bias			man_mi1_biasSimSman_mimanic		man_mi1	
					Decreased n man_mi2			+ rbinom(n=	0.1)		Decreased n man_mi2			man_mi2 = rbinSman_mimanic			
					More talkativ man_mi3			man_mi3 = man_mi3_1 + man_mi3_2 + man_mi3_bias			More talkativ man_mi3			man_mi3 = 1*(man_mi3_manic		man_mi3	
					More talkativ	man_mi3_1		+ rbinom(n=	0.1)		More talkativ	man_mi3_1		man_mi3_1 =simSman_mimanic		man_mi3	
					Pressure to k	man_mi3_2		+ rbinom(n=	0.1)		Pressure to k	man_mi3_2		man_mi3_2 =simSman_mimanic		man_mi3	
														simSman_mi3 = 1*((simSman_mi3_1 + simSman_mi3_2) > 0)			
		Information cman_mi3_bias									Information cman_mi3_bias			man_mi3_biasSimSman_mimanic		man_mi3	
					Flight of ideas man_mi4			man_mi4 = man_mi4_1 + man_mi4_2 + man_mi4_bias			Flight of idea man_mi4			man_mi4 = 1*(man_mi4_manic		man_mi4	
					Flight of idea	man_mi4_1		+ rbinom(n=	0.1)		Flight of idea	man_mi4_1		man_mi4_1 =simSman_mimanic		man_mi4	
					Subjective e man_mi4_2			+ rbinom(n=	0.1)		Subjective e man_mi4_2			man_mi4_2 =simSman_mimanic		man_mi4	
														simSman_mi4 = 1*((simSman_mi4_1 + simSman_mi4_2) > 0)			
		Information cman_mi4_bias									Information cman_mi4_bias			man_mi4_biasSimSman_mimanic		man_mi4	
					Distractibility man_mi5			+ rbinom(n=	0.1)		Distractibility man_mi5			man_mi5 = rbinSman_mimanic			
					Increase in g man_mi6			man_mi6 = man_mi6_1 + man_mi6_2 + man_mi6_bias			Increase in g man_mi6			man_mi6 = 1*(man_mi6_manic		man_mi6	
					Increase in g	man_mi6_1		+ rbinom(n=	0.1)		Increase in g	man_mi6_1		man_mi6_1 =simSman_mimanic		man_mi6	
					Psychomotor	man_mi6_2		+ rbinom(n=	0.1)		Psychomotor	man_mi6_2		man_mi6_2 =simSman_mimanic		man_mi6	
														simSman_mi6 = 1*((simSman_mi6_1 + simSman_mi6_2) >0)			
		Information cman_mi6_bias									Information cman_mi6_bias			man_mi6_biasSimSman_mimanic		man_mi6	
					Excessive inv man_mi7			+ rbinom(n=	0.1)		Excessive inv man_mi7			man_mi7 = rbinSman_mimanic			
														simSmanic = (1- simSman_ma1 * simSman_ma2) * (simSman_ma1 + simSman_ma2) * simSman_ma3 * (simSman_mi1 + simSman_mi2 + simSman_mi3 + simSman_mi4 + simSman_mi5 + simSman_mi6 + simSman_mi7 + simSman_bias3) + [1 - [1 - simSman_ma1 * simSman_ma2](simSman_ma1 + simSman_ma2)] * simSman_ma3 * (simSman_mi1 + simSman_mi2 + simSman_mi3 + simSman_mi4 + simSman_mi5 + simSman_mi6 + simSman_mi7 + simSman_bias2)			
	Information cman_bias1										Information cman_bias1			man_bias1 = simSman_biamanic			
	Information cman_bias2										Information cman_bias2			man_bias2 = simSman_biamanic			
	Information cman_bias										Information cman_bias			man_bias = nsimSman_biamanic			
	female										female			female = rbinsimSfemale +mde			
	age										age			age = sample.simSage + samde			
	edu										edu			edu = rnorm(simSedu + mmde			
														edu[which(sasimSedu[which(simSedu <= 0)] = 0			

```
---
title: "2019_09_06 simulated mental illnesses"
author: "Yi-Sheng Chao"
date: "November 22, 2018"
output: pdf_document
editor_options:
  chunk_output_type: inline
---


##Adding correlations to the random variables

```{r}
library(bindata)

library(openxlsx)
resu = read.xlsx("A simulation study to demonstrate the biases in three
diagnoses of mental illnesses.xlsx", sheet = "Prob 1")
names(resu)
unique(resu$variable)
memory.limit(size = 10^13)
ssize = 10^5
times = 10^2

prevalence = c(0.05, 0.1, 0.3, 0.5, 0.7)
rho = c(0, 0.1, 0.4, 0.7, 0.9)#correlation coefficients of the input
symptoms

collect = c("mean", "max",
"min","derivedprevalence","coef","coefse","p","intercept",
"interceptp","r2", "subcoef","subcoefse","subp","subintercept",
"subinterceptp","subr2", "appbyownr2", "appbybiasr2", "appbyallr2",
"appbyownvar", "appbybiasvar", "appbyallvar", "appbyownn", "appbybiasn",
"appbyalln")


set.seed(1)


##Create a simulated data set to extract variables
for(preval in 1:length(prevalence)){
  for(rh in 1:length(rho)){


  library(openxlsx)
resu = read.xlsx("A simulation study to demonstrate the biases in three
diagnoses of mental illnesses.xlsx", sheet = "Prob 1")

    # foreach(c = 1:times) %dopar% {
    for(c in 1:times){

library(bindata)
bindata = as.data.frame(rmvbin(ssize, rep(prevalence[preval], 40),
bincorr=(1 - rho[rh])*diag(40) + rho[rh]))
bindata2 = as.data.frame(rmvbin(ssize, rep(prevalence[preval], 20),
bincorr=(1 - rho[rh])*diag(20) + rho[rh]))

##demographic characteristics
```
```

```
sim = data.frame(1:ssize)
names(sim) = "id"
sim$female = rbinom(n = ssize, size = 1, prob = 0.51)
sim$age = sample(30:60, ssize, replace = TRUE)
sim$edu = rnorm(ssize, mean = 12, sd = 5)
sim$edu[which(sim$edu <= 0)] = 0
sim$id = NULL


sim$mde_ma1 = bindata[,1]
sim$mde_ma2 = bindata[,2]

sim$mde_mi3_1 = bindata[,3]
sim$mde_mi3_2 = bindata[,4]
sim$mde_mi3 = 1*((sim$mde_mi3_1 + sim$mde_mi3_2) > 0)
sim$mde_mi3_bias = sim$mde_mi3 - sim$mde_mi3_1 - sim$mde_mi3_2

sim$mde_mi4_1 = bindata[,5]
sim$mde_mi4_2 = bindata[,6]
sim$mde_mi4 = 1*((sim$mde_mi4_1 + sim$mde_mi4_2) > 0)
sim$mde_mi4_bias = sim$mde_mi4 - sim$mde_mi4_1 - sim$mde_mi4_2

sim$mde_mi5_1 = bindata[,7]
sim$mde_mi5_2 = bindata[,8]
sim$mde_mi5 = 1*((sim$mde_mi5_1 + sim$mde_mi5_2) > 0)
sim$mde_mi5_bias = sim$mde_mi5 - sim$mde_mi5_1 - sim$mde_mi5_2

sim$mde_mi6_1 = bindata[,9]
sim$mde_mi6_2 = bindata[,10]
sim$mde_mi6 = 1*((sim$mde_mi6_1 + sim$mde_mi6_2) > 0)
sim$mde_mi6_bias = sim$mde_mi6 - sim$mde_mi6_1 - sim$mde_mi6_2

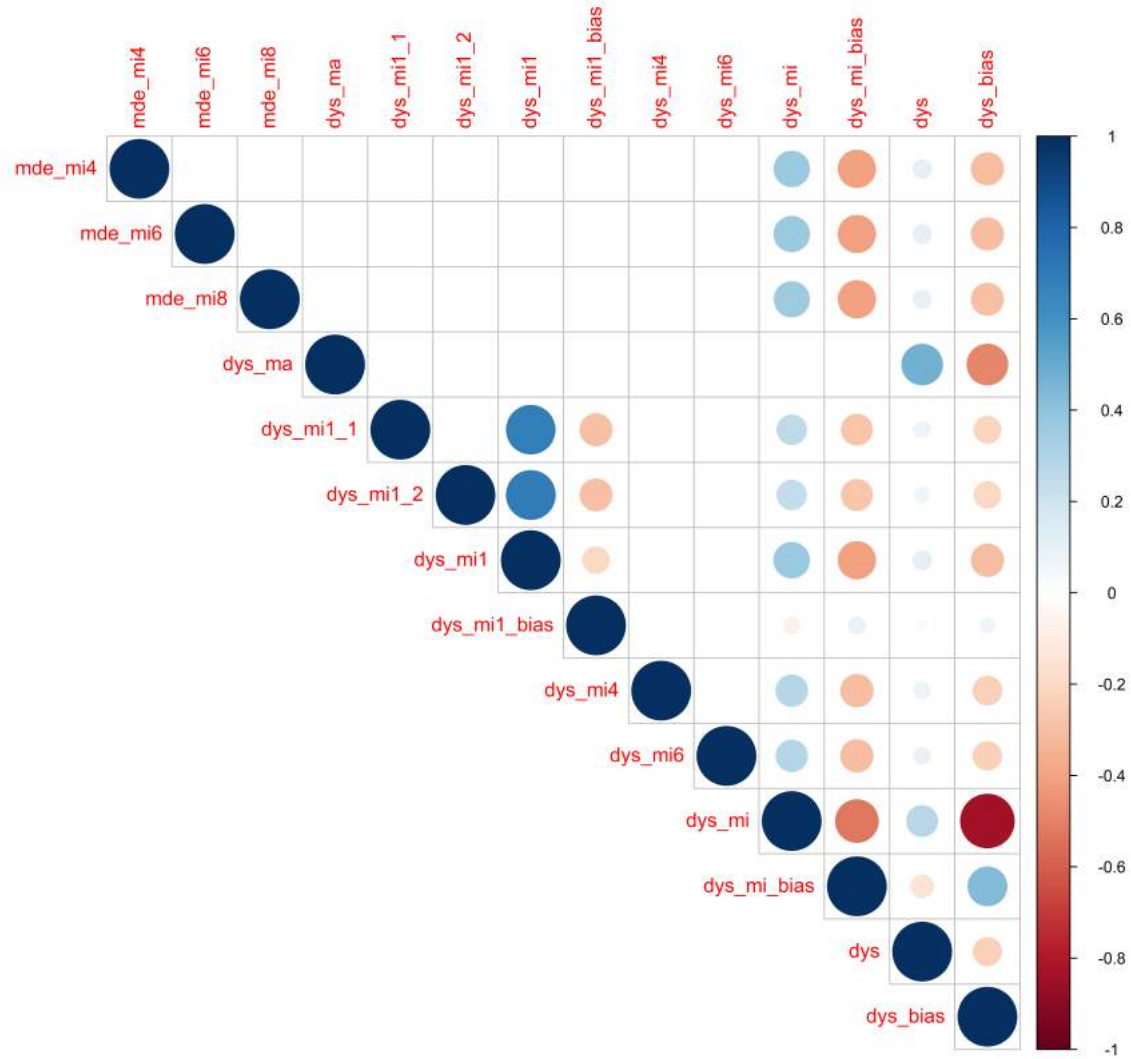
sim$mde_mi7_1 = bindata[,11]
sim$mde_mi7_2 = bindata[,12]
sim$mde_mi7 = 1*((sim$mde_mi7_1 + sim$mde_mi7_2) > 0)
sim$mde_mi7_bias = sim$mde_mi7 - sim$mde_mi7_1 - sim$mde_mi7_2

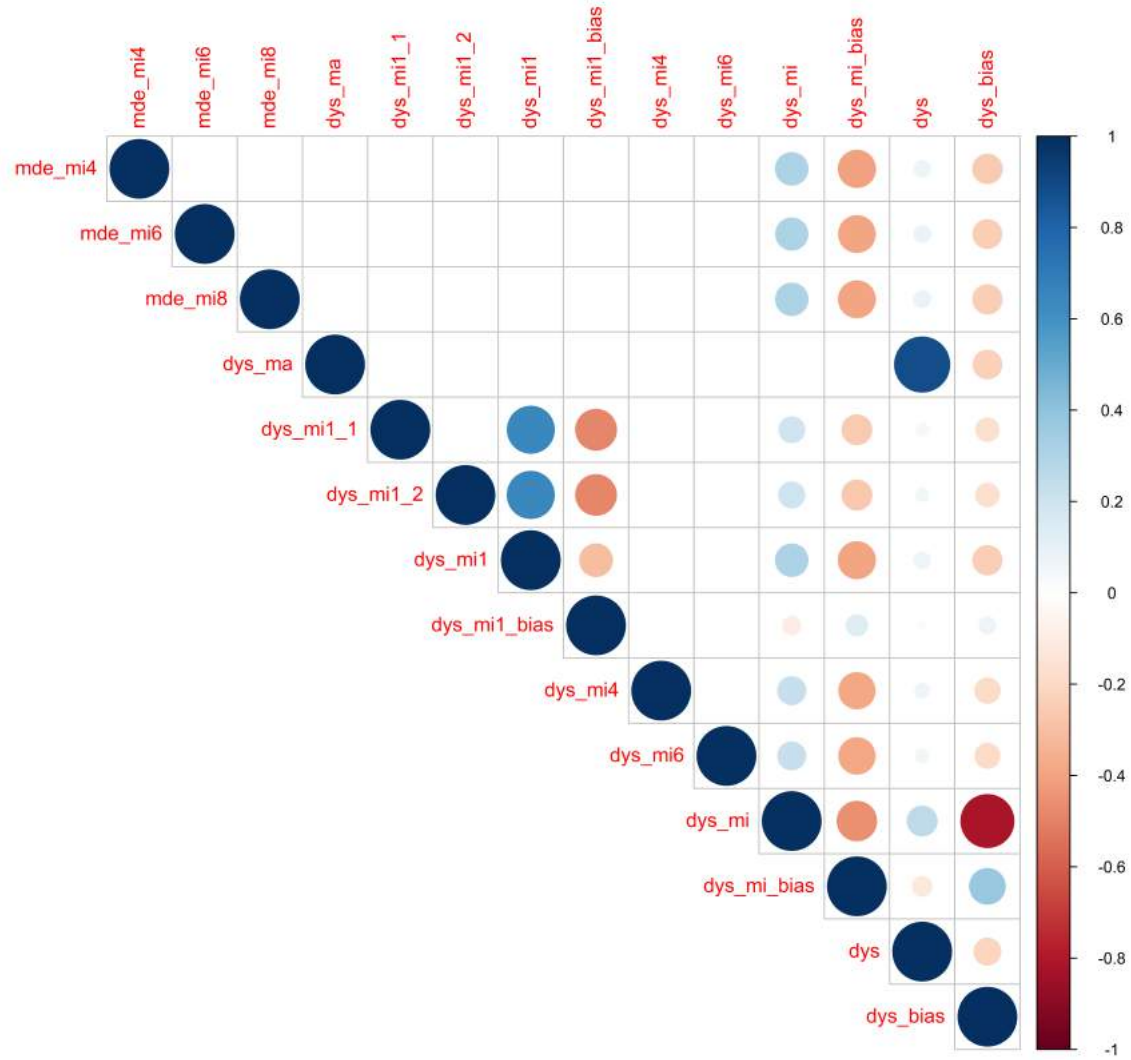
sim$mde_mi8_1 = bindata[,13]
sim$mde_mi8_2 = bindata[,14]
sim$mde_mi8 = 1*((sim$mde_mi8_1 + sim$mde_mi8_2) > 0)
sim$mde_mi8_bias = sim$mde_mi8 - sim$mde_mi8_1 - sim$mde_mi8_2

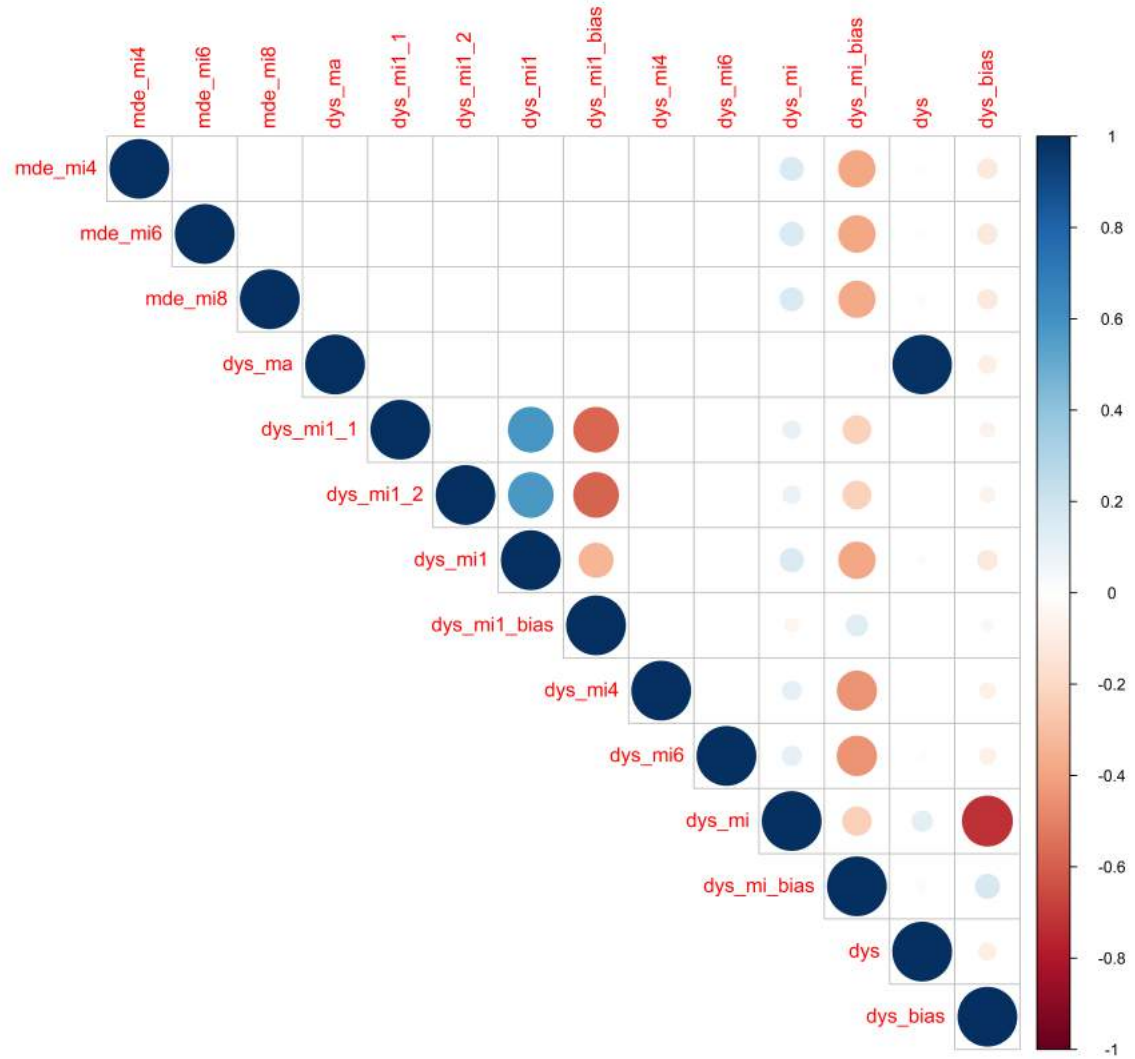
sim$mde_mi9 = bindata[,15]

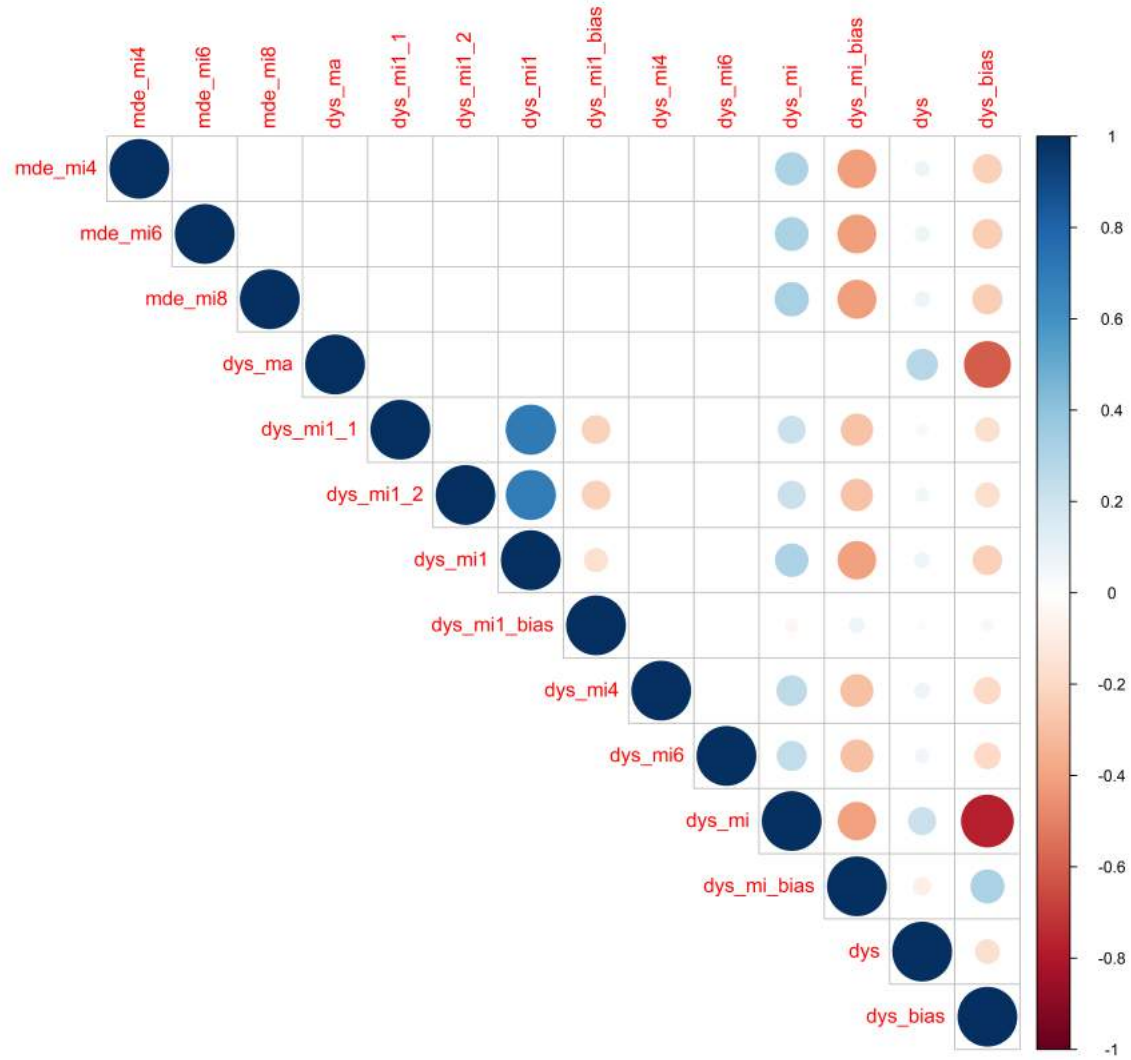
sim$mde_bias1 = 1 * ((sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 +
sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9)>2) - (sim$mde_mi3
+ sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9)
sim$mde_bias2 = 1 * ((sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 +
sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9)>3) - (sim$mde_mi3
+ sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9)

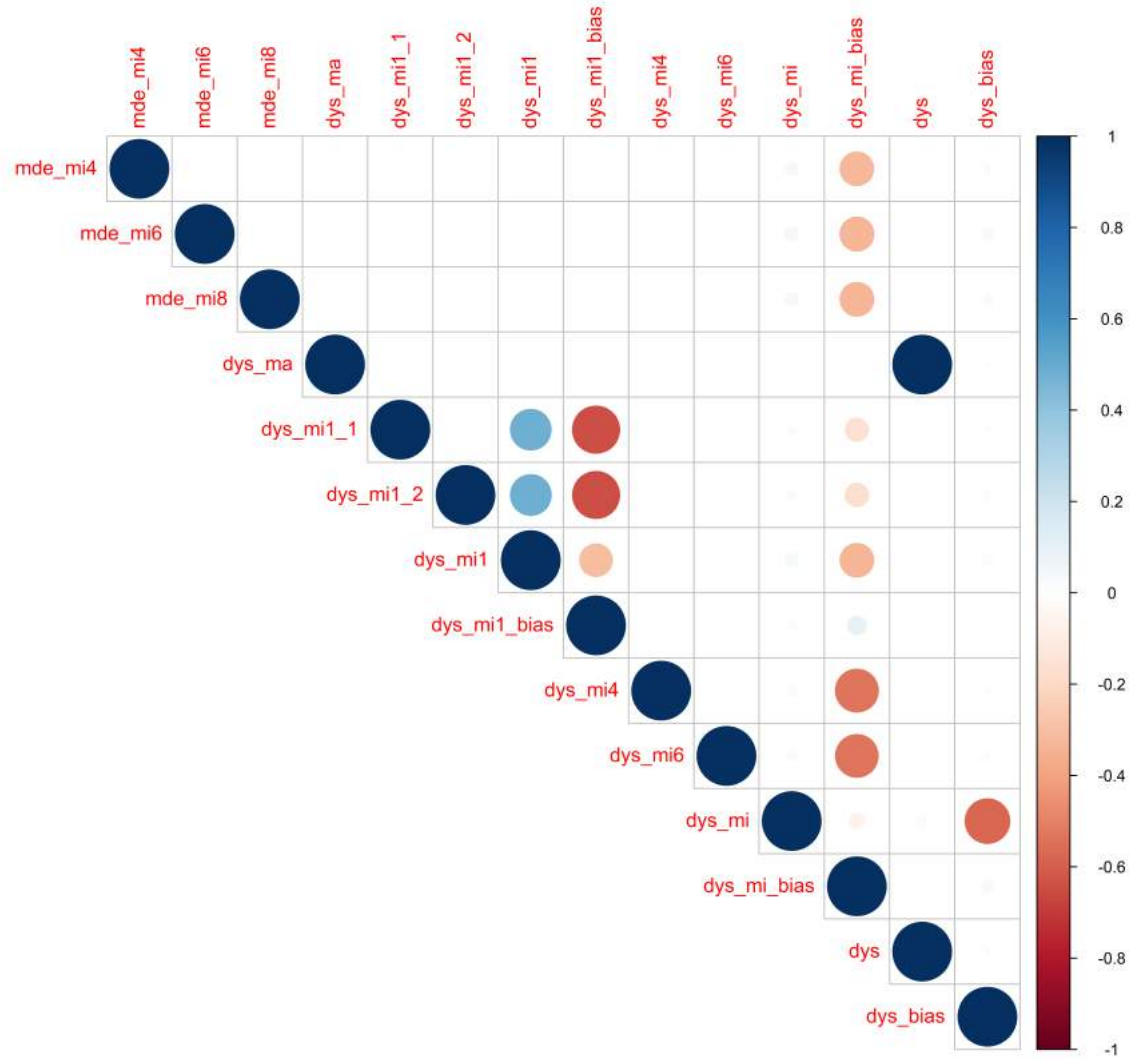
sim$mde = sim$mde_ma1 * sim$mde_ma2 * (sim$mde_mi3 + sim$mde_mi4 +
sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9 +
sim$mde_bias1) + (1- sim$mde_ma1 * sim$mde_ma2) * (sim$mde_ma1 *
sim$mde_ma2) * (sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 +
sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9 + sim$mde_bias2)
```

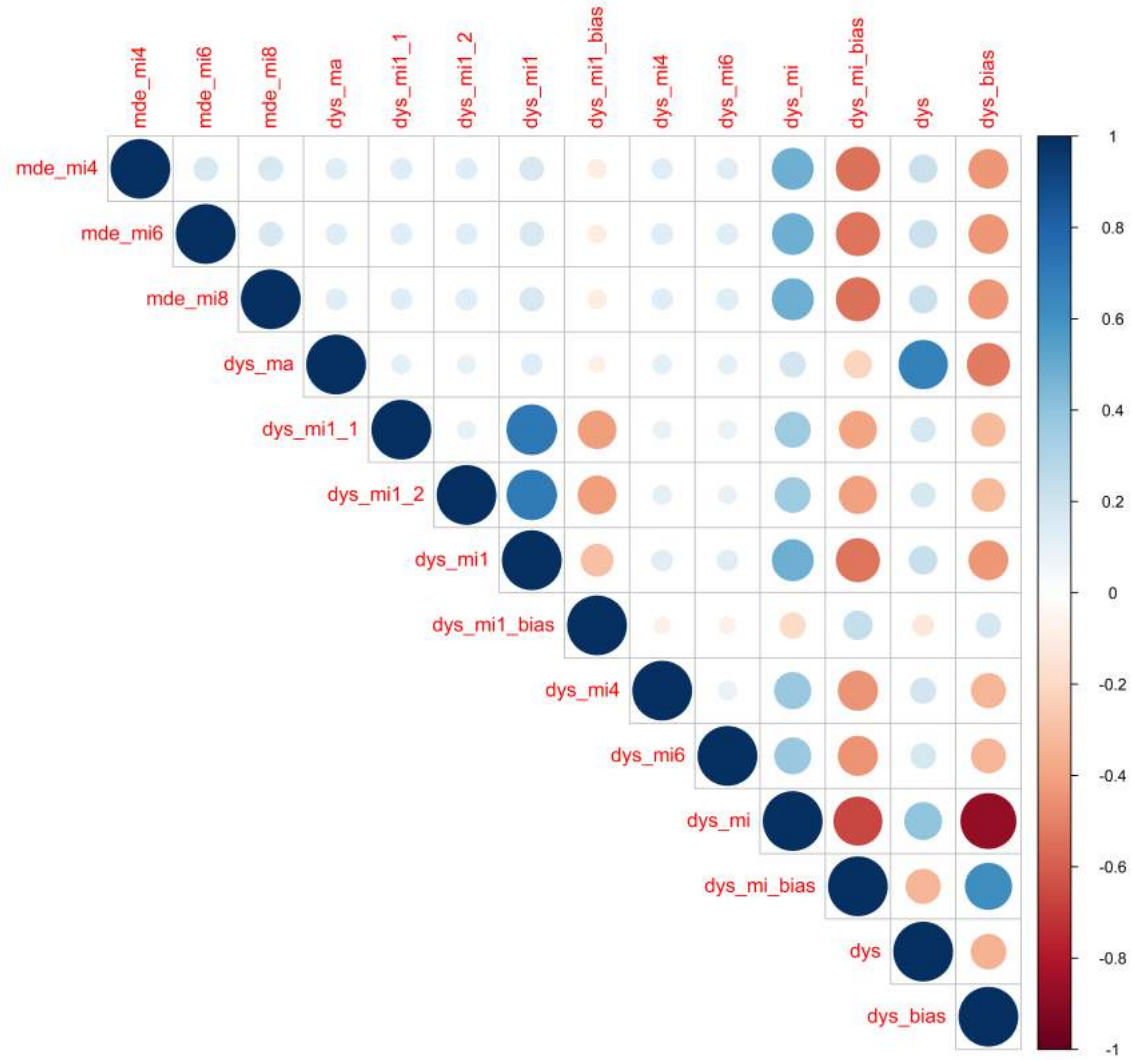

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

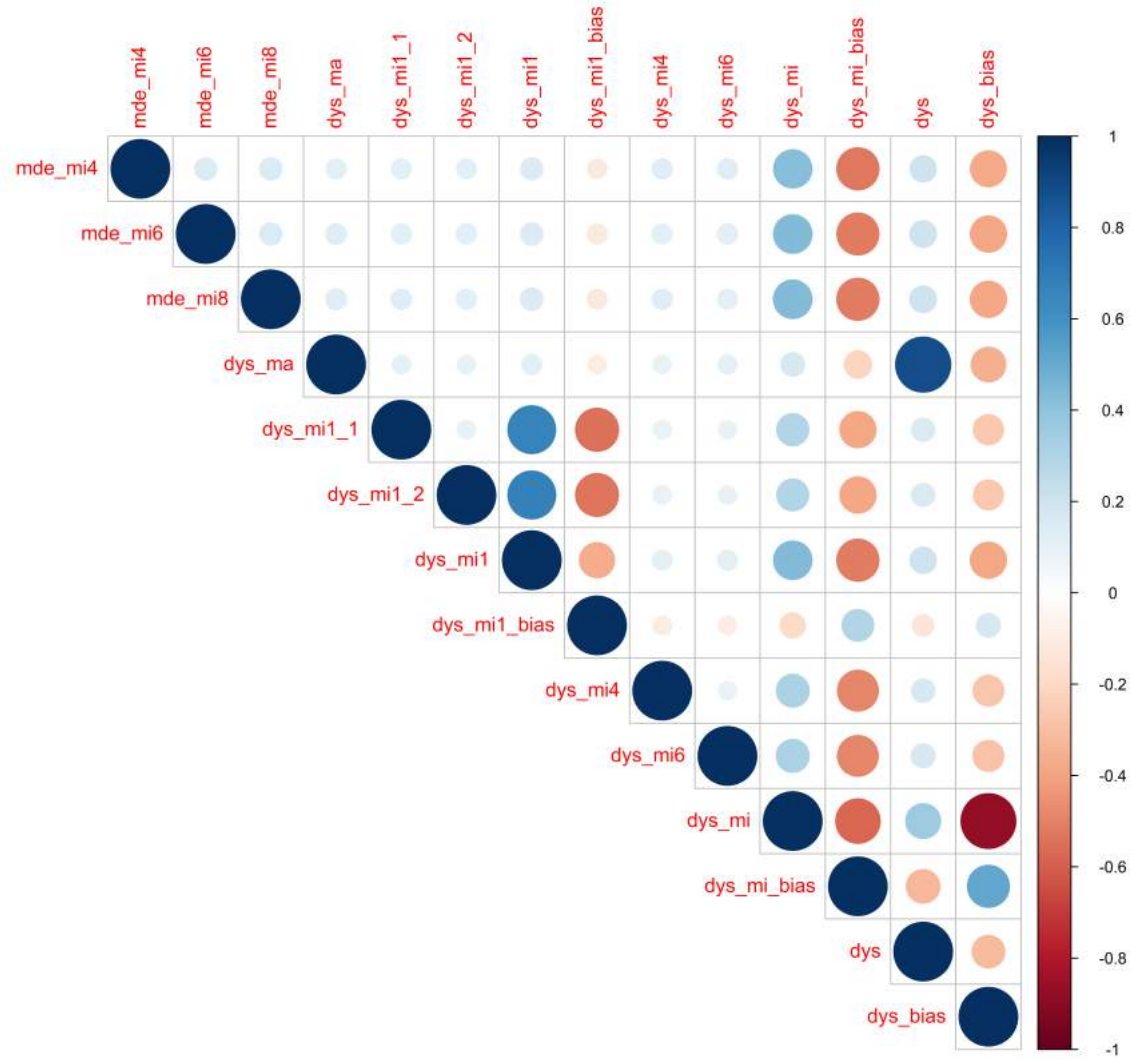

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

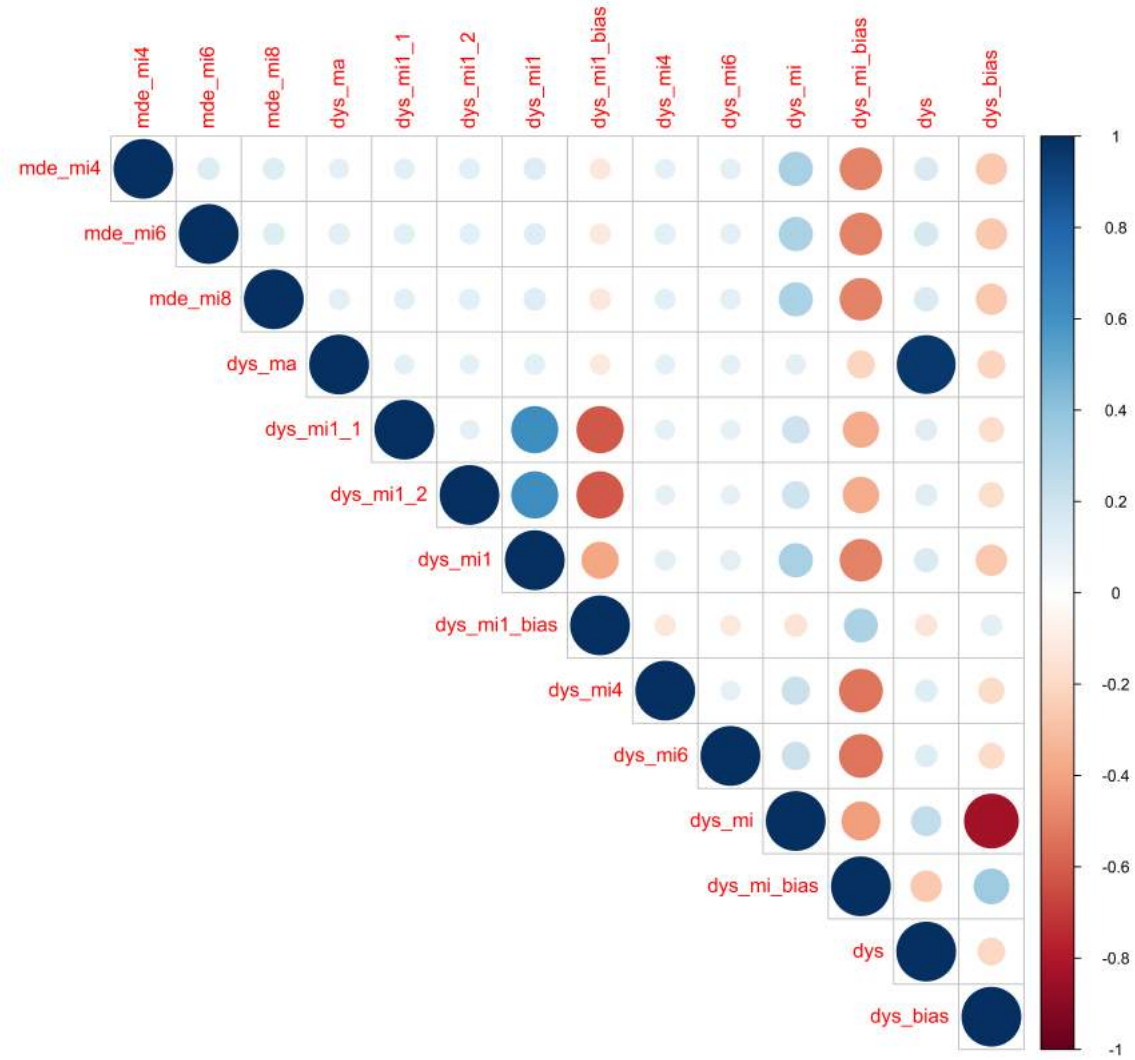

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

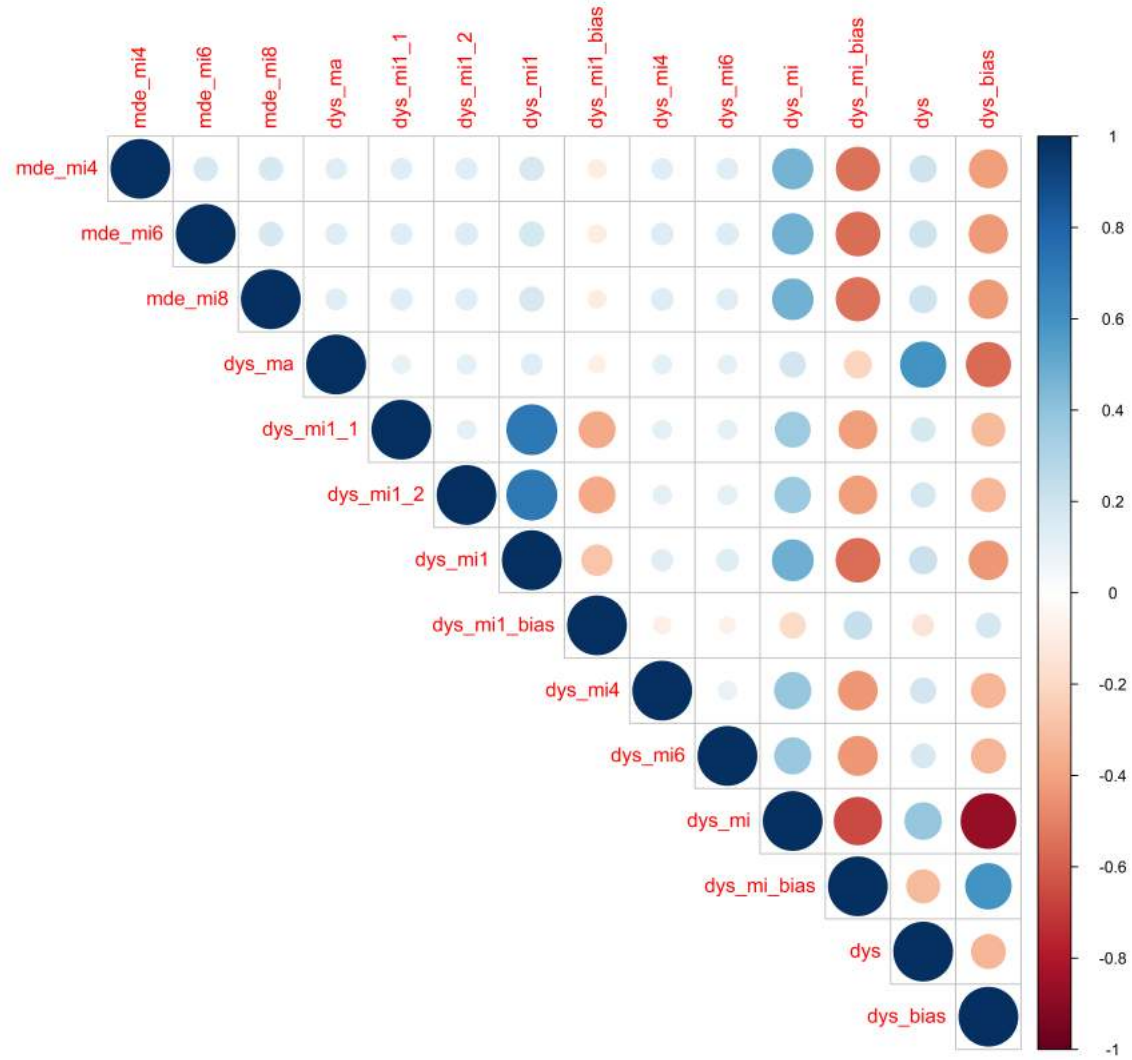

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

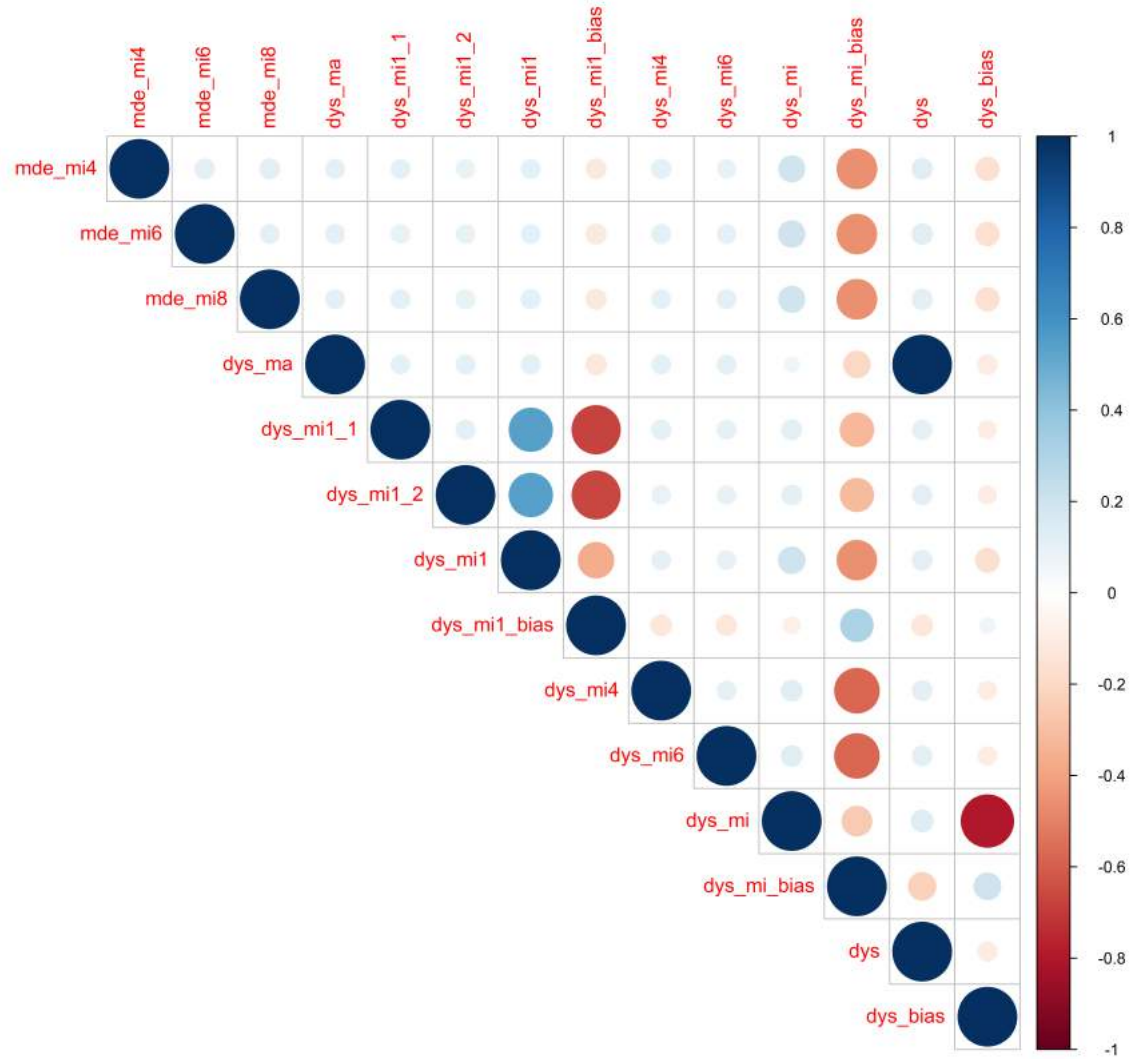

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

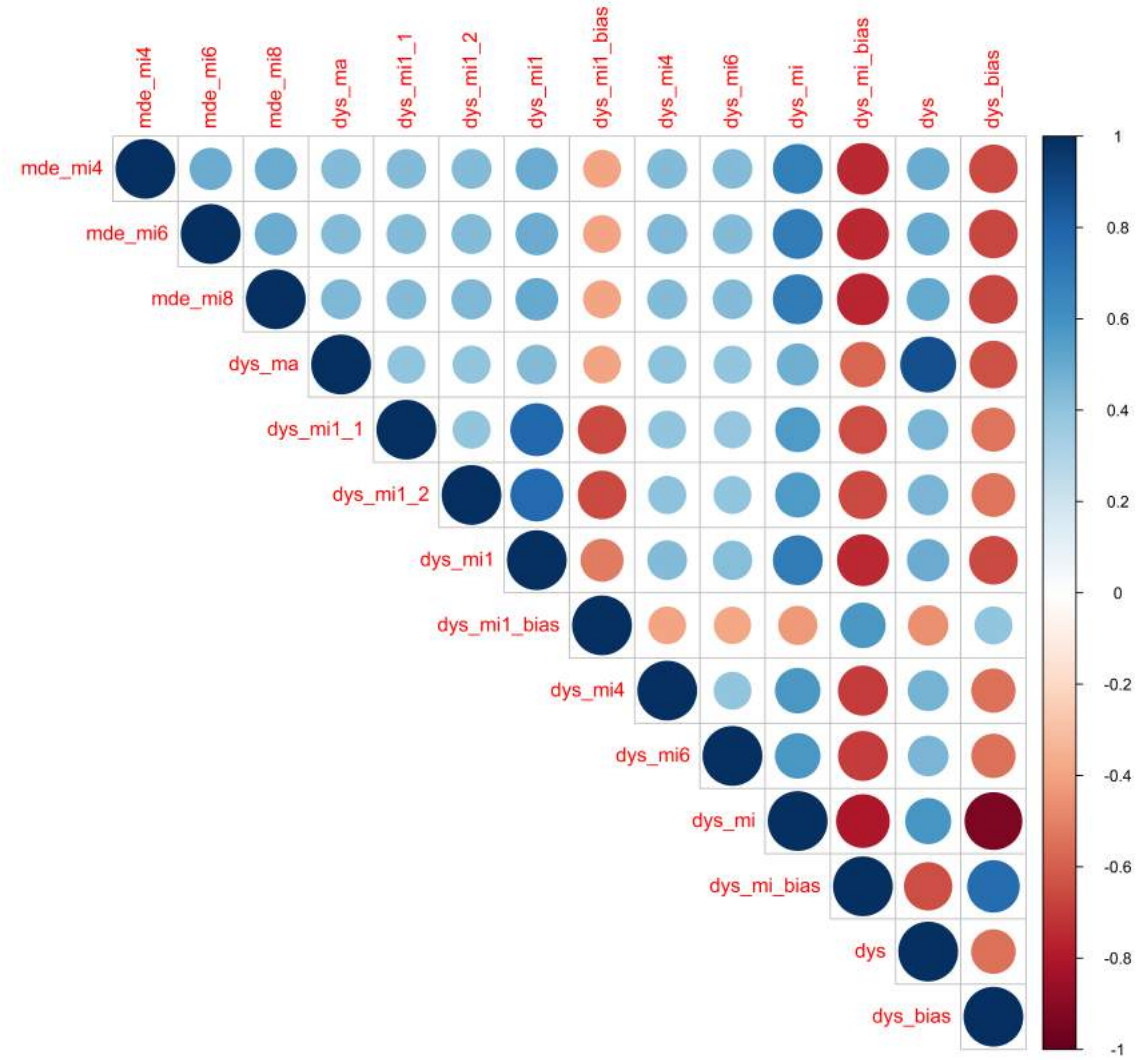

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

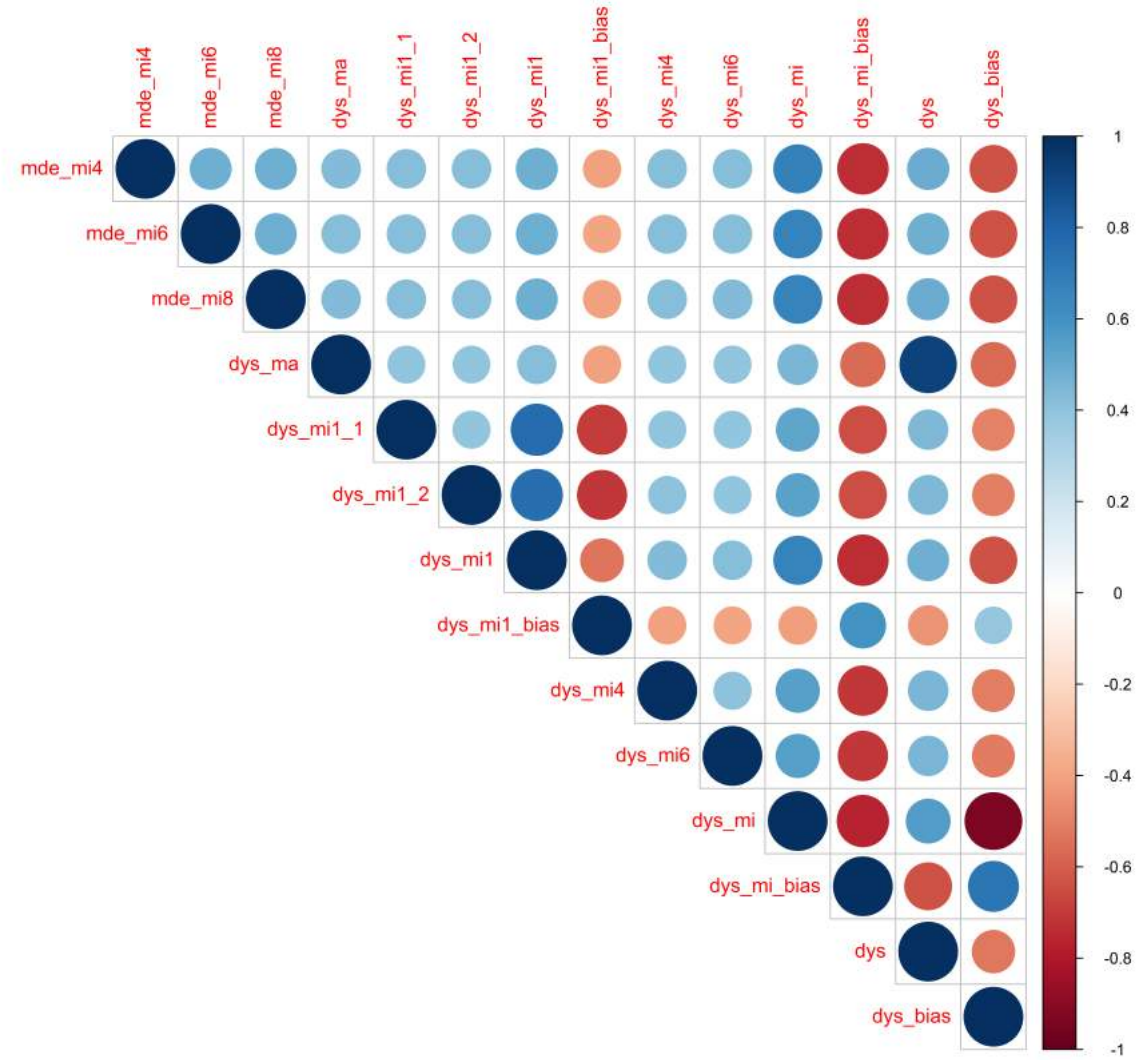

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

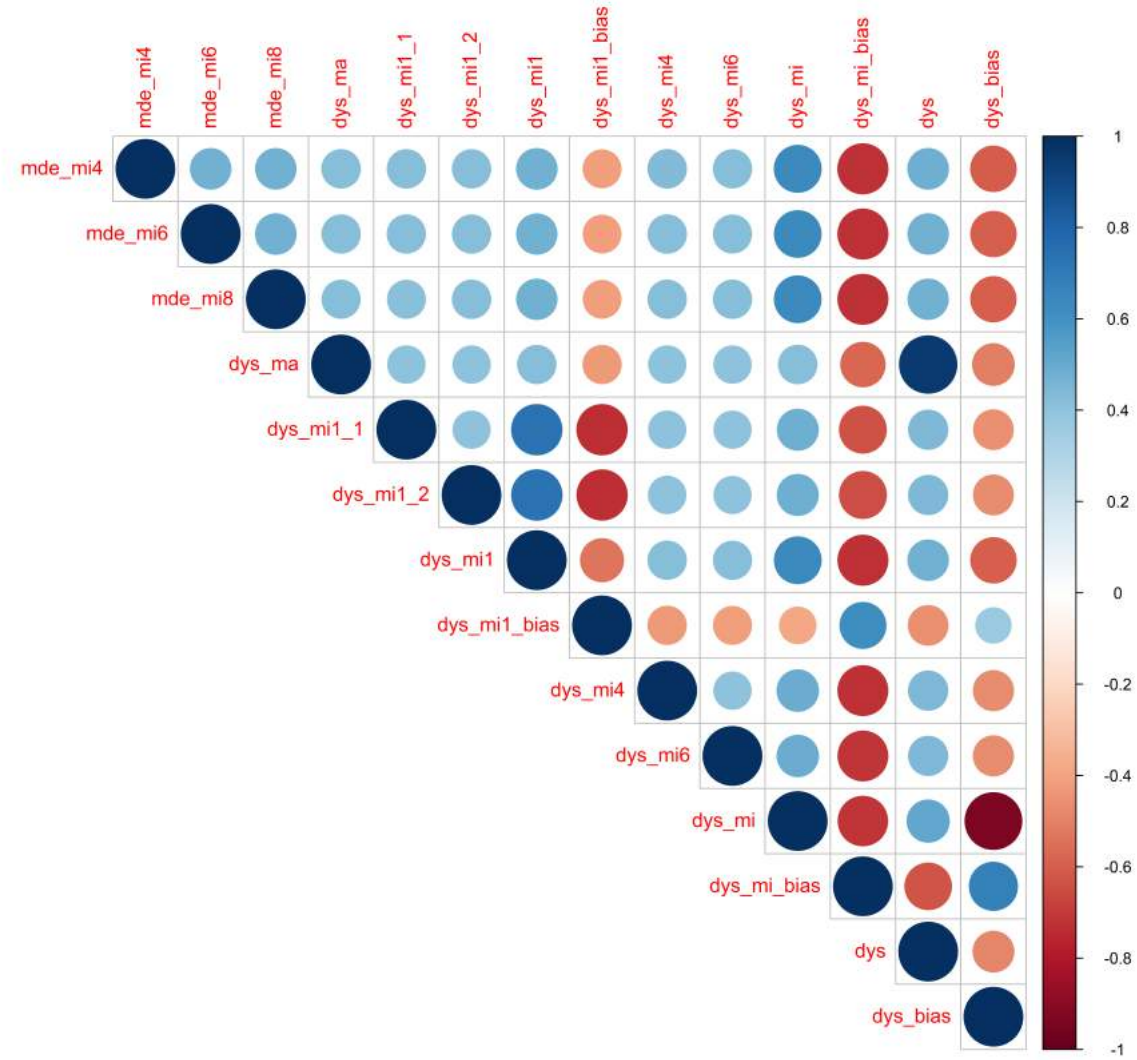

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

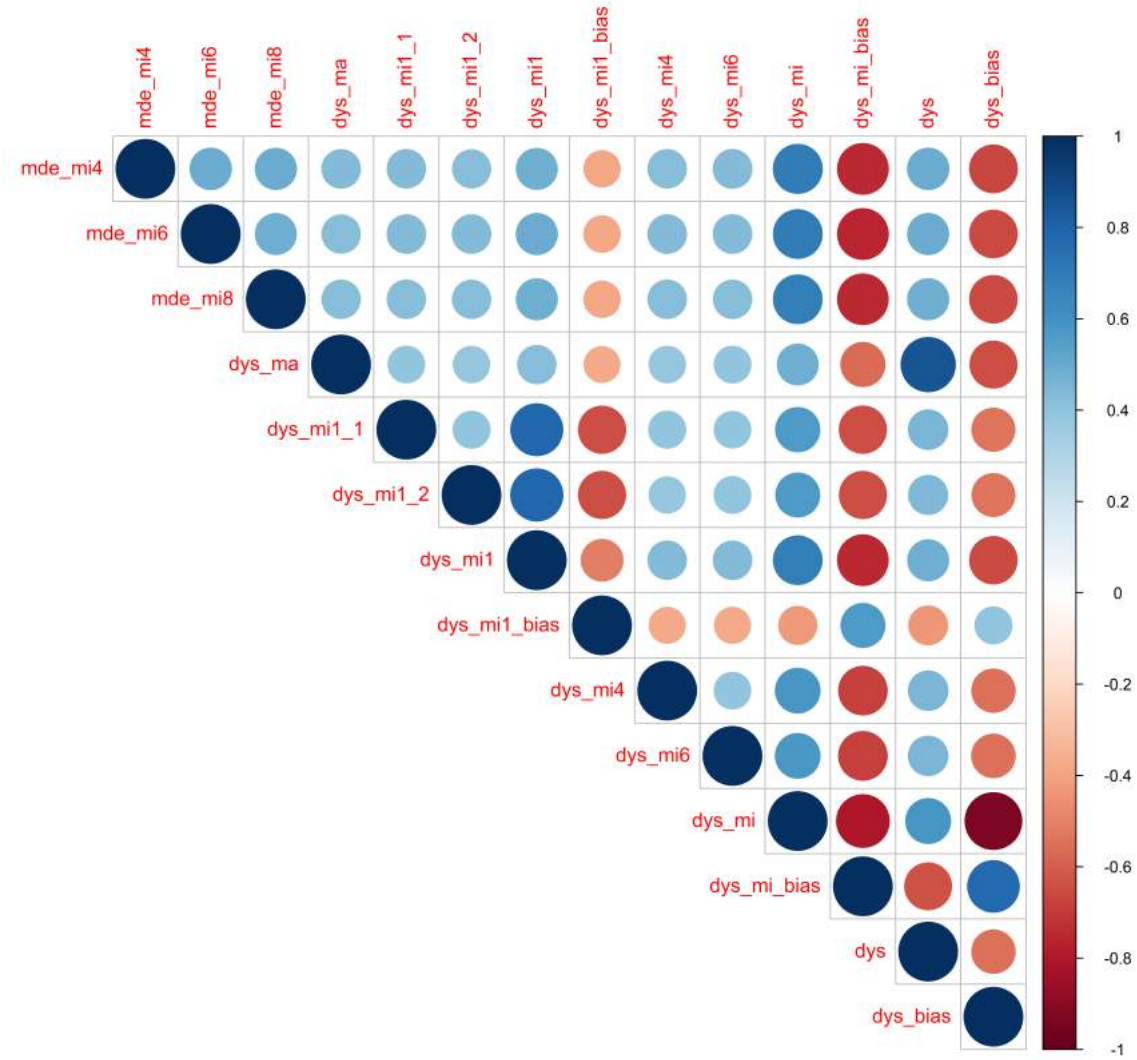

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

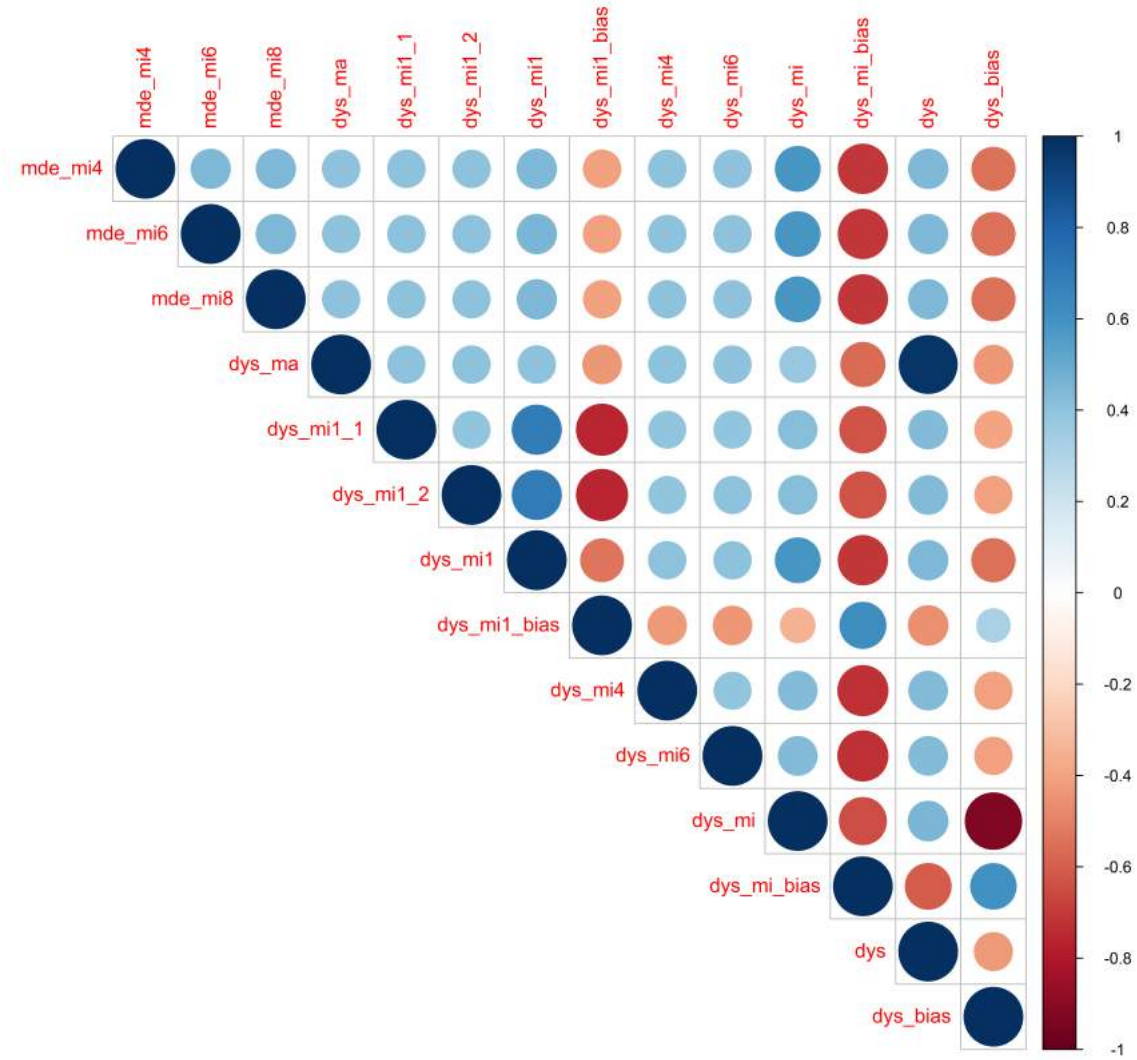


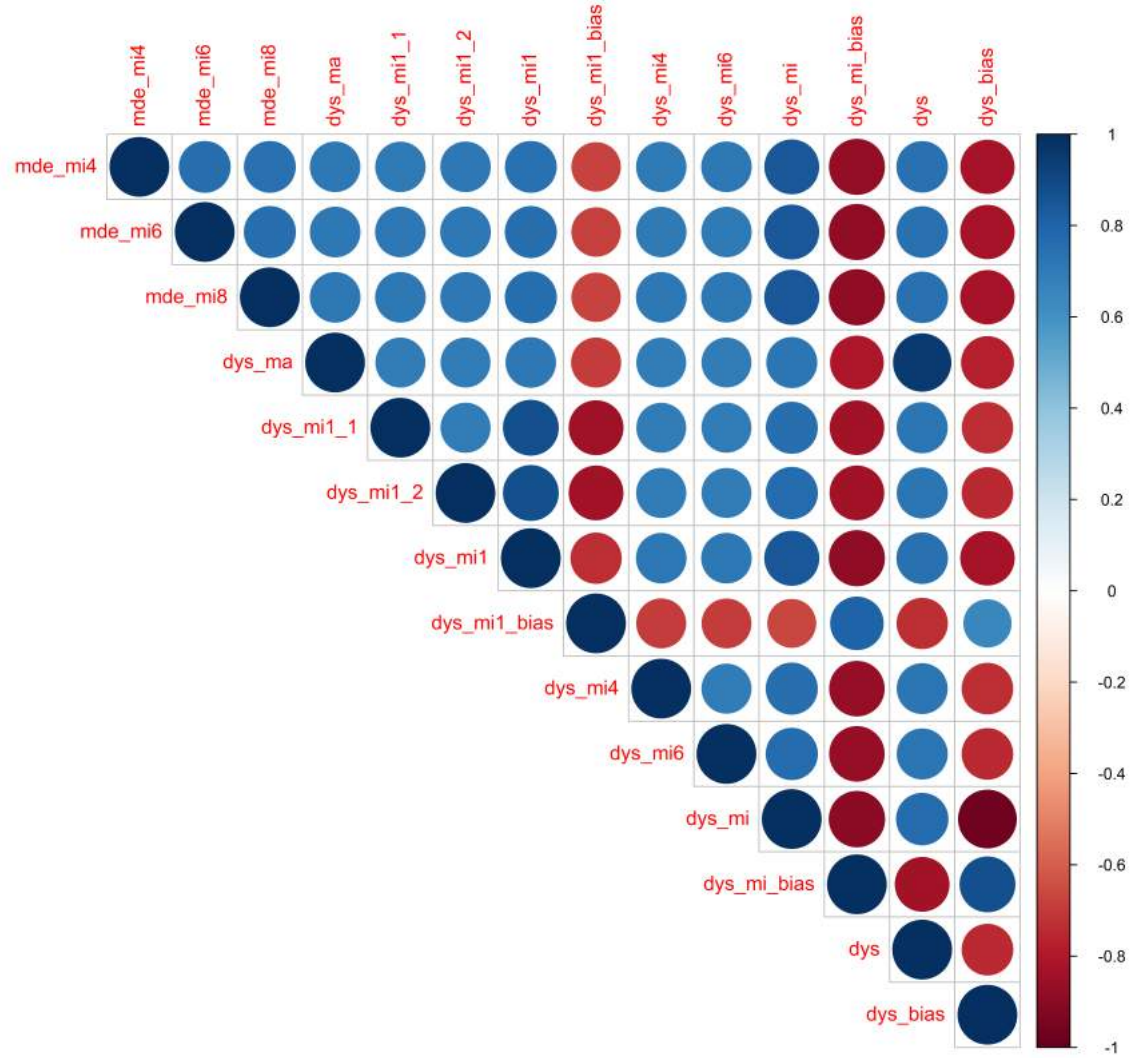


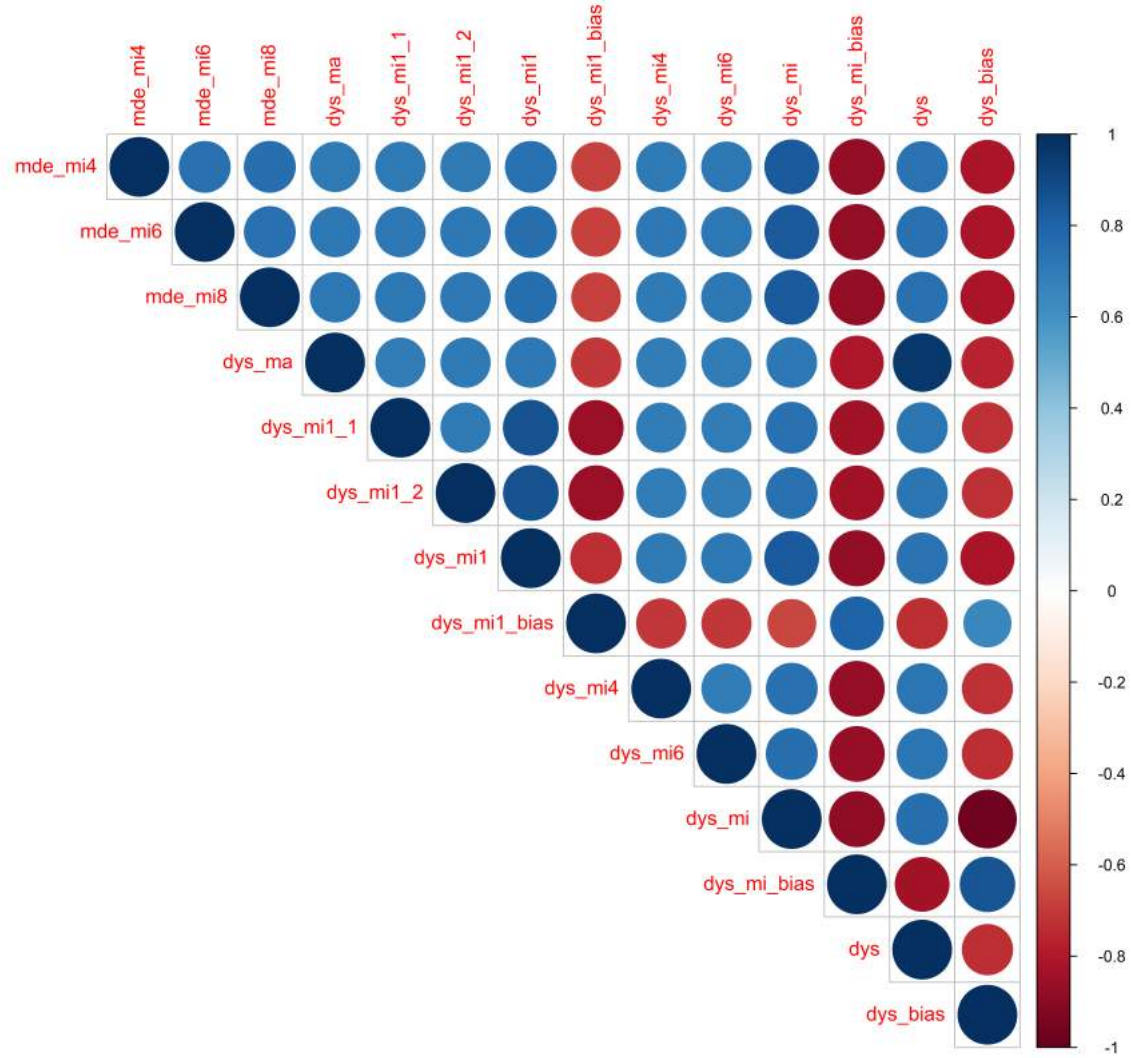


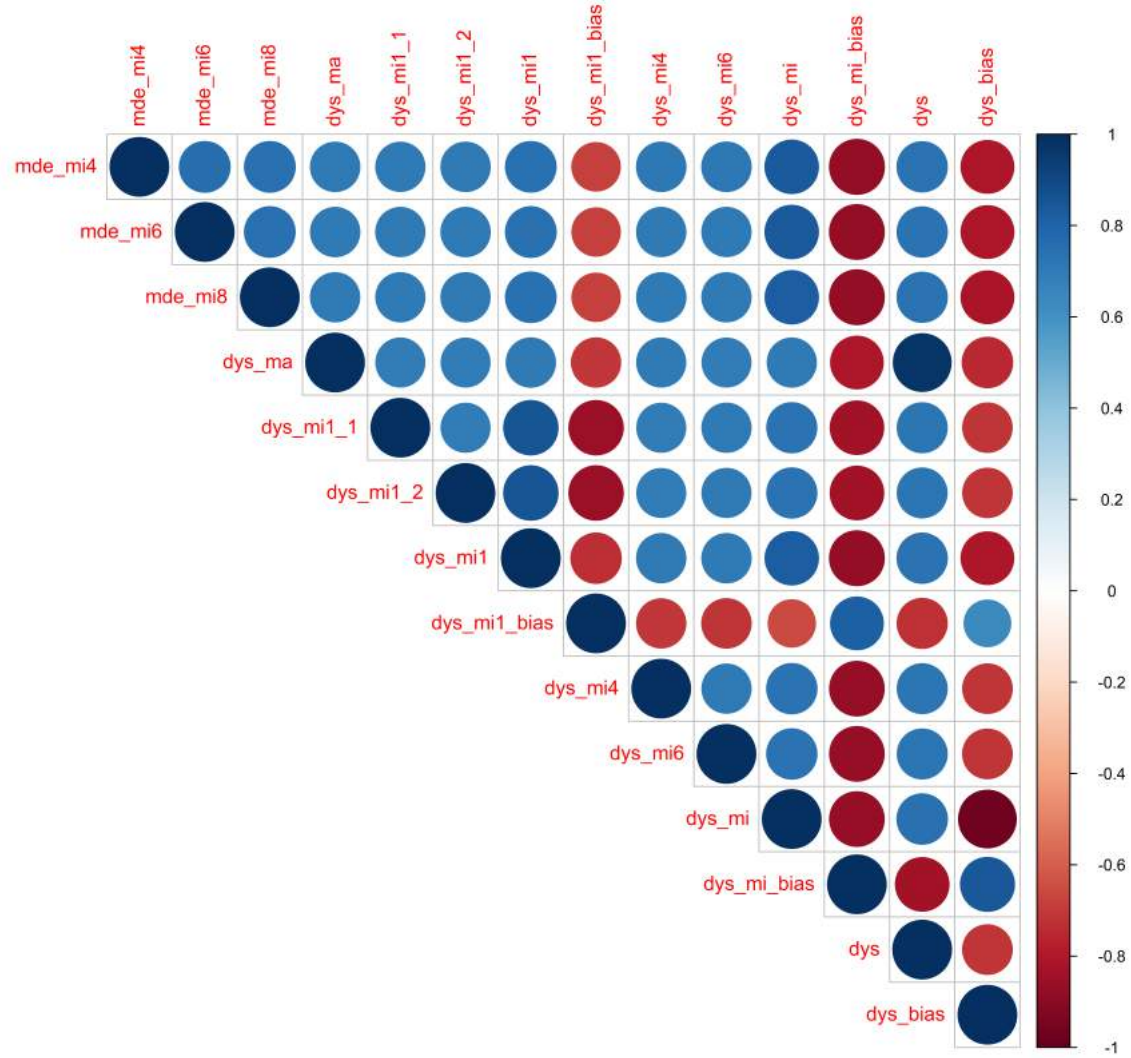

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

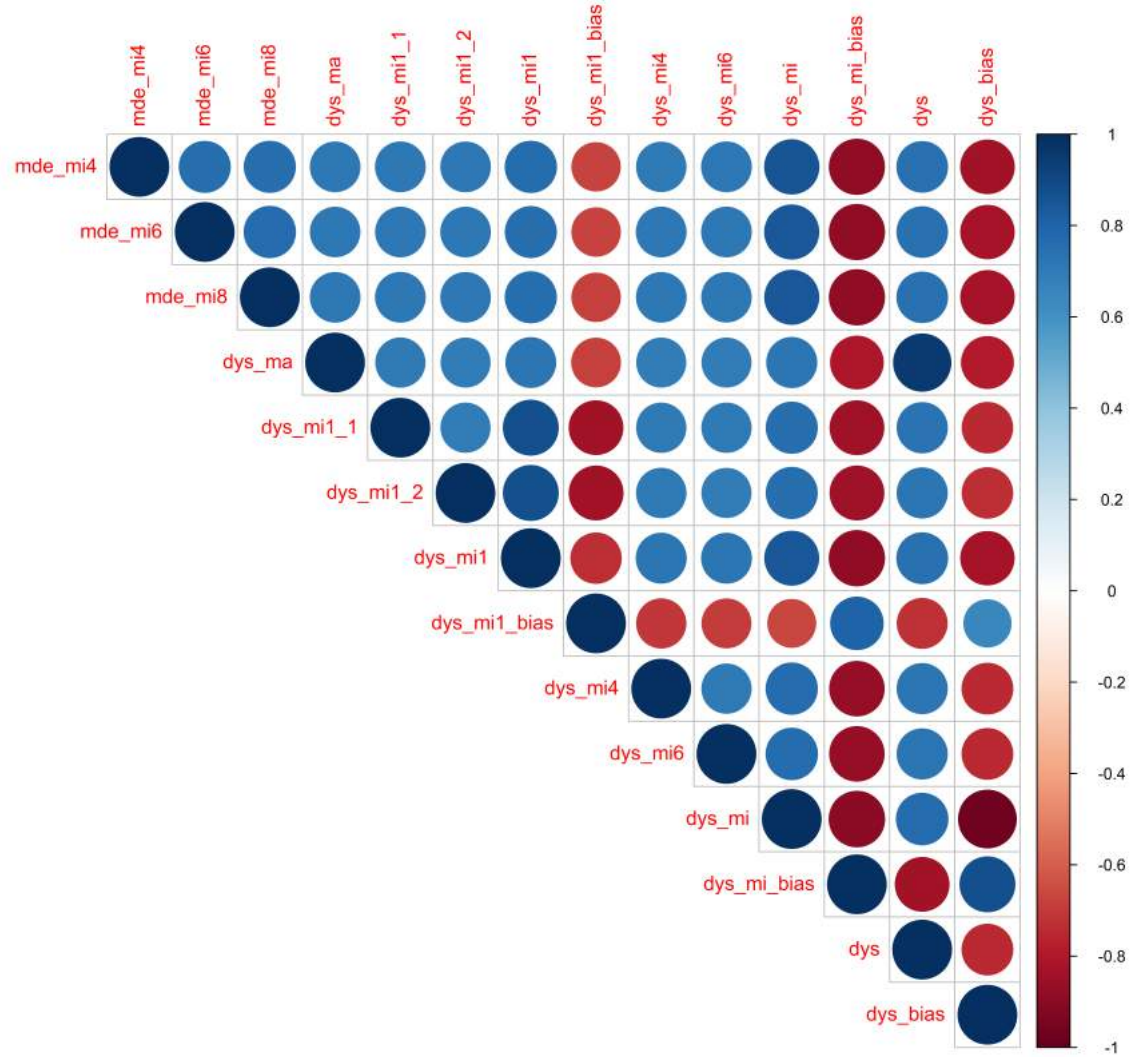


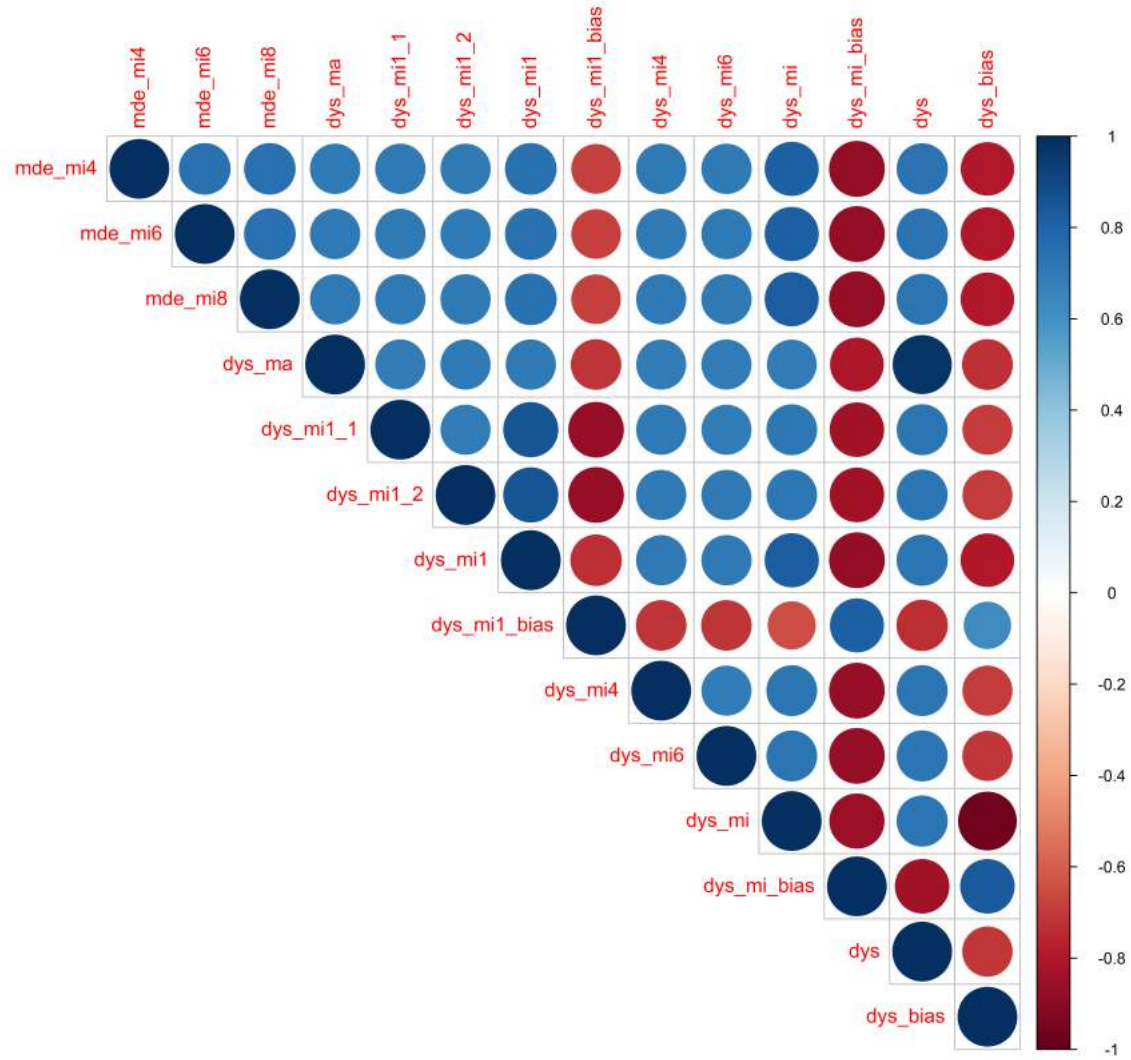


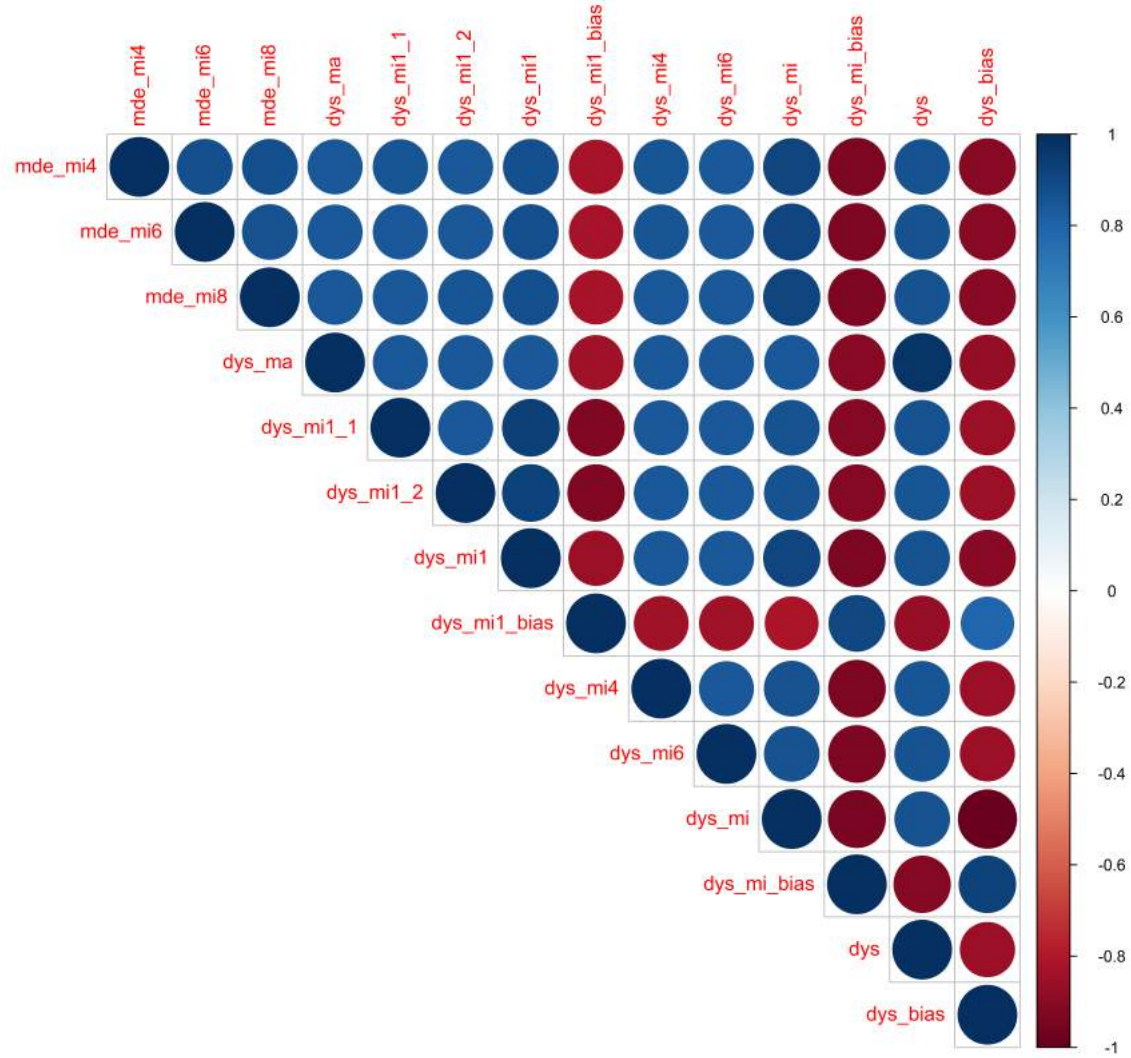


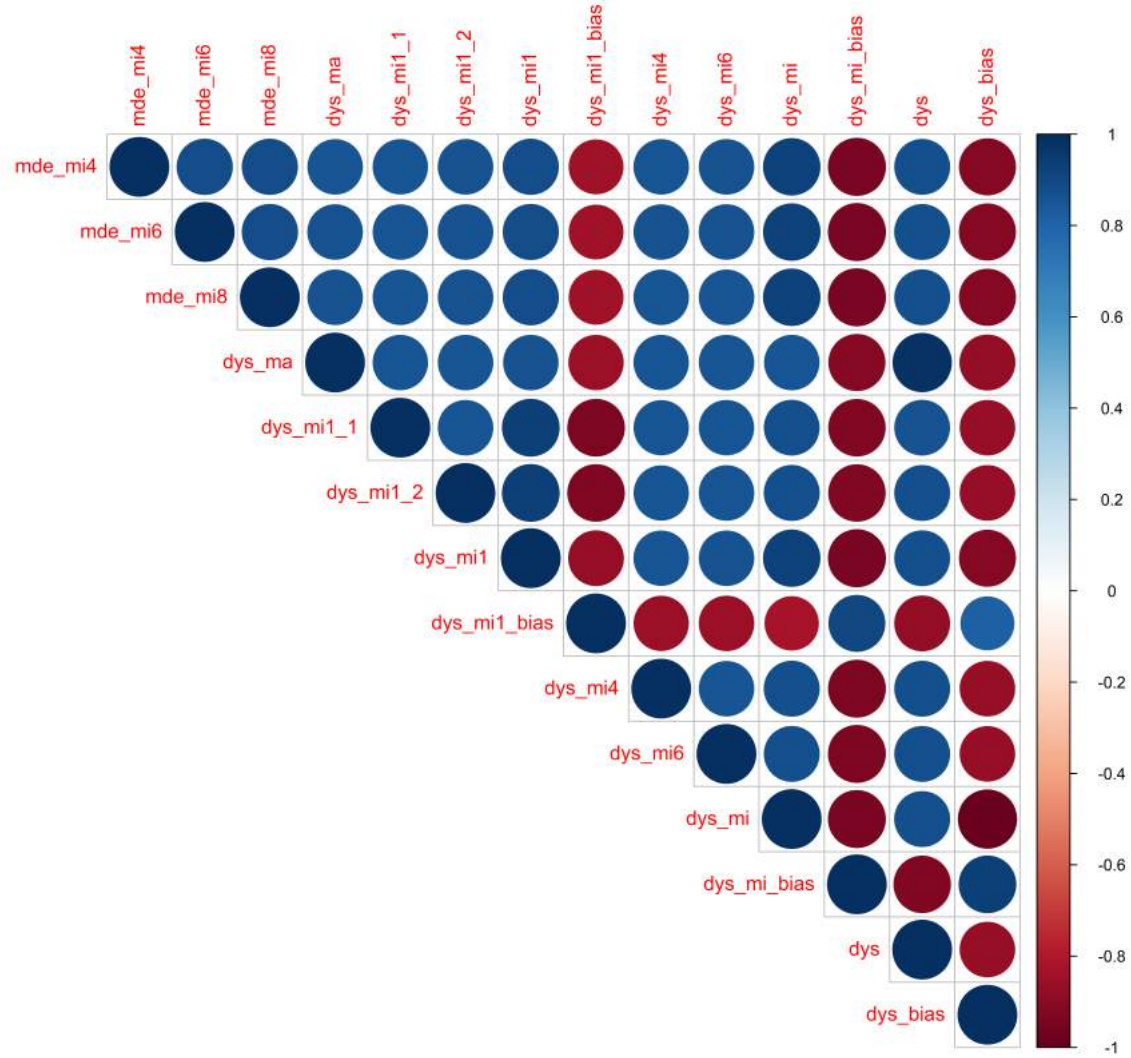

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

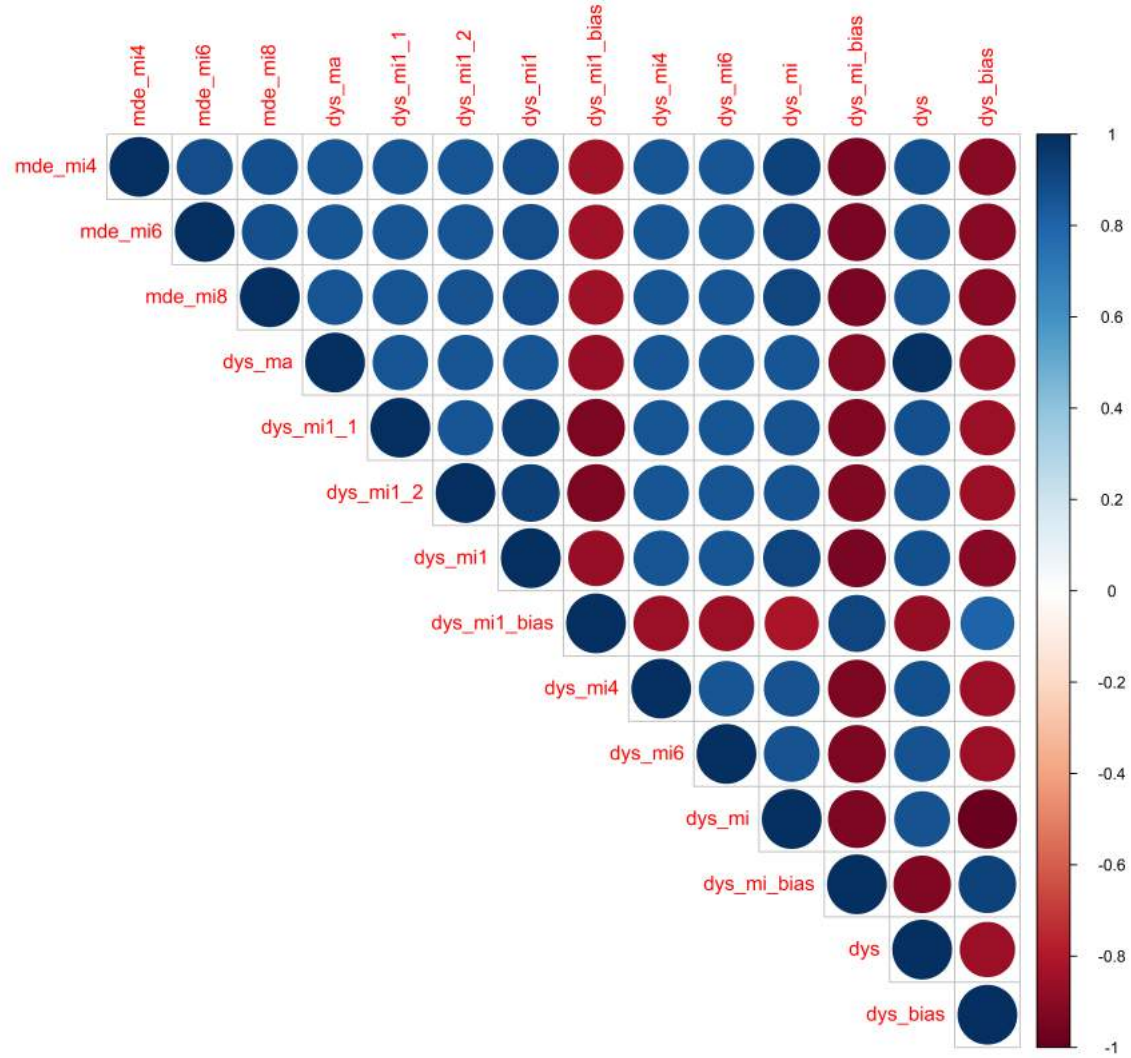


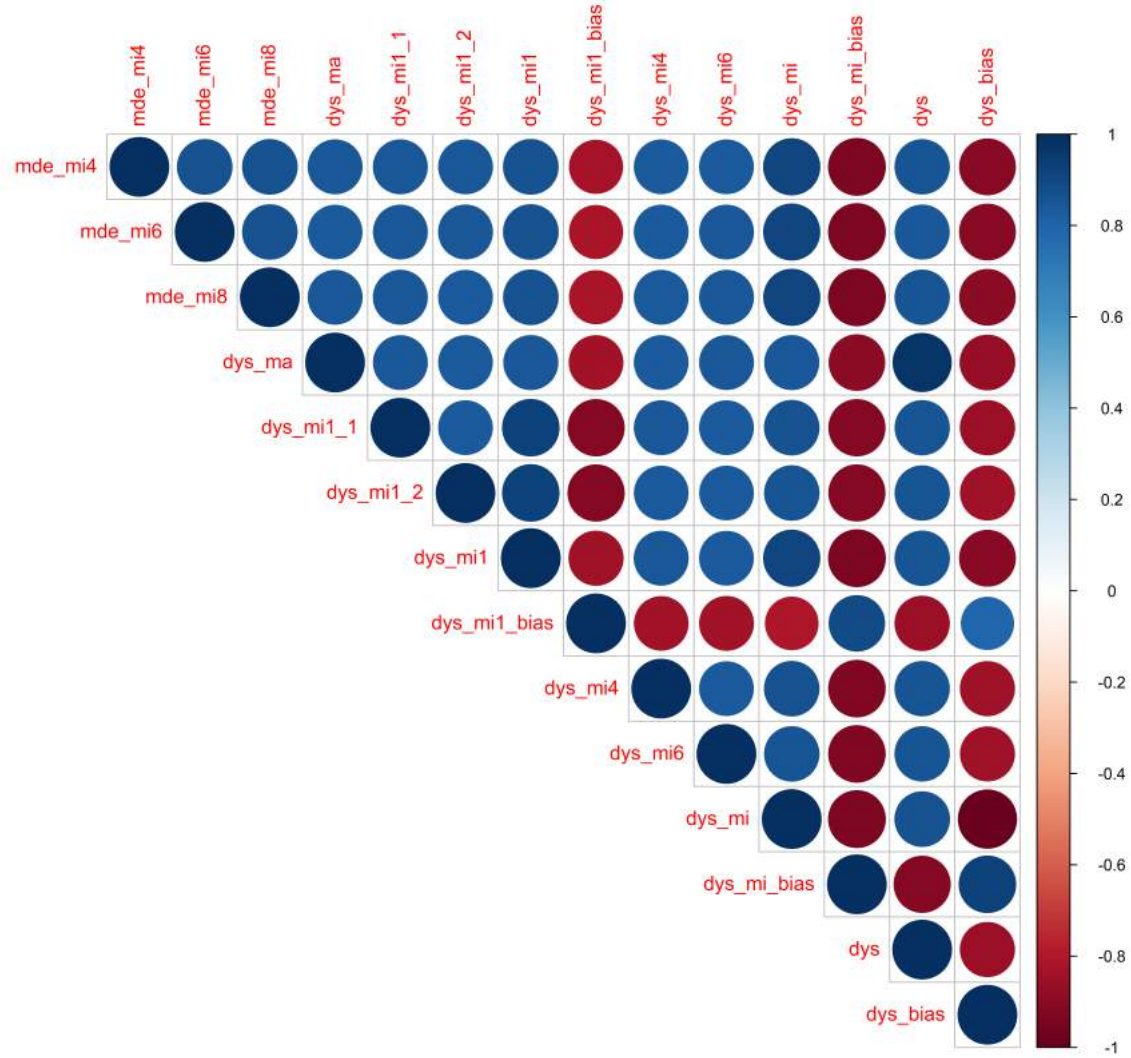

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

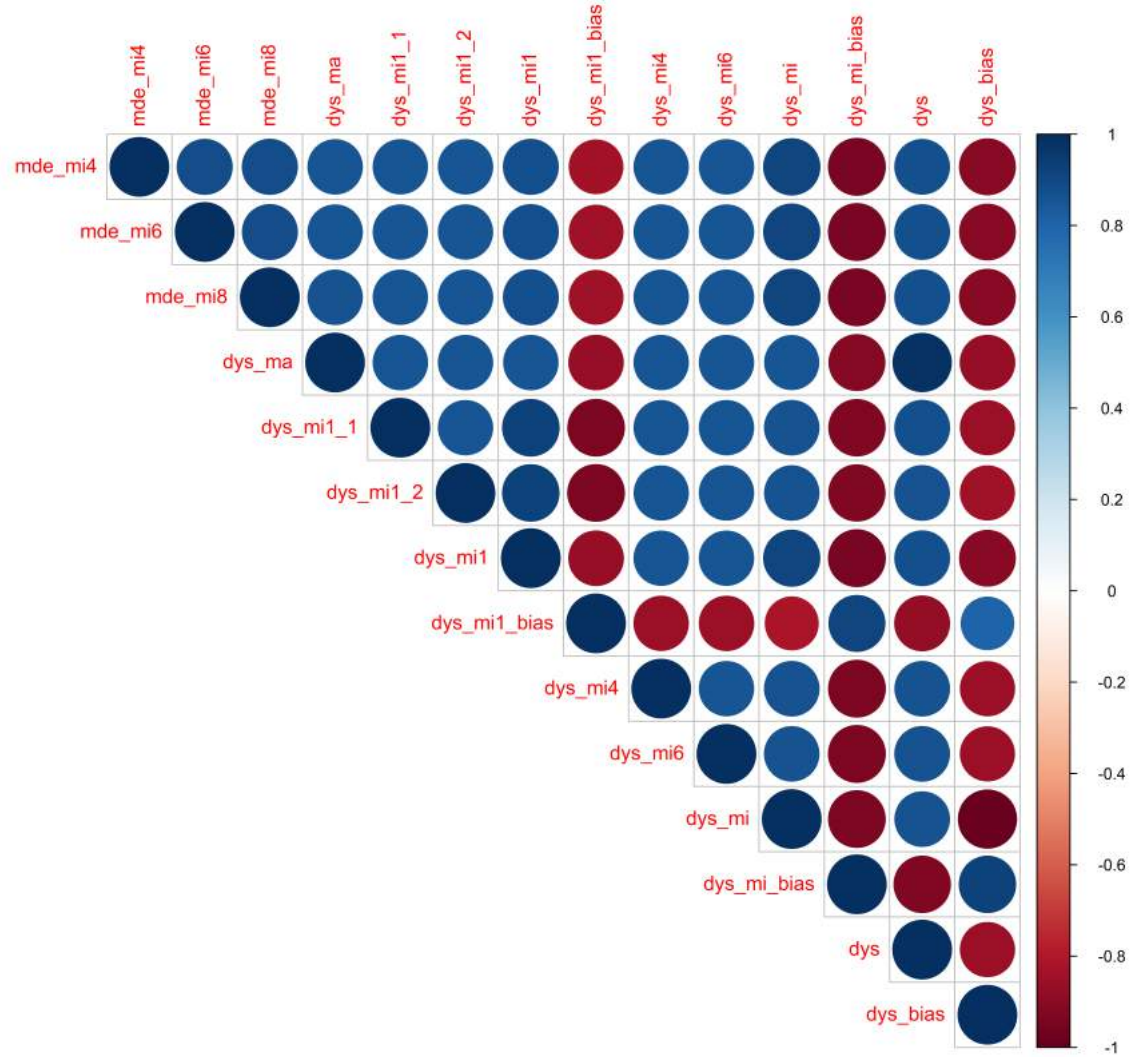


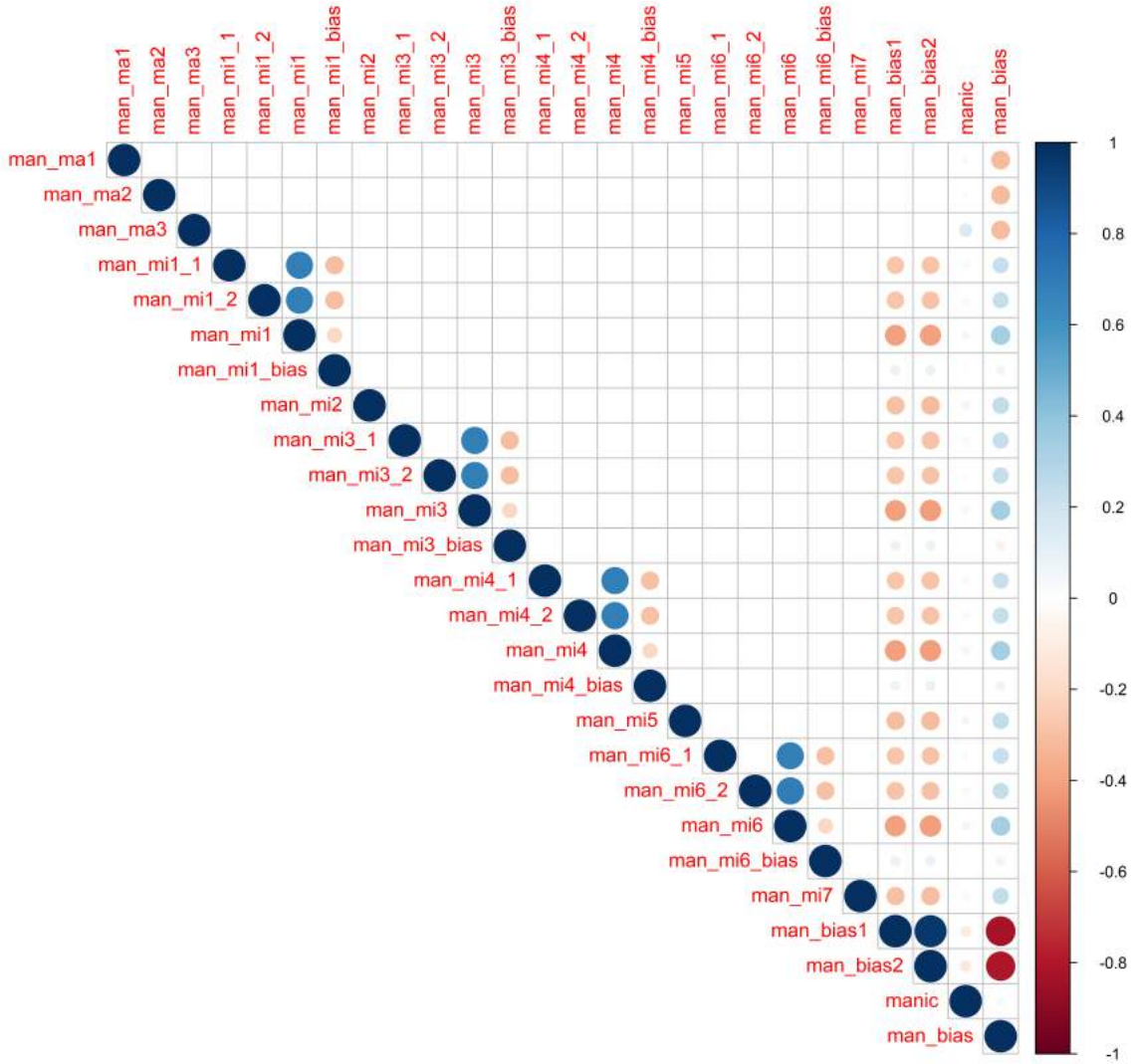

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

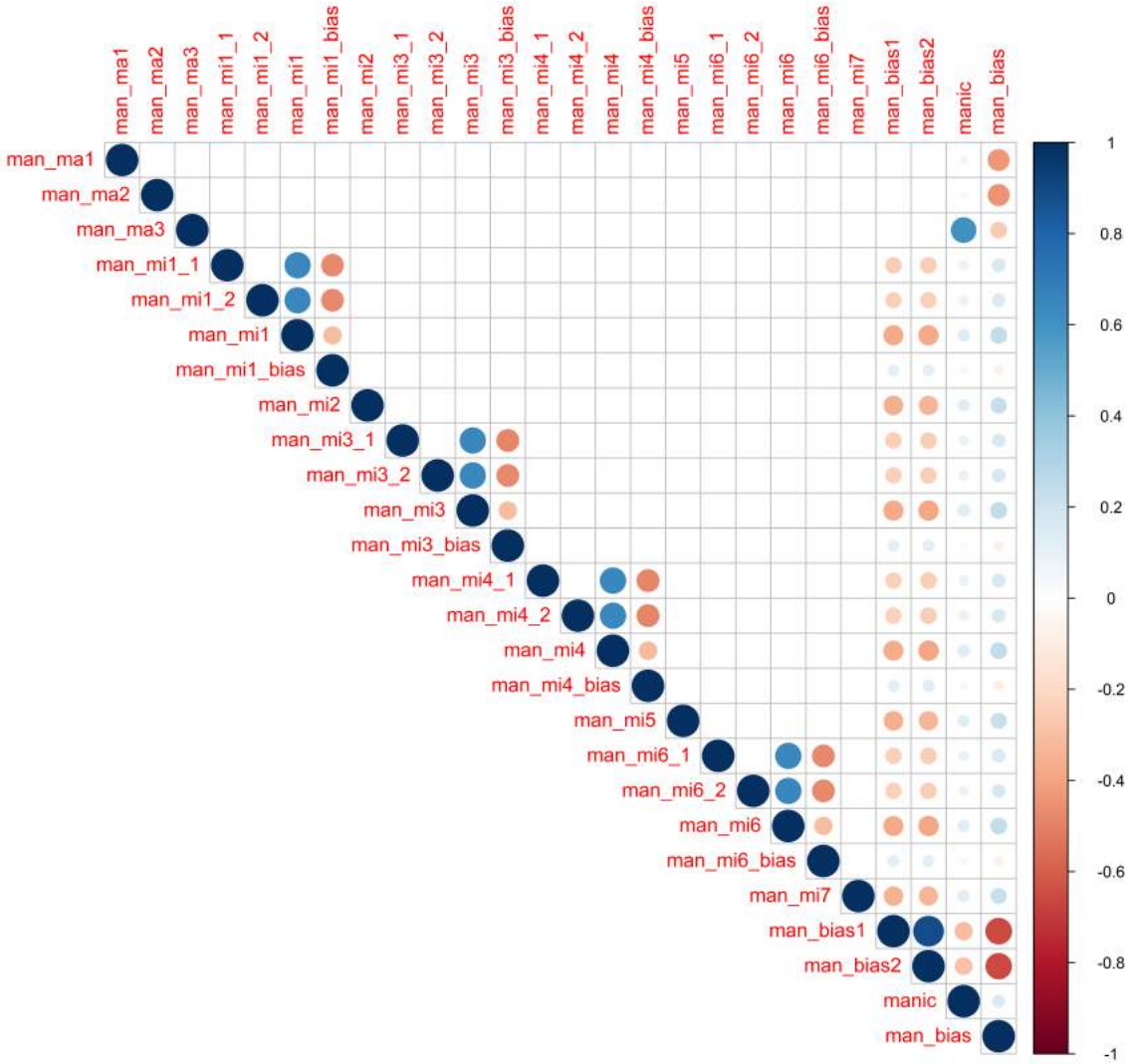


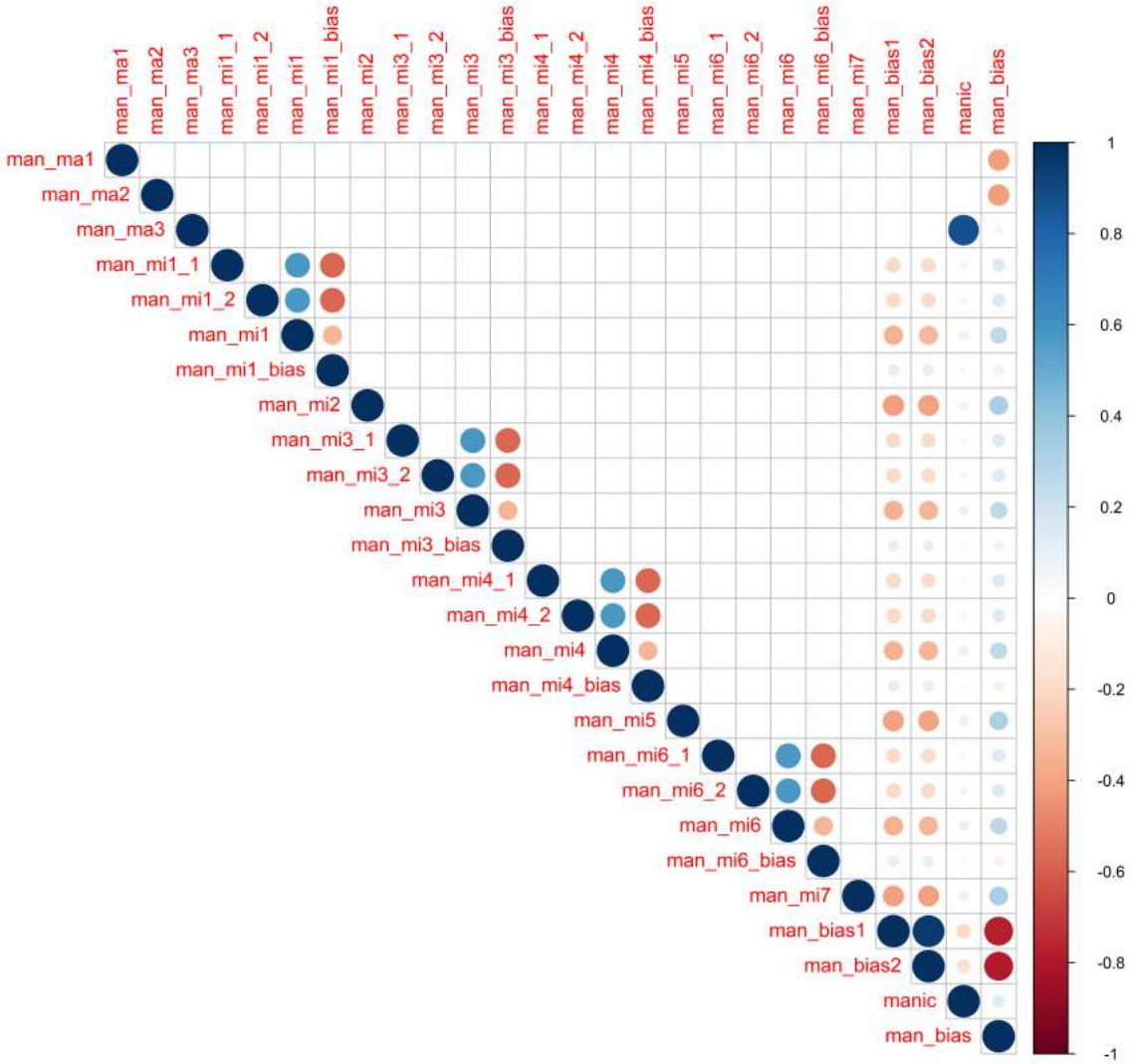

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

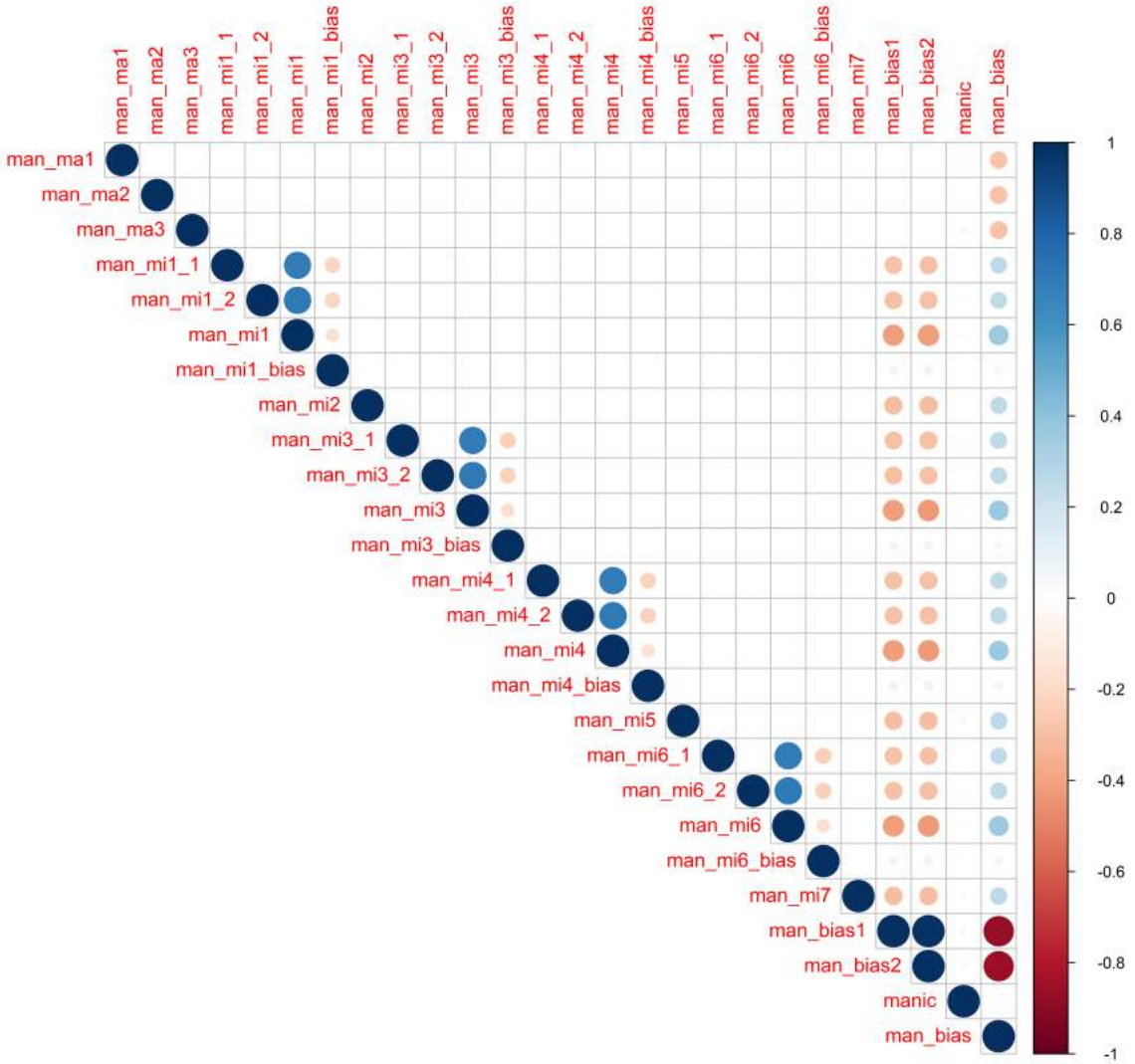


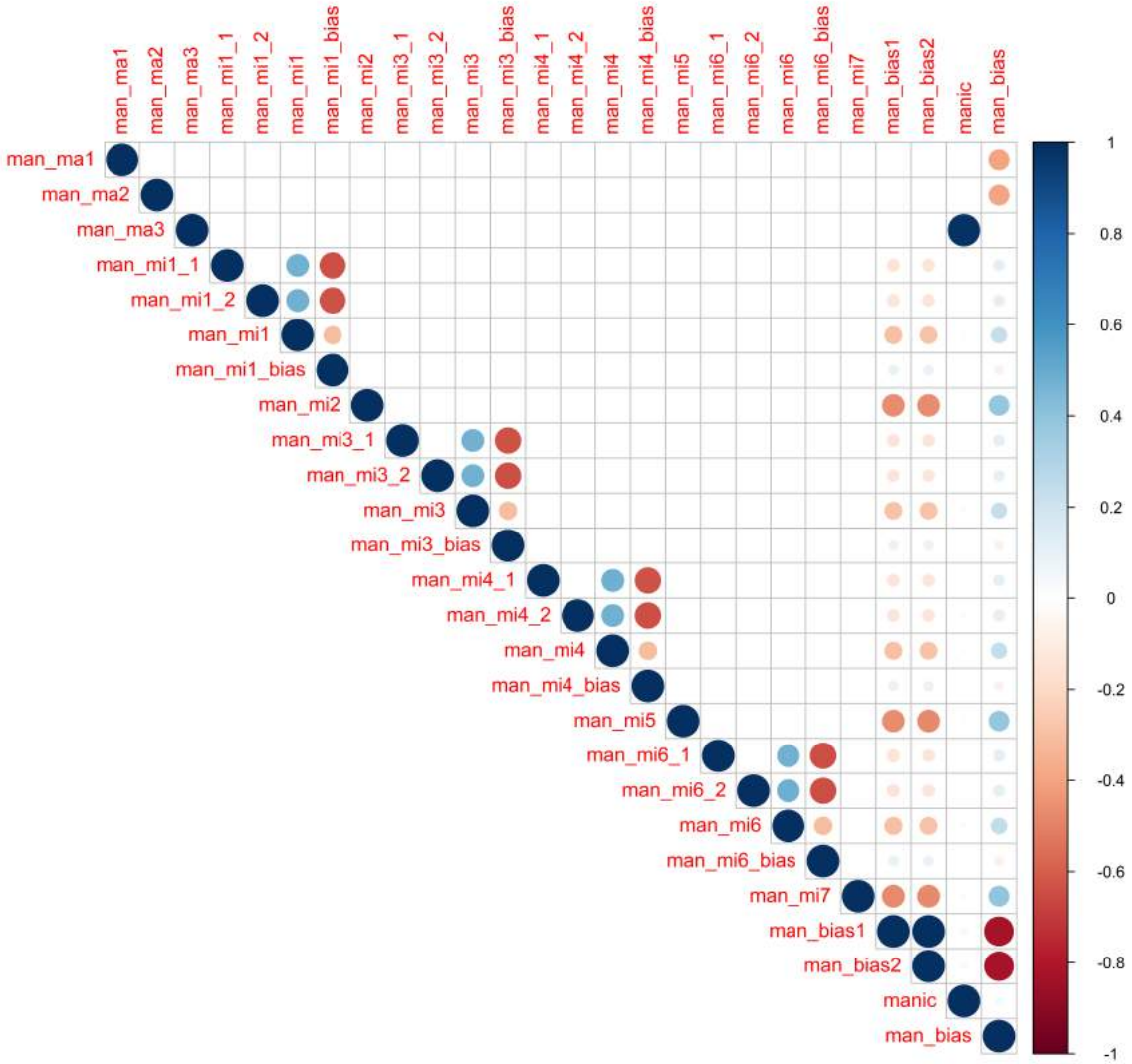

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

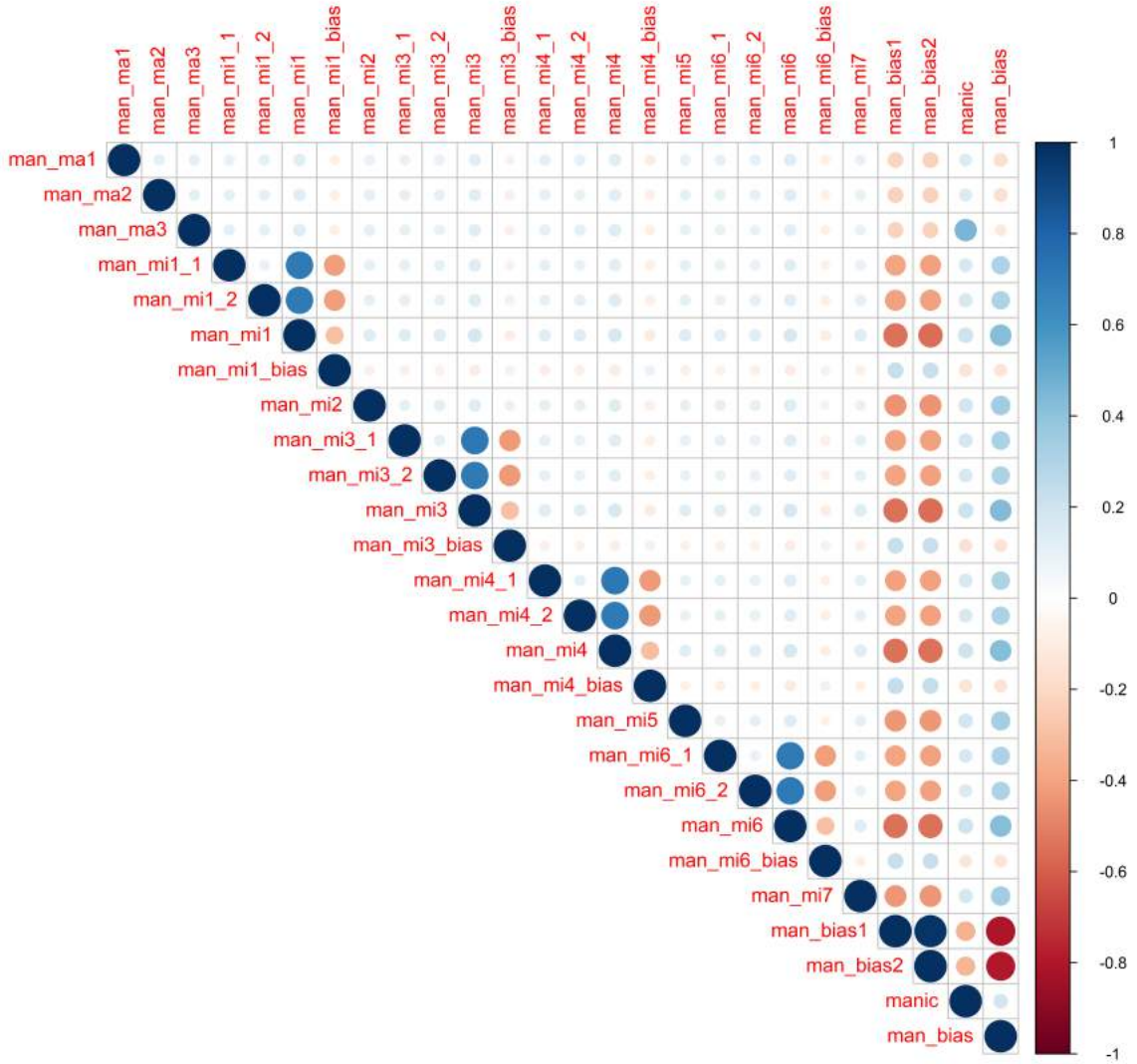

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

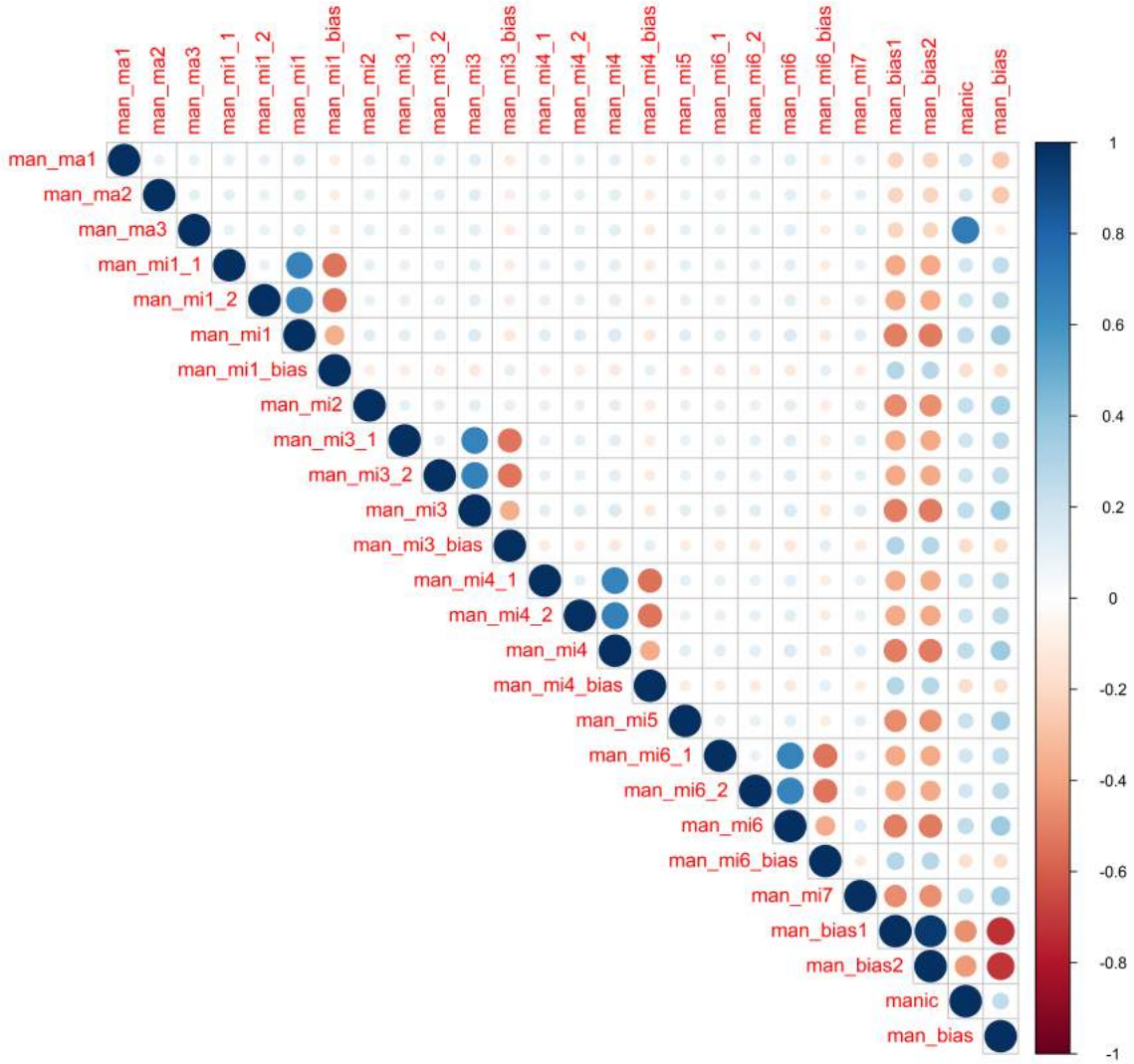

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

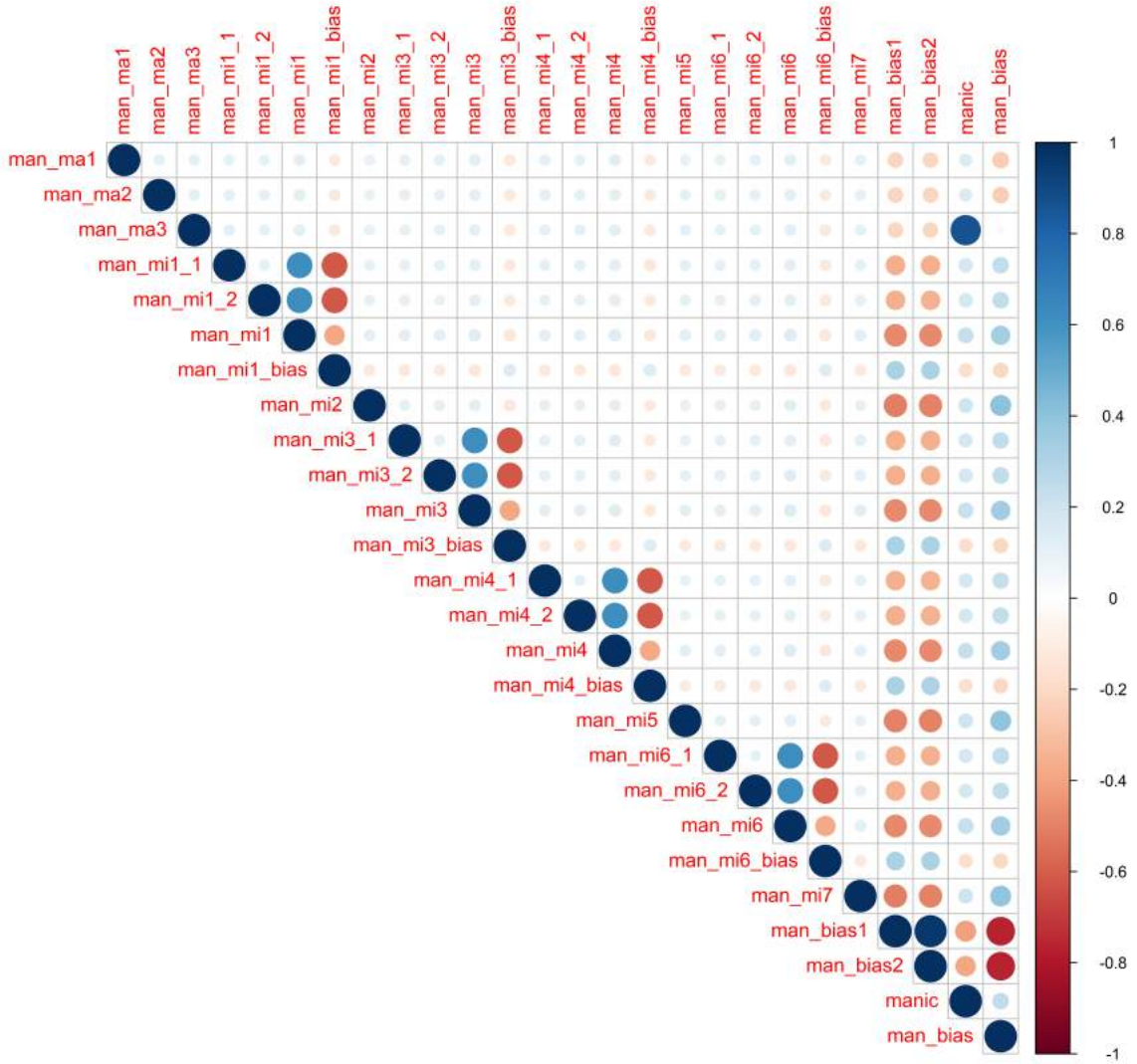


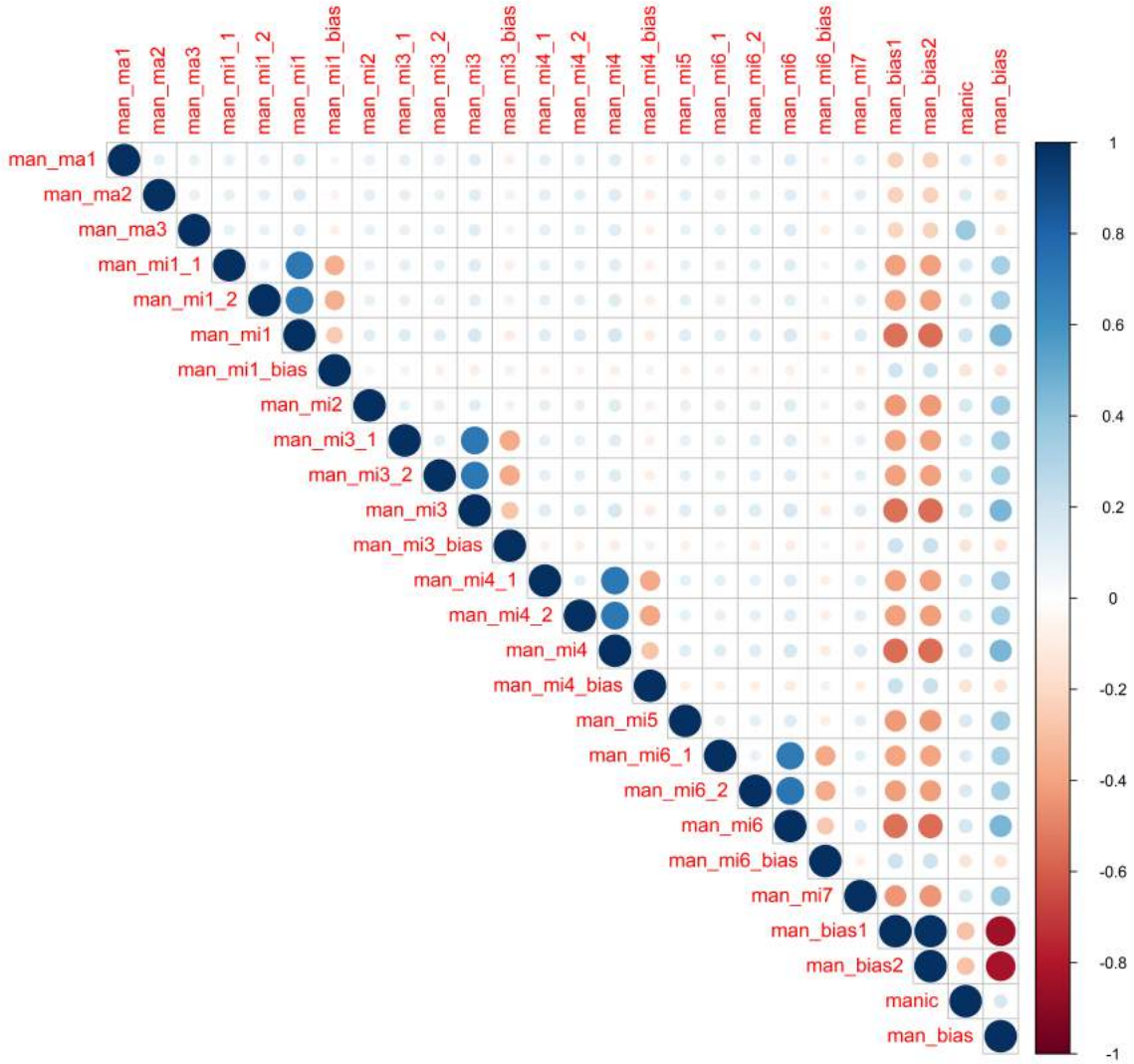

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

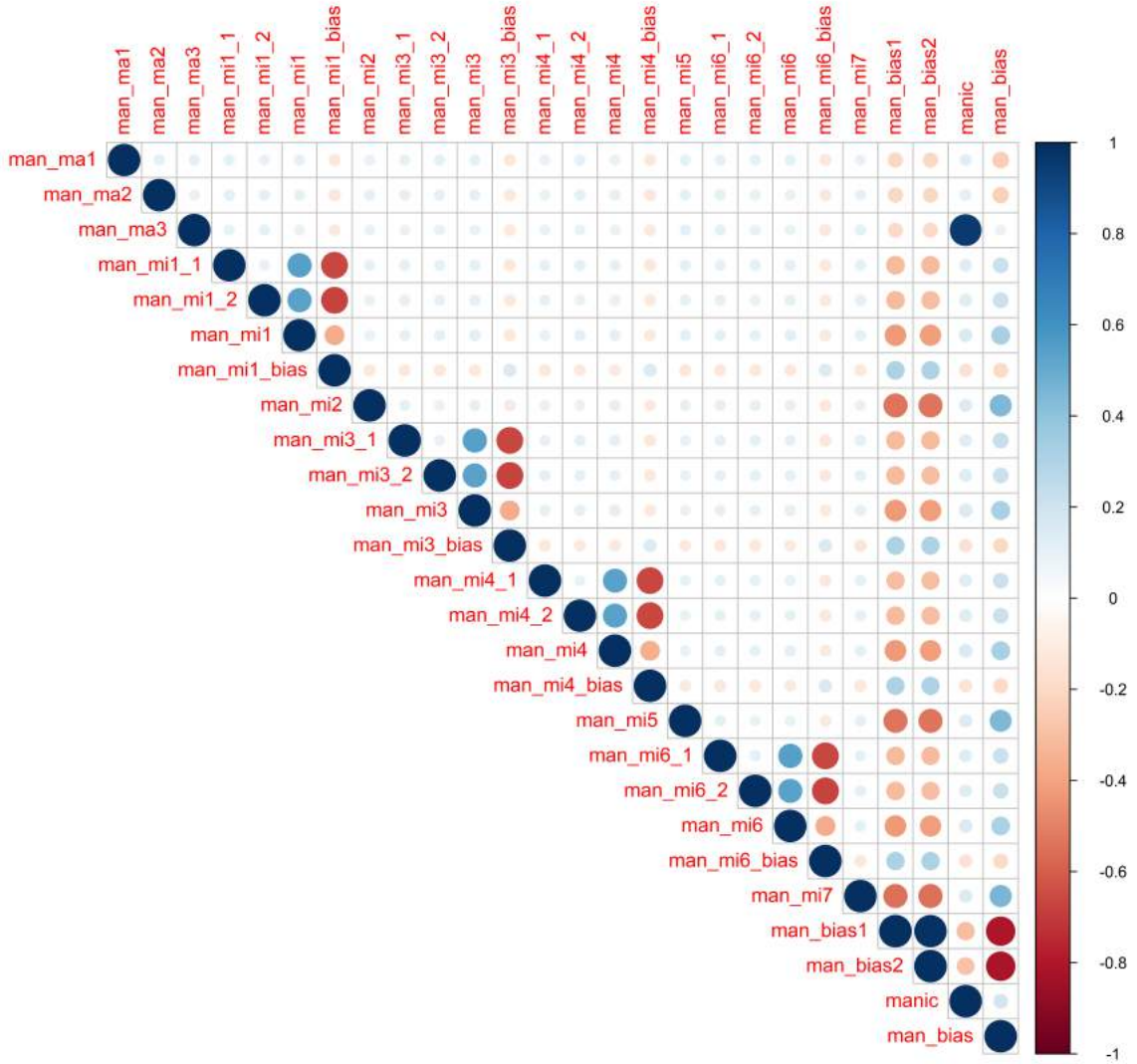

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

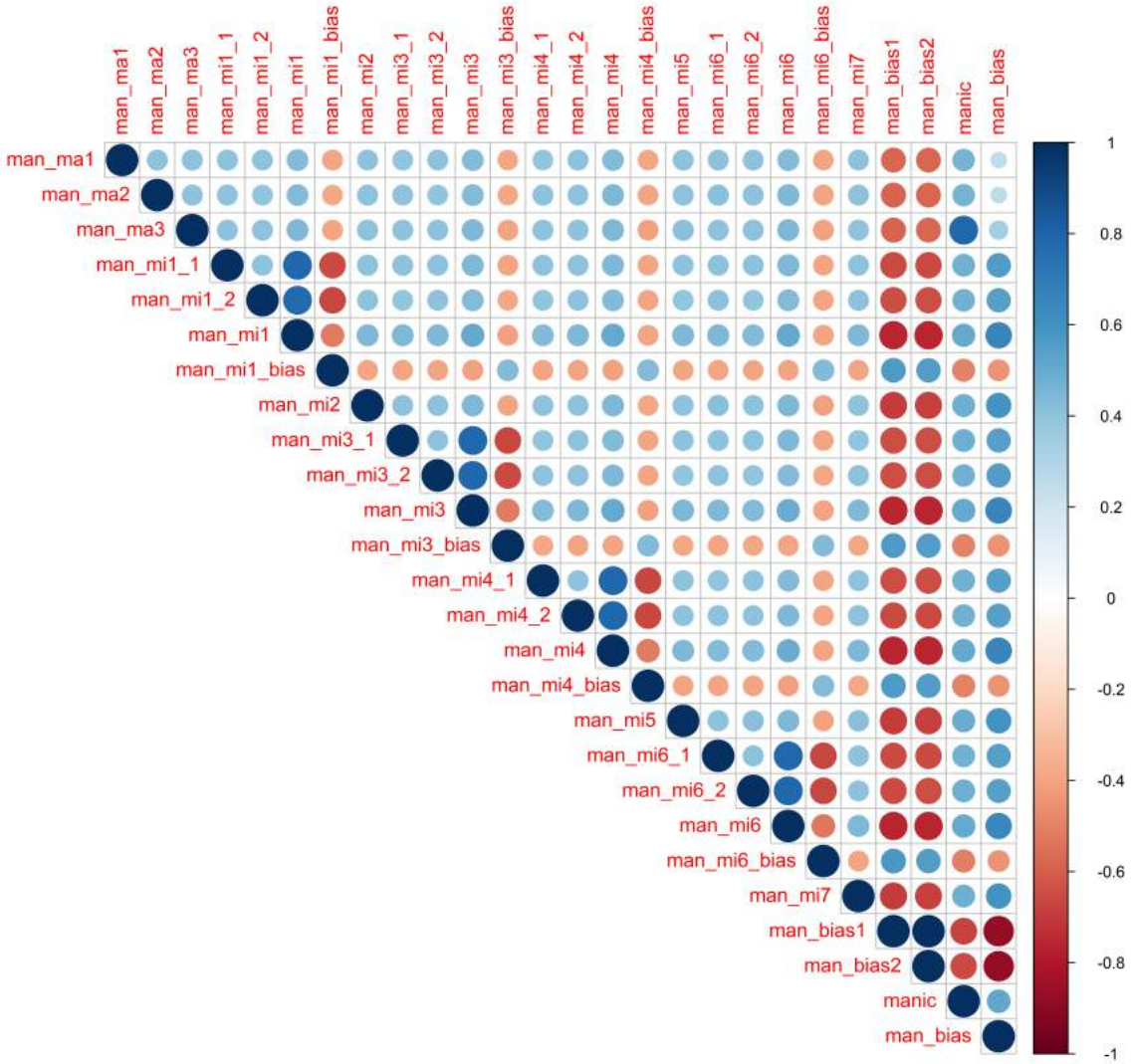


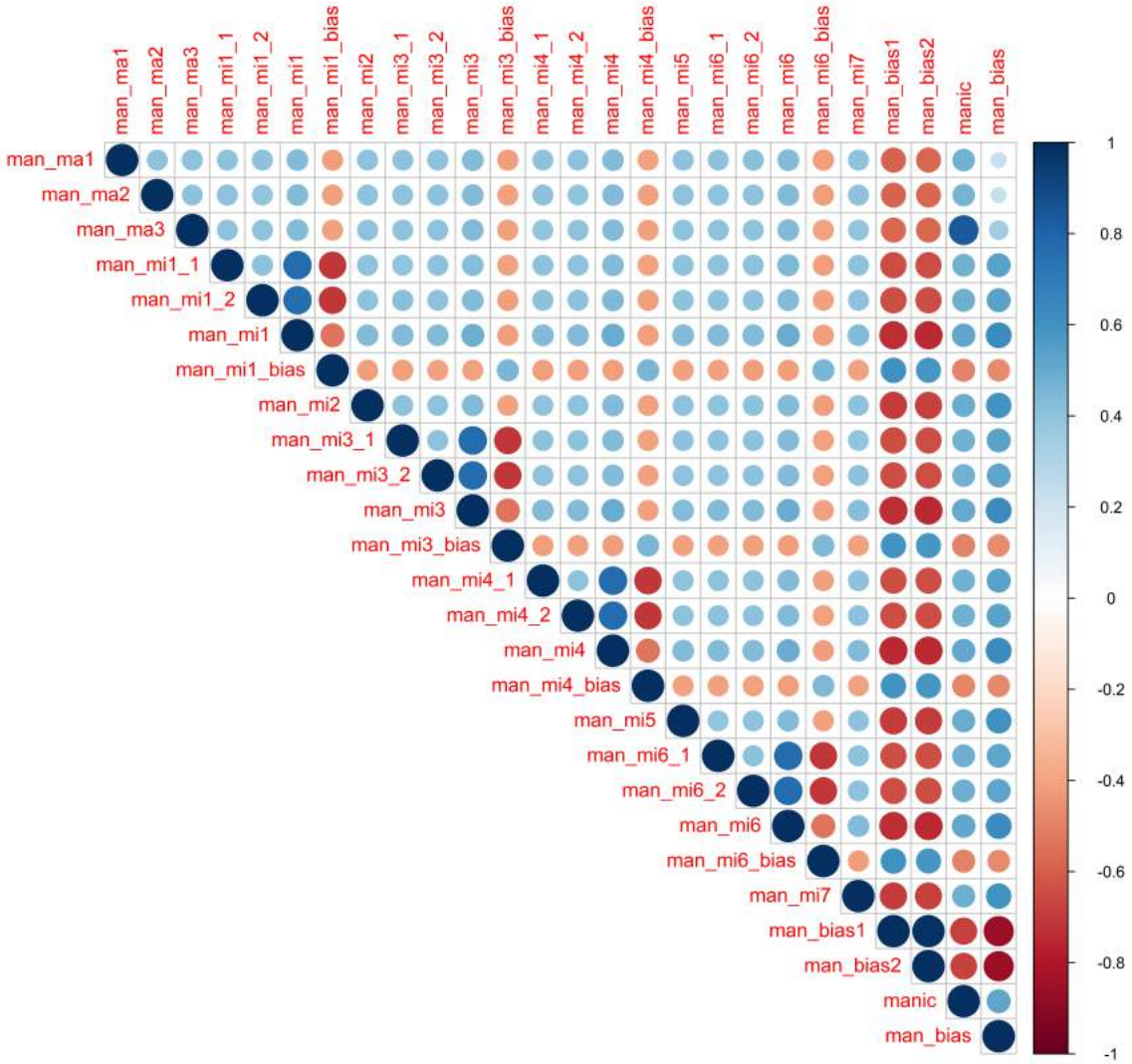

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

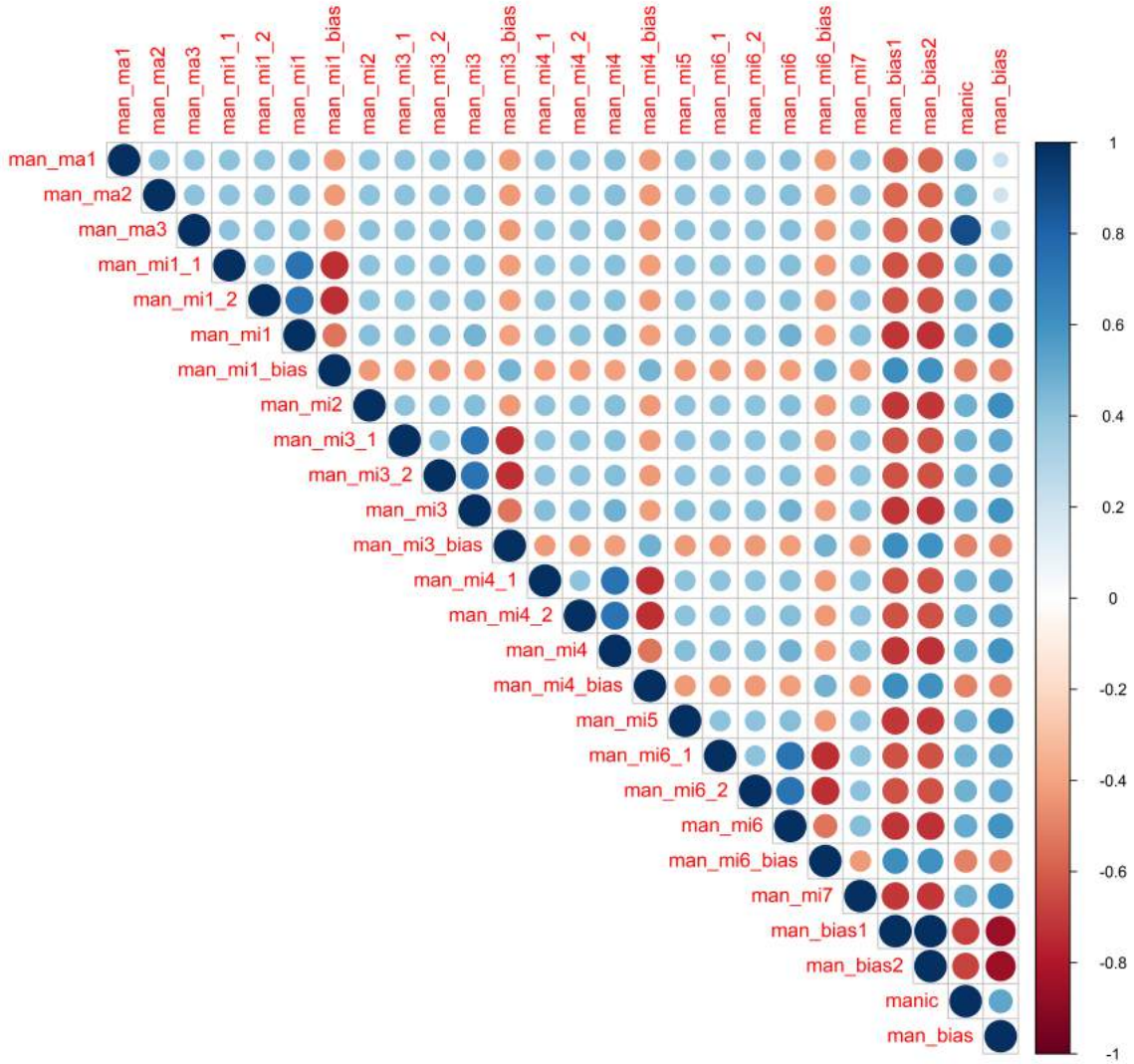

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

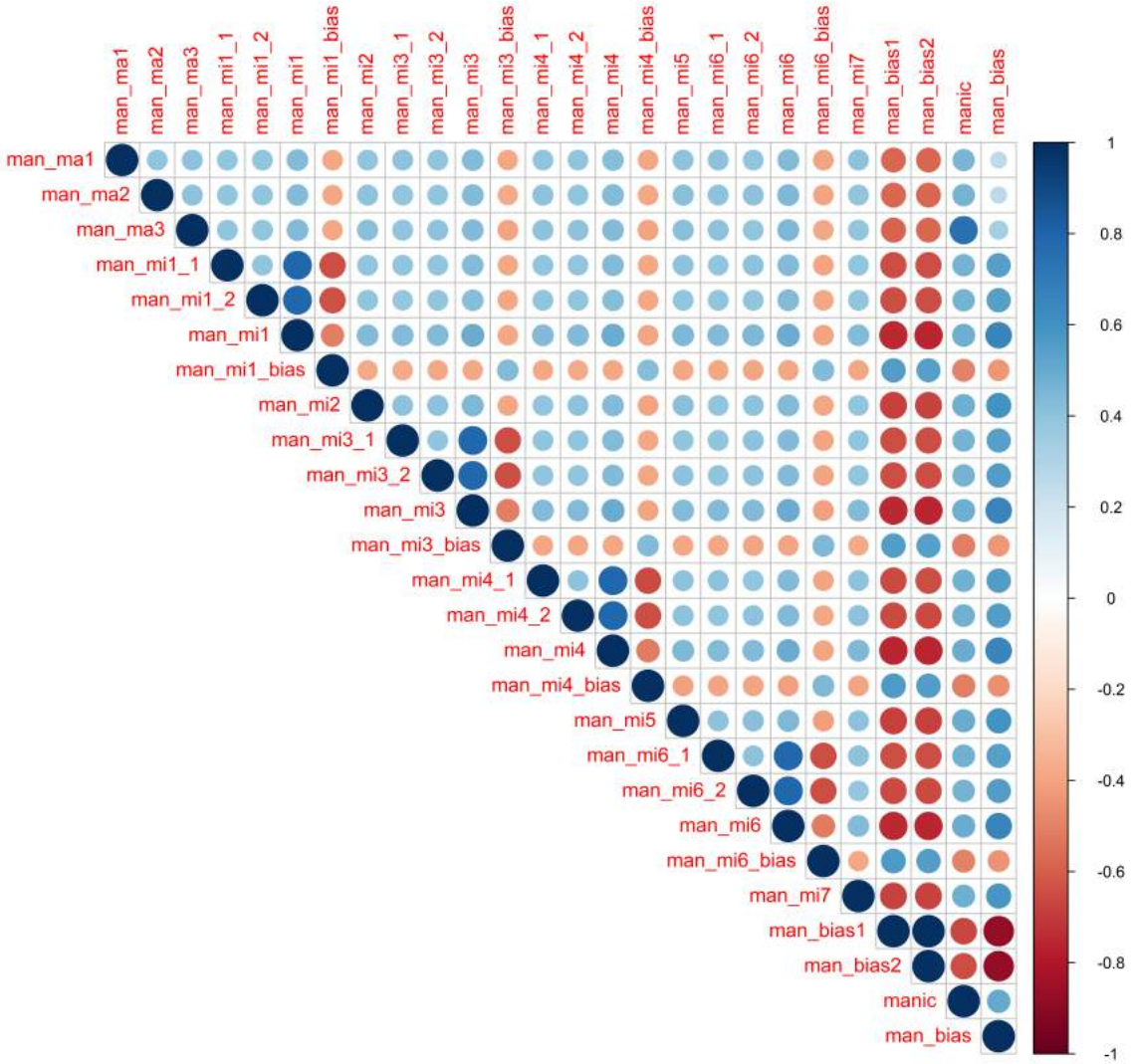

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

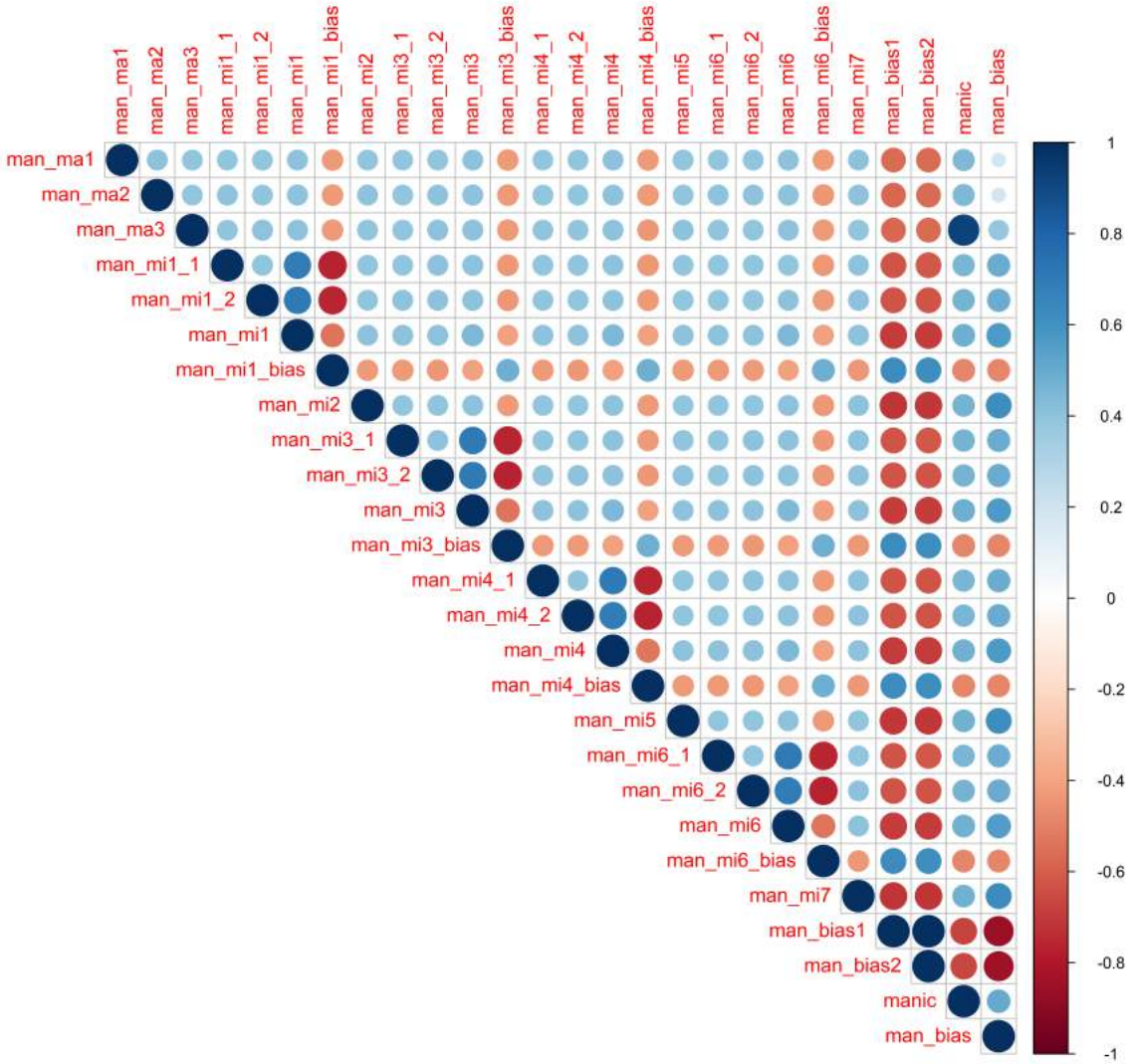

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1









BMJ Open

A simulation study to demonstrate biases created by diagnostic criteria of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037022.R1
Article Type:	Original research
Date Submitted by the Author:	24-Jun-2020
Complete List of Authors:	Chao, Yi-Sheng; Independent researcher Lin, Kuan-Fu; National Taiwan University Hospital Yun-Lin Branch, Psychiatry Wu, Chao-Jung; UQAM, Département d'informatique Wu, Hsing-Chien; Taipei Hospital, Internal Medicine Hsu, Hui-Ting; Changhua Christian Healthcare System, Pathology Tsao, Lien-Cheng; Changhua Christian Healthcare System, Surgery Cheng, Yen-Po; Changhua Christian Healthcare System, Surgery Lai, Yi-Chun; National Yang Ming University Hospital, Chest Medicine Chen, Wei-Chih; Taipei Veterans General Hospital, Chest Medicine; National Yang-Ming University, Institute of Emergency and Critical Care Medicine
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Epidemiology, Research methods, Diagnostics
Keywords:	MENTAL HEALTH, Depression & mood disorders < PSYCHIATRY, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A simulation study to demonstrate biases created by diagnostic criteria of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Yi-Sheng Chao[1], Kuan-Fu Lin,[2] Chao-Jung Wu[3], Hsing-Chien Wu[4], Hui-Ting Hsu[5], Lien-Cheng Tsao[5], Yen-Po Cheng[5], Yi-Chun Lai[6], Wei-Chih Chen[7,8*]

[1]Independent researcher, Montréal, H2X 0A8 Canada, [2]National Taiwan University Hospital Yun-Lin Branch, Yunlin County, 640 Taiwan, [3]Département d'informatique Université du Québec à Montréal, Montréal H3B 1B4 Canada, [4]Taipei Hospital Ministry of Health and Welfare New Taipei city, 242 Taiwan, [5]Changhua Christian Hospital, Changhua County 526, Taiwan, [6]National Yang-Ming University Hospital, Yilan 260 Taiwan, [7]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 112, Taiwan, [8]Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei 112, Taiwan

[*]Corresponding author: Wei-Chih Chen
Department of Chest Medicine, Taipei Veterans General Hospital
No. 201, Section 2, Shih-Pai Road, Taipei 11217, Taiwan
Telephone: +886-2-28757456
Fax: +886-2-28757610
Email address: wiji.chen@gmail.com

Keywords: frailty; bias; forward-stepwise regression; the Health and Retirement Study; index mining

Abstract

Objectives

Composite diagnostic criteria alone are likely to create and introduce biases into diagnoses that subsequently have poor relationships with input symptoms. This study aims to understand the magnitudes of biases created by diagnostic criteria alone and introduced into the diagnoses of mental illnesses with large disease burdens (major depressive episodes, dysthymic disorder, and manic episodes) and the relationships between the diagnoses and the input symptoms.

Settings

General psychiatric care.

Participants

Without real-world data available to the public, 100,000 subjects were simulated and the input symptoms were assigned based on the assumed prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations between symptoms (0, 0.1, 0.4, 0.7, and 0.9). The input symptoms were extracted from the diagnostic criteria. The diagnostic criteria were transformed into mathematical equations to demonstrate the sources of biases and convert the input symptoms into diagnoses.

Primary and secondary outcomes

Biases due to data censoring or categorization introduced into the intermediate variables, and the three diagnoses were measured. The relationships between the input symptoms and diagnoses were interpreted using forward stepwise linear regressions.

Results

The prevalence rates of the diagnoses were lower than those of the input symptoms and proportional to the assumed prevalence rates and the correlations between the input symptoms. Certain input or bias variables consistently explained the diagnoses better than the others. Except for zero correlations and 0.7 prevalence rates of the input symptoms for the diagnosis of dysthymic disorder, the input symptoms could not fully explain the diagnoses.

Conclusions

There are biases created due to composite diagnostic criteria and introduced into the diagnoses. The design of the diagnostic criteria determines the prevalence of the diagnoses, the relationships between the input symptoms, the diagnoses, and the biases. The importance of the input symptoms has been distorted largely by the diagnostic criteria.

Trial registration

Not applicable

Strength and limitation

1. The prevalence of three mental illnesses was determined by the prevalence of the input symptoms and modified by the diagnostic criteria and correlations between the input variables in simulated populations.

2. Biases due to data censoring or categorization were created by the diagnostic criteria and introduced into the intermediate variables and the three diagnoses of mental illnesses in simulated populations.
3. The diagnostic criteria modified the importance of the input symptoms; certain input symptoms or bias variables were weighted more than expected in simulated populations.
4. The design of diagnostic criteria influenced the diagnosis prevalence. With the same input symptom prevalence, dysthymic disorder was the most prevalent among three illnesses. Major depressive episodes were the least prevalent.
5. This study is based on simulated data and needs to be verified with real-world data.

Background

The diagnoses of several mental illnesses in patients are often made based on a variety of criteria. These criteria often involve symptoms reported by the patients.[1, 2] For example, the diagnosis of major depressive disorder defined in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) requires at least one major depressive episode.[1, 2] For each major depressive episode, the major criteria are "*depressive mood and/or loss of interest or pleasure in life activities for at least 2 weeks*".[1, 2] In addition to the major criteria, the patients need to report at least five of the nine symptoms that "*cause clinically significant impairment in social, work, or other important areas of functioning almost every day,*" including insomnia or hypersomnia and fatigue or loss of interest.[1, 2] In other words, patients need to match both the major and minor criteria before being diagnosed with a major depressive episode.

Historically this symptom-based diagnostic approach developed by Feighner et al. has been widely accepted.[3, 4] Since then, mental illnesses can be diagnosed through different sets of criteria. This approach is important because clinicians become capable of screening important symptoms before diagnosing and treating patients accordingly. In fact, these criteria can also be seen as composite measures that use multiple measures to capture disorders that may not be quantified with single variables.[5, 6] Recent studies on composite measures have found them problematic because biases can be introduced while aggregating information from input variables.[6] The biases emerge while the sums of input variables are censored or while input variables are transformed inadequately.[6, 7] In other words, biases can be created when there is information in the composite measures that is not explained by and unrelated to the input variables.[6] For example, categorizing continuous variables considers individuals in the same group the same and disregards the heterogeneity between those in the same categories.[6] Such practices induce biases and decrease measurement precision.[6, 7]

Currently there is no extensive review on the existence of these biases created by composite measures or medical diagnoses, and only selected diagnoses have been studied for such biases. These biases have been proven vital to another symptom-based composite measure, the diagnosis of frailty, a condition that often occurs in the elderly and is significantly associated with health outcomes, such as mortality, falls, and morbidity.[6] Frailty is diagnosed based on several symptoms and characterized by weakness and vulnerability to adverse health events.[6] While using one of the most widely used diagnostic criteria, the Biological Syndrome Model scores, to diagnose frailty,[8] biases alone can explain more than 71% of the variances of the frailty diagnosis.[6] The biases introduced by data censoring and data categorization can better explain the frailty diagnosis than the input symptoms.[6]

Mostly designed as symptom-based composite measures, it is possible that the diagnostic criteria of mental illnesses also create and introduce biases into diagnoses so that the diagnoses could not be fully explained by the input symptoms. In concern of the biases created by the diagnostic criteria alone, this study aims first to understand the relationships between mental symptoms and diagnoses and then to quantify the potential role of the biases regarding the diagnoses by simulating populations with different prevalence rates and between-variable correlations of mental symptoms.

Methods

Assumptions and simulation parameters

A file containing R codes to reproduce the simulations was attached in the supplementary file. Simulated populations with mental symptoms of different prevalence rates and between-variable correlations were created to interpret the diagnoses and understand the potential magnitudes of biases that could be introduced via data processing implied by the diagnostic criteria (reproducible using R codes and data in the supplementary file). Three diagnoses of mental illnesses were chosen for the leading associated disease burdens:[2] major depressive episodes for the diagnosis of major depressive disorder, dysthymic disorder, and manic episodes for the diagnosis of bipolar disorder.[1]

There were assumptions made to simulate the populations (Table 1). First, for each simulation, the prevalence rates of the input symptoms were assumed to be similar for the three diagnoses in this study. Second, the input symptoms for the diagnoses of major depressive episodes and dysthymic disorder correlated with the same correlation coefficients.[9] The symptoms for the diagnosis of manic episodes correlated to one another. Third, the input symptoms for the diagnosis of manic episodes were created independently of those for the diagnosis of the other two mental illnesses. The assumptions of the prevalence rates and between-variable correlations were made because there was no acceptable-quality data on the symptoms of mental illnesses published. There were studies on the prevalence of mental illnesses,[10, 11] but the information on the prevalence of mental symptoms was very limited. There were variables about depression or anxiety collected in national surveys, such as the items collected through the Center for Epidemiologic Studies Depression Scale.[6, 12-18] However, these variables were not the symptoms used in the DSM-IV-TR. Lastly, we assumed that the diagnoses were made accurately based on the input symptoms reported precisely by patients. The diagnostic criteria in the DSM-IV-TR were strictly followed. However, these assumptions did not hold in the real world.[19] For simplicity and practicality reasons, we assumed perfect diagnostic quality by physicians and accurate reporting of the input symptoms by patients in the simulated populations.

Diagnostic criteria as mathematical functions

The input symptoms were extracted from the major and minor criteria of the diagnoses and listed in Table 2 to Table 4. The input symptoms, major and minor criteria, and the diagnoses were assigned variable names. All input symptoms, items or domains in the major or minor criteria, and the diagnoses were binomial variables, presenting zero and one for the absence and presence of the symptoms, criteria, and the diagnoses respectively. For example, "*insomnia*" and "*hypersomnia*" were extracted from one of the minor criteria for the diagnosis of major depressive episodes.[1] "*More talkative than usual*" and "*pressure to keep talking*" were extracted from one of the minor criteria for the diagnosis of manic episodes.[1]

Mathematical functions were generated based on the diagnostic criteria to convert input symptoms into diagnoses. For example, one of the minor criteria of dysthymic disorder was "*poor appetite or overeating.*" This required two input symptoms and one bias variable to generate the criterion.[6] In other words, "*poor appetite or overeating*" equaling the sum of two input variables, "*poor appetite*" and "*overeating,*" and a bias variable to achieve censoring of the sum of both variables.[6] The sum of two binomial variables could be zero,

one and two for the subjects. However, to derive a binomial variable (having at least one symptom) based on a distribution of 0 to 2, the bias variable had values of -1 for subjects with both symptoms to obtain values less than or equal to one in all subjects.[6] Therefore, the bias variable had values of -1 for the subject with both symptoms and 0 for the other subjects. In addition to adding variables together to derive an intermediate variable or a diagnosis, multiplication, categorization, and other more complicated methods were used in the diagnostic criteria to generate diagnosis variables and domain variables in the major or minor criteria.

For example, the diagnosis of dysthymic disorder required the confirmation of both the major criteria, "*depressed mood most of the day for more days than not, for at least 2 years*" and the minor criteria, "*the presence of two or more of the following symptoms*," at the same time.[1] The diagnosis based on whether subjects meeting both the major and minor criteria of dysthymic disorder is the same as identifying those with a multiplicative product of 1 of two binomial variables (0 and 1 for absence and presence of the major or minor criteria). In the equations, two binomial variables were multiplied to obtain the diagnosis of dysthymic disorder among those with a multiplicative product of 1. I2ndividuals could be assigned zero or one for whether they met both criteria, while the sum of major and minor criteria were zero, one or two for the individuals. Linearly, a bias variable with values of -1 or zero was created and those meeting the major or minor criteria were assigned -1.[6] For categorization of continuous variables, bias variables were required to remove the variations between the subjects in the same categories.[6] Other equations to generate the intermediate variables and the diagnoses were listed and explained in Table 2 to Table 4.

Generation of bias variables

Bias variables could be generated while binomial input symptoms were summed or multiplied to obtain binomial intermediate or diagnosis variables (see the example in the previous two paragraphs).[6] A visual presentation of how bias variables were generated was published.[6] Therefore the number of bias variables depended on the complexity of how the diagnoses were made. For example, six of the nine items or domains in the minor criteria for the diagnosis of major depressive episodes were the censored sums of the input symptoms and six bias variables were derived along with the intermediate variables that represented the items in the minor criteria. The other bias variables were described in Table 2 to Table 4.

Simulation parameters and simulated populations

We simulated populations of 100,000 subjects. There were five prevalence rates to simulate the input symptoms for the diagnosis of major depressive episodes, dysthymic disorder, and manic episodes: 0.05, 0.1, 0.3, 0.5, and 0.7. The correlations between the input symptoms were hypothesized to be 0, 0.1, 0.4, 0.7, and 0.9. There were 25 combinations of the assumed prevalence rates and between-variable correlations. The presence of the input symptoms was randomly assigned to the subjects after specifying the prevalence rates and between-variable correlations between the input symptoms.[20, 21] The intermediate and diagnosis variables were derived according to the equations in Table 2 to Table 4. For each combination of prevalence rates and between-variable correlations, the populations were simulated for 100 times to obtain the mean values and 95% confidence intervals (CIs) of derived prevalence rates, as well as the adjusted R squared and p values to approximate the diagnosis variables.

Diagnosis approximation

Due to the existence of the biases, the input symptoms were not likely to fully explain the diagnoses.[6] Therefore, the diagnoses were approximated by the input, bias, and intermediate variables individually and collectively.[6, 12, 14, 16] The approximation was conducted using forward-stepwise linear regressions.[6, 12, 14, 16, 22] The interpretability of the diagnoses by the input symptoms and bias variables was assessed via adjusted R square: zero suggested that the input symptoms were unrelated to the diagnosis, and one suggested that the input symptoms perfectly explained the diagnosis.[14, 15, 23-26]

All statistical analyses were conducted within the R environment (v3.4.1)[27] and RStudio (v1.0.153).[28] Two-tailed P values less than 0.05 were considered statistical significant.

Patient and Public Involvement

This is a simulation study that did not involve patients or human subjects.

Results

The derived prevalence rates of the input symptoms for the three mental illnesses matched the assumed rates in the supplementary file. The derived correlations between the input symptoms were close to assumed levels in the supplementary file. The simulations were successful and accurate based on the assumed prevalence rates and correlations.

Prevalence of intermediate variables

The items in the major and minor criteria were the intermediate variables necessary to create the diagnoses. The methods used to generate the intermediate variables were important for the prevalence rates of the intermediate variables and the derived diagnoses in Figure 1. For example, an intermediate variable, "*significant unintentional weight loss or gain,*" was created by summing and censoring two binomial variables with values of zero and one (significant unintentional weight loss; significant unintentional weight gain). The prevalence rates of the intermediate variables were larger than those of the two input symptoms regardless of the assumed prevalence rates or between-variable correlations of the input symptoms.

In contrast, the diagnosis of dysthymic disorder was a multiplication product of two intermediate binomial variables, the major and minor criteria, and the prevalence rates of dysthymic disorder were lower than those of the major or minor criteria under all combinations of assumed correlations and prevalence rates in Figure 2.

Prevalence of mental illnesses

The derived prevalence rates of three diagnoses were plotted against the assumed prevalence rates and correlations of the input symptoms in Figure 2 to Figure 4 and listed in Table 5. None of the three diagnoses had prevalence rates exceeding those of the input symptoms. In general, higher prevalence rates or between-variable correlations of the input symptoms were associated with higher prevalence rates in the three diagnoses, except for manic episodes that had higher prevalence rates (0.692) assuming zero correlations and 0.7 prevalence rates than the prevalence rate (0.679) assuming 0.1 correlations and 0.7 prevalence rates of the input symptoms. When compared across Figure 2 to Figure 4, given the same assumed prevalence rates and between-variable correlations of the input

symptoms, the diagnostic criteria of dysthymic disorder consistently generated diagnoses of the highest prevalence rates and the criteria of major depressive episodes created diagnoses of the least prevalence rates (see Table 5 for details).

Associations between the diagnoses and individual input symptoms or bias variables

The diagnoses were first interpreted with the input symptoms (including intermediate variables) and the bias variables individually. The diagnosis of dysthymic disorder, for example, was interpreted with the input symptoms, the bias variables, and both in Figure 5. For each simulation, the diagnosis of dysthymic disorder was approximated with an increasing number of the input symptoms, the bias variables, or both. After selecting the variables that best approximated the diagnosis based on adjusted R-squared, the input symptoms could explain a proportion of 0.955 of the diagnosis variance and the bias variables could explain at most a proportion of 0.405 of the diagnosis variance in Figure 5. With all variables used in the regression, the diagnosis could be perfectly explained by the input symptoms and bias variables (adjusted R-squared = 1). The individual input symptoms and the bias variables that individually best explained the diagnoses are listed in Table 6 and Table 7, respectively.

For the diagnosis of major depressive episodes, the first and second items in the major criteria (variable names: mde_ma1 for or mde_ma2 in Table 2) individually best explained the diagnosis depending on the assumed prevalence rates and correlations in Table 6. For the diagnosis of dysthymic disorder, the major criteria (dys_ma in Table 3) consistently and individually explained the diagnosis the best. For the diagnosis of manic episodes, the third item of the major criteria (man_ma3 in Table 4) individually best explained the diagnosis in all combinations of assumed prevalence rates and correlations. However, the proportions of diagnosis variances best explained by individual input symptoms varied widely between 0.001 to 0.974, depending on the assumed prevalence rates and between-variable correlations. Based on the adjusted R-squared for individual input symptoms, certain input variables were more important than other symptoms due to a high correlation with the diagnoses, such as the major criteria for the diagnosis of dysthymic disorder. The prevalence rates and between-variable correlations were important to determine the relationships between input symptoms and diagnoses.

Similarly, there were bias variables that consistently best explained the diagnoses in Table 7. For the diagnosis of major depressive episodes, the biases due to categorization of the numbers of confirmed input symptoms (mde_bias1 and mde_bias2 in Table 2) were the leading bias variable. The diagnosis of major depressive episodes not explained by the input symptoms or information censoring (mde_bias in Table 2) was the leading bias variable in two combinations of the assumed prevalence rates and correlations. For the diagnosis of dysthymic disorder, the residual of the diagnosis not explained by the major and minor criteria (dys_bias in Table 3) and the bias due to the categorization of the confirmed input symptoms in the minor criteria (dys_mi_bias) were the leading bias variables. For the diagnosis of manic episodes, the bias due to the categorization of the number of confirmed input symptoms in the minor criteria up to three (man_bias1 in Table 4) was the leading bias variables, except for two combinations of the assumed prevalence rates and correlations, in which the bias due to categorization of the confirmed input symptoms in the minor criteria up to four (man_bias2 in Table 4) best explained the diagnosis. However, the proportions of

diagnosis variances explained by individual bias variables varied widely from zero to 0.87. Depending on the assumed prevalence rates and between-variable correlations of the input symptoms, certain bias variables were more important than other bias variables and even some input variables. The assumed prevalence rates and between-variable correlations were important factors for the relationships between the bias variables and the diagnoses.

In general, the proportions of the diagnosis variance that could be explained by either individual input symptoms or single bias variables were low when the prevalence rates and between-variable correlations of the input symptoms were assumed to be low. With higher assumed prevalence rates or correlations, the proportions of the diagnoses explained by the single input symptoms or bias variables were higher. Across three diagnoses, the diagnosis of dysthymic disorder could be better explained by single input variables (higher adjusted R-squared), and the diagnosis of major depressive episodes was associated with the least adjusted R-squared. The bias variables of the diagnosis of manic episodes could explain the diagnosis individually better than the bias variables of the other two diagnoses.

Approximating the diagnoses with input symptoms

When the diagnoses were approximated by their own input symptoms (Table 8), there were always some diagnosis variances that could not be explained by the input symptoms. In other words, the input symptoms could not fully explain the diagnoses, except for the diagnosis of dysthymic disorder that could be fully explained by the input symptoms (adjusted R-squared = 1) assuming zero between-variable correlations and 0.7 prevalence rates for the input symptoms. In Table 8, the proportions of diagnosis variances explained by input symptoms increased with higher assumed prevalence rates or between-variable correlations of the input symptoms in general. The input symptoms of dysthymic disorder explained the diagnosis better than those of the other two diagnoses under all combinations of assumed prevalence rates and between-variable correlations. However, the proportion of diagnosis variance explained by own input symptoms varied widely from 0.003 to 1.0. The assumed prevalence rates and between-variable correlations of the input symptoms and the design of the diagnostic criteria were all important for the relationships between input symptoms and diagnoses.

Approximating the diagnoses with bias variables

The diagnoses were approximated with the bias variables of their own. The bias variables always explained some of the diagnosis variances, except for the diagnosis of dysthymic disorder assuming zero between-variable correlations and 0.7 prevalence rates for the input symptoms (adjusted R-squared = 0). With increasing assumed between-variable correlations for the input symptoms, the adjusted R-squared increased. However, given the same assumed between-variable correlations, the proportions of diagnosis variances explained by the bias variables might increase or decrease with the assumed prevalence rates. Compared to the adjusted R-squared in Table 8, the proportion of the diagnosis variances explained by the bias variables was always smaller than that explained by the input symptoms in Table 9. However, the proportions of the diagnosis variance explained by bias variables also varied widely from zero to 0.89. The assumed prevalence rates and between-variable correlations of input symptoms and the design of the diagnostic criteria were important for the relationship between the bias variables and the diagnoses. Only when the input symptoms for the diagnosis of dysthymic disorder were randomly and

independently prevalent to 70% of the simulated populations, the bias variables became irrelevant to the diagnosis.

Discussion

This study is a first attempt to assess the biases created by mental illness diagnostic criteria, as well as understand the relationships between input symptoms and the diagnoses of three mental illnesses: major depressive episodes (at least one episode required for the diagnosis of major depressive disorder), dysthymic disorder, and manic episodes. The diagnostic criteria of these three mental illnesses have been reviewed and rewritten as mathematical functions. Simulated populations of 100,000 for each of 100 simulations, with input symptoms of the three diagnoses, were created. For simplicity and practicality, the presence of the input symptoms was randomly assigned, and the input symptoms were assumed to have uniform prevalence rates and between-variable correlations. There were 25 combinations of assumed prevalence rates and between-variable correlations simulated.

Mathematically, the diagnostic criteria are functions and composite measures to transform information from the input symptoms to diagnoses. There are bias variables created by the diagnostic criteria due to data processing.[6] There are three major mechanisms of introducing biases: censoring, data categorization,[7] and multiplication of input symptoms.[6] These mechanisms introduce information or biases that cannot be fully explained by the input symptoms.[6] The introduced biases can sometimes explain more than half of the variance in the diagnoses depending on the prevalence rates and between-variable correlations of the input symptoms. The findings show that the design of the diagnostic criteria is important for bias introduction and significant for the prevalence of the diagnoses in populations, the relationships between the input symptoms and the diagnoses, and the relationships between the bias variables and the diagnoses.

The role of the diagnostic criteria

With the same assumptions in the prevalence rates and between-variable correlations of the input symptoms, the design of the diagnostic criteria of three mental illnesses can be compared to each other. The design of diagnostic criteria transform input symptoms into various diagnosis prevalence rates with implicit upper limits (i.e. no more prevalent than the input symptoms), unacknowledged differential weights on the input symptoms (i.e., certain input symptoms better explaining the diagnoses), and the introduction of biases (i.e., due to censoring, data categorization, or multiplication).

We were the first to notice that the prevalence rates of the three diagnoses were lower than those of the input symptoms if randomly distributed with uniform prevalence rates and correlations. Given similar assumed input symptom prevalence and correlations, dysthymic disorder is the most prevalent, and major depressive episodes are the least. The diagnosis of dysthymic disorder can be better explained by its input symptoms individually or collectively than the other two diagnoses. The diagnosis of major depressive episodes is least explained by own input symptoms individually or collectively. As expected, the diagnosis of the three mental illness is similar to composite measures or indices and is subject to the biases introduced by data processing, given all combinations of the assumed prevalence rates and between-variable correlations of the input symptoms.[6] There is only one exception: dysthymic disorder with the input symptoms that are randomly and

independently present in 70% of the population. This is because the diagnosis of dysthymic disorder is a multiplicative product of the major and minor criteria. Without correlations, everyone in the population is certain to qualify for the minor criteria (probability of 100% because having at least two out of the six items in the minor criteria: mathematically [C(2,6) + C(3,6) + C(4,6) + C(5,6) + C(6,6)] X $(0.7)^6$ = 37 X 0.117 = 4.35 > 100%). If 70% of the population were also randomly assigned with the major criteria and 100% were assigned with the minor criteria, 70% would be diagnosed with dysthymic disorder, and the diagnosis of dysthymic disorder can be fully explained by the major criteria. In fact, without correlations between input symptoms, it only requires each of the six items in the minor criteria to be randomly assigned to 54.8% [$(1/37)^{(1/6)}$] of the population for everyone to qualify for the minor criteria, and the diagnosis can be fully explained by the minor and major criteria.

Distortion of the input symptoms

The importance of the input symptoms has been distorted due to the diagnostic criteria for the three mental illnesses. The same phenomenon has been proven in the diagnosis of frailty based on three of the most commonly used scoring methods.[6] In other words, based on the functions to generate the diagnoses, the input symptoms are differentially weighted, and weights are not explicitly acknowledged. The most prominent is the diagnosis of dysthymic disorder; more than 90% of the variance can be explained by its major criteria assuming 0.7 or 0.9 between-variable correlations for the input symptoms in Table 6. Another example is that the third item of the major criteria for the diagnosis of manic episodes, "*irritable mood,*" individually predicts the diagnosis better than any other input symptoms or intermediate variables. The input symptom has been given more weight than others and can explain more than 91.8% of the diagnosis variance, assuming 0.9 correlations between input symptoms. Based on the texts in the DSM-IV-TR, we do not think this symptom should be emphasized to this degree. However, the diagnostic criteria impose implicit and unequal weights to the input symptoms, and introduce biases into the diagnoses.

Future directions

We think it important to rethink the role and importance of the diagnostic system. Current approaches are embedded with implicit assumptions of the prevalence rates of the diagnoses (no higher than input symptoms if similar symptom prevalence), unacknowledged weights to input symptoms (certain input symptoms explaining the diagnoses much better), and biases that were induced by data processing and could not be explained by the input symptoms. It is unclear whether the diagnosis of dysthymic disorder was intentionally designed to be more prevalent than those of major depressive episodes or manic episodes assuming input symptoms with the same prevalence rates.

In the real world, there are other important issues related to the diagnostic criteria. For example, diagnoses are not closely linked to treatment,[19, 29] diagnoses are not well made particularly by non-psychiatrists,[30] and there are two diagnostic systems (the DSM and the International Classification of Disease) that require efforts to harmonize.[31] Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved to be easier to understand and use without introducing biases, and closely linked to clinical decisions. Certain measures and biomarkers have been proven useful to identify mental illnesses.[32, 33] We are developing methods to detect symptom-based conditions better and propose methods to search for neglected mental symptoms.

Limitations

The strength of this study is the use of simple assumptions in simulated populations that enables the comparison of the diagnostic criteria of three mental illnesses. However, the assumptions in the prevalence rates and between-variable correlations for the input symptoms might not be realistic. Some of the assumptions are unlikely to hold in the real world. However, simulations are the only option for us due to the lack of real-world data on the prevalence of the input symptoms. In addition, the translation from symptoms to diagnoses was assumed to be perfect based on the diagnostic criteria. The simulations in this study only reflect the problems in the design of the diagnostic criteria and are not designed to review the impact of how they are used in the real world.

Conclusion

To the best of our knowledge, there is no study on the relationships between the input symptoms and diagnoses. The input symptoms were extracted from the diagnostic criteria and the diagnostic criteria were transformed into mathematical functions. Without mental illness data available to the public, 100,000 subjects were simulated with different assumptions on the prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations (0, 0.1, 0.4, 0.7, and 0.9) of the input symptoms. We found that biases were introduced into the diagnoses of three mental illnesses: major depressive episodes, dysthymic disorder, and manic episodes. The prevalence rates of the diagnoses were proportional to the assumed prevalence rates and between-variable correlations of the input symptoms. Certain input symptoms were more important than the others in explaining the diagnoses. However, the input symptoms could not fully explain the diagnoses, except when the input symptoms independent of each other with 0.7 symptom prevalence rates were used for the diagnosis of dysthymic disorder. In conclusion, the criteria used to diagnose these three mental illnesses may fail to represent the concepts they are based on, in a similar manner to three of the most commonly used scoring methods to diagnose frailty.

Declarations

Acknowledgments

Not applicable.

Ethical Statement

Not applicable.

Funding Statement

The authors received no specific funding for this work.

Consent to participate

Not applicable

Consent for publication

Not applicable

Data Availability

No real-world data used - all analysis are based on simulations reproducible with the files in the supplemental materials.

Competing Interests

YSC is currently employed by the Canadian Agency for Drugs and Technologies in Health. The other authors have declared that no competing interests exist.

Authors' Contributions

YSC conceptualized and designed this study, managed and analyzed data and drafted the manuscript. KFL assisted in the interpretation of the diagnostic criteria. CJW assisted in data management and computation. HCW, HTH, LCT, YPC, YCL, and WCC participated in the design of this study. All authors reviewed and approved the manuscript.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR®). Washington, DC: American Psychiatric Association Publishing; 2010.
2. Center for Substance Abuse Treatment. Managing Depressive Symptoms in Substance Abuse Clients During Early Recovery. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2008.
3. Feighner JP, Robins E, Guze SB, Woodruff RA, Jr., Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry. 1972;26(1):57-63. Epub 1972/01/01.
4. Kendler KS, Munoz RA, Murphy G. The development of the Feighner criteria: a historical perspective. The American journal of psychiatry. 2010;167(2):134-42. Epub 2009/12/17. doi: 10.1176/appi.ajp.2009.09081155.
5. Chao Y-S, Wu C-J. PP46 When Composite Measures Or Indices Fail: Data Processing Lessons. International Journal of Technology Assessment in Health Care. 2019;34(S1):83-. Epub 01/03. doi: 10.1017/S0266462318002088.
6. Chao Y-S, Wu H-C, Wu C-J, Chen W-C. Index or illusion: The case of frailty indices in the Health and Retirement Study. PLOS ONE. 2018;13(7):e0197859. doi: 10.1371/journal.pone.0197859.
7. Barnwell-Menard JL, Li Q, Cohen AA. Effects of categorization method, regression type, and variable distribution on the inflation of Type-I error rate when categorizing a confounding variable. Stat Med. 2015;34(6):936-49. Epub 2014/12/17. doi: 10.1002/sim.6387.
8. Cigolle CT, Ofstedal MB, Tian Z, Blaum CS. Comparing models of frailty: the Health and Retirement Study. J Am Geriatr Soc. 2009;57(5):830-9. Epub 2009/05/21. doi: 10.1111/j.1532-5415.2009.02225.x.
9. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behaviour research and therapy. 1997;35(1):79-89.
10. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. Scientific reports. 2018;8(1):2861-. doi: 10.1038/s41598-018-21243-x. PubMed PMID: 29434331.
11. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. PLOS ONE. 2013;8(11):e75362. doi: 10.1371/journal.pone.0075362.
12. Chao Y-S, Wu C-J. PD26 Principal Component Approximation: Canadian Health Measures Survey. International Journal of Technology Assessment in Health Care. 2019;34(S1):138-9. Epub 01/03. doi: 10.1017/S026646231800301X.
13. Chao Y-S, Wu C-J, Wu H-C, Chen W-C. Trend analysis for national surveys: Application to all variables from the Canadian Health Measures Survey cycle 1 to 4. PLOS ONE. 2018;13(8):e0200127. doi: 10.1371/journal.pone.0200127.
14. Chao Y-S, Wu H-C, Wu C-J, Chen W-C. Principal Component Approximation and Interpretation in Health Survey and Biobank Data. Frontiers in Digital Humanities. 2018;5(11). doi: 10.3389/fdigh.2018.00011.
15. Chao YS, Wu HC, Wu CJ, Chen WC. Stages of Biological Development across Age: An Analysis of Canadian Health Measure Survey 2007-2011. Front Public Health. 2018;5(2296-2565 (Print)). doi: 10.3389/fpubh.2017.00355. eCollection 2017.
16. Chao Y-S, Wu C-J. PD25 Principal Component Approximation: Medical Expenditure Panel Survey. International Journal of Technology Assessment in Health Care. 2019;34(S1):138-. Epub 01/03. doi: 10.1017/S0266462318003008.
17. Chao Y-S, Wu C-J, Chen T-S. Risk adjustment and observation time: comparison between cross-sectional and 2-year panel data from the Medical Expenditure Panel Survey (MEPS). Health Information Science and Systems. 2014;2:5. doi: 10.1186/2047-2501-2-5. PubMed PMID: PMC4340859.

18. Chao YS, Wu HT, Scutari M, Chen TS, Wu CJ, Durand M, et al. A network perspective on patient experiences and health status: the Medical Expenditure Panel Survey 2004 to 2011. BMC health services research. 2017;17(1472-6963 (Electronic)). doi: 10.1186/s12913-017-2496-5.
19. Bonnin JE. Treating without diagnosis: psychoanalysis in medical settings in Argentina. 2015.
20. Leisch F, Weingessel A, Hornik K. On the generation of correlated artificial binary data. 1998.
21. Leisch F, Weingessel A, Hornik K. bindata: Generation of Artificial Binary Data, 2012. URL http://CRAN R-project org/package= bindata R package version 09-19.
22. Lumley T, Lumley MT. Package 'leaps'. Regression Subset Selection Thomas Lumley Based on Fortran Code by Alan Miller Available online: http://CRAN R-project org/package= leaps (accessed on 18 March 2018). 2013.
23. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition: Springer New York; 2009.
24. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R: Springer New York; 2013.
25. Chao Y-S, Wu C-J. Principal component-based weighted indices and a framework to evaluate indices: Results from the Medical Expenditure Panel Survey 1996 to 2011. PLoS ONE. 2017;12(9):e0183997. doi: 10.1371/journal.pone.0183997. PubMed PMID: PMC5590867.
26. Chao YS, Wu HT, Wu CJ. Feasibility of Classifying Life Stages and Searching for the Determinants: Results from the Medical Expenditure Panel Survey 1996-2011. Front Public Health. 2017;5:247(2296-2565 (Print)). doi: 10.3389/fpubh.2017.00247. eCollection 2017.
27. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
28. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2016.
29. Demyttenaere K, Bonnewyn A, Bruffaerts R, De Girolamo G, Gasquet I, Kovess V, et al. Clinical factors influencing the prescription of antidepressants and benzodiazepines:: Results from the European study of the epidemiology of mental disorders (ESEMeD). Journal of affective disorders. 2008;110(1-2):84-93.
30. Margolis RL. Nonpsychiatrist house staff frequently misdiagnose psychiatric disorders in general hospital inpatients. Psychosomatics. 1994;35(5):485-91.
31. First MB. Harmonisation of ICD–11 and DSM–V: opportunities and challenges. The British Journal of Psychiatry. 2009;195(5):382-90.
32. Husain SF, Tang T-B, Yu R, Tam WW, Tran B, Quek TT, et al. Cortical haemodynamic response measured by functional near infrared spectroscopy during a verbal fluency task in patients with major depression and borderline personality disorder. EBioMedicine. 2020;51:102586.
33. Ho CSH, Zhang MWB, Ho R. Optical topography in psychiatry: a chip off the old block or a new look beyond the mind–brain frontiers? Frontiers in psychiatry. 2016;7:74.

Table 1. The assumptions and parameters in the simulations

	Assumptions	
1	Equal prevalence rates for the input symptoms of the same diagnosis; presence of input symptoms assigned randomly	
2	Same correlations between the input symptoms of the diagnoses of major depressive episodes and dysthymic disorder; same correlations between the input symptoms of manic episodes	
3	The input symptoms of manic episodes created independent of those of major depressive episodes and dysthymic disorder	
4	Diagnoses made accurately based on the diagnostic criteria and symptoms reported accurately by patients	
	Parameters of input symptoms of the same diagnosis for each simulation	
1	Population sizes	10,000
2	Prevalence rates (uniform for all input symptoms in a simulation)	0.05, 0.1, 0.3, 0.5, and 0.7
3	Correlations (uniform between all input symptoms of the same diagnosis in a simulation)	0, 0.1, 0.4, 0.7, and 0.9
4	Number of simulations for each combination of the assumed prevalence rates and between-variable correlations of the input symptoms	100

Table 2. The input symptoms, intermediate variables, and bias variables for the diagnosis of major depressive episodes.

Classification of symptoms	Criterion variable	Domains in the major or minor criteria	Domain variables	Symptoms	Symptom variables	Equations to derive diagnosis or domain variables	Approximation by linear regression	Mechanisms related to introducing biases
Major depressive episode (variable = mde)						mde = mde_ma1 x mde_ma2 x (mde_mi3 + mde_mi4 + mde_mi5 + mde_mi6 + mde_mi7 + mde_mi8 + mde_mi9 + mde_bias1) + (1-mde_ma1 x mde_ma2) x (me_ma1 x mde_ma2) x (mde_mi3 + mde_mi4 + mde_mi5 + mde_mi6 + mde_mi7 + mde_mi8 + mde_mi9 + mde_bias2)	mde = intercept + coef1 x mde_ma1 + coef2 x mde_ma2 + coef3 x mde_mi3 + coef4 x mde_mi4 + coef5 x mde_mi5 + coef6 x mde_mi6 + coef7 x mde_mi7 + coef8 x mde_mi8 + coef9 x mde_mi9 + coef10 x mde_bias	1) Multiplication to create the situations when one or two symptoms in the major criteria confirmed and the bias (mde_bias) calculated by extracting the information of the diagnosis not explained by the input symptoms and two bias variables generated by censoring (mde_bias1 and mde_bias2) 2) Categorizing of the sum of the input symptoms in the minor criteria at the threshold of three or four (mde_bias1 and mde_bias2)
Major criteria, essential for diagnosis								
		Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.						
		Depressed mood for more than two weeks.	mde_ma1					
		Loss of interest or pleasure in daily activities for more than two weeks.	mde_ma2					
Minor criteria (at least 5 of the symptoms including the two in major criteria)	mde_mi							
		Significant unintentional weight loss or gain	mde_mi3			mde_mi3 = mde_mi3_1 + mde_mi3_2 + mde_mi3_bias		Censoring of the sum of multiple input variables
				Significant unintentional weight gain	mde_mi3_1			
				Significant unintentional weight loss	mde_mi3_2			
				Information of the domain not explained by the input variables	mde_mi3_bias			
		Insomnia or sleeping too much	mde_mi4			mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias		Censoring of the sum of multiple input variables
				Insomnia	mde_mi4_1			
				Sleeping too much	mde_mi4_2			
				Information of the domain not explained by the input variables	mde_mi4_bias			
		Agitation or psychomotor retardation noticed by others	mde_mi5			mde_mi5 = mde_mi5_1 + mde_mi5_2 + mde_mi5_bias		Censoring of the sum of multiple input variables
				Agitation	mde_mi5_1			

				Psychomotor retardation noticed by others	mde_mi5_2		
				Information of the domain not explained by the input variables	mde_mi5_bias		
		Fatigue or loss of energy	mde_mi6			mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias	Censoring of the sum of multiple input variables
				Fatigue	mde_mi6_1		
				Loss of energy	mde_mi6_2		
				Information of the domain not explained by the input variables	mde_mi6_bias		
		Feelings of worthlessness or excessive guilt	mde_mi7			mde_mi7 = mde_mi7_1 + mde_mi7_2 + mde_mi7_bias	Censoring of the sum of multiple input variables
				Feelings of worthlessness	mde_mi7_1		
				Feelings of excessive guilt	mde_mi7_2		
				Information of the domain not explained by the input variables	mde_mi7_bias		
		Diminished ability to think or concentrate, or indecisiveness+	mde_mi8			mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias	Censoring of the sum of multiple input variables
				Diminished ability to think or concentrate	mde_mi8_1		
				Indecisiveness	mde_mi8_2		
				Information of the domain not explained by the input variables	mde_mi8_bias		
		Recurrent thoughts of death	mde_mi9				
Information due to categorization (choosing three domains in minor criteria)	mde_bias1						Bias introduced to categorize the sum of the number of confirmed symptoms in the minor criteria
Information due to categorization (choosing four domains in minor criteria)	mde_bias2						Bias introduced to categorize the sum of the number of confirmed symptoms in the minor criteria
Information of diagnosis not explained by the domains	mde_bias						Information of the diagnosis not explained by the input variables and two bias variables generated due to data categorization

577

578

Table 3. The input symptoms, intermediate variables, and bias variables for the diagnosis of dysthymic disorder.

Classification of symptoms	Criterion variable	Major or minor criteria (domains)	Intermediate variables	Symptoms	Symptom variables	Equations to generate diagnosis or domain variables	Approximation	Mechanisms related to introducing biases
Dysthymia (variable = dys)						dys = dys_ma x dys_mi	dys = intercept + coef1 x dys_ma + coef2 x dys_mi + coef3 x dys_bias	Multiplication to create the situations where both the major and minor criteria met (union of two binomial variables, mde_ma x mde_mi) and the bias variable (dys_bias) equivalent to the residual of the diagnosis not explained by the input symptoms and the bias variables due to censoring and categorization
Major criteria, essential for diagnosis								
		Depressed mood most of the day for more days than not, for at least 2 years	dys_ma					
Minor criteria (at least 2 items)			dys_mi			dys_mi = dys_mi1 + dys_mi2 + dys_mi3 + dys_mi4 + dys_mi5 + dys_mi6 + dys_mi_bias		Categorizing of the sum of multiple input variables
		Poor appetite or overeating	dys_mi1			dys_mi1 = dys_mi1_1 + dys_mi1_2 + dys_mi1_bias		Censoring of the sum of multiple input variables
				Poor appetite	dys_mi1_1			
				Overeating	dys_mi1_2			
				Information of the domain not explained by the input variables	dys_mi1_bias			
		Insomnia or sleeping too much*	dys_mi2/mde_mi4			dys_mi2 = mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias		Censoring of the sum of multiple input variables
				Insomnia	mde_mi4_1			
				Sleeping too much	mde_mi4_2			
				Information of the domain not explained by the input variables	mde_mi4_bias			
		Low energy or fatigue*	dys_mi3/mde_mi6			dys_mi3 = mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias		Censoring of the sum of multiple input variables
				Fatigue	mde_mi6_1			
				Loss of energy (low energy)	mde_mi6_2			
				Information of the domain not explained by the input variables	mde_mi6_bias			
		Low self-esteem	dys_mi4					
		Poor concentration or difficulty making decisions*	dys_mi5/mde_mi8			dys_mi5 = mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias		Censoring of the sum of multiple input variables
				Diminished ability to think or concentrate (Poor concentration)	mde_mi8_1			
				difficulty making decisions (indecisiveness)	mde_mi8_2			
				Information of the domain not explained by the input variables	mde_mi8_bias			
		Feelings of hopelessness	dys_mi6					

		Information of minor criteria not explained by input variables	dys_mi_bias	Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis not explained by major or minor criteria	dys_bias			Information of the diagnosis not explained by the input symptoms and the bias variables generated due to data categorization (dys_mi_bias)

*The same input symptoms used for the diagnosis of major depressive episodes.

584 Table 4. The input symptoms, intermediate variables, and bias variables for the diagnosis of manic episodes.

Classification of symptoms	Criterion variable	Major or minor criteria (domains)	Domain variables	Symptoms	Symptom variables	Equations	Approximation	Mechanisms related to introducing biases
Manic episode (variable = manic)						manic = (1- man_ma1 x man_ma2) x (man_ma1 + man_ma2) x man_ma3 x (man_mi1 + man_mi2 + man_mi3 + man_mi4 + man_mi5 + man_mi6 + man_mi7 + man_bias1) + [1 - (1 - man_ma1 x man_ma2)(man_ma1 + man_ma2)] x man_ma3 x (man_mi1 + man_mi2 + man_mi3 + man_mi4 + man_mi5 + man_mi6 + man_mi7 + man_bias2)	manic = intercept + coef1 x man_ma1 + coef2 x man_ma2 + coef3 x man_ma3 + coef4 x man_mi1 + coef5 x man_mi2 + coef6 x man_mi3 + coef7 x man_mi4 + coef8 x man_mi5 + coef9 x man_mi6 + coef10 x man_mi7 + coef11 x man_bias	1) Multiplication to create the situations where one of the symptom in the major criteria met (union of three binomial variables, such as man_ma1 + man_ma2 and man_ma1 x man_ma2), \n 2) multiplication for the condition of presenting irritable mood (... x man_ma3), and 3) the bias variable (man_bias) equivalent to the residual of the diagnosis not explained by the input symptoms and the bias variables due to censoring; 4) the bias variables introduced by categorizing the number of input symptoms confirmed in the minor criteria (man_bias1 and man_bias2)
Major criteria, essential for the diagnosis of a manic episode (more than one bipolar episode required to diagnose bipolar disorder)								
		A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary)						
				Elevated mood, lasting at least 1 week	man_ma1			
				Expansive mood, lasting at least 1 week	man_ma2			
				Irritable mood, lasting at least 1 week	man_ma3			
Minor criteria (3 or more of the following symptoms have persisted; 4 if the mood is only irritable)								
		Increased self-esteem or grandiosity	man_mi1			man_mi1 = man_mi1_1 + man_mi1_2 + man_mi1_bias		Censoring of the sum of multiple input variables
				Increased self-esteem	man_mi1_1			
				Grandiosity	man_mi1_2			

				Information of the domain not explained by the input variables	man_mi1_bias		
		Decreased need for sleep (e.g., feels rested after only 3 hours of sleep)	man_mi2				
		More talkative than usual or pressure to keep talking	man_mi3			man_mi3 = man_mi3_1 + man_mi3_2 + man_mi3_bias	Censoring of the sum of multiple input variables
				More talkative than usual	man_mi3_1		
				Pressure to keep talking	man_mi3_2		
				Information of the domain not explained by the input variables	man_mi3_bias		
		Flight of ideas or subjective experience that thoughts are racing	man_mi4			man_mi4 = man_mi4_1 + man_mi4_2 + man_mi4_bias	Censoring of the sum of multiple input variables
				Flight of ideas	man_mi4_1		
				Subjective experience that thoughts are racing	man_mi4_2		
				Information of the domain not explained by the input variables	man_mi4_bias		
		Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)	man_mi5				
		Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation	man_mi6			man_mi6 = man_mi6_1 + man_mi6_2 + man_mi6_bias	Censoring of the sum of multiple input variables
				Increase in goal-directed activity	man_mi6_1		
				Psychomotor agitation	man_mi6_2		
				Information of the domain not explained by the input variables	man_mi6_bias		
		Excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)	man_mi7				
Information of diagnosis due to categorization (choosing at least three symptoms)	man_bias1						Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis due to categorization (choosing at least four symptoms)	man_bias2						Bias introduced by categorizing the number of input symptoms confirmed in the minor criteria
Information of diagnosis not	man_bias						Information of the diagnosis not explained by the input symptoms and

explained by symptoms	the bias variables generated due to data categorization, man_bias1 and man_bias2

Table 5. The derived prevalence rates of the diagnoses of major depressive episodes, dysthymic disorder, and manic episodes based on the assumed prevalence rates and between-variable correlations of the input symptoms

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0 (95% CI = 0 to 0)	0.004 (95% CI = 0.004 to 0.004)	0 (95% CI = 0 to 0)
0	0.1	0.001 (95% CI = 0.001 to 0.001)	0.025 (95% CI = 0.025 to 0.025)	0.002 (95% CI = 0.002 to 0.002)
0	0.3	0.067 (95% CI = 0.067 to 0.067)	0.249 (95% CI = 0.249 to 0.249)	0.136 (95% CI = 0.135 to 0.136)
0	0.5	0.245 (95% CI = 0.244 to 0.245)	0.493 (95% CI = 0.493 to 0.493)	0.436 (95% CI = 0.436 to 0.436)
0	0.7	0.49 (95% CI = 0.49 to 0.49)	0.7 (95% CI = 0.7 to 0.7)	0.692 (95% CI = 0.692 to 0.693)
0.1	0.05	0.004 (95% CI = 0.004 to 0.004)	0.018 (95% CI = 0.018 to 0.018)	0.007 (95% CI = 0.007 to 0.007)
0.1	0.1	0.011 (95% CI = 0.011 to 0.011)	0.049 (95% CI = 0.049 to 0.049)	0.022 (95% CI = 0.021 to 0.022)
0.1	0.3	0.094 (95% CI = 0.094 to 0.094)	0.25 (95% CI = 0.25 to 0.25)	0.172 (95% CI = 0.171 to 0.172)
0.1	0.5	0.267 (95% CI = 0.267 to 0.268)	0.482 (95% CI = 0.482 to 0.482)	0.425 (95% CI = 0.425 to 0.425)
0.1	0.7	0.51 (95% CI = 0.509 to 0.51)	0.697 (95% CI = 0.697 to 0.697)	0.679 (95% CI = 0.679 to 0.679)
0.4	0.05	0.019 (95% CI = 0.019 to 0.019)	0.037 (95% CI = 0.037 to 0.037)	0.029 (95% CI = 0.029 to 0.029)
0.4	0.1	0.042 (95% CI = 0.042 to 0.042)	0.078 (95% CI = 0.078 to 0.078)	0.062 (95% CI = 0.062 to 0.062)
0.4	0.3	0.166 (95% CI = 0.166 to 0.167)	0.267 (95% CI = 0.267 to 0.267)	0.231 (95% CI = 0.231 to 0.231)
0.4	0.5	0.344 (95% CI = 0.344 to 0.344)	0.476 (95% CI = 0.476 to 0.476)	0.44 (95% CI = 0.44 to 0.441)
0.4	0.7	0.57 (95% CI = 0.57 to 0.57)	0.689 (95% CI = 0.688 to 0.689)	0.666 (95% CI = 0.666 to 0.666)
0.7	0.05	0.035 (95% CI = 0.035 to 0.035)	0.046 (95% CI = 0.046 to 0.046)	0.042 (95% CI = 0.042 to 0.042)
0.7	0.1	0.071 (95% CI = 0.071 to 0.071)	0.092 (95% CI = 0.092 to 0.092)	0.085 (95% CI = 0.085 to 0.085)
0.7	0.3	0.233 (95% CI = 0.233 to 0.234)	0.285 (95% CI = 0.285 to 0.285)	0.27 (95% CI = 0.27 to 0.27)
0.7	0.5	0.422 (95% CI = 0.421 to 0.422)	0.486 (95% CI = 0.485 to 0.486)	0.469 (95% CI = 0.468 to 0.469)
0.7	0.7	0.635 (95% CI = 0.635 to 0.635)	0.69 (95% CI = 0.69 to 0.691)	0.678 (95% CI = 0.677 to 0.678)
0.9	0.05	0.042 (95% CI = 0.042 to 0.042)	0.048 (95% CI = 0.048 to 0.048)	0.046 (95% CI = 0.046 to 0.046)
0.9	0.1	0.085 (95% CI = 0.085 to 0.085)	0.096 (95% CI = 0.096 to 0.097)	0.093 (95% CI = 0.093 to 0.093)
0.9	0.3	0.268 (95% CI = 0.268 to 0.268)	0.293 (95% CI = 0.293 to 0.293)	0.286 (95% CI = 0.286 to 0.287)
0.9	0.5	0.463 (95% CI = 0.463 to 0.463)	0.493 (95% CI = 0.492 to 0.493)	0.485 (95% CI = 0.485 to 0.486)
0.9	0.7	0.669 (95% CI = 0.669 to 0.669)	0.695 (95% CI = 0.694 to 0.695)	0.688 (95% CI = 0.688 to 0.688)

Table 6. The individual input symptoms that best explained the diagnoses: major depressive episodes, dysthymic disorder, and manic episodes

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	mde_ma1	dys_ma	man_ma3
0	0.05	0.001 (95% CI = 0.001 to 0.001)	0.076 (95% CI = 0.075 to 0.077)	0.002 (95% CI = 0.002 to 0.002)
0	0.1	mde_ma1	dys_ma	man_ma3
0	0.1	0.01 (95% CI = 0.01 to 0.01)	0.228 (95% CI = 0.227 to 0.229)	0.021 (95% CI = 0.02 to 0.021)
0	0.3	mde_ma1	dys_ma	man_ma3
0	0.3	0.167 (95% CI = 0.167 to 0.167)	0.774 (95% CI = 0.773 to 0.774)	0.366 (95% CI = 0.366 to 0.367)
0	0.5	mde_ma2	dys_ma	man_ma3
0	0.5	0.324 (95% CI = 0.324 to 0.325)	0.971 (95% CI = 0.971 to 0.971)	0.773 (95% CI = 0.772 to 0.773)
0	0.7	mde_ma2	dys_ma	man_ma3
0	0.7	0.412 (95% CI = 0.412 to 0.412)	0.999 (95% CI = 0.999 to 0.999)	0.964 (95% CI = 0.964 to 0.964)
0.1	0.05	mde_ma2	dys_ma	man_ma3
0.1	0.05	0.07 (95% CI = 0.07 to 0.071)	0.353 (95% CI = 0.352 to 0.355)	0.136 (95% CI = 0.135 to 0.137)
0.1	0.1	mde_ma1	dys_ma	man_ma3
0.1	0.1	0.101 (95% CI = 0.1 to 0.101)	0.462 (95% CI = 0.461 to 0.463)	0.199 (95% CI = 0.198 to 0.199)
0.1	0.3	mde_ma2	dys_ma	man_ma3
0.1	0.3	0.242 (95% CI = 0.242 to 0.243)	0.777 (95% CI = 0.777 to 0.778)	0.483 (95% CI = 0.483 to 0.484)
0.1	0.5	mde_ma2	dys_ma	man_ma3
0.1	0.5	0.365 (95% CI = 0.365 to 0.366)	0.932 (95% CI = 0.931 to 0.932)	0.74 (95% CI = 0.74 to 0.741)
0.1	0.7	mde_ma2	dys_ma	man_ma3
0.1	0.7	0.445 (95% CI = 0.445 to 0.446)	0.986 (95% CI = 0.986 to 0.986)	0.906 (95% CI = 0.906 to 0.907)
0.4	0.05	mde_ma1	dys_ma	man_ma3
0.4	0.05	0.375 (95% CI = 0.373 to 0.376)	0.731 (95% CI = 0.729 to 0.732)	0.561 (95% CI = 0.559 to 0.562)
0.4	0.1	mde_ma1	dys_ma	man_ma3
0.4	0.1	0.395 (95% CI = 0.394 to 0.396)	0.763 (95% CI = 0.762 to 0.764)	0.595 (95% CI = 0.594 to 0.596)
0.4	0.3	mde_ma1	dys_ma	man_ma3
0.4	0.3	0.465 (95% CI = 0.465 to 0.466)	0.851 (95% CI = 0.85 to 0.851)	0.701 (95% CI = 0.701 to 0.702)
0.4	0.5	mde_ma2	dys_ma	man_ma3
0.4	0.5	0.525 (95% CI = 0.524 to 0.525)	0.908 (95% CI = 0.908 to 0.908)	0.787 (95% CI = 0.786 to 0.787)
0.4	0.7	mde_ma2	dys_ma	man_ma3
0.4	0.7	0.568 (95% CI = 0.568 to 0.569)	0.946 (95% CI = 0.946 to 0.947)	0.855 (95% CI = 0.854 to 0.855)
0.7	0.05	mde_ma2	dys_ma	man_ma3
0.7	0.05	0.688 (95% CI = 0.687 to 0.69)	0.909 (95% CI = 0.908 to 0.909)	0.831 (95% CI = 0.83 to 0.832)
0.7	0.1	mde_ma1	dys_ma	man_ma3
0.7	0.1	0.688 (95% CI = 0.687 to 0.689)	0.912 (95% CI = 0.911 to 0.913)	0.836 (95% CI = 0.835 to 0.836)
0.7	0.3	mde_ma2	dys_ma	man_ma3
0.7	0.3	0.71 (95% CI = 0.709 to 0.711)	0.93 (95% CI = 0.93 to 0.93)	0.862 (95% CI = 0.861 to 0.862)
0.7	0.5	mde_ma2	dys_ma	man_ma3
0.7	0.5	0.729 (95% CI = 0.728 to 0.729)	0.944 (95% CI = 0.943 to 0.944)	0.882 (95% CI = 0.882 to 0.883)
0.7	0.7	mde_ma1	dys_ma	man_ma3
0.7	0.7	0.745 (95% CI = 0.744 to 0.745)	0.954 (95% CI = 0.954 to 0.955)	0.9 (95% CI = 0.9 to 0.9)
0.9	0.05	mde_ma1	dys_ma	man_ma3
0.9	0.05	0.828 (95% CI = 0.827 to 0.829)	0.958 (95% CI = 0.957 to 0.958)	0.918 (95% CI = 0.917 to 0.919)
0.9	0.1	mde_ma2	dys_ma	man_ma3
0.9	0.1	0.838 (95% CI = 0.838 to 0.839)	0.961 (95% CI = 0.961 to 0.961)	0.925 (95% CI = 0.924 to 0.925)
0.9	0.3	mde_ma2	dys_ma	man_ma3
0.9	0.3	0.856 (95% CI = 0.856 to 0.857)	0.969 (95% CI = 0.968 to 0.969)	0.937 (95% CI = 0.936 to 0.937)
0.9	0.5	mde_ma2	dys_ma	man_ma3
0.9	0.5	0.862 (95% CI = 0.862 to 0.863)	0.972 (95% CI = 0.972 to 0.972)	0.942 (95% CI = 0.942 to 0.943)
0.9	0.7	mde_ma2	dys_ma	man_ma3
0.9	0.7	0.865 (95% CI = 0.865 to 0.866)	0.974 (95% CI = 0.974 to 0.974)	0.946 (95% CI = 0.946 to 0.946)

Table 7. The individual bias variables that best explained the diagnoses: major depressive episodes, dysthymic disorder, and manic episodes

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	mde_bias2	dys_bias	man_bias2
0	0.05	0 (95% CI = 0 to 0)	0.028 (95% CI = 0.028 to 0.028)	0.001 (95% CI = 0.001 to 0.001)
0	0.1	mde_bias2	dys_bias	man_bias2
0	0.1	0.004 (95% CI = 0.004 to 0.004)	0.053 (95% CI = 0.053 to 0.054)	0.011 (95% CI = 0.011 to 0.011)
0	0.3	mde_bias2	dys_bias	man_bias1
0	0.3	0.015 (95% CI = 0.015 to 0.015)	0.045 (95% CI = 0.045 to 0.045)	0.089 (95% CI = 0.089 to 0.09)
0	0.5	mde_bias	dys_bias	man_bias1
0	0.5	0.013 (95% CI = 0.013 to 0.014)	0.007 (95% CI = 0.007 to 0.007)	0.035 (95% CI = 0.034 to 0.035)
0	0.7	mde_bias	dys_bias	man_bias1
0	0.7	0.01 (95% CI = 0.01 to 0.01)	0 (95% CI = 0 to 0)	0.002 (95% CI = 0.002 to 0.002)
0.1	0.05	mde_bias2	dys_bias	man_bias1
0.1	0.05	0.037 (95% CI = 0.037 to 0.037)	0.113 (95% CI = 0.113 to 0.114)	0.083 (95% CI = 0.083 to 0.084)
0.1	0.1	mde_bias2	dys_bias	man_bias1
0.1	0.1	0.047 (95% CI = 0.047 to 0.048)	0.122 (95% CI = 0.121 to 0.122)	0.116 (95% CI = 0.115 to 0.116)
0.1	0.3	mde_bias2	dys_mi_bias	man_bias1
0.1	0.3	0.077 (95% CI = 0.077 to 0.077)	0.105 (95% CI = 0.105 to 0.106)	0.198 (95% CI = 0.197 to 0.198)
0.1	0.5	mde_bias2	dys_mi_bias	man_bias1
0.1	0.5	0.079 (95% CI = 0.079 to 0.08)	0.073 (95% CI = 0.073 to 0.073)	0.166 (95% CI = 0.166 to 0.167)
0.1	0.7	mde_bias2	dys_mi_bias	man_bias1
0.1	0.7	0.065 (95% CI = 0.065 to 0.065)	0.047 (95% CI = 0.046 to 0.047)	0.094 (95% CI = 0.093 to 0.094)
0.4	0.05	mde_bias1	dys_mi_bias	man_bias1
0.4	0.05	0.294 (95% CI = 0.293 to 0.295)	0.415 (95% CI = 0.413 to 0.416)	0.432 (95% CI = 0.431 to 0.433)
0.4	0.1	mde_bias1	dys_mi_bias	man_bias1
0.4	0.1	0.304 (95% CI = 0.303 to 0.304)	0.419 (95% CI = 0.418 to 0.42)	0.445 (95% CI = 0.444 to 0.445)
0.4	0.3	mde_bias1	dys_mi_bias	man_bias1
0.4	0.3	0.335 (95% CI = 0.334 to 0.335)	0.411 (95% CI = 0.411 to 0.412)	0.473 (95% CI = 0.472 to 0.473)
0.4	0.5	mde_bias1	dys_mi_bias	man_bias1
0.4	0.5	0.354 (95% CI = 0.354 to 0.355)	0.395 (95% CI = 0.395 to 0.396)	0.475 (95% CI = 0.474 to 0.475)
0.4	0.7	mde_bias1	dys_mi_bias	man_bias1
0.4	0.7	0.356 (95% CI = 0.355 to 0.356)	0.367 (95% CI = 0.366 to 0.367)	0.451 (95% CI = 0.45 to 0.451)
0.7	0.05	mde_bias1	dys_mi_bias	man_bias1
0.7	0.05	0.616 (95% CI = 0.615 to 0.617)	0.705 (95% CI = 0.704 to 0.706)	0.723 (95% CI = 0.722 to 0.724)
0.7	0.1	mde_bias1	dys_mi_bias	man_bias1
0.7	0.1	0.611 (95% CI = 0.611 to 0.612)	0.699 (95% CI = 0.698 to 0.699)	0.72 (95% CI = 0.72 to 0.721)
0.7	0.3	mde_bias1	dys_mi_bias	man_bias1
0.7	0.3	0.623 (95% CI = 0.623 to 0.624)	0.699 (95% CI = 0.699 to 0.7)	0.728 (95% CI = 0.728 to 0.729)
0.7	0.5	mde_bias1	dys_mi_bias	man_bias1
0.7	0.5	0.632 (95% CI = 0.632 to 0.633)	0.696 (95% CI = 0.696 to 0.697)	0.731 (95% CI = 0.731 to 0.732)
0.7	0.7	mde_bias1	dys_mi_bias	man_bias1
0.7	0.7	0.639 (95% CI = 0.638 to 0.639)	0.693 (95% CI = 0.692 to 0.693)	0.732 (95% CI = 0.731 to 0.732)
0.9	0.05	mde_bias1	dys_mi_bias	man_bias1
0.9	0.05	0.777 (95% CI = 0.776 to 0.778)	0.835 (95% CI = 0.834 to 0.835)	0.847 (95% CI = 0.847 to 0.848)
0.9	0.1	mde_bias1	dys_mi_bias	man_bias1
0.9	0.1	0.788 (95% CI = 0.788 to 0.789)	0.842 (95% CI = 0.841 to 0.843)	0.855 (95% CI = 0.854 to 0.855)
0.9	0.3	mde_bias1	dys_mi_bias	man_bias1
0.9	0.3	0.807 (95% CI = 0.806 to 0.807)	0.854 (95% CI = 0.853 to 0.854)	0.867 (95% CI = 0.867 to 0.868)
0.9	0.5	mde_bias1	dys_mi_bias	man_bias1
0.9	0.5	0.811 (95% CI = 0.811 to 0.811)	0.855 (95% CI = 0.855 to 0.856)	0.87 (95% CI = 0.87 to 0.871)
0.9	0.7	mde_bias1	dys_mi_bias	man_bias1
0.9	0.7	0.812 (95% CI = 0.811 to 0.812)	0.853 (95% CI = 0.853 to 0.853)	0.869 (95% CI = 0.869 to 0.87)

Table 8. Approximating the diagnoses using input symptoms and derived adjusted R-squared

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0.003 (95% CI = 0.002 to 0.003)	0.122 (95% CI = 0.121 to 0.123)	0.004 (95% CI = 0.004 to 0.005)
0	0.1	0.024 (95% CI = 0.023 to 0.024)	0.305 (95% CI = 0.304 to 0.306)	0.039 (95% CI = 0.038 to 0.039)
0	0.3	0.348 (95% CI = 0.348 to 0.349)	0.842 (95% CI = 0.841 to 0.842)	0.483 (95% CI = 0.482 to 0.483)
0	0.5	0.649 (95% CI = 0.649 to 0.649)	0.986 (95% CI = 0.986 to 0.986)	0.817 (95% CI = 0.817 to 0.817)
0	0.7	0.823 (95% CI = 0.823 to 0.823)	1 (95% CI = 1 to 1)	0.967 (95% CI = 0.967 to 0.967)
0.1	0.05	0.143 (95% CI = 0.141 to 0.144)	0.435 (95% CI = 0.433 to 0.436)	0.212 (95% CI = 0.211 to 0.213)
0.1	0.1	0.198 (95% CI = 0.197 to 0.199)	0.539 (95% CI = 0.538 to 0.54)	0.29 (95% CI = 0.289 to 0.291)
0.1	0.3	0.45 (95% CI = 0.45 to 0.451)	0.826 (95% CI = 0.826 to 0.827)	0.588 (95% CI = 0.588 to 0.589)
0.1	0.5	0.663 (95% CI = 0.663 to 0.664)	0.952 (95% CI = 0.952 to 0.952)	0.799 (95% CI = 0.799 to 0.799)
0.1	0.7	0.809 (95% CI = 0.809 to 0.809)	0.991 (95% CI = 0.991 to 0.991)	0.922 (95% CI = 0.922 to 0.922)
0.4	0.05	0.587 (95% CI = 0.585 to 0.588)	0.782 (95% CI = 0.781 to 0.783)	0.675 (95% CI = 0.674 to 0.676)
0.4	0.1	0.607 (95% CI = 0.606 to 0.608)	0.807 (95% CI = 0.807 to 0.808)	0.698 (95% CI = 0.697 to 0.698)
0.4	0.3	0.688 (95% CI = 0.688 to 0.689)	0.878 (95% CI = 0.877 to 0.878)	0.775 (95% CI = 0.774 to 0.775)
0.4	0.5	0.761 (95% CI = 0.761 to 0.762)	0.925 (95% CI = 0.924 to 0.925)	0.838 (95% CI = 0.838 to 0.838)
0.4	0.7	0.821 (95% CI = 0.821 to 0.822)	0.956 (95% CI = 0.956 to 0.956)	0.887 (95% CI = 0.887 to 0.888)
0.7	0.05	0.813 (95% CI = 0.812 to 0.814)	0.925 (95% CI = 0.925 to 0.926)	0.877 (95% CI = 0.877 to 0.878)
0.7	0.1	0.826 (95% CI = 0.826 to 0.827)	0.928 (95% CI = 0.927 to 0.928)	0.881 (95% CI = 0.881 to 0.882)
0.7	0.3	0.86 (95% CI = 0.86 to 0.86)	0.942 (95% CI = 0.942 to 0.942)	0.9 (95% CI = 0.9 to 0.9)
0.7	0.5	0.88 (95% CI = 0.88 to 0.88)	0.953 (95% CI = 0.953 to 0.953)	0.913 (95% CI = 0.913 to 0.913)
0.7	0.7	0.895 (95% CI = 0.895 to 0.895)	0.962 (95% CI = 0.962 to 0.962)	0.925 (95% CI = 0.925 to 0.925)
0.9	0.05	0.903 (95% CI = 0.903 to 0.904)	0.965 (95% CI = 0.965 to 0.966)	0.941 (95% CI = 0.94 to 0.941)
0.9	0.1	0.91 (95% CI = 0.91 to 0.911)	0.968 (95% CI = 0.968 to 0.968)	0.945 (95% CI = 0.945 to 0.945)
0.9	0.3	0.923 (95% CI = 0.923 to 0.923)	0.974 (95% CI = 0.974 to 0.974)	0.954 (95% CI = 0.953 to 0.954)
0.9	0.5	0.928 (95% CI = 0.928 to 0.928)	0.976 (95% CI = 0.976 to 0.977)	0.958 (95% CI = 0.957 to 0.958)
0.9	0.7	0.932 (95% CI = 0.932 to 0.932)	0.978 (95% CI = 0.978 to 0.978)	0.96 (95% CI = 0.96 to 0.96)

617 Table 9. Approximating the diagnoses using bias variables and derived R-squared

Assumed correlations between input symptoms	Assumed prevalence of input symptoms	Major depressive episodes	Dysthymic disorder	Manic episodes
0	0.05	0.003 (95% CI = 0.002 to 0.003)	0.029 (95% CI = 0.029 to 0.03)	0.004 (95% CI = 0.004 to 0.004)
0	0.1	0.013 (95% CI = 0.012 to 0.013)	0.056 (95% CI = 0.056 to 0.056)	0.017 (95% CI = 0.017 to 0.017)
0	0.3	0.083 (95% CI = 0.083 to 0.083)	0.047 (95% CI = 0.047 to 0.047)	0.098 (95% CI = 0.098 to 0.099)
0	0.5	0.111 (95% CI = 0.111 to 0.112)	0.007 (95% CI = 0.007 to 0.007)	0.039 (95% CI = 0.038 to 0.039)
0	0.7	0.095 (95% CI = 0.095 to 0.095)	0 (95% CI = 0 to 0)	0.012 (95% CI = 0.012 to 0.013)
0.1	0.05	0.083 (95% CI = 0.082 to 0.084)	0.145 (95% CI = 0.144 to 0.146)	0.126 (95% CI = 0.125 to 0.127)
0.1	0.1	0.096 (95% CI = 0.095 to 0.097)	0.156 (95% CI = 0.155 to 0.156)	0.154 (95% CI = 0.153 to 0.154)
0.1	0.3	0.145 (95% CI = 0.144 to 0.145)	0.139 (95% CI = 0.138 to 0.139)	0.216 (95% CI = 0.216 to 0.216)
0.1	0.5	0.172 (95% CI = 0.172 to 0.173)	0.097 (95% CI = 0.097 to 0.097)	0.182 (95% CI = 0.181 to 0.182)
0.1	0.7	0.175 (95% CI = 0.175 to 0.175)	0.065 (95% CI = 0.064 to 0.065)	0.115 (95% CI = 0.115 to 0.116)
0.4	0.05	0.421 (95% CI = 0.419 to 0.423)	0.455 (95% CI = 0.453 to 0.456)	0.505 (95% CI = 0.504 to 0.506)
0.4	0.1	0.422 (95% CI = 0.421 to 0.423)	0.454 (95% CI = 0.453 to 0.455)	0.507 (95% CI = 0.506 to 0.508)
0.4	0.3	0.435 (95% CI = 0.434 to 0.435)	0.442 (95% CI = 0.442 to 0.443)	0.512 (95% CI = 0.512 to 0.513)
0.4	0.5	0.452 (95% CI = 0.452 to 0.453)	0.427 (95% CI = 0.427 to 0.427)	0.506 (95% CI = 0.505 to 0.506)
0.4	0.7	0.46 (95% CI = 0.459 to 0.46)	0.403 (95% CI = 0.402 to 0.403)	0.481 (95% CI = 0.481 to 0.482)
0.7	0.05	0.728 (95% CI = 0.727 to 0.729)	0.729 (95% CI = 0.728 to 0.731)	0.764 (95% CI = 0.763 to 0.765)
0.7	0.1	0.722 (95% CI = 0.721 to 0.723)	0.723 (95% CI = 0.722 to 0.724)	0.76 (95% CI = 0.759 to 0.761)
0.7	0.3	0.726 (95% CI = 0.726 to 0.727)	0.722 (95% CI = 0.722 to 0.723)	0.761 (95% CI = 0.761 to 0.762)
0.7	0.5	0.732 (95% CI = 0.731 to 0.732)	0.72 (95% CI = 0.719 to 0.72)	0.76 (95% CI = 0.76 to 0.761)
0.7	0.7	0.737 (95% CI = 0.736 to 0.737)	0.717 (95% CI = 0.716 to 0.717)	0.758 (95% CI = 0.758 to 0.759)
0.9	0.05	0.852 (95% CI = 0.851 to 0.853)	0.85 (95% CI = 0.849 to 0.851)	0.871 (95% CI = 0.871 to 0.872)
0.9	0.1	0.86 (95% CI = 0.859 to 0.861)	0.857 (95% CI = 0.856 to 0.857)	0.876 (95% CI = 0.876 to 0.877)
0.9	0.3	0.872 (95% CI = 0.871 to 0.872)	0.867 (95% CI = 0.867 to 0.868)	0.886 (95% CI = 0.886 to 0.886)
0.9	0.5	0.874 (95% CI = 0.874 to 0.875)	0.869 (95% CI = 0.868 to 0.869)	0.888 (95% CI = 0.887 to 0.888)
0.9	0.7	0.874 (95% CI = 0.874 to 0.875)	0.867 (95% CI = 0.866 to 0.867)	0.886 (95% CI = 0.886 to 0.886)

Figure 1. The prevalence rates of an intermediate variable for the diagnosis of major depressive episodes.

Note: The intermediate variable is "*significant unintentional weight loss or gain*" and the input symptoms are "*significant unintentional weight loss*" and "*significant unintentional weight gain.*" The black line represents the situation where the prevalence rates of the input symptoms are the same as that of the intermediate variable. Lines above the black lines have prevalence rates larger than those of the input symptoms.

Figure 2. The prevalence rates of dysthymic disorder.

Note: Dysthymic disorder is diagnosed when both the major (depressed mood most of the day for more days than not, for at least 2 years) and minor criteria (at least two of the six items) are confirmed. The black line represents the situation where the prevalence rates of the input symptoms are the same as those of the intermediate variable. Lines below the black lines have prevalence rates lower than those of the input symptoms.

Figure 3. The prevalence rates of major depressive episodes.

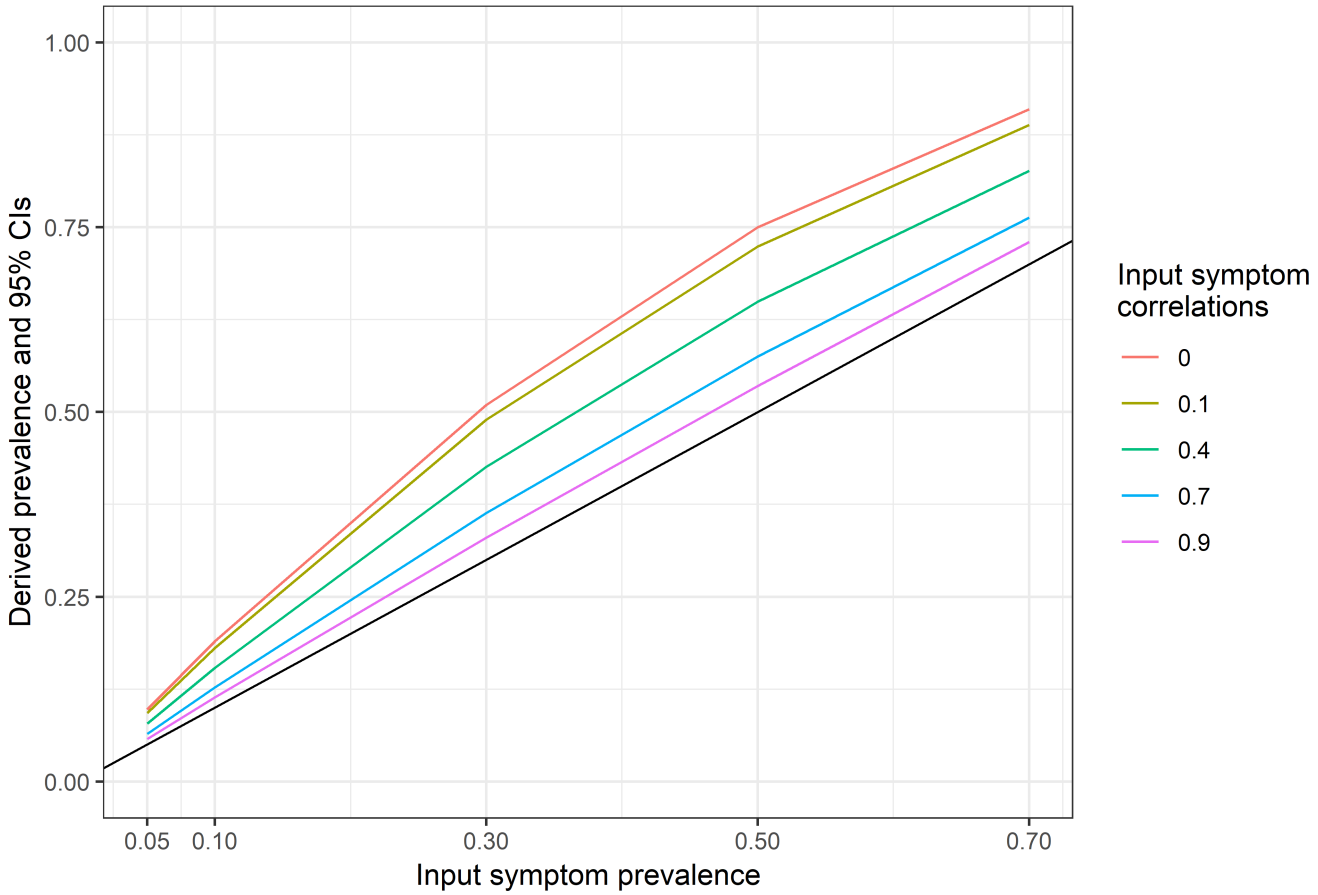
Note: Major depressive episodes are diagnosed when both major and minor criteria are confirmed. The black line represents the situation where the prevalence rates of the input symptoms are the same as that of the intermediate variable. Lines below the black lines have prevalence rates lower than those of the input symptoms.

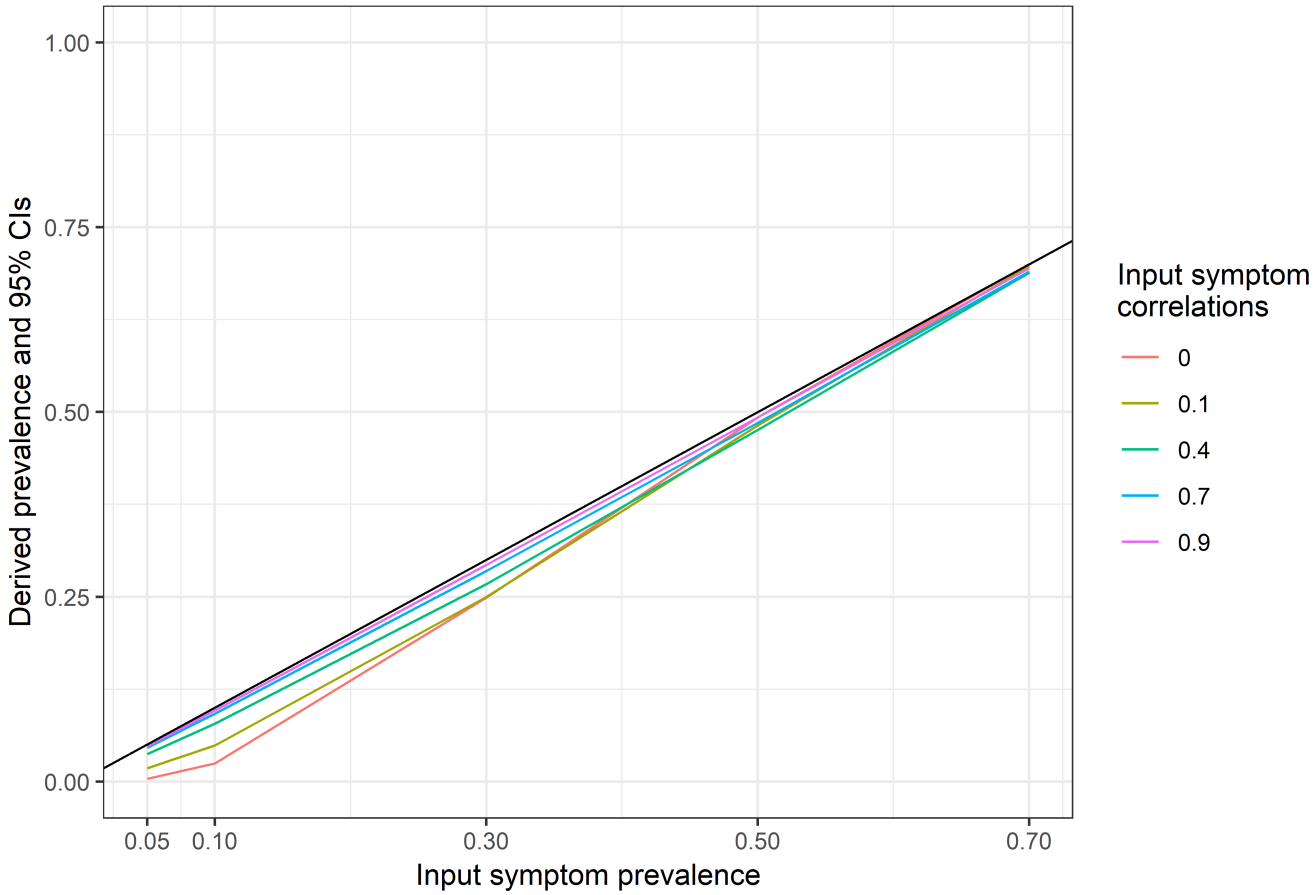
Figure 4. The prevalence rates of manic episodes

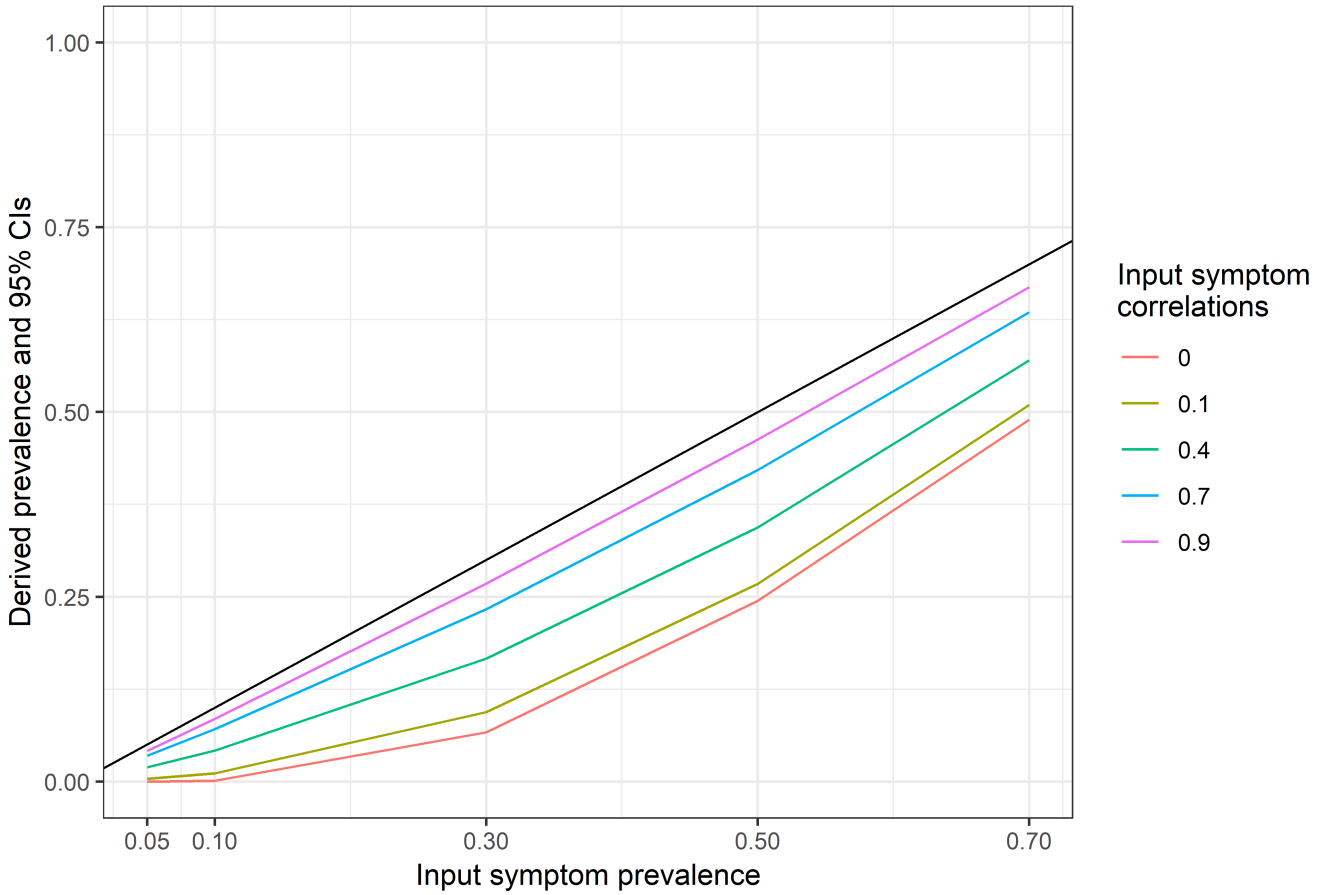
Note: Manic episodes are diagnosed when the symptoms present as described in the diagnostic manual. The black line represents the situation where the prevalence rates of the input symptoms are the same as those of the input symptoms. Lines below the black lines have prevalence rates lower than those of the input symptoms.

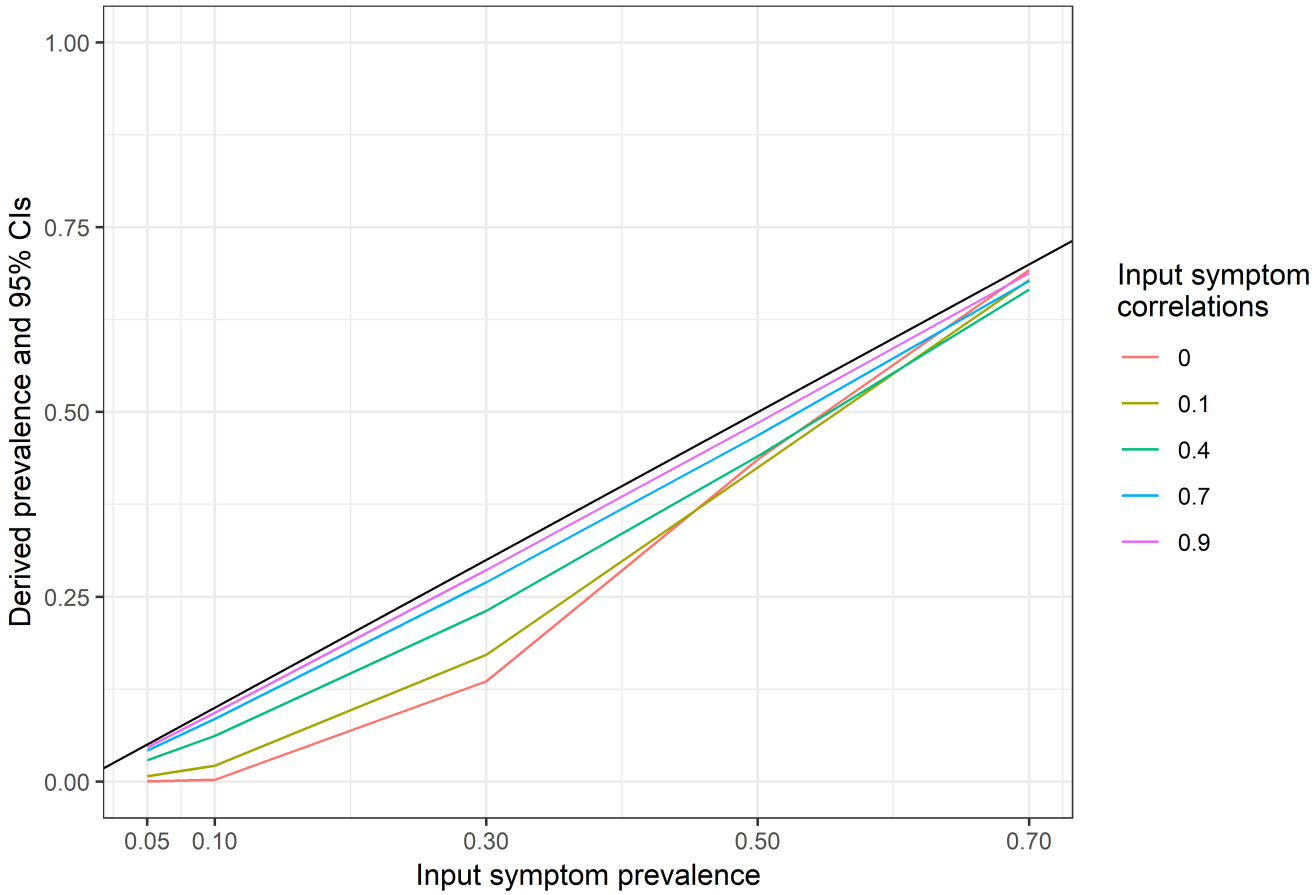
Figure 5. The approximation of the diagnosis of dysthymic disorder by the input symptoms, the bias variables, and both, measured by R-squared

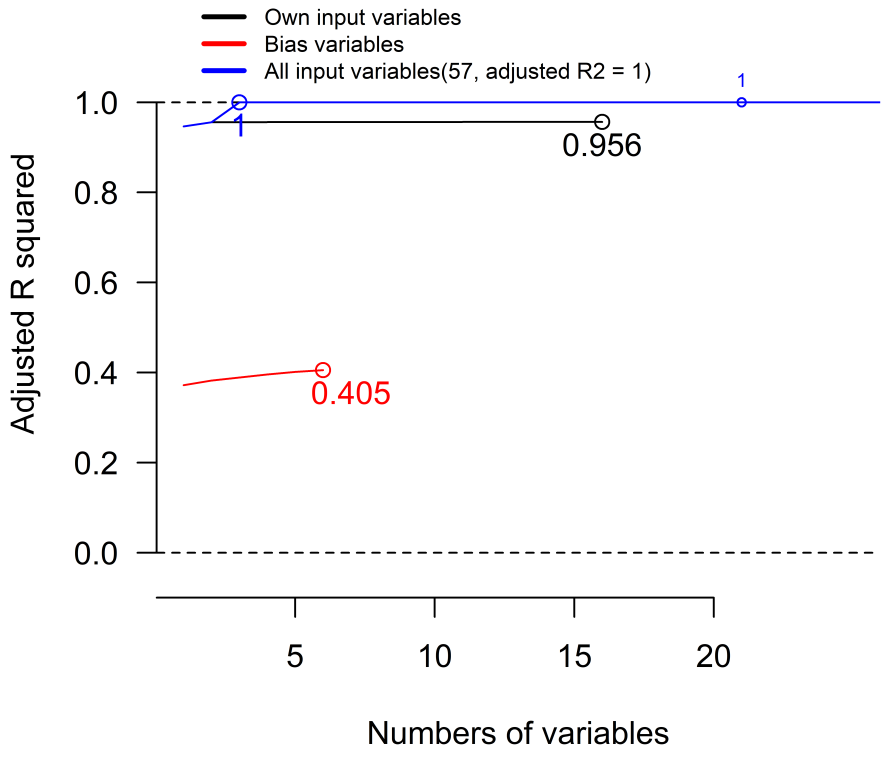
Note: the diagnosis of dysthymic disorder is approximated by the input symptoms, the bias variables, and both using forward-stepwise regression. The selection of the variables was determined by adjusted R-squared. See Table 4 for the details in the input symptoms and the bias variables. The assumed correlations between the input symptoms are 0.4 and the assumed prevalence rates of the input symptoms are 0.7.


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


1.00
0.75
0.50
0.25
0.00
Derived prevalence and 95% CIs
0.05
0.10
0.30
0.50
0.70
Input symptom prevalence
Input symptom correlations
0
0.1
0.4
0.7
0.9


Own input variables
Bias variables
All input variables(57, adjusted R2 = 1)
Adjusted R squared
1.0
0.8
0.6
0.4
0.2
0.0
1
1
0.956
0.405
5
10
15
20
Numbers of variables

A simulation study to demonstrate the biases in the diagnoses of mental illnesses: major depressive episodes, dysthymia, and manic episodes

Yi-Sheng Chao[1*], Kuan-Fu Lin,[2] Chao-Jung Wu[3], Hsing-Chien Wu[4], Hui-Ting Hsu[5], Lien-Cheng Tsao[5], Yen-Po Cheng[5], Yi-Chun Lai[6], Wei-Chih Chen[7,8]

[1]Independent researcher, Montréal, H2X 0A8 Canada, [2]National Taiwan University Hospital Yun-Lin Branch, Yunlin County, 640 Taiwan, [3]Département d'informatique Université du Québec à Montréal, Montréal H3B 1B4 Canada, [4]Taipei Hospital Ministry of Health and Welfare New Taipei city, 242 Taiwan, [5]Changhua Christian Hospital, Changhua County 526, Taiwan, [6]National Yang-Ming University Hospital, Yilan 260 Taiwan, [7]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 112, Taiwan, [8]Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei 112, Taiwan

[*]chaoyisheng@post.harvard.edu

Keywords: Frailty; bias; forward-stepwise regression; the Health and Retirement Study; index mining

Mental illness	Illness variable	Classification	Criterion var	Major or min	Domain	Symptoms	Symptom var	Equations	Approximation	Assumed pre	Derived prev	definition	variable	R	formula	code	outcome	suboutcome
Major Depre	mde							mde = mde_nmde + intercept + coef1 x mde_ma1 + coef2 x mde_ma2 + coef3 x				Major Depre	mde		mde = mde_nsimSmde + simde			
		Major criteria, essential for diagnosis										Major criteria						
			Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.									Depressed mood or a loss of interest or pleasure in daily activities for more than two weeks.						
			Depressed m	mde_ma1					+ rbinom(n=	0.4)		Depressed m	mde_ma1		= nsimSmde_mamde			
			Loss of interes	mde_ma2					+ rbinom(n=	0.4)		Loss of interes	mde_ma2		= nsimSmde_mamde			
		Minor criteria	mde_mi									Minor criteria (at least 5 of the symptoms including the two in major criteria)						
			Significant un	mde_mi3				mde_mi3 = mde_mi3_1 + mde_mi3_2 + mde_mi3_bias				Significant un	mde_mi3		= 1*((mde_mi3_mde		mde_mi3	
					Significant un	mde_mi3_1			+ rbinom(n=	0.4)		Significant un	mde_mi3_1		= nsimSmde_mimde		mde_mi3	
					Significant un	mde_mi3_2			+ rbinom(n=	0.4)		Significant un	mde_mi3_2		= nsimSmde_mimde		mde_mi3	
															simSmde_mi3 = 1*((simSmde_mi3_1 + simSmde_mi3_2) > 0)			
			Information	mde_mi3_bias								Information	mde_mi3_bias		= nsimSmde_mimde		mde_mi3	
			Insomnia or	mde_mi4				mde_mi4 = mde_mi4_1 + mde_mi4_2 + mde_mi4_bias				Insomnia or	mde_mi4		= 1*((mde_mi4_mde		mde_mi4	
					Insomnia	mde_mi4_1			+ rbinom(n=	0.4)		Insomnia	mde_mi4_1		= nsimSmde_mimde		mde_mi4	
					Sleeping too	mde_mi4_2			+ rbinom(n=	0.4)		Sleeping too	mde_mi4_2		= nsimSmde_mimde		mde_mi4	
															simSmde_mi4 = 1*((simSmde_mi4_1 + simSmde_mi4_2) > 0)			
			Information	mde_mi4_bias								Information	mde_mi4_bias		= nsimSmde_mimde		mde_mi4	
			Agitation or	mde_mi5				mde_mi5 = mde_mi5_1 + mde_mi5_2 + mde_mi5_bias				Agitation or	mde_mi5		= 1*((mde_mi5_mde		mde_mi5	
					Agitation	mde_mi5_1			+ rbinom(n=	0.4)		Agitation	mde_mi5_1		= nsimSmde_mimde		mde_mi5	
					Psychomotor	mde_mi5_2			+ rbinom(n=	0.4)		Psychomotor	mde_mi5_2		= nsimSmde_mimde		mde_mi5	
															simSmde_mi5 = 1*((simSmde_mi5_1 + simSmde_mi5_2) > 0)			
			Information	mde_mi5_bias								Information	mde_mi5_bias		= nsimSmde_mimde		mde_mi5	
			Fatigue or lo	mde_mi6				mde_mi6 = mde_mi6_1 + mde_mi6_2 + mde_mi6_bias				Fatigue or lo	mde_mi6		= 1*((mde_mi6_mde		mde_mi6	
					Fatigue	mde_mi6_1			+ rbinom(n=	0.4)		Fatigue	mde_mi6_1		= nsimSmde_mimde		mde_mi6	
					Loss of energ	mde_mi6_2			+ rbinom(n=	0.4)		Loss of energ	mde_mi6_2		= nsimSmde_mimde		mde_mi6	
															simSmde_mi6 = 1*((simSmde_mi6_1 + simSmde_mi6_2) > 0)			
			Information	mde_mi6_bias								Information	mde_mi6_bias		= nsimSmde_mimde		mde_mi6	
			Feelings of w	mde_mi7				mde_mi7 = mde_mi7_1 + mde_mi7_2 + mde_mi7_bias				Feelings of w	mde_mi7		= 1*((mde_mi7_mde		mde_mi7	
					Feelings of w	mde_mi7_1			+ rbinom(n=	0.4)		Feelings of w	mde_mi7_1		= nsimSmde_mimde		mde_mi7	
					Feelings of w	mde_mi7_2			+ rbinom(n=	0.4)		Feelings of w	mde_mi7_2		= nsimSmde_mimde		mde_mi7	
															simSmde_mi7 = 1*((simSmde_mi7_1 + simSmde_mi7_2) > 0)			
			Information	mde_mi7_bias								Information	mde_mi7_bias		= nsimSmde_mimde		mde_mi7	
			Diminished a	mde_mi8				mde_mi8 = mde_mi8_1 + mde_mi8_2 + mde_mi8_bias				Diminished a	mde_mi8		= 1*((mde_mi8_mde		mde_mi8	
					Diminished a	mde_mi8_1			+ rbinom(n=	0.4)		Diminished a	mde_mi8_1		= nsimSmde_mimde		mde_mi8	
					Indecisivenes	mde_mi8_2			+ rbinom(n=	0.4)		Indecisivenes	mde_mi8_2		= nsimSmde_mimde		mde_mi8	
															simSmde_mi8 = 1*((simSmde_mi8_1 + simSmde_mi8_2) > 0)			
			Information	mde_mi8_bias								Information	mde_mi8_bias		= nsimSmde_mimde		mde_mi8	
			Recurrent th	mde_mi9					+ rbinom(n=	0.4)		Recurrent th	mde_mi9		= rbinSmde_mimde			
			Information of the minor criteria not expained by the major or minir criteria									Information of the minor criteria not expained by the major or minir criteria						
															simSmde = simSmde_ma1 * simSmde_ma2 * (simSmde_mi3 + simSmde_mi4 + simSmde_mi5 + simSmde_mi6 + simSmde_mi7 + simSmde_mi8 + simSmde_mi9 + simSmde_bias1) + (1- simSmde_ma1 * simSmde_ma2) * (mde_ma1 * simSmde_ma2) * (simSmde_mi3 + simSmde_mi4 + simSmde_mi5 + simSmde_mi6 + simSmde_mi7 + simSmde_mi8 + simSmde_mi9 + simSmde_bias2)			
		Information	mde_bias1									Information	mde_bias1		= nsimSmde_biamde			
		Information	mde_bias2									Information	mde_bias2		= nsimSmde_biamde			
		Information	mde_bias									Information	mde_bias		= nsimSmde_biamde			
Dysthymic Di	dys							dys = dys_madys + intercept + coef1 x dys_ma + coef2 x dys_mi + coef3 x dys_b				Dysthymic Di	dys		dys = dys_maxsimSdys + sindys			
		Major criteria, essential for diagnosis										Major criteria						
			dys_ma	Depressed mood most of the day for more days than not, for at least 2 years					+ rbinom(n=	0.4)		Depressed m	dys_ma		= rbinsimSdys_ma	dys		
		Minor criteria	dys_mi					dys_mi = dys_mi1 + dys_mi2 + dys_mi3 + dys_mi4 + dys_mi5 + dys_mi6 + dys_mi8 + dys_m				Minor criteria	dys_mi		= nsimSdys_mi	dys		
			Poor appetite	dys_mi1				dys_mi1 = dys_mi1_1 + dys_mi1_2 + dys_mi1_bias				Poor appetite	dys_mi1		= 1*((dys_mi1_1	dys	dys_mi1	
					Poor appetite	dys_mi1_1			+ rbinom(n=	0.4)		Poor appetite	dys_mi1_1		= nsimSdys_mi1	dys	dys_mi1	
					Overeating	dys_mi1_2			+ rbinom(n=	0.4)		Overeating	dys_mi1_2		= simSdys_mi1	dys	dys_mi1	
															simSdys_mi1 = 1*((simSdys_mi1_1 + simSdys_mi1_2) > 0)			
			Information	dys_mi1_bias								Information	dys_mi1_bias		= nsimSdys_mi1	dys	dys_mi1	
			Insomnia or	sdys_mi2				dys_mi2 = mde_mi4 + mde_mi4_1 + mde_mi4_2 + mde_mi4_bias				Insomnia or	mde_mi4		= 1*((mde_mi4_dys			
					Insomnia	mde_mi4_1			+ rbinom(n=	0.4)		Insomnia	mde_mi4_1		= nsimSmde_midys		mde_mi4	
					Sleeping too	mde_mi4_2			+ rbinom(n=	0.4)		Sleeping too	mde_mi4_2		= nsimSmde_midys		mde_mi4	
															simSmde_mi4 = 1*((simSmde_mi4_1 + simSmde_mi4_2) > 0)			
			Information	mde_mi4_bias								Information	mde_mi4_bias		= nsimSmde_midys		mde_mi4	
			Low energy c	dys_mi3				dys_mi3 = mde_mi6 + mde_mi6_1 + mde_mi6_2 + mde_mi6_bias				Low energy c	mde_mi6		= 1*((mde_mi6_dys			
					Fatigue	mde_mi6_1			+ rbinom(n=	0.4)		Fatigue	mde_mi6_1		= nsimSmde_midys		mde_mi6	
					Loss of energ	mde_mi6_2			+ rbinom(n=	0.4)		Loss of energ	mde_mi6_2		= nsimSmde_midys		mde_mi6	
															simSmde_mi6 = 1*((simSmde_mi6_1 + simSmde_mi6_2) > 0)			
			Information	mde_mi6_bias								Information	mde_mi6_bias		= nsimSmde_midys		mde_mi6	
			Low self-este	dys_mi4					+ rbinom(n=	0.4)		Low self-este	dys_mi4		= rbinSdys_mi4dys			
			Poor concentr	dys_mi5				dys_mi5 = mde_mi8 + mde_mi8_1 + mde_mi8_2 + mde_mi8_bias				Poor concentr	mde_mi8		= 1*((mde_mi8_dys			
					Diminished a	mde_mi8_1			+ rbinom(n=	0.4)		Diminished a	mde_mi8_1		= nsimSmde_midys		mde_mi8	
					difficulty ma	mde_mi8_2			+ rbinom(n=	0.4)		difficulty ma	mde_mi8_2		= nsimSmde_midys		mde_mi8	
															simSmde_mi8 = 1*((simSmde_mi8_1 + simSmde_mi8_2) > 0)			
			Information	mde_mi8_bias								Information	mde_mi8_bias		= nsimSmde_midys		mde_mi8	
			Feelings of h	dys_mi6					+ rbinom(n=	0.4)		Feelings of h	dys_mi6		= rbinSdys_mi6dys			
			Information	dys_mi_bias								Information	dys_mi_bias		simSdys = simSdys_ma * simSdys_mi			
		Information	dys_bias									Information	dys_bias		= dysimSdys_biasdys			
Manic episod	manic							manic = (1- mmanic + intercept + coef1 x man_ma1 + coef2 x man_ma2 + coef3				Manic episod	manic		= (1- nsimSmanic = dys			
		Major criteria, essential for the diagnosis of a manic episode (more than one bipolar episode required to diagnose bipolar disorder)										Major criteria						
			A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hosp									A distinct per						
					Elevated mo	man_ma1			+ rbinom(n=	0.1)		Elevated mo	man_ma1		= rsimSman_mamanic			
					Expansive m	man_ma2			+ rbinom(n=	0.1)		Expansive m	man_ma2		= rsimSman_mamanic			
					Irritable moo	man_ma3			+ rbinom(n=	0.1)		Irritable moo	man_ma3		= rsimSman_mamanic			
		Minor criteria (3 or more of the following symptoms have persisted; 4 if the mood is only irritable)										Minor criteria						
			Increased sel	man_mi1				man_mi1 = man_mi1_1 + man_mi1_2 + man_mi1_bias				Increased sel	man_mi1		= 1*((man_mi1_manic		man_mi1	
					Increased sel	man_mi1_1			+ rbinom(n=	0.1)		Increased sel	man_mi1_1		= nsimSman_mimanic		man_mi1	
					Grandiosity	man_mi1_2			+ rbinom(n=	0.1)		Grandiosity	man_mi1_2		= nsimSman_mimanic		man_mi1	
															simSman_mi1 = 1*((simSman_mi1_1 + simSman_mi1_2) >0)			
			Information	man_mi1_bias								Information	man_mi1_bias		= nsimSman_mimanic		man_mi1	
			Decreased n	man_mi2					+ rbinom(n=	0.1)		Decreased n	man_mi2		= rbinSman_mimanic			
			More talkativ	man_mi3				man_mi3 = man_mi3_1 + man_mi3_2 + man_mi3_bias				More talkativ	man_mi3		= 1*((man_mi3_manic		man_mi3	
					More talkativ	man_mi3_1			+ rbinom(n=	0.1)		More talkativ	man_mi3_1		= nsimSman_mimanic		man_mi3	
					Pressure to k	man_mi3_2			+ rbinom(n=	0.1)		Pressure to k	man_mi3_2		= nsimSman_mimanic		man_mi3	
															simSman_mi3 = 1*((simSman_mi3_1 + simSman_mi3_2) > 0)			
			Information	man_mi3_bias								Information	man_mi3_bias		= nsimSman_mimanic		man_mi3	
			Flight of ide	man_mi4				man_mi4 = man_mi4_1 + man_mi4_2 + man_mi4_bias				Flight of idea	man_mi4		= 1*((man_mi4_manic		man_mi4	
					Flight of idea	man_mi4_1			+ rbinom(n=	0.1)		Flight of idea	man_mi4_1		= nsimSman_mimanic		man_mi4	
					Subjective e	man_mi4_2			+ rbinom(n=	0.1)		Subjective e	man_mi4_2		= nsimSman_mimanic		man_mi4	
															simSman_mi4 = 1*((simSman_mi4_1 + simSman_mi4_2) > 0)			
			Information	man_mi4_bias								Information	man_mi4_bias		= nsimSman_mimanic		man_mi4	
			Distractibili	man_mi5					+ rbinom(n=	0.1)		Distractibili	man_mi5		= rbinSman_mimanic			
			Increase in g	man_mi6				man_mi6 = man_mi6_1 + man_mi6_2 + man_mi6_bias				Increase in g	man_mi6		= 1*((man_mi6_manic		man_mi6	
					Increase in g	man_mi6_1			+ rbinom(n=	0.1)		Increase in g	man_mi6_1		= nsimSman_mimanic		man_mi6	
					Psychomotor	man_mi6_2			+ rbinom(n=	0.1)		Psychomotor	man_mi6_2		= nsimSman_mimanic		man_mi6	
															simSman_mi6 = 1*((simSman_mi6_1 + simSman_mi6_2) >0)			
			Information	man_mi6_bias								Information	man_mi6_bias		= nsimSman_mimanic		man_mi6	
			Excessive inv	man_mi7					+ rbinom(n=	0.1)		Excessive inv	man_mi7		= rbinSman_mimanic			
															simSmanic = (1- simSman_ma1 * simSman_ma2) * (simSman_ma1 + simSman_ma2) * simSman_ma3 * (simSman_mi1 + simSman_mi2 + simSman_mi3 + simSman_mi4 + simSman_mi5 + simSman_mi6 + simSman_mi7 + simSman_bias3) + (1 - (1 - simSman_ma1 * simSman_ma2)(simSman_ma1 + simSman_ma2)) * simSman_ma3 * (simSman_mi1 + simSman_mi2 + simSman_mi3 + simSman_mi4 + simSman_mi5 + simSman_mi6 + simSman_mi7 + simSman_bias2)			
		Information	man_bias1									Information	man_bias1		= simSman_biamanic			
		Information	man_bias2									Information	man_bias2		= simSman_biamanic			
		Information	man_bias									Information	man_bias		= nsimSman_biamanic			
			female										female		female = rbinsimSfemale >mde			
			age										age		age = sampleimSage > sexmde			
			edu										edu		edu = rnormSimSedu = mmde			
															edu[whichSsexmSedu[whichSimSedu <= 0]] = 0			

```
---
title: "2019_09_06 simulated mental illnesses"
author: "Yi-Sheng Chao"
date: "November 22, 2018"
output: pdf_document
editor_options:
  chunk_output_type: inline
---


##Adding correlations to the random variables

```{r}
library(bindata)

library(openxlsx)
resu = read.xlsx("A simulation study to demonstrate the biases in three
diagnoses of mental illnesses.xlsx", sheet = "Prob 1")
names(resu)
unique(resu$variable)
memory.limit(size = 10^13)
ssize = 10^5
times = 10^2

prevalence = c(0.05, 0.1, 0.3, 0.5, 0.7)
rho = c(0, 0.1, 0.4, 0.7, 0.9)#correlation coefficients of the input
symptoms

collect = c("mean", "max",
"min","derivedprevalence","coef","coefse","p","intercept",
"interceptp","r2", "subcoef","subcoefse","subp","subintercept",
"subinterceptp","subr2", "appbyownr2", "appbybiasr2", "appbyallr2",
"appbyownvar", "appbybiasvar", "appbyallvar", "appbyownn", "appbybiasn",
"appbyalln")


set.seed(1)


##Create a simulated data set to extract variables
for(preval in 1:length(prevalence)){
  for(rh in 1:length(rho)){


  library(openxlsx)
resu = read.xlsx("A simulation study to demonstrate the biases in three
diagnoses of mental illnesses.xlsx", sheet = "Prob 1")

    # foreach(c = 1:times) %dopar% {
    for(c in 1:times){

library(bindata)
bindata = as.data.frame(rmvbin(ssize, rep(prevalence[preval], 40),
bincorr=(1 - rho[rh])*diag(40) + rho[rh]))
bindata2 = as.data.frame(rmvbin(ssize, rep(prevalence[preval], 20),
bincorr=(1 - rho[rh])*diag(20) + rho[rh]))

##demographic characteristics
```
```

```
sim = data.frame(1:ssize)
names(sim) = "id"
sim$female = rbinom(n = ssize, size = 1, prob = 0.51)
sim$age = sample(30:60, ssize, replace = TRUE)
sim$edu = rnorm(ssize, mean = 12, sd = 5)
sim$edu[which(sim$edu <= 0)] = 0
sim$id = NULL


sim$mde_ma1 = bindata[,1]
sim$mde_ma2 = bindata[,2]

sim$mde_mi3_1 = bindata[,3]
sim$mde_mi3_2 = bindata[,4]
sim$mde_mi3 = 1*((sim$mde_mi3_1 + sim$mde_mi3_2) > 0)
sim$mde_mi3_bias = sim$mde_mi3 - sim$mde_mi3_1 - sim$mde_mi3_2

sim$mde_mi4_1 = bindata[,5]
sim$mde_mi4_2 = bindata[,6]
sim$mde_mi4 = 1*((sim$mde_mi4_1 + sim$mde_mi4_2) > 0)
sim$mde_mi4_bias = sim$mde_mi4 - sim$mde_mi4_1 - sim$mde_mi4_2

sim$mde_mi5_1 = bindata[,7]
sim$mde_mi5_2 = bindata[,8]
sim$mde_mi5 = 1*((sim$mde_mi5_1 + sim$mde_mi5_2) > 0)
sim$mde_mi5_bias = sim$mde_mi5 - sim$mde_mi5_1 - sim$mde_mi5_2

sim$mde_mi6_1 = bindata[,9]
sim$mde_mi6_2 = bindata[,10]
sim$mde_mi6 = 1*((sim$mde_mi6_1 + sim$mde_mi6_2) > 0)
sim$mde_mi6_bias = sim$mde_mi6 - sim$mde_mi6_1 - sim$mde_mi6_2

sim$mde_mi7_1 = bindata[,11]
sim$mde_mi7_2 = bindata[,12]
sim$mde_mi7 = 1*((sim$mde_mi7_1 + sim$mde_mi7_2) > 0)
sim$mde_mi7_bias = sim$mde_mi7 - sim$mde_mi7_1 - sim$mde_mi7_2

sim$mde_mi8_1 = bindata[,13]
sim$mde_mi8_2 = bindata[,14]
sim$mde_mi8 = 1*((sim$mde_mi8_1 + sim$mde_mi8_2) > 0)
sim$mde_mi8_bias = sim$mde_mi8 - sim$mde_mi8_1 - sim$mde_mi8_2

sim$mde_mi9 = bindata[,15]

sim$mde_bias1 = 1 * ((sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 +
sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9)>2) - (sim$mde_mi3
+ sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9)
sim$mde_bias2 = 1 * ((sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 +
sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9)>3) - (sim$mde_mi3
+ sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9)

sim$mde = sim$mde_ma1 * sim$mde_ma2 * (sim$mde_mi3 + sim$mde_mi4 +
sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9 +
sim$mde_bias1) + (1- sim$mde_ma1 * sim$mde_ma2) * (sim$mde_ma1 *
sim$mde_ma2) * (sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 +
sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9 + sim$mde_bias2)
```

```
sim$mde_bias = sim$mde - (sim$mde_ma1 + sim$mde_ma2) - (sim$mde_mi3 +
sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9 + sim$mde_bias1) - (sim$mde_bias2)

##Definition Below: even the bias and own input variables could not fully
explain the diagnosis
# sim$mde_bias = sim$mde - (sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 +
sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 + sim$mde_mi9 + sim$mde_bias1) -
(sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 +
sim$mde_mi8 + sim$mde_mi9 + sim$mde_bias2)

# sim$mde_bias = sim$mde - (sim$mde_ma1 + sim$mde_ma2 + sim$mde_mi3 +
sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 + sim$mde_mi8 +
sim$mde_mi9)


# sim$mde_bias = resid(lm(sim$mde ~ sim$mde_ma1 + sim$mde_ma2 +
sim$mde_mi3 + sim$mde_mi4 + sim$mde_mi5 + sim$mde_mi6 + sim$mde_mi7 +
sim$mde_mi8 + sim$mde_mi9, data=sim))

##DYS

sim$dys_ma = bindata[,16]

sim$dys_mi1_1 = bindata[,17]
sim$dys_mi1_2 = bindata[,18]
sim$dys_mi1 = 1*((sim$dys_mi1_1 + sim$dys_mi1_2) > 0)
sim$dys_mi1_bias = sim$dys_mi1 - sim$dys_mi1_1 - sim$dys_mi1_2

sim$dys_mi4 = bindata[,19]

sim$dys_mi6 = bindata[,20]

sim$dys_mi = 1*((sim$dys_mi1 + sim$mde_mi4 + sim$mde_mi6 + sim$dys_mi4 +
sim$mde_mi8 + sim$dys_mi6)>1)

sim$dys_mi_bias = sim$dys_mi - (sim$dys_mi1 + sim$mde_mi4 + sim$mde_mi6 +
sim$dys_mi4 + sim$mde_mi8 + sim$dys_mi6)

sim$dys = sim$dys_ma * sim$dys_mi


sim$dys_bias = sim$dys - (sim$dys_ma + sim$dys_mi)

# sim$dys_bias = resid(lm(sim$dys ~ sim$dys_ma + sim$dys_mi, data=sim))


##Manic
sim$man_ma1 = bindata2[,1]
sim$man_ma2 = bindata2[,2]
sim$man_ma3 = bindata2[,3]

sim$man_mi1_1 = bindata2[,4]
sim$man_mi1_2 = bindata2[,5]
sim$man_mi1 = 1*((sim$man_mi1_1 + sim$man_mi1_2) >0)
sim$man_mi1_bias = sim$man_mi1 - (sim$man_mi1_1 + sim$man_mi1_2)
sim$man_mi2 = bindata2[,6]
```

```
sim$man_mi3_1 = bindata2[,7]
sim$man_mi3_2 = bindata2[,8]
sim$man_mi3 = 1*((sim$man_mi3_1 + sim$man_mi3_2) > 0)
sim$man_mi3_bias = sim$man_mi3 - (sim$man_mi3_1 + sim$man_mi3_2)
sim$man_mi4_1 = bindata2[,9]
sim$man_mi4_2 = bindata2[,10]
sim$man_mi4 = 1*((sim$man_mi4_1 + sim$man_mi4_2) > 0)
sim$man_mi4_bias = sim$man_mi4 - (sim$man_mi4_1 + sim$man_mi4_2)
sim$man_mi5 = bindata2[,11]
sim$man_mi6_1 = bindata2[,12]
sim$man_mi6_2 = bindata2[,13]
sim$man_mi6 = 1*((sim$man_mi6_1 + sim$man_mi6_2) > 0)
sim$man_mi6_bias = sim$man_mi6 - (sim$man_mi6_1 + sim$man_mi6_2)
sim$man_mi7 = bindata2[,14]
sim$man_bias1 = 1*((sim$man_mi1 + sim$man_mi2 + sim$man_mi3 + sim$man_mi4
+ sim$man_mi5 + sim$man_mi6 + sim$man_mi7) > 2) - (sim$man_mi1 +
sim$man_mi2 + sim$man_mi3 + sim$man_mi4 + sim$man_mi5 + sim$man_mi6 +
sim$man_mi7)
sim$man_bias2 = 1*((sim$man_mi1 + sim$man_mi2 + sim$man_mi3 + sim$man_mi4
+ sim$man_mi5 + sim$man_mi6 + sim$man_mi7) > 3) - (sim$man_mi1 +
sim$man_mi2 + sim$man_mi3 + sim$man_mi4 + sim$man_mi5 + sim$man_mi6 +
sim$man_mi7)


sim$manic = (1- sim$man_ma1 * sim$man_ma2) * (sim$man_ma1 + sim$man_ma2)
* sim$man_ma3 * (sim$man_mi1 + sim$man_mi2 + sim$man_mi3 + sim$man_mi4 +
sim$man_mi5 + sim$man_mi6 + sim$man_mi7 + sim$man_bias1) + (1 - (1 -
sim$man_ma1 * sim$man_ma2) * (sim$man_ma1 + sim$man_ma2)) * sim$man_ma3 *
(sim$man_mi1 + sim$man_mi2 + sim$man_mi3 + sim$man_mi4 + sim$man_mi5 +
sim$man_mi6 + sim$man_mi7 + sim$man_bias2)


sim$man_bias = sim$manic - (sim$man_ma1 + sim$man_ma2 + sim$man_ma3) -
(sim$man_mi1 + sim$man_mi2 + sim$man_mi3 + sim$man_mi4 + sim$man_mi5 +
sim$man_mi6 + sim$man_mi7 + sim$man_bias1) - (sim$man_bias2)

  ##end of generate data


  resu[, paste(collect, "_", c, sep = "")] = NA
    for(r in 1:nrow(resu)){
      #variable characteristics
      if(is.na(resu$variable[r]) == FALSE){
        resu[r, paste0("derivedprevalence_", c, collapse = "")] =
nrow(sim[which(sim[, resu$variable[r]] == 1),])/ssize
        resu[r, paste0("mean_", c, collapse = "")] =
mean(sim[,resu$variable[r]])
        resu[r, paste0("max_", c, collapse = "")] =
max(sim[,resu$variable[r]])
        resu[r, paste0("min_", c, collapse = "")] =
min(sim[,resu$variable[r]])
      }
      ##regression for the diagnosis
      if(is.na(resu$variable[r]) == FALSE & resu$variable[r] !=
resu$outcome[r]){
        eval(parse(text = paste0("templm = summary(lm(", resu$outcome[r],
" ~ ", resu$variable[r], ", data = sim))", collpase = "")))
```

```
        resu[r, paste0("coef_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Estimate"]
        resu[r, paste0("coefse_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Std. Error"]
        resu[r, paste0("p_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Pr(>|t|)"]
        resu[r, paste0("intercept_", c, collapse = "")] =
templm$coefficients["(Intercept)", "Estimate"]
        resu[r, paste0("interceptp_", c, collapse = "")] =
templm$coefficients["(Intercept)", "Pr(>|t|)"]
        resu[r, paste0("r2_", c, collapse = "")] = templm$r.squared
      }
            ##regression for the suboutcome/domain variables
      if(is.na(resu$variable[r]) == FALSE & is.na(resu$suboutcome[r]) ==
FALSE & resu$variable[r] != resu$outcome[r] & resu$variable[r] !=
resu$suboutcome[r]){
        eval(parse(text = paste0("templm = summary(lm(",
resu$suboutcome[r], " ~ ", resu$variable[r], ", data = sim))", collpase =
"")))
        resu[r, paste0("subcoef_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Estimate"]
        resu[r, paste0("subcoefse_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Std. Error"]
        resu[r, paste0("subp_", c, collapse = "")] =
templm$coefficients[resu$variable[r], "Pr(>|t|)"]
        resu[r, paste0("subintercept_", c, collapse = "")] =
templm$coefficients["(Intercept)", "Estimate"]
        resu[r, paste0("subinterceptp_", c, collapse = "")] =
templm$coefficients["(Intercept)", "Pr(>|t|)"]
        resu[r, paste0("subr2_", c, collapse = "")] = templm$r.squared
      }


            if(r %in% as.character(1:100*50)){print(c("r:", r))}
    }#r = rows of the variable list


        ##Approximation by own, bias or all variables


          #ploting area_start: only the last simulation data set used for
plotting
          library(leaps)
          #MDE
          #own variables only
          mdeown = NA
          library(car)
          sim.new = sim[,c("mde",
names(summary((lm(as.formula(paste0("mde ~ ", paste0(names(sim)
[grepl("mde_", names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE],
collapse = " + "), collapse = "")), data = sim)))$aliased)
[summary((lm(as.formula(paste0("mde ~ ", paste0(names(sim)[grepl("mde_",
names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE], collapse = " +
"), collapse = "")), data = sim)))$aliased == FALSE &
names(summary((lm(as.formula(paste0("mde ~ ", paste0(names(sim)
[grepl("mde_", names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE],
collapse = " + "), collapse = "")), data = sim)))$aliased) !=
"(Intercept)"])]
```

```
          for(repe in 1:40){
            tempvif = vif(lm(mde~., data = sim.new))
            if(any(tempvif > 10)){
            sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
            }
          }


          try(
            (mdeown = regsubsets(mde~., data = sim.new, really.big=T,
method = "forward", nvmax = ncol(sim.new))), silent = F
          )

          mdeownsummary = NA
          if(any(is.na(mdeown)) == FALSE){
            mdeownsummary = summary(mdeown)
          }

          mdeownsummary$adjr2



          ##own and bias variables
          mdebias = NA

          mdebias = regsubsets(as.formula(paste0("mde ~ ",
paste0(names(sim)[grepl("mde_", names(sim)) == TRUE & grepl("bias",
names(sim)) == TRUE], collapse = " + "), collapse = "")), data = sim,
nvmax = 100, really.big=T, method = "forward")
          mdebiassummary = summary(mdebias)
          mdebiassummary$adjr2


          ##all variables
          ###in case of collinearity
          mdeall = NA


          ##Deal with collinearity
          library(car)
          sim.new = sim[,c("mde", names(summary((lm(mde~., data = sim)))
$aliased)[summary((lm(mde~., data = sim)))$aliased == FALSE &
names(summary((lm(mde~., data = sim)))$aliased) != "(Intercept)"])]


          for(repe in 1:40){
            tempvif = vif(lm(mde~., data = sim.new))
            if(any(tempvif > 10)){
            sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
            }
          }

          ##Somehow there are problems in executing regsubsets even after
removing collinear variables
```

```
          try(
            (mdeall = regsubsets(mde~., data = sim.new, really.big=T,
method = "forward", nvmax = ncol(sim.new))), silent = F
          )

          mdeallsummary = NA
          if(any(is.na(mdeall)) == FALSE){
            mdeallsummary = summary(mdeall)
          }

          # mdeallsummary$adjr2


          #DYS
          #dys
          #own variables only
          dysown = NA
                library(car)
          sim.new = sim[,c("dys",
names(summary((lm(as.formula(paste0("dys ~ ", paste0(c(names(sim)
[(grepl("dys_", names(sim)) == TRUE | grepl("mde_mi4", names(sim)) ==
TRUE | grepl("mde_mi6", names(sim)) == TRUE | grepl("mde_mi8",
names(sim)) == TRUE) & grepl("bias", names(sim)) == FALSE]), collapse = "
+ "), collapse = "")), data = sim)))$aliased)
[summary((lm(as.formula(paste0("dys ~ ", paste0(c(names(sim)
[(grepl("dys_", names(sim)) == TRUE | grepl("mde_mi4", names(sim)) ==
TRUE | grepl("mde_mi6", names(sim)) == TRUE | grepl("mde_mi8",
names(sim)) == TRUE) & grepl("bias", names(sim)) == FALSE]), collapse = "
+ "), collapse = "")), data = sim)))$aliased == FALSE &
names(summary((lm(as.formula(paste0("dys ~ ", paste0(c(names(sim)
[(grepl("dys_", names(sim)) == TRUE | grepl("mde_mi4", names(sim)) ==
TRUE | grepl("mde_mi6", names(sim)) == TRUE | grepl("mde_mi8",
names(sim)) == TRUE) & grepl("bias", names(sim)) == FALSE]), collapse = "
+ "), collapse = "")), data = sim)))$aliased) != "(Intercept)"])]

          for(repe in 1:40){
                      tempvif = vif(lm(dys~., data = sim.new))
              if(any(tempvif > 10)){
              sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
              }
          }
          ##Somehow there are problems in executing regsubsets even after
removing collinear variables
          try(
            (dysown = regsubsets(dys~., data = sim.new, really.big=T,
method = "forward", nvmax = 100)), silent = T
          )

          if(any(is.na(dysown)) == FALSE){
            dysownsummary = summary(dysown)
          }

          ##own and bias variables
          dysbias = NA
```

```
          dysbias = regsubsets(as.formula(paste0("dys ~ ",
paste0(names(sim)[(grepl("dys_", names(sim)) == TRUE | grepl("mde_mi4",
names(sim)) == TRUE | grepl("mde_mi6", names(sim)) == TRUE |
grepl("mde_mi8", names(sim)) == TRUE) & grepl("bias", names(sim)) ==
TRUE], collapse = " + "), collapse = "")), data = sim, nvmax = 100,
really.big=T, method = "forward")
          dysbiassummary = summary(dysbias)
          dysbiassummary$adjr2


          ##all variables
          ###in case of collinearity
          dysall = NA
          library(car)
          sim.new = sim[,c("dys", names(summary((lm(dys~., data = sim)))
$aliased)[summary((lm(dys~., data = sim)))$aliased == FALSE &
names(summary((lm(dys~., data = sim)))$aliased) != "(Intercept)"])]

          for(repe in 1:40){
                    tempvif = vif(lm(dys~., data = sim.new))
              if(any(tempvif > 10)){
              sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
              }
          }
          ##Somehow there are problems in executing regsubsets even after
removing collinear variables
          try(
            (dysall = regsubsets(dys~., data = sim.new, really.big=T,
method = "forward", nvmax = 100)), silent = T
          )

          if(any(is.na(dysall)) == FALSE){
            dysallsummary = summary(dysall)
          }
          # dysallsummary$adjr2




          #manic
          #own variables only
          manown = NA
                    library(car)
          sim.new = sim[,c("manic",
names(summary((lm(as.formula(paste0("manic ~ ", paste0(names(sim)
[grepl("man_", names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE],
collapse = " + "), collapse = "")), data = sim)))$aliased)
[summary((lm(as.formula(paste0("manic ~ ", paste0(names(sim)
[grepl("man_", names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE],
collapse = " + "), collapse = "")), data = sim)))$aliased == FALSE &
names(summary((lm(as.formula(paste0("manic ~ ", paste0(names(sim)
[grepl("man_", names(sim)) == TRUE & grepl("bias", names(sim)) == FALSE],
collapse = " + "), collapse = "")), data = sim)))$aliased) !=
"(Intercept)"])]
          for(repe in 1:40){
                    tempvif = vif(lm(manic~., data = sim.new))
              if(any(tempvif > 10)){
```

```
              sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
              }
          }
          try(
            (manown = regsubsets(manic~., data = sim.new, really.big=T,
method = "forward", nvmax = 100)), silent = T
          )
          manownsummary = NA
          if(any(is.na(manown)) == FALSE){
            manownsummary = summary(manown)
          }

          ##own and bias variables
          manbias = NA

          manbias = regsubsets(as.formula(paste0("manic ~ ",
paste0(names(sim)[grepl("man_", names(sim)) == TRUE & grepl("bias",
names(sim)) == TRUE], collapse = " + "), collapse = "")), data = sim,
nvmax = 100, really.big=T, method = "forward")
          manbiassummary = summary(manbias)
          manbiassummary$adjr2


          ##all variables
          ###in case of collinearity
          manall = NA
                    library(car)
          sim.new = sim[,c("manic", names(summary((lm(manic~., data =
sim)))$aliased)[summary((lm(manic~., data = sim)))$aliased == FALSE &
names(summary((lm(manic~., data = sim)))$aliased) != "(Intercept)"])]
          for(repe in 1:40){
                      tempvif = vif(lm(manic~., data = sim.new))
              if(any(tempvif > 10)){
              sim.new = sim.new[,which(names(sim.new) != names(tempvif)
[which(tempvif == tempvif[order(-tempvif)][1])])]
              }
          }
          ##Somehow there are problems in executing regsubsets even after
removing collinear variables
          try(
            (manall = regsubsets(manic~., data = sim.new, really.big=T,
method = "forward", nvmax = 100)), silent = T
          )
          manallsummary = NA
          if(any(is.na(manall)) == FALSE){
            manallsummary = summary(manall)
          }

  ##extract information from the outmat
       #MDE

          resu[which(resu$variable == "mde"), paste0("appbyownr2_", c,
collapse = "")] = mdeownsummary$adjr2[which.max(mdeownsummary$adjr2)]
          resu[which(resu$variable == "mde"), paste0("appbyownn_", c,
collapse = "")] = which.max(mdeownsummary$adjr2)
```

```
          resu[which(resu$variable == "mde"), paste0("appbyownvar_", c,
collapse = "")] = paste0(dimnames(mdeownsummary$outmat)[[2]]
[which(mdeownsummary$outmat[which.max(mdeownsummary$adjr2),] == "*")],
collapse = ",")

          resu[which(resu$variable == "mde"), paste0("appbybiasr2_", c,
collapse = "")] = mdebiassummary$adjr2[which.max(mdebiassummary$adjr2)]
          resu[which(resu$variable == "mde"), paste0("appbybiasn_", c,
collapse = "")] = which.max(mdebiassummary$adjr2)
          resu[which(resu$variable == "mde"), paste0("appbybiasvar_", c,
collapse = "")] = paste0(dimnames(mdebiassummary$outmat)[[2]]
[which(mdebiassummary$outmat[which.max(mdebiassummary$adjr2),] == "*")],
collapse = ",")

          if(any(is.na(mdeall)) == FALSE){
              resu[which(resu$variable == "mde"), paste0("appbyallr2_",
c, collapse = "")] = mdeallsummary$adjr2[which.max(mdeallsummary$adjr2)]
              resu[which(resu$variable == "mde"), paste0("appbyalln_", c,
collapse = "")] = which.max(mdeallsummary$adjr2)
              resu[which(resu$variable == "mde"), paste0("appbyallvar_",
c, collapse = "")] = paste0(dimnames(mdeallsummary$outmat)[[2]]
[which(mdeallsummary$outmat[which.max(mdeallsummary$adjr2),] == "*")],
collapse = ",")

          }


       #DYS
          resu[which(resu$variable == "dys"), paste0("appbyownr2_", c,
collapse = "")] = dysownsummary$adjr2[which.max(dysownsummary$adjr2)]
          resu[which(resu$variable == "dys"), paste0("appbyownn_", c,
collapse = "")] = which.max(dysownsummary$adjr2)
          resu[which(resu$variable == "dys"), paste0("appbyownvar_", c,
collapse = "")] = paste0(dimnames(dysownsummary$outmat)[[2]]
[which(dysownsummary$outmat[which.max(dysownsummary$adjr2),] == "*")],
collapse = ",")

          resu[which(resu$variable == "dys"), paste0("appbybiasr2_", c,
collapse = "")] = dysbiassummary$adjr2[which.max(dysbiassummary$adjr2)]
          resu[which(resu$variable == "dys"), paste0("appbybiasn_", c,
collapse = "")] = which.max(dysbiassummary$adjr2)
          resu[which(resu$variable == "dys"), paste0("appbybiasvar_", c,
collapse = "")] = paste0(dimnames(dysbiassummary$outmat)[[2]]
[which(dysbiassummary$outmat[which.max(dysbiassummary$adjr2),] == "*")],
collapse = ",")

          if(any(is.na(dysall)) == FALSE){
            resu[which(resu$variable == "dys"), paste0("appbyallr2_", c,
collapse = "")] = dysallsummary$adjr2[which.max(dysallsummary$adjr2)]
          resu[which(resu$variable == "dys"), paste0("appbyalln_", c,
collapse = "")] = which.max(dysallsummary$adjr2)
          resu[which(resu$variable == "dys"), paste0("appbyallvar_", c,
collapse = "")] = paste0(dimnames(dysallsummary$outmat)[[2]]
[which(dysallsummary$outmat[which.max(dysallsummary$adjr2),] == "*")],
collapse = ",")
          }
```

```
       #MANIC
          resu[which(resu$variable == "manic"), paste0("appbyownr2_", c,
collapse = "")] = manownsummary$adjr2[which.max(manownsummary$adjr2)]
          resu[which(resu$variable == "manic"), paste0("appbyownn_", c,
collapse = "")] = which.max(manownsummary$adjr2)
          resu[which(resu$variable == "manic"), paste0("appbyownvar_", c,
collapse = "")] = paste0(dimnames(manownsummary$outmat)[[2]]
[which(manownsummary$outmat[which.max(manownsummary$adjr2),] == "*")],
collapse = ",")

          resu[which(resu$variable == "manic"), paste0("appbybiasr2_", c,
collapse = "")] = manbiassummary$adjr2[which.max(manbiassummary$adjr2)]
          resu[which(resu$variable == "manic"), paste0("appbybiasn_", c,
collapse = "")] = which.max(manbiassummary$adjr2)
          resu[which(resu$variable == "manic"), paste0("appbybiasvar_",
c, collapse = "")] = paste0(dimnames(manbiassummary$outmat)[[2]]
[which(manbiassummary$outmat[which.max(manbiassummary$adjr2),] == "*")],
collapse = ",")


          if(any(is.na(manall)) == FALSE){
            resu[which(resu$variable == "manic"), paste0("appbyallr2_",
c, collapse = "")] = manallsummary$adjr2[which.max(manallsummary$adjr2)]
            resu[which(resu$variable == "manic"), paste0("appbyalln_", c,
collapse = "")] = which.max(manallsummary$adjr2)
            resu[which(resu$variable == "manic"), paste0("appbyallvar_",
c, collapse = "")] = paste0(dimnames(manallsummary$outmat)[[2]]
[which(manallsummary$outmat[which.max(manallsummary$adjr2),] == "*")],
collapse = ",")
          }



            print(c("c:", c))
            print(c("cor:", rho[rh]))
            print(c("Prevalence: ", prevalence[preval]))

    }#c



    ##adding summary statistics to the result data frame
    resu[, paste(collect, "_mean", sep = "")] = NA
resu[, paste(collect, "_sd", sep = "")] = NA
resu[, paste(collect, "_se", sep = "")] = NA
resu[, paste(collect, "_95up", sep = "")] = NA
resu[, paste(collect, "_95lo", sep = "")] = NA
resu[, paste(collect, "_rangeup", sep = "")] = NA
resu[, paste(collect, "_rangelo", sep = "")] = NA

for(co in 1:length(collect)){
  for(r in 1:nrow(resu)){
    if((collect[co] %in% c("appbyownvar", "appbybiasvar", "appbyallvar"))
== FALSE){
          resu[r,paste0(collect[co], "_mean", collapse = "")] =
mean(unlist(resu[r, paste(collect[co], "_", 1:times, sep = "")])[which(!
is.na(unlist(resu[r, paste(collect[co], "_", 1:times, sep = "")])))])
```

```
    resu[r,paste0(collect[co], "_sd", collapse = "")] = sd(unlist(resu[r,
paste(collect[co], "_", 1:times, sep = "")])[which(!is.na(unlist(resu[r,
paste(collect[co], "_", 1:times, sep = ""))))])
    resu[r,paste0(collect[co], "_se", collapse = "")] = sd(unlist(resu[r,
paste(collect[co], "_", 1:times, sep = "")])[which(!is.na(unlist(resu[r,
paste(collect[co], "_", 1:times, sep = ""))))])/(times^0.5)

    #95% CIs
    resu[r,paste0(collect[co], "_95up", collapse = "")] =
resu[r,paste0(collect[co], "_mean", collapse = "")] +
1.96*resu[r,paste0(collect[co], "_se", collapse = "")]
    resu[r,paste0(collect[co], "_95lo", collapse = "")] =
resu[r,paste0(collect[co], "_mean", collapse = "")] -
1.96*resu[r,paste0(collect[co], "_se", collapse = "")]

    #range
        resu[r,paste0(collect[co], "_rangelo", collapse = "")] =
min(resu[r,paste(collect[co], "_", 1:times, sep = "")])
        resu[r,paste0(collect[co], "_rangeup", collapse = "")] =
max(resu[r,paste(collect[co], "_", 1:times, sep = "")])
  }#r

    ##Add information about the aliased variables



    ##save in another data set
    eval(parse(text = paste0("resu_cor", rho[rh], "_preval",
prevalence[preval], " = resu", collapse = "")))


    }

#export results
write.csv(cbind(resu[,c("definition", "variable", "mean_mean",
"mean_95up","mean_95lo", "max_rangeup",
"min_rangelo","derivedprevalence_mean", "derivedprevalence_95up",
"derivedprevalence_95lo", "coef_mean", "coef_95up", "coef_95lo",
"p_mean","p_95up","p_95lo", "r2_mean", "r2_95up",
"r2_95lo","subcoef_mean", "subcoef_95up", "subcoef_95lo",
"subp_mean","subp_95up","subp_95lo", "subr2_mean", "subr2_95up",
"subr2_95lo", "appbyownr2_mean", "appbyownr2_95up", "appbyownr2_95lo",
"appbyownn_mean", "appbyownn_95up", "appbyownn_95lo", "appbybiasr2_mean",
"appbybiasr2_95up", "appbybiasr2_95lo", "appbybiasn_mean",
"appbybiasn_95up", "appbybiasn_95lo", "appbyallr2_mean",
"appbyallr2_95up", "appbyallr2_95lo", "appbyalln_mean", "appbyalln_95up",
"appbyalln_95lo"

)], resu), file = paste0("simulation results_cor", rho[rh], "_preval",
prevalence[preval], ".csv"))



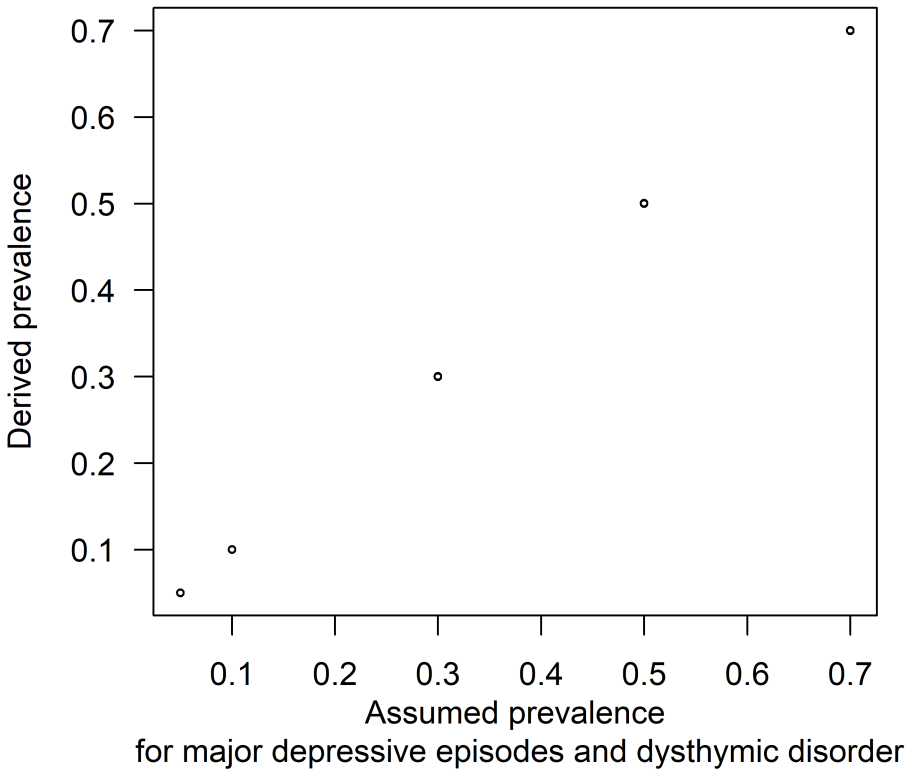
  }#co
```

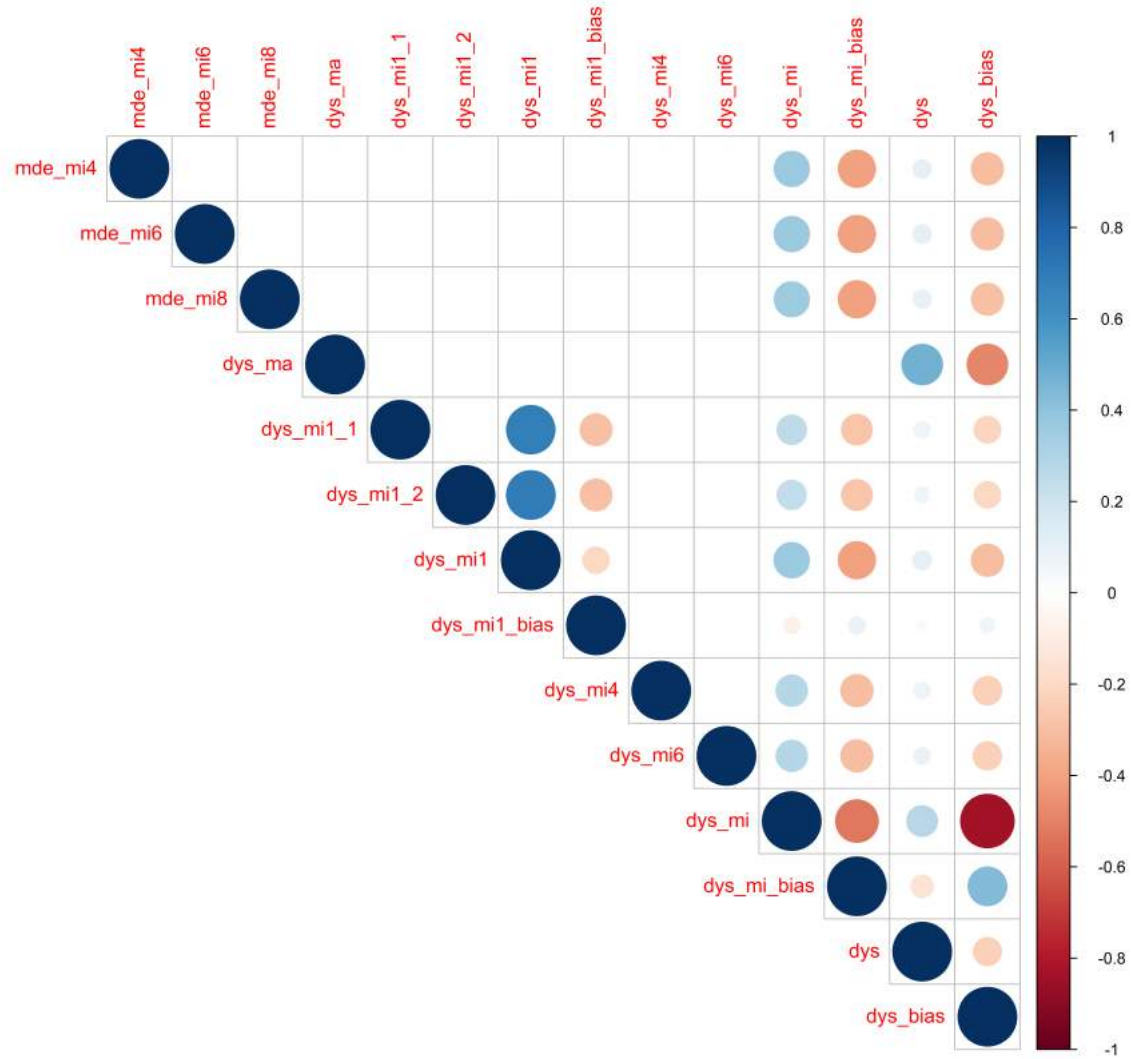
```
print(c("cor:", rho[rh]))
print(c("Prevalence: ", prevalence[preval]))

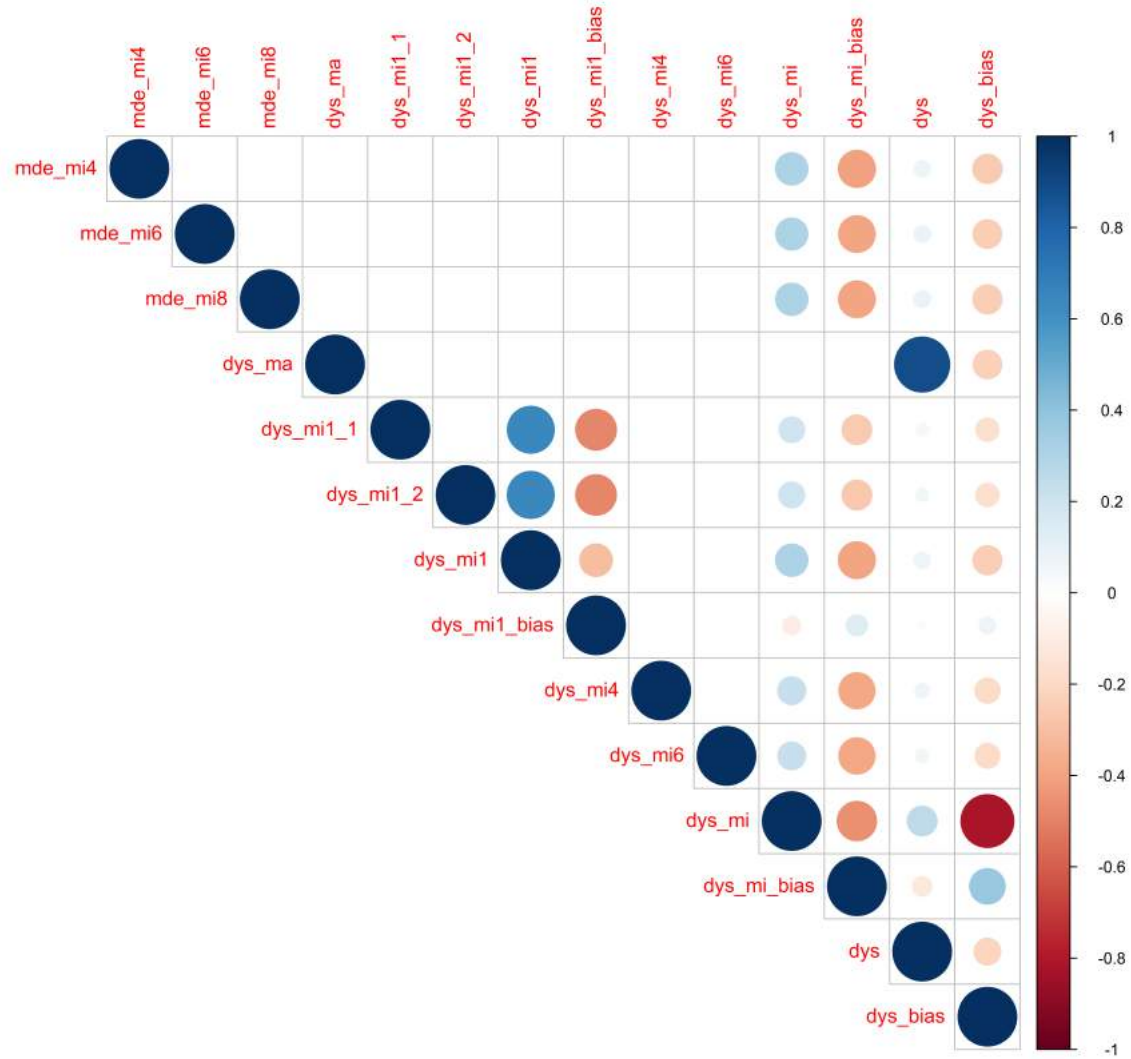
  }#rho
#store data

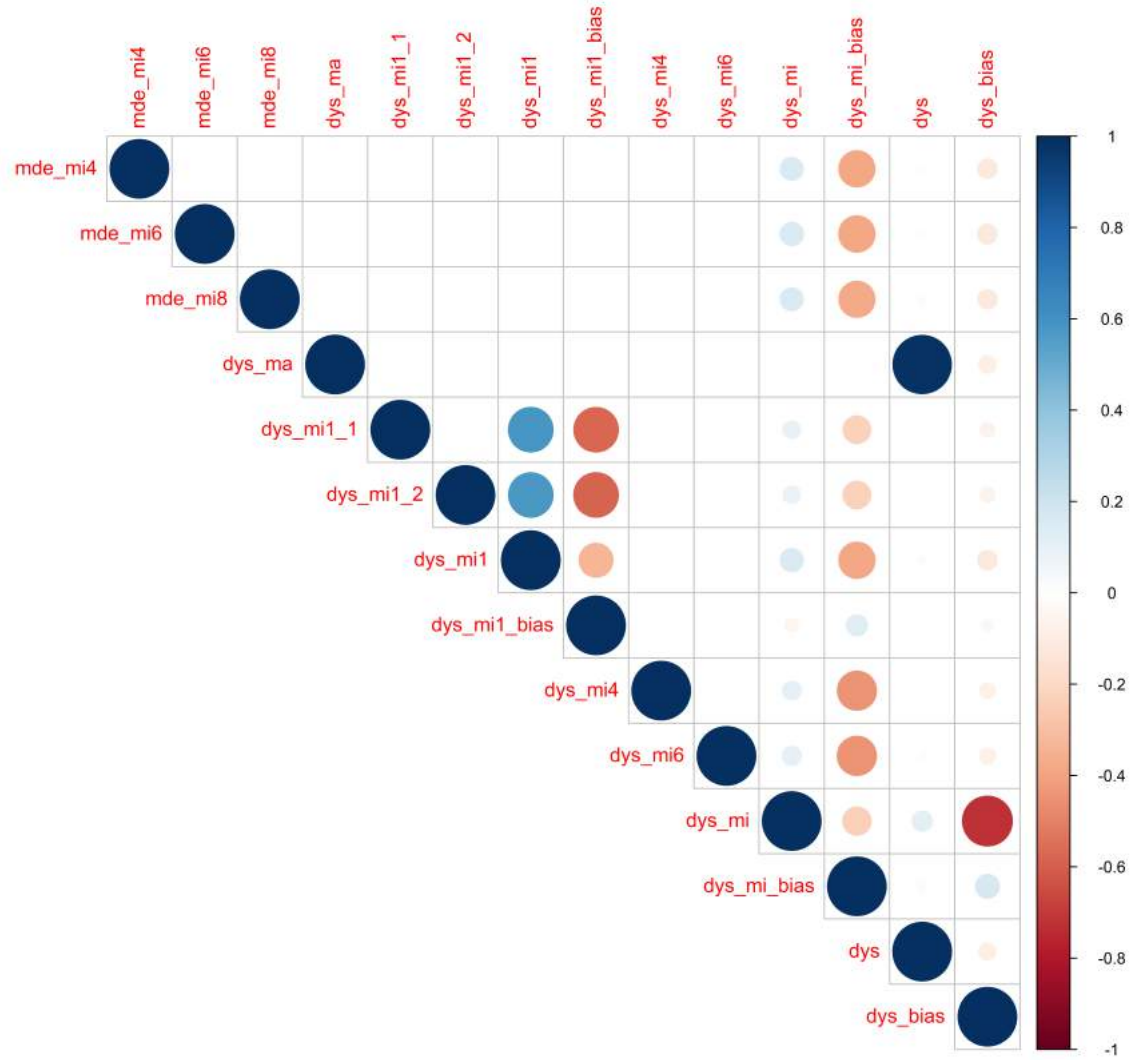
print(c("Prevalence: ", prevalence[preval]))
  }#prevalence

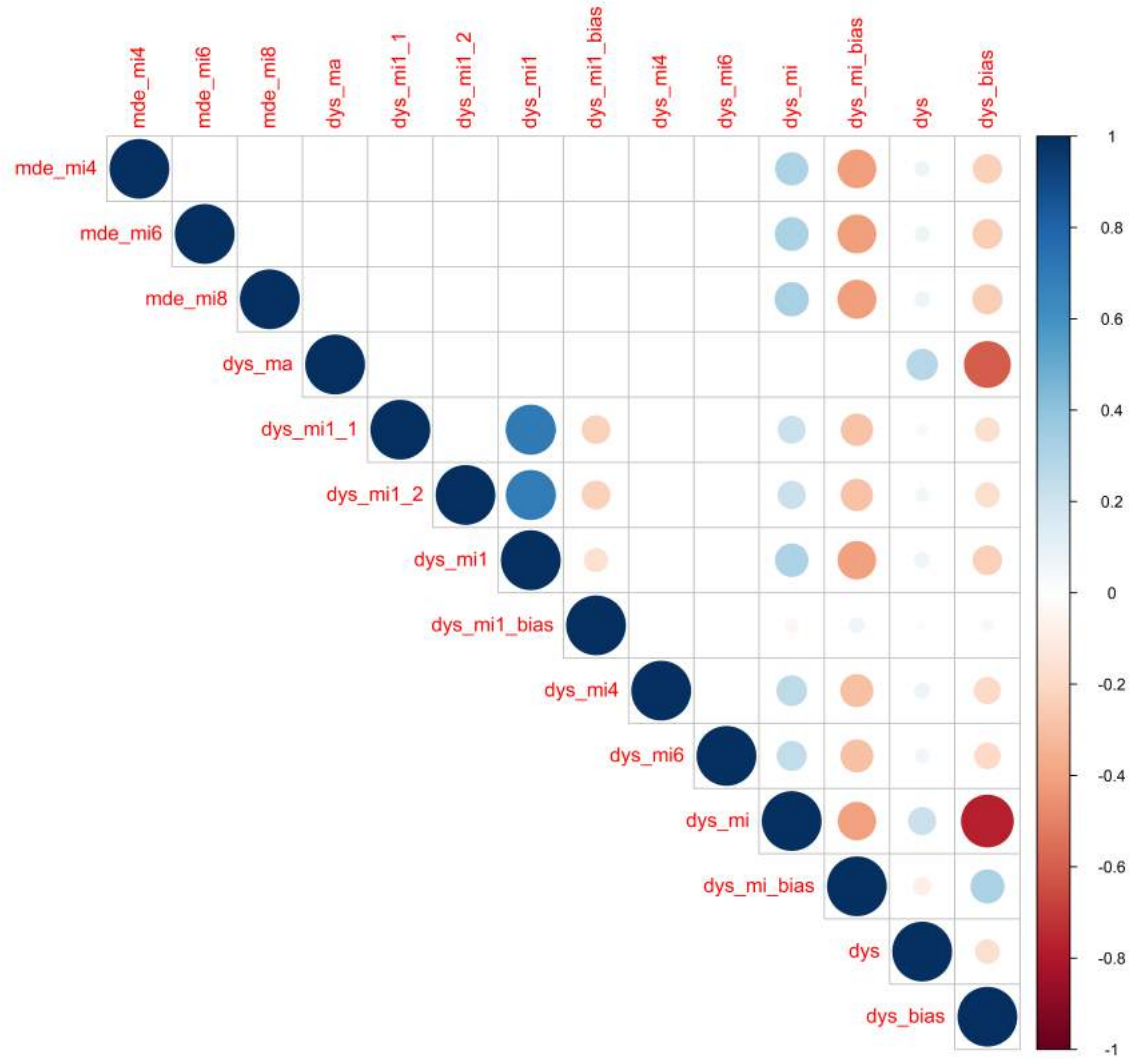

```

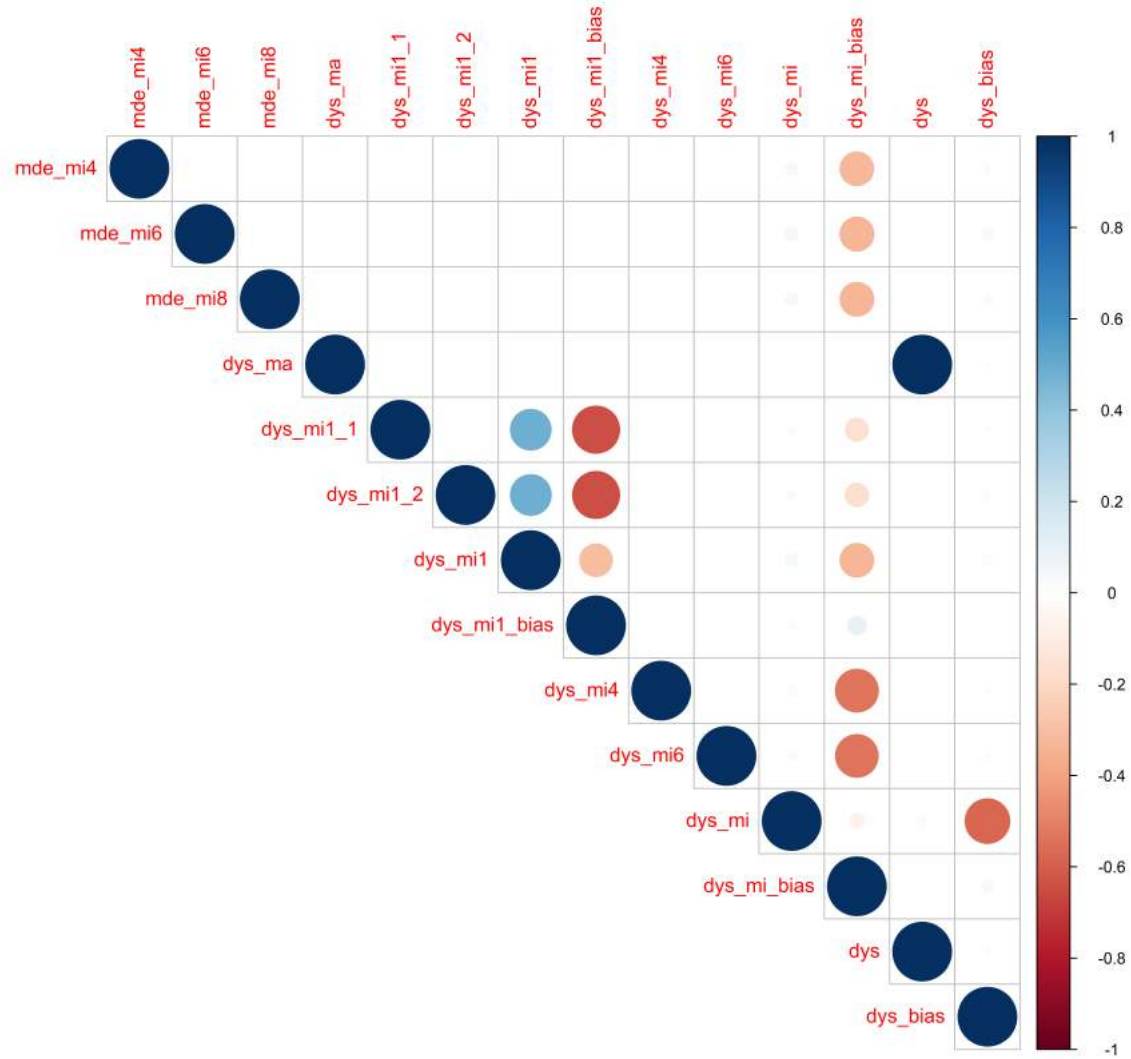

Derived prevalence
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Assumed prevalence
for major depressive episodes and dysthymic disorder

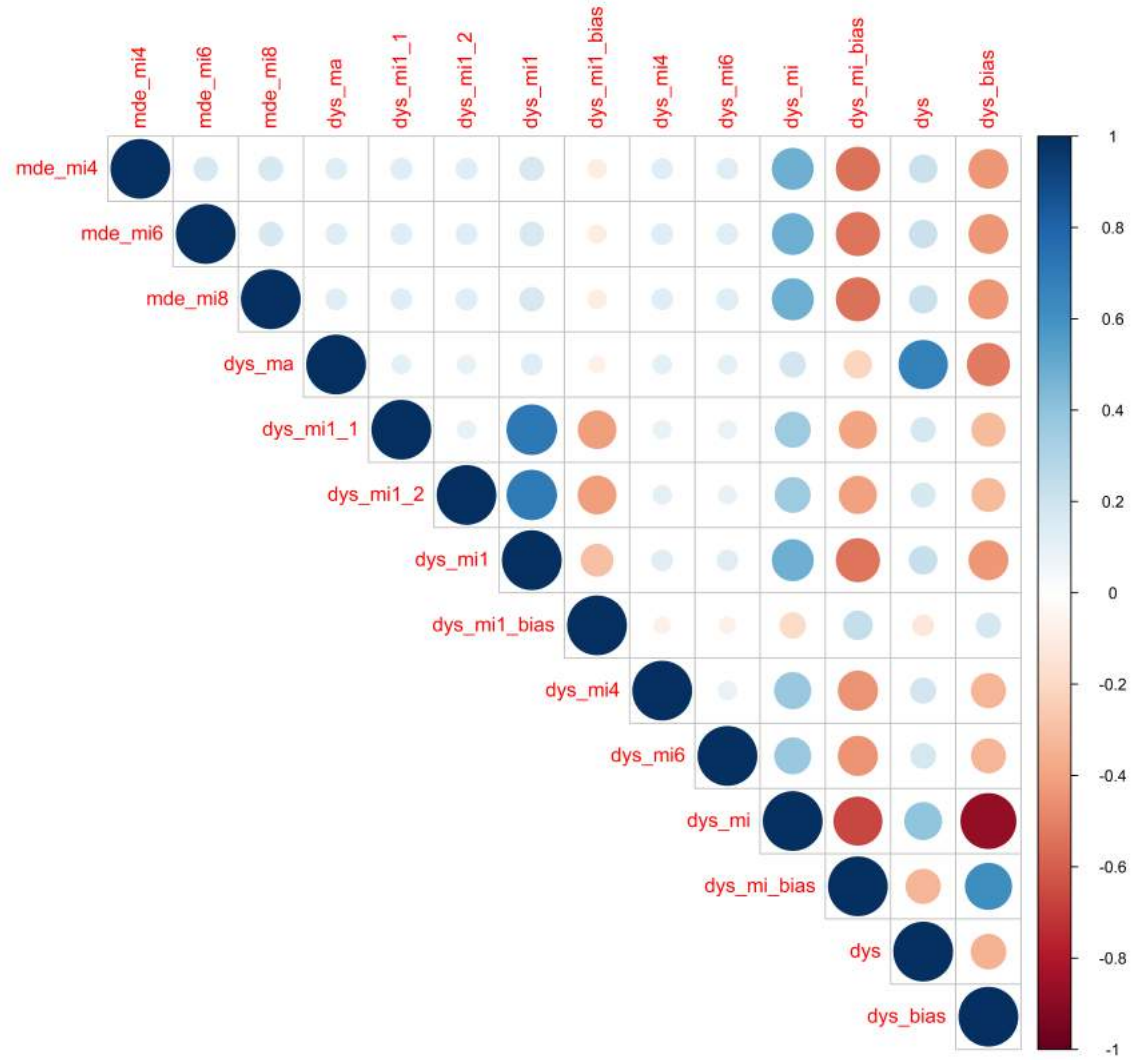


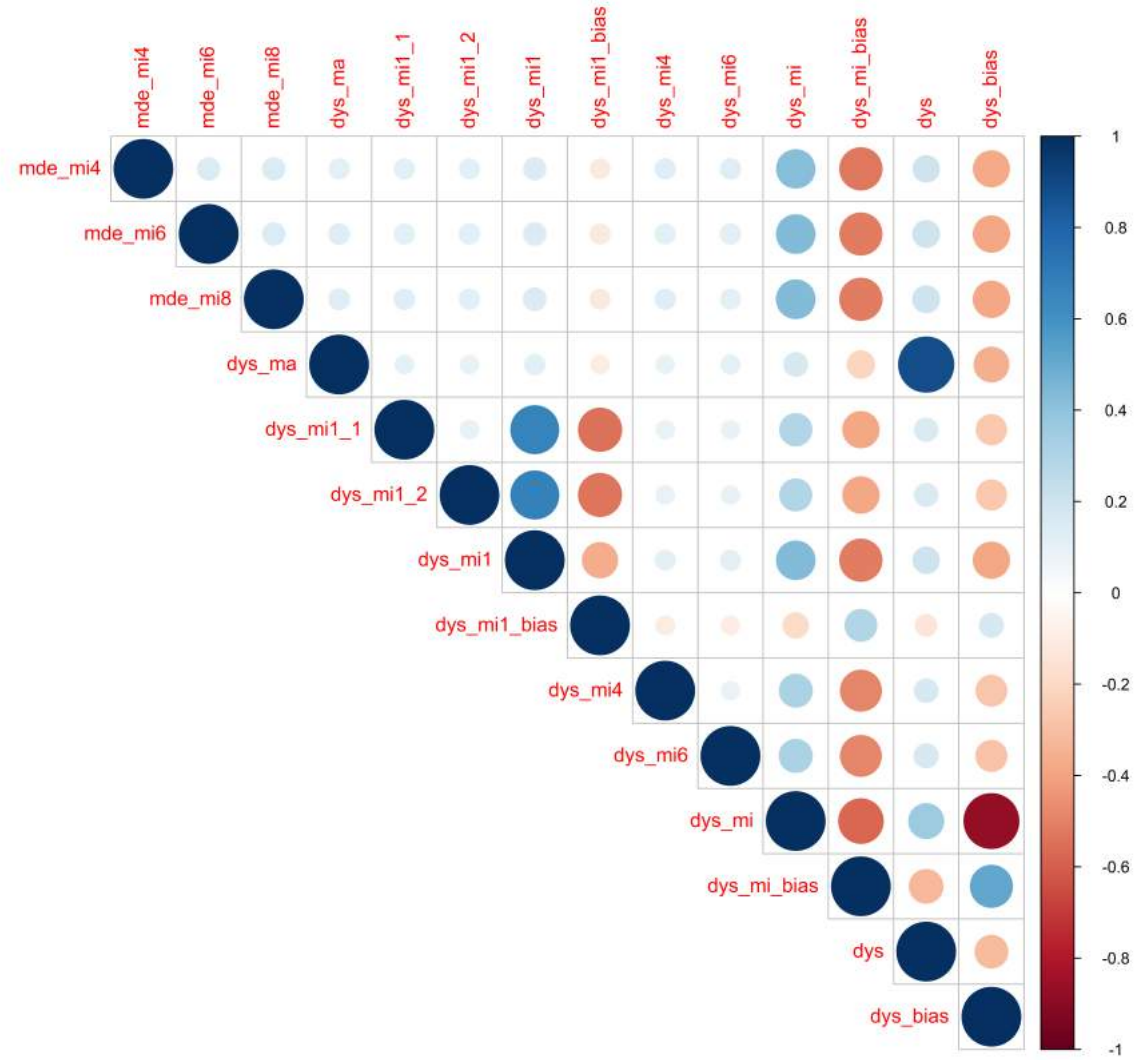

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

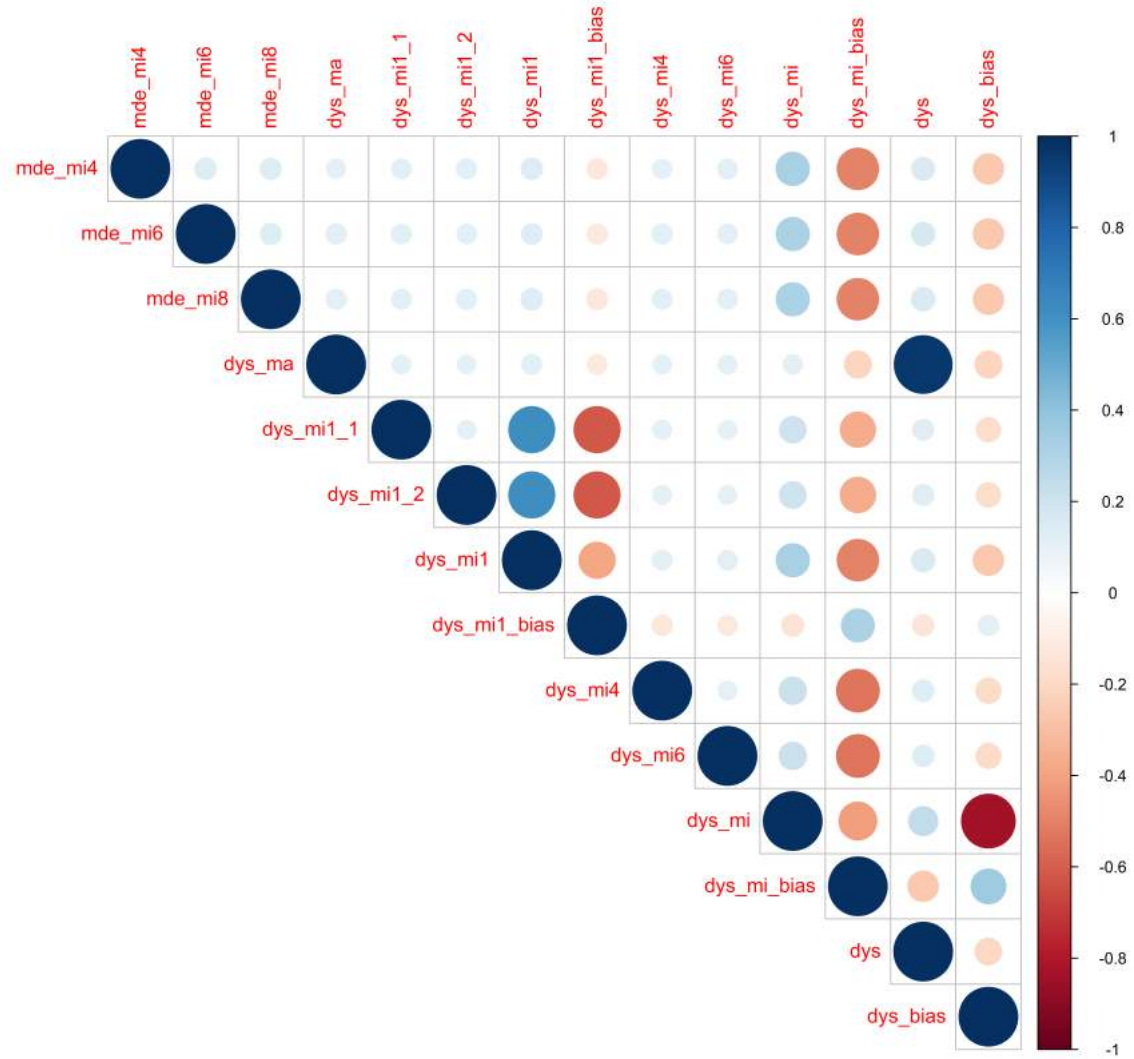

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

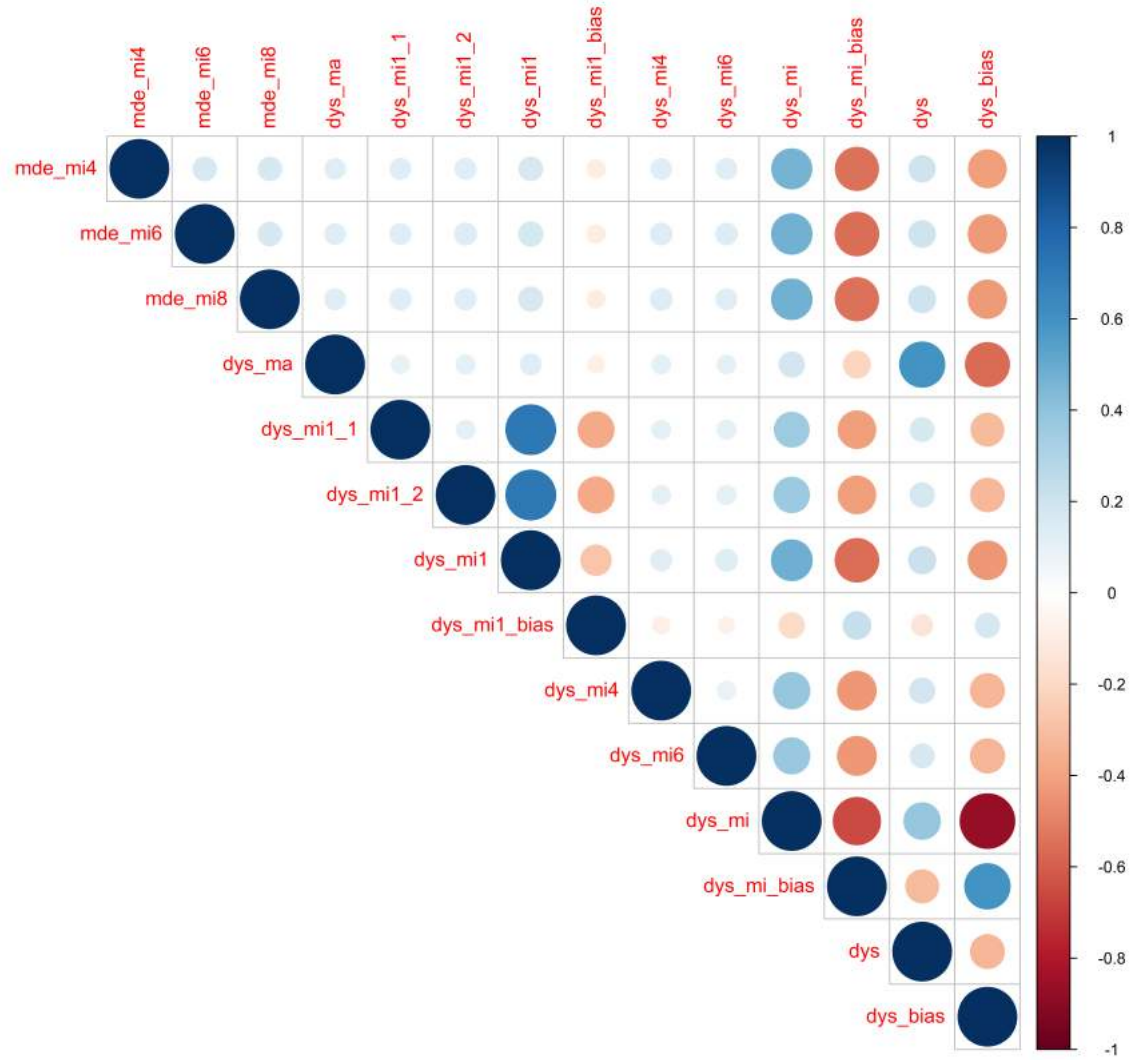

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

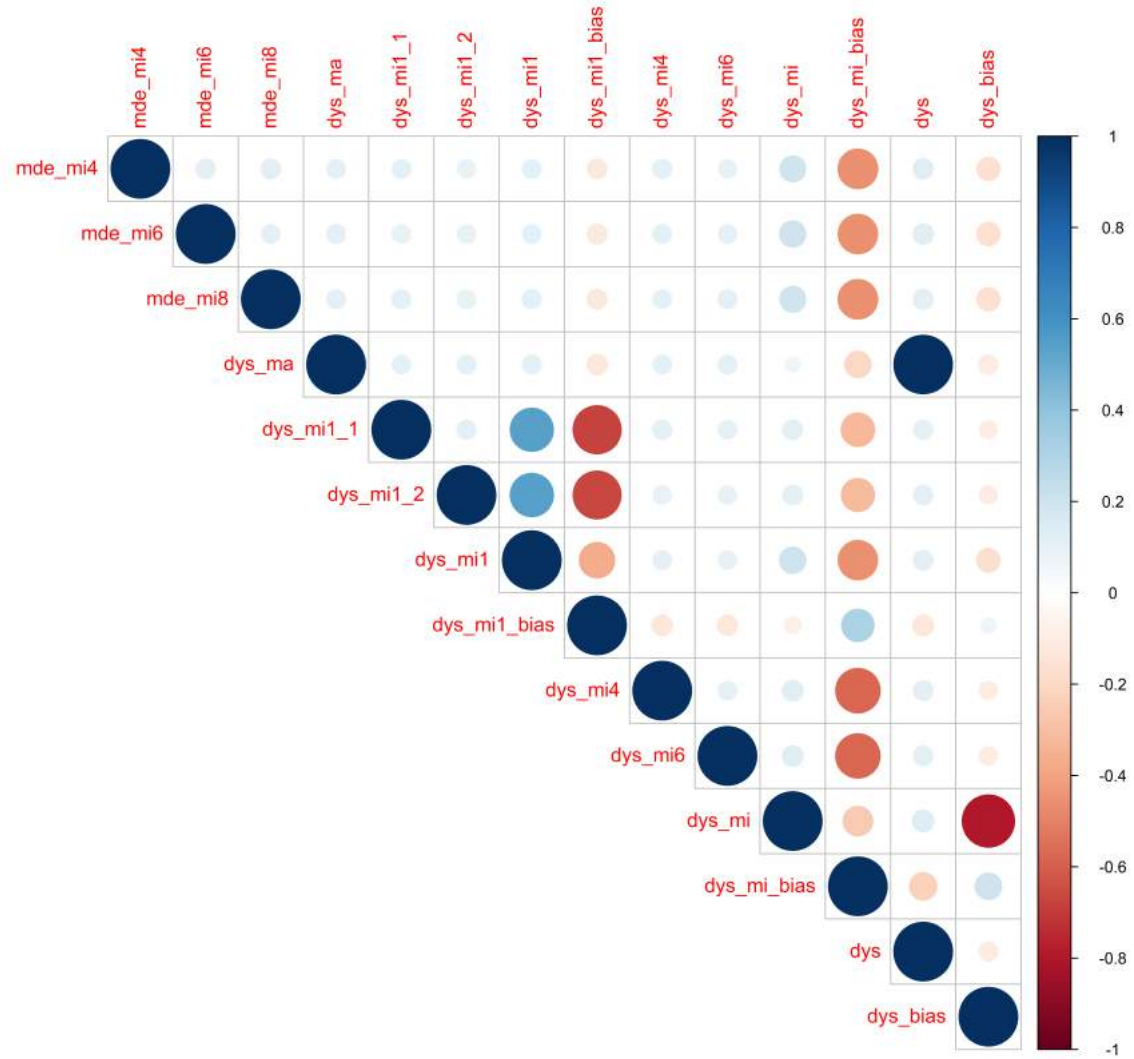

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

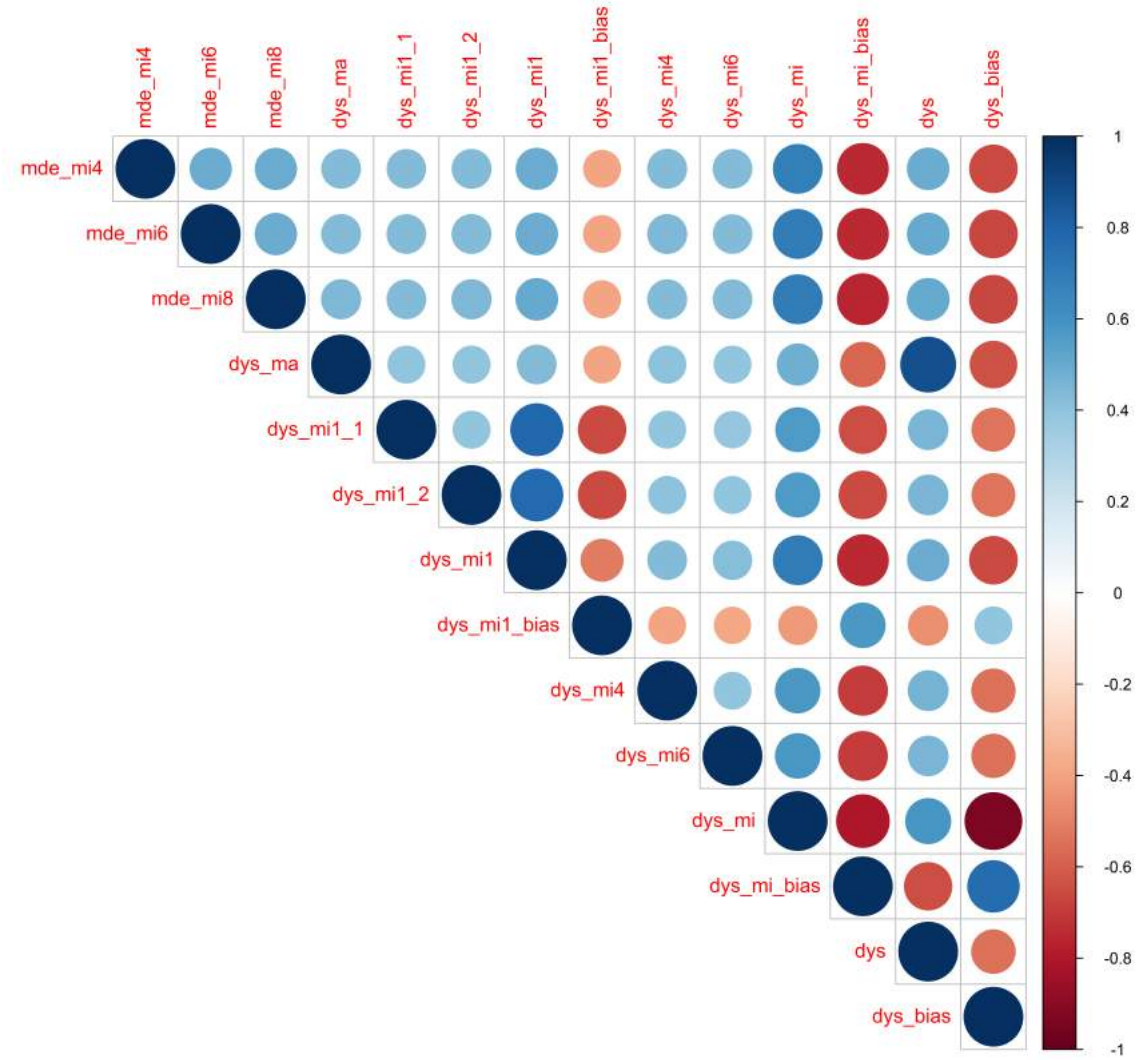

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

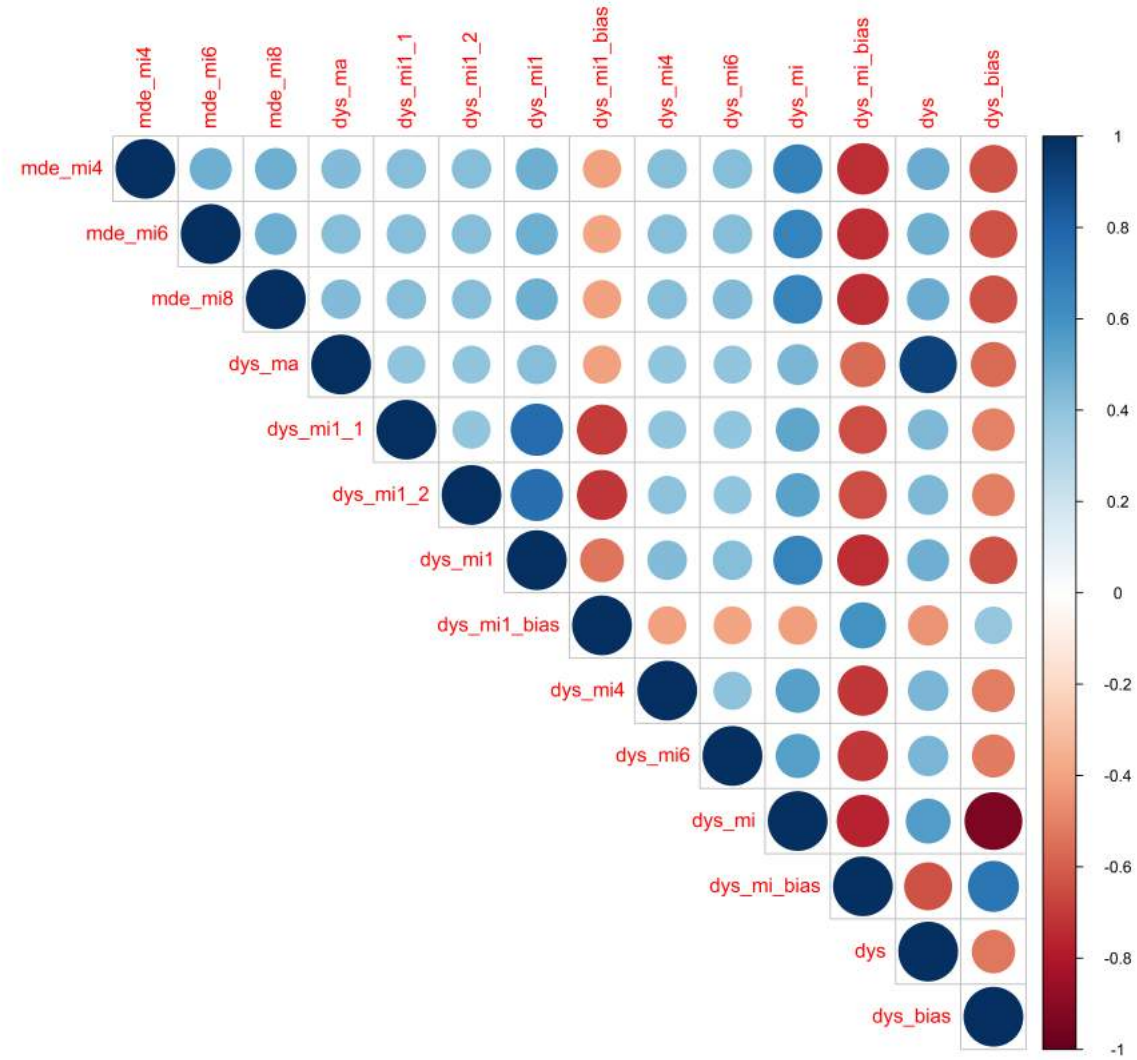


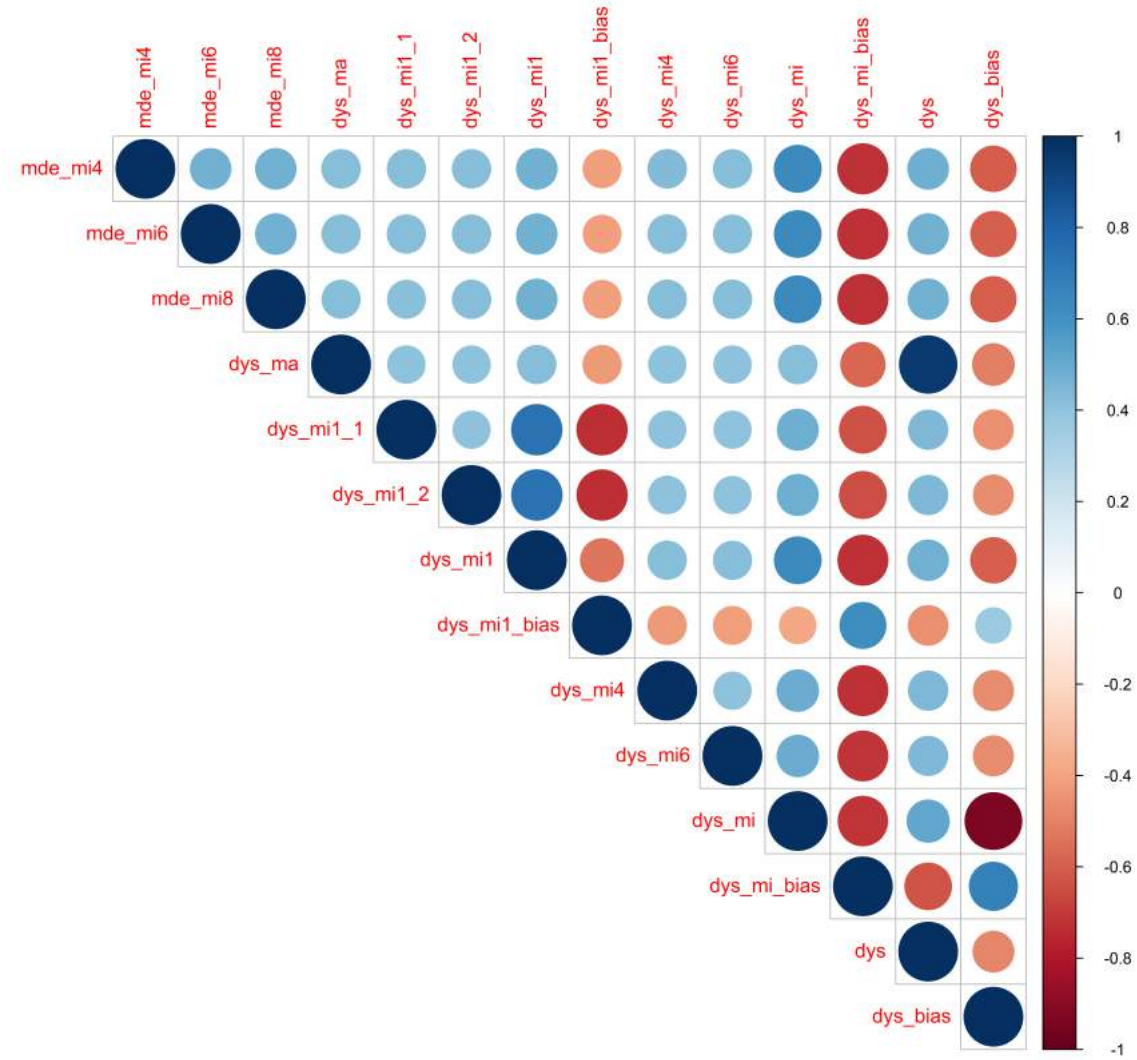

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

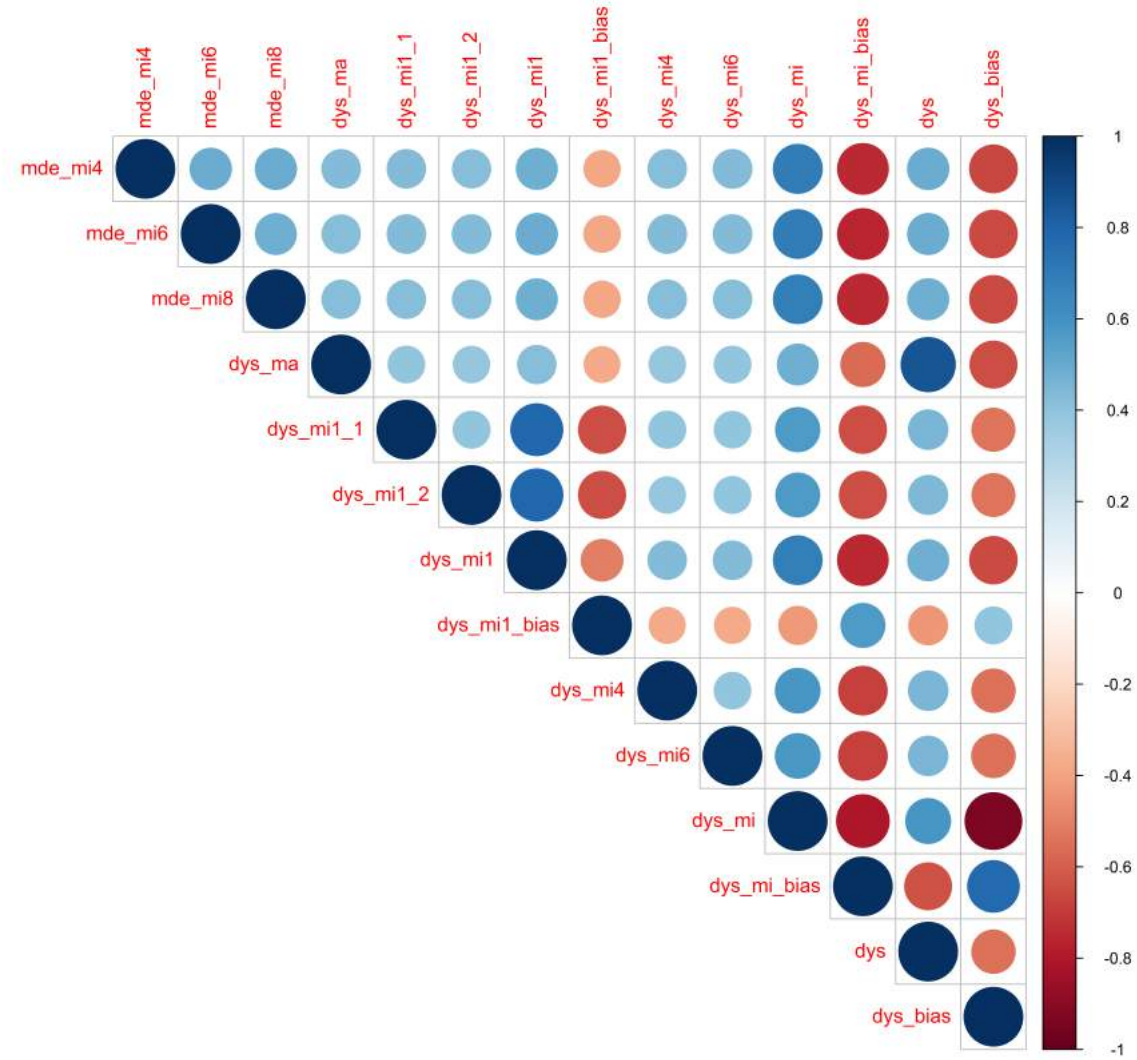


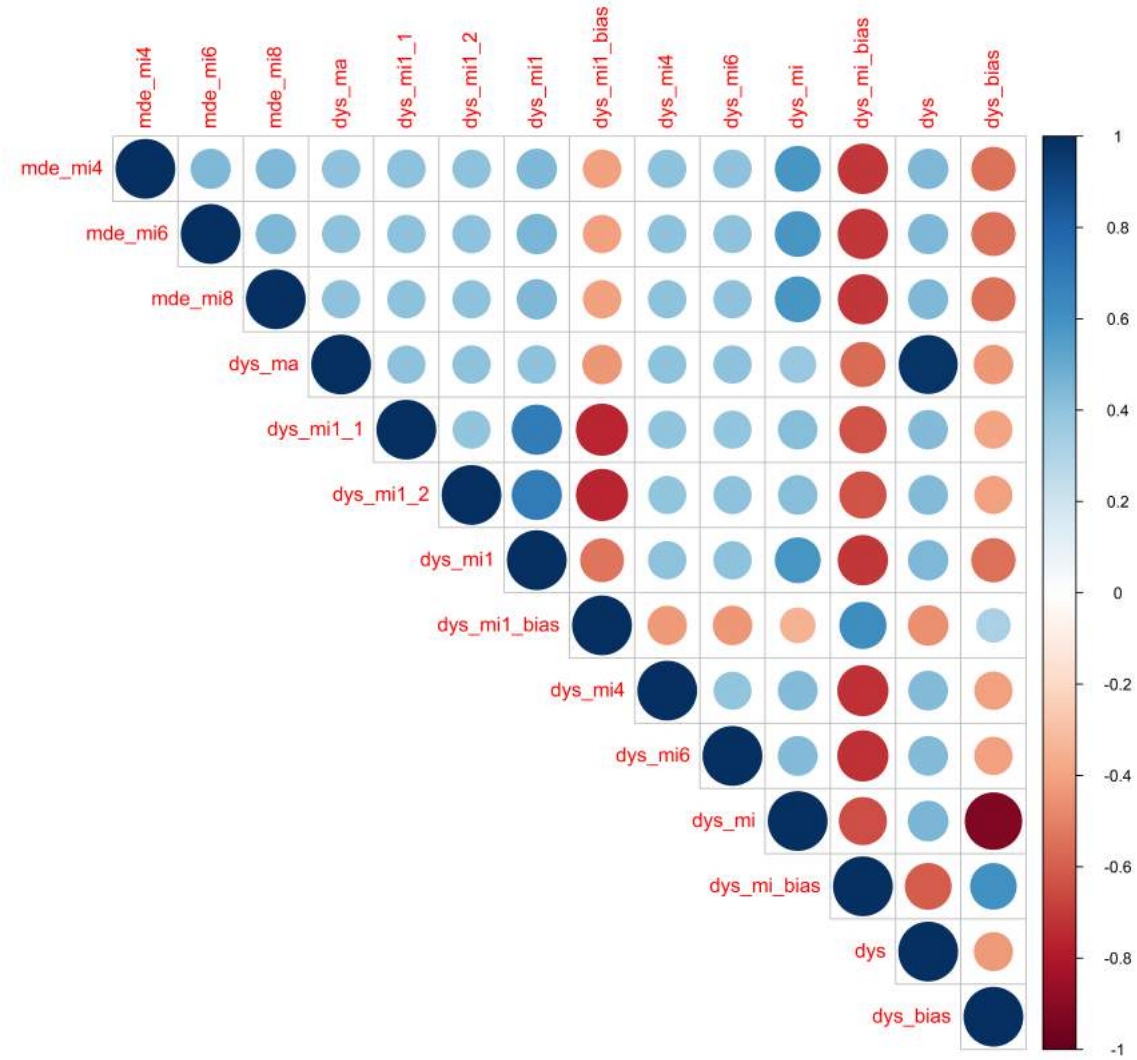

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

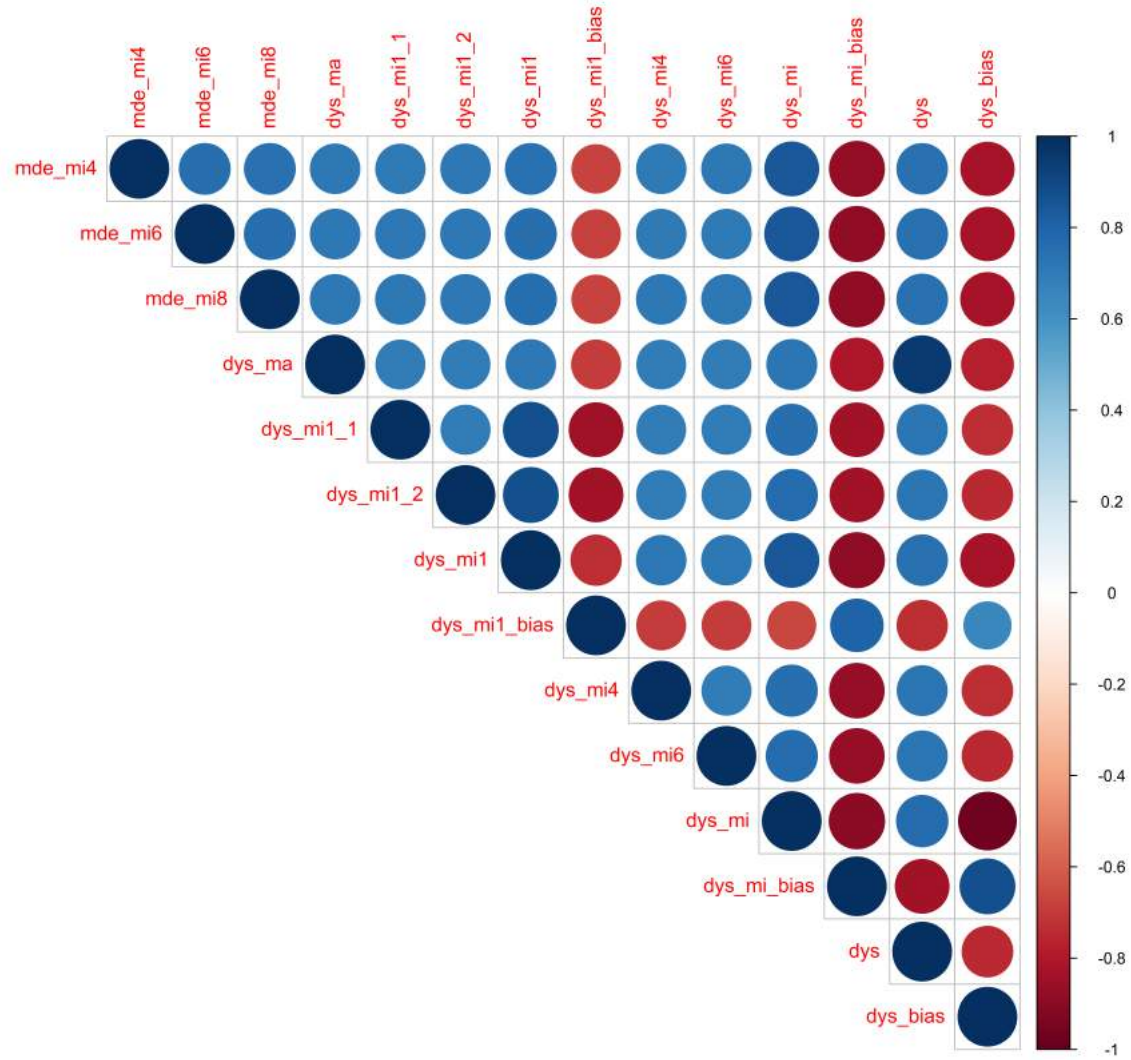


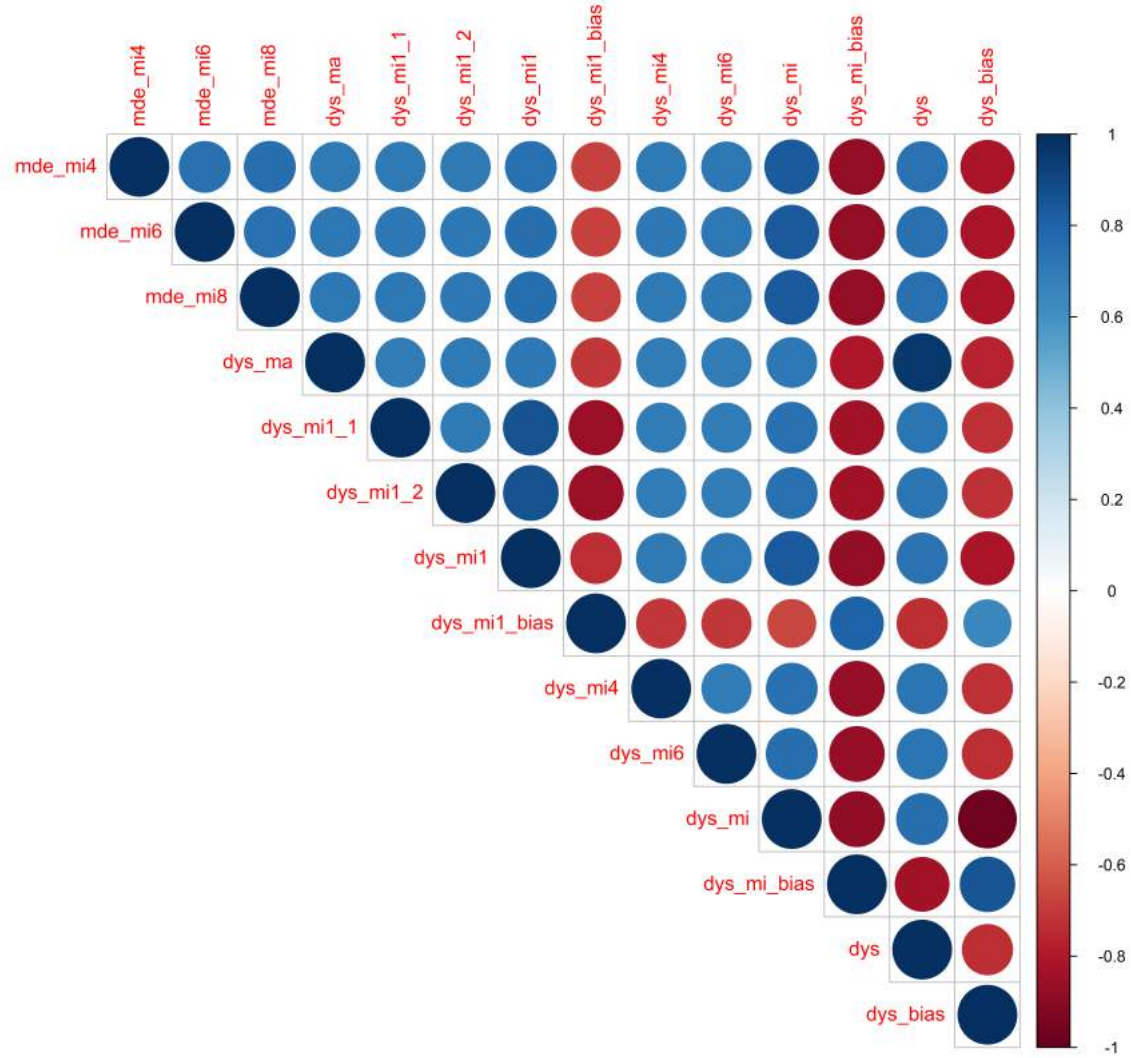


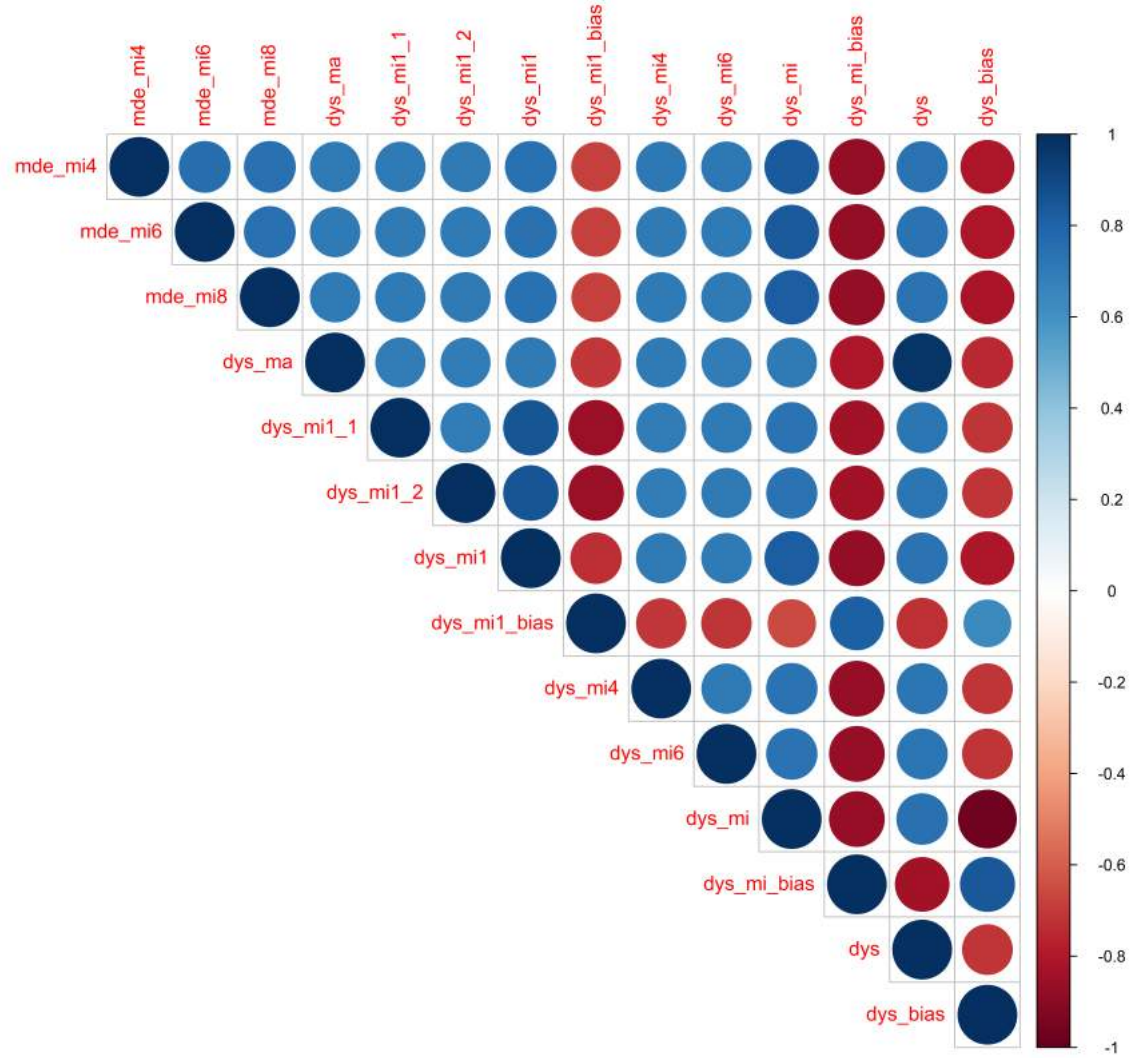


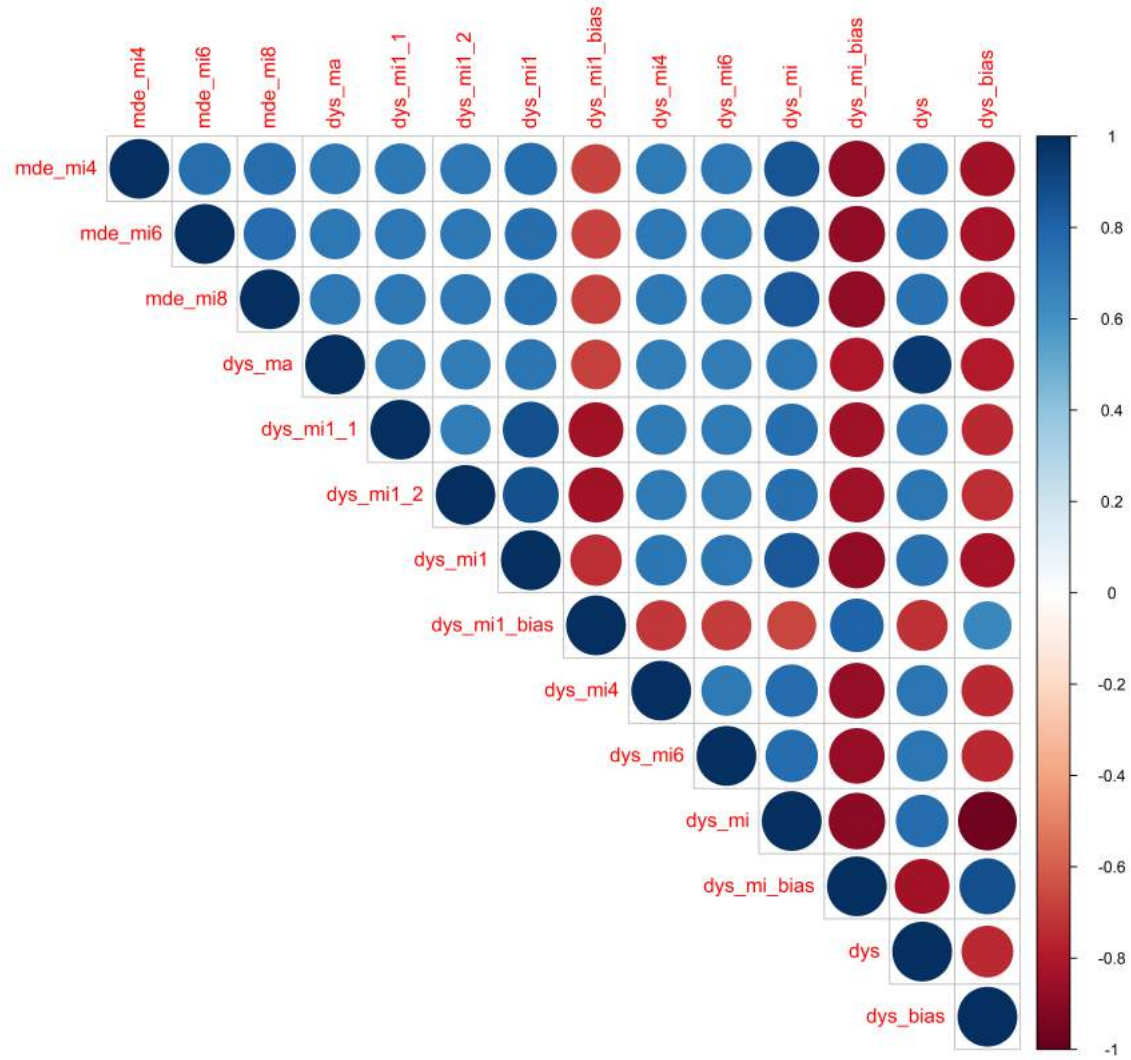


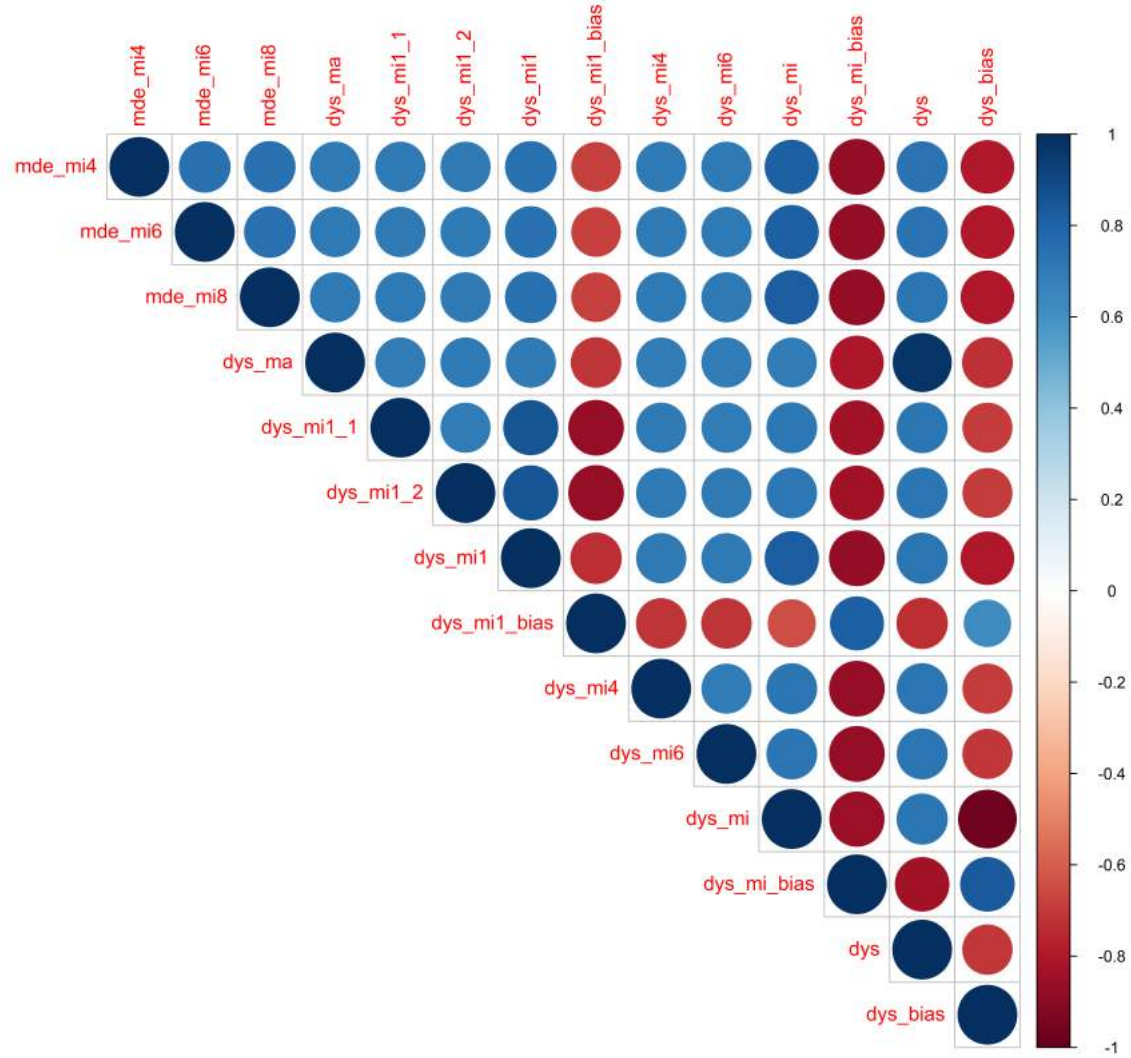


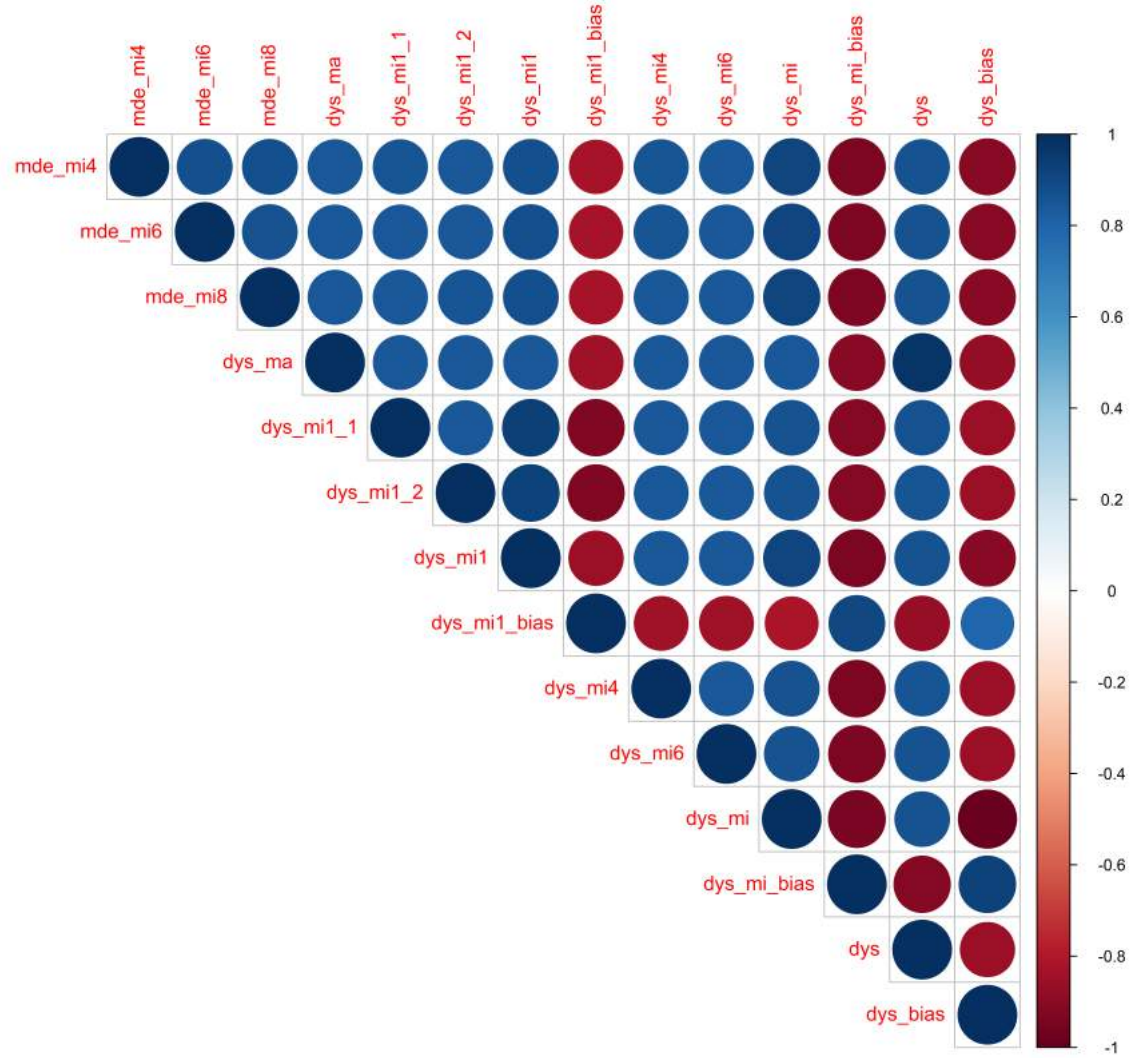

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

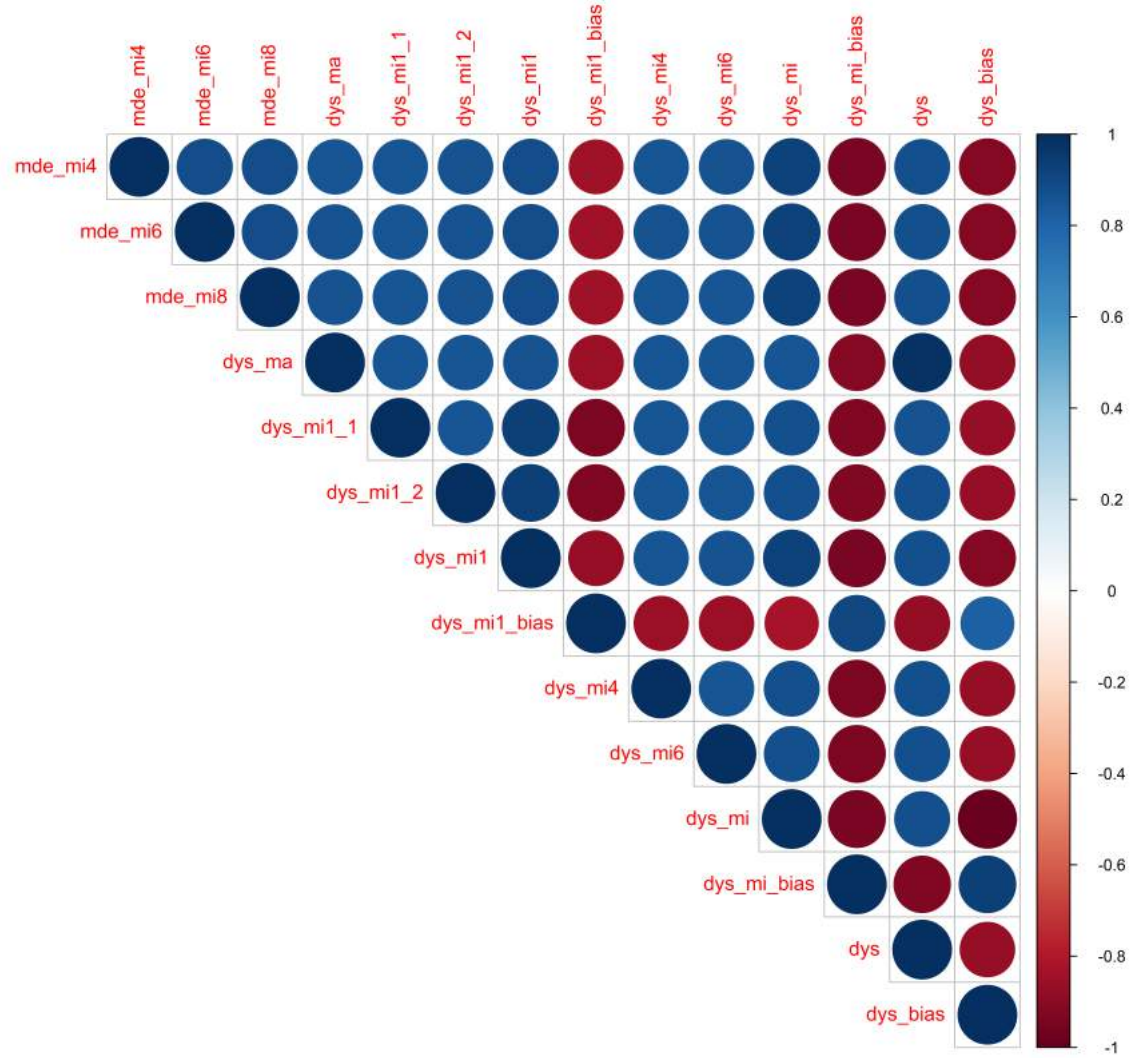


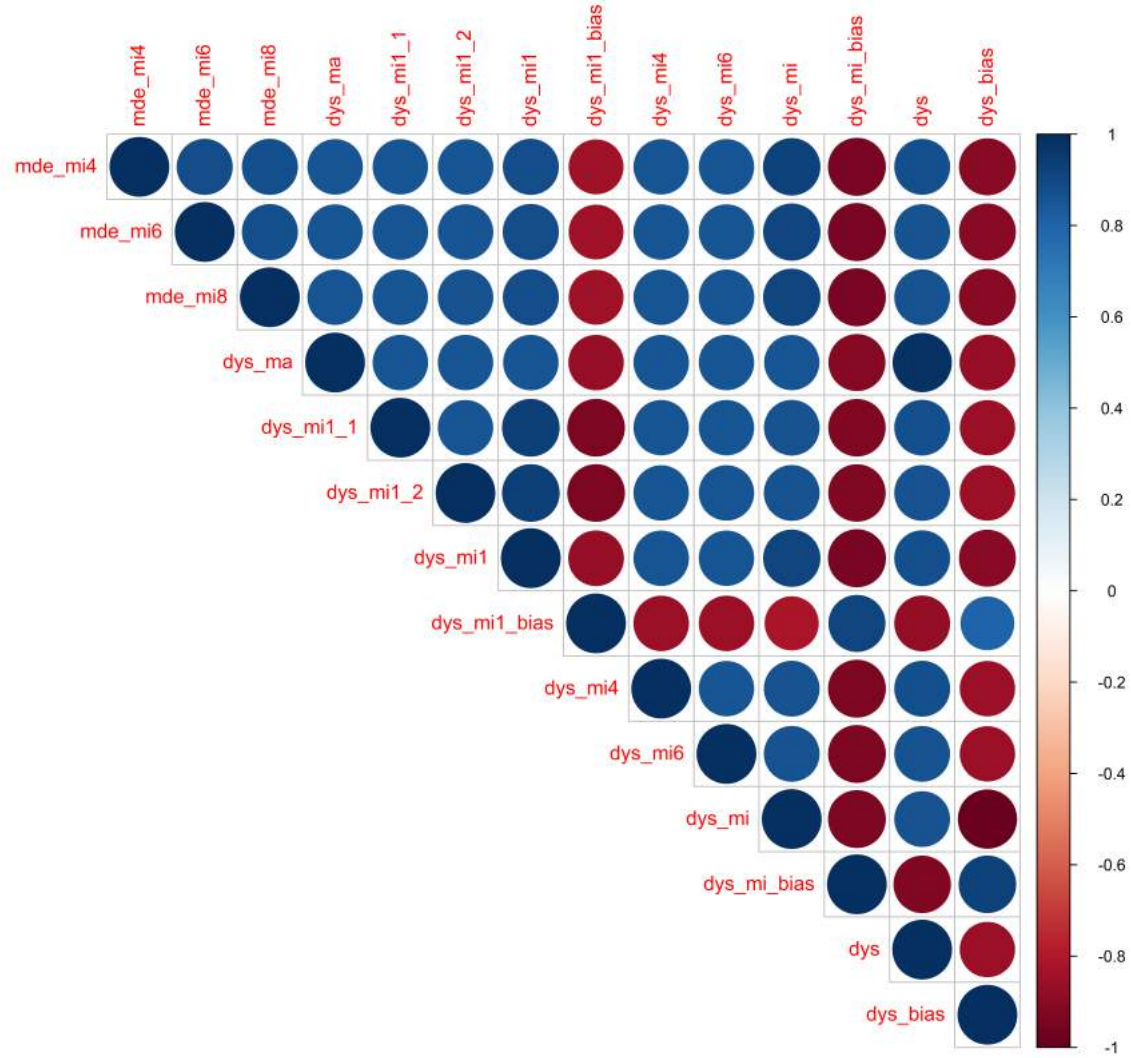

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

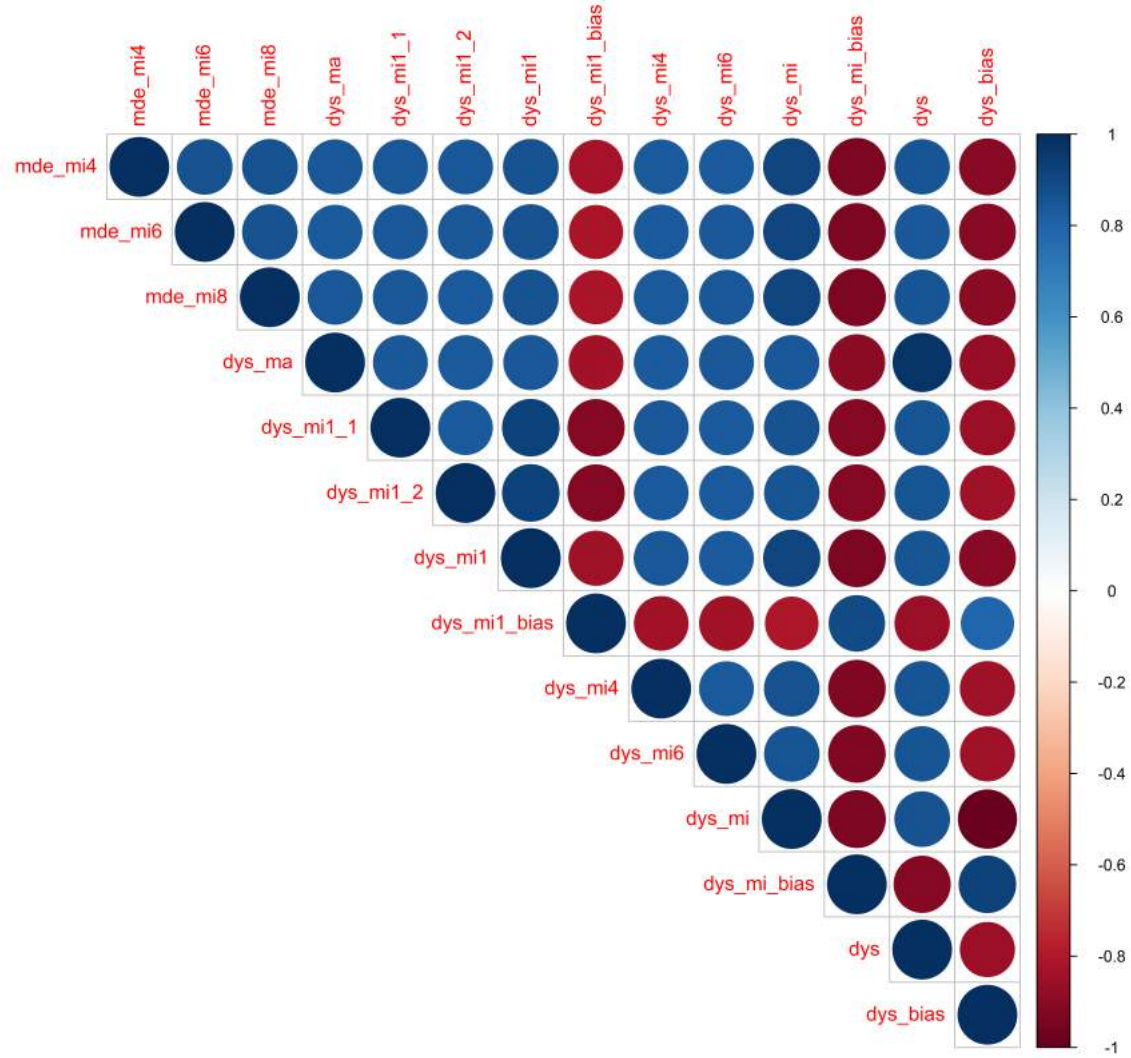


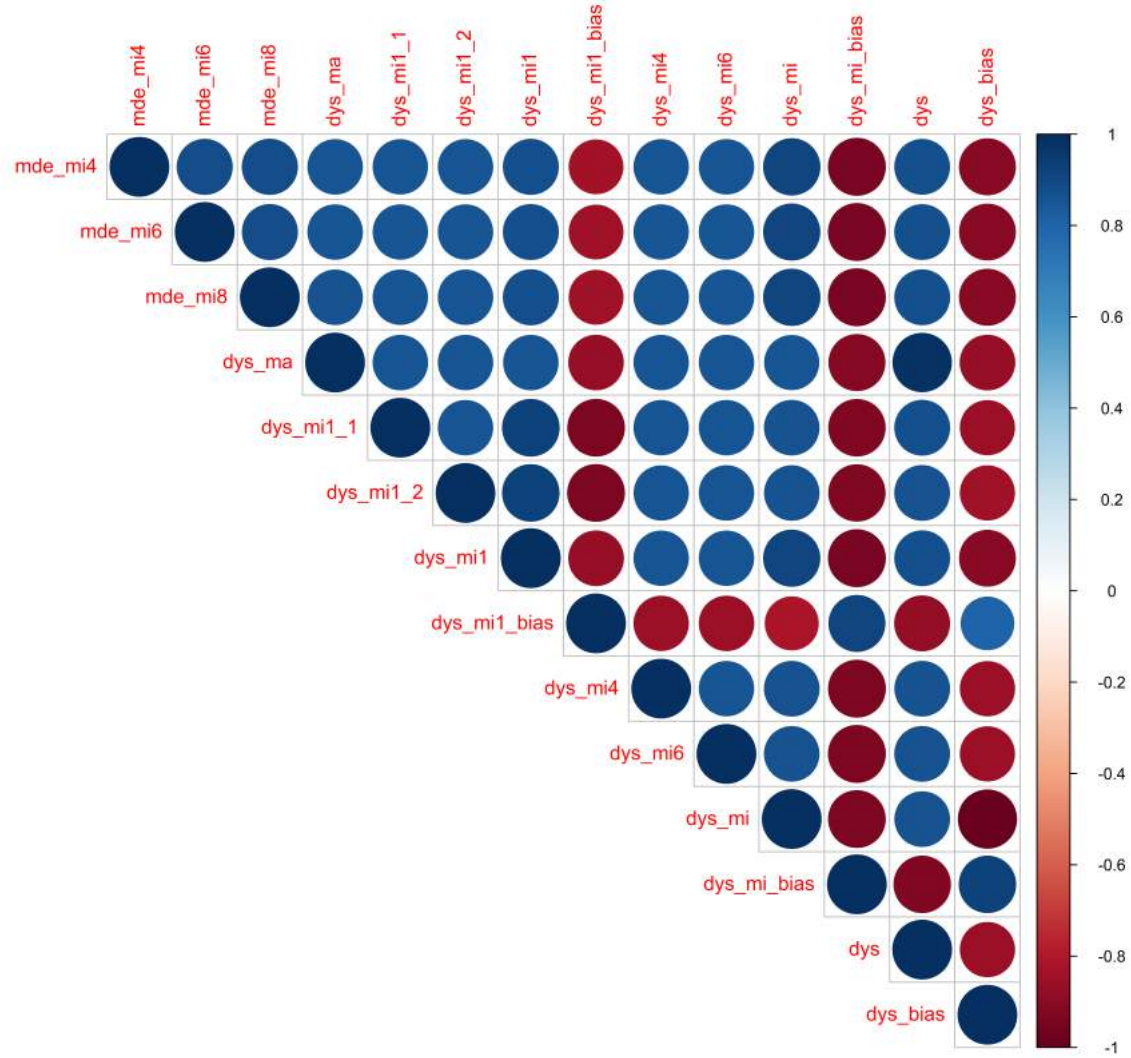

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

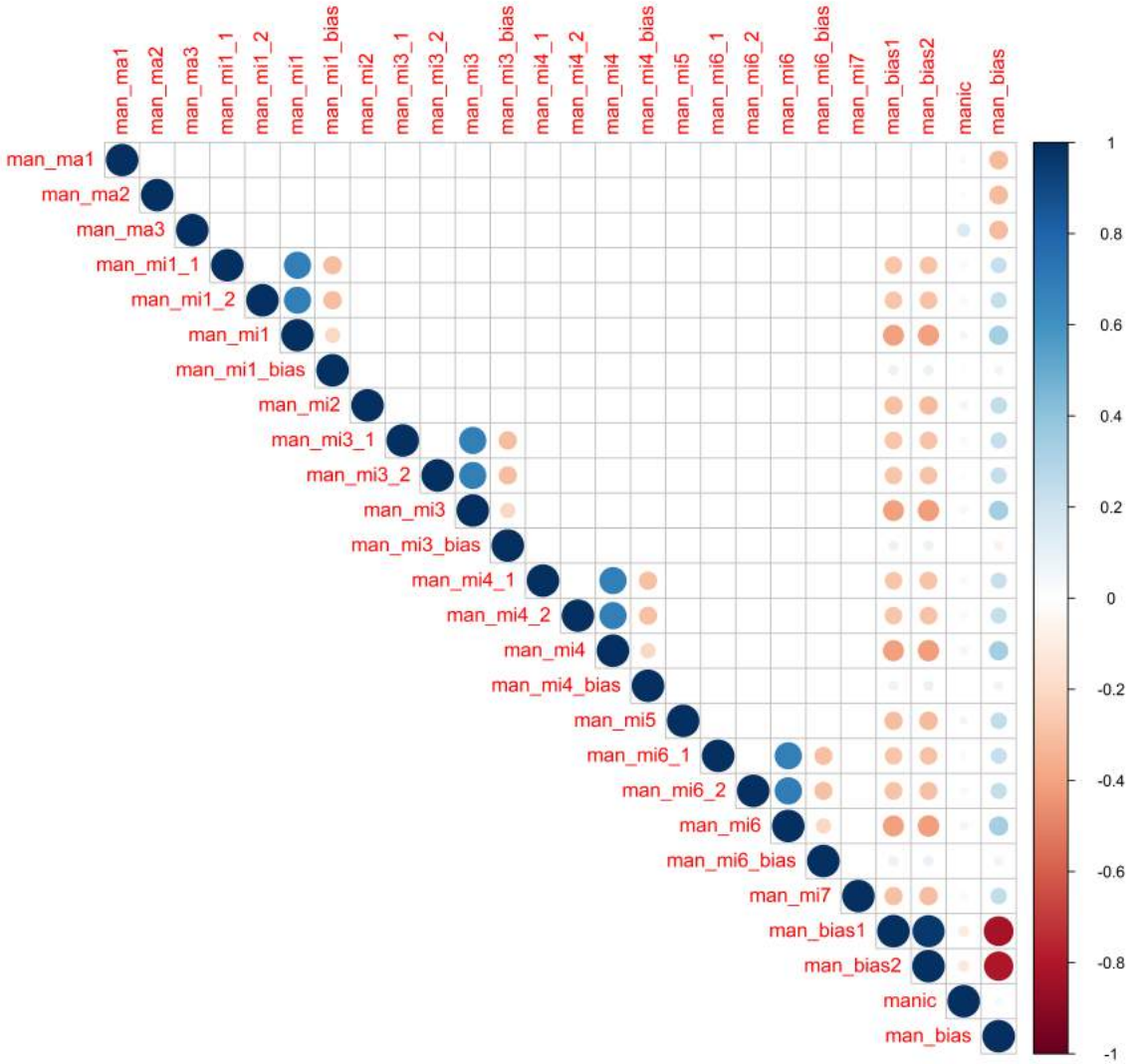

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

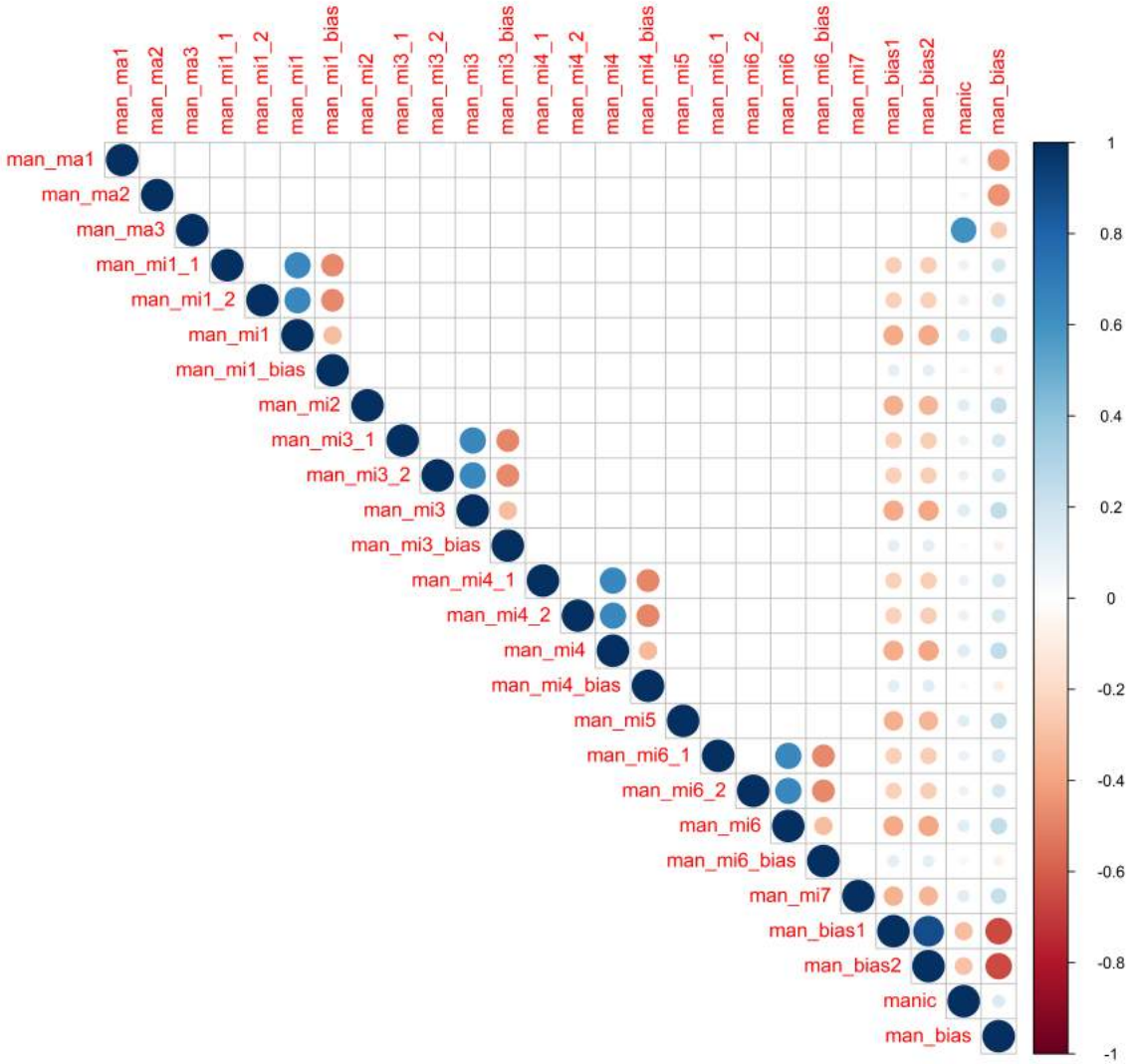

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

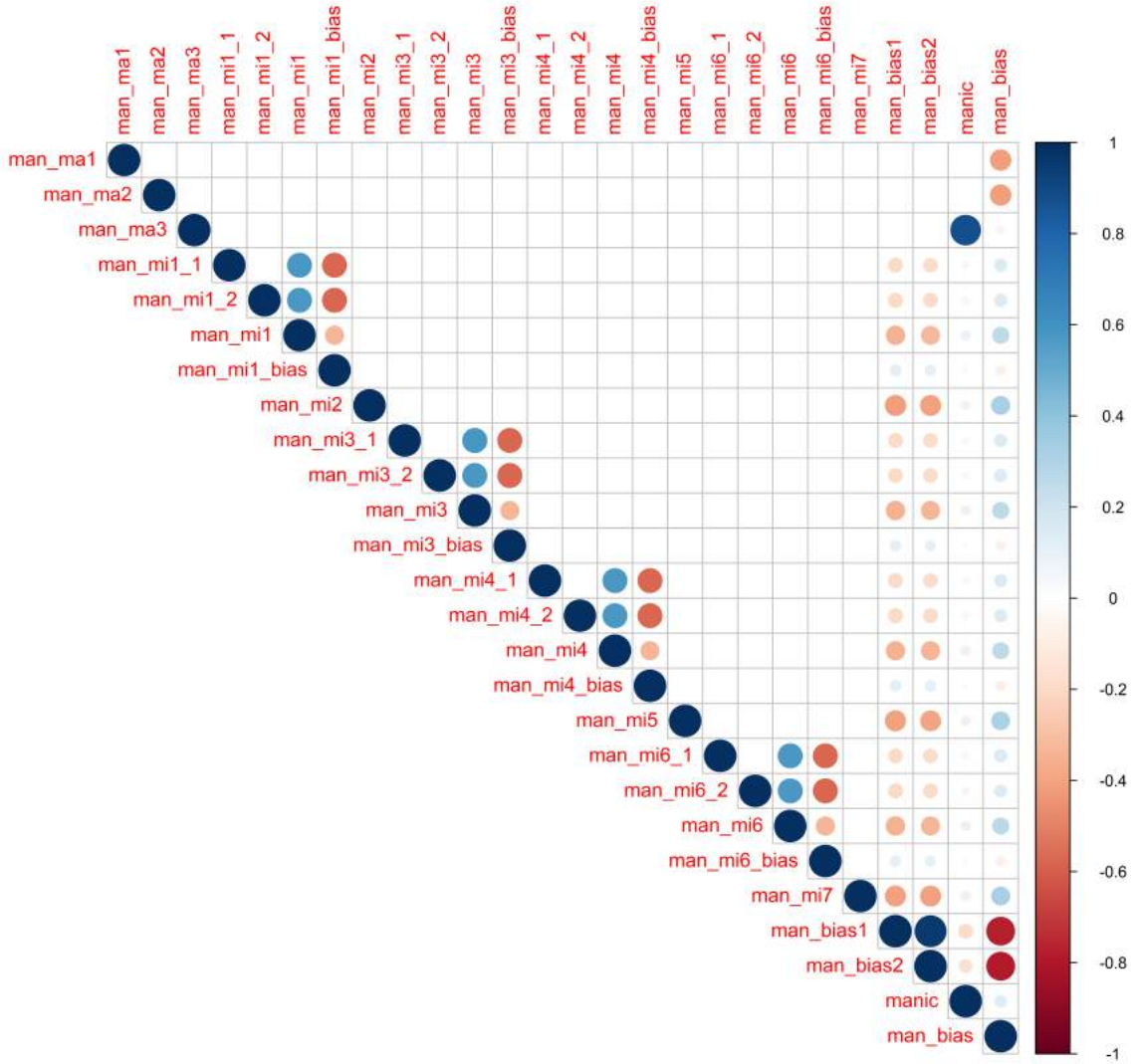

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

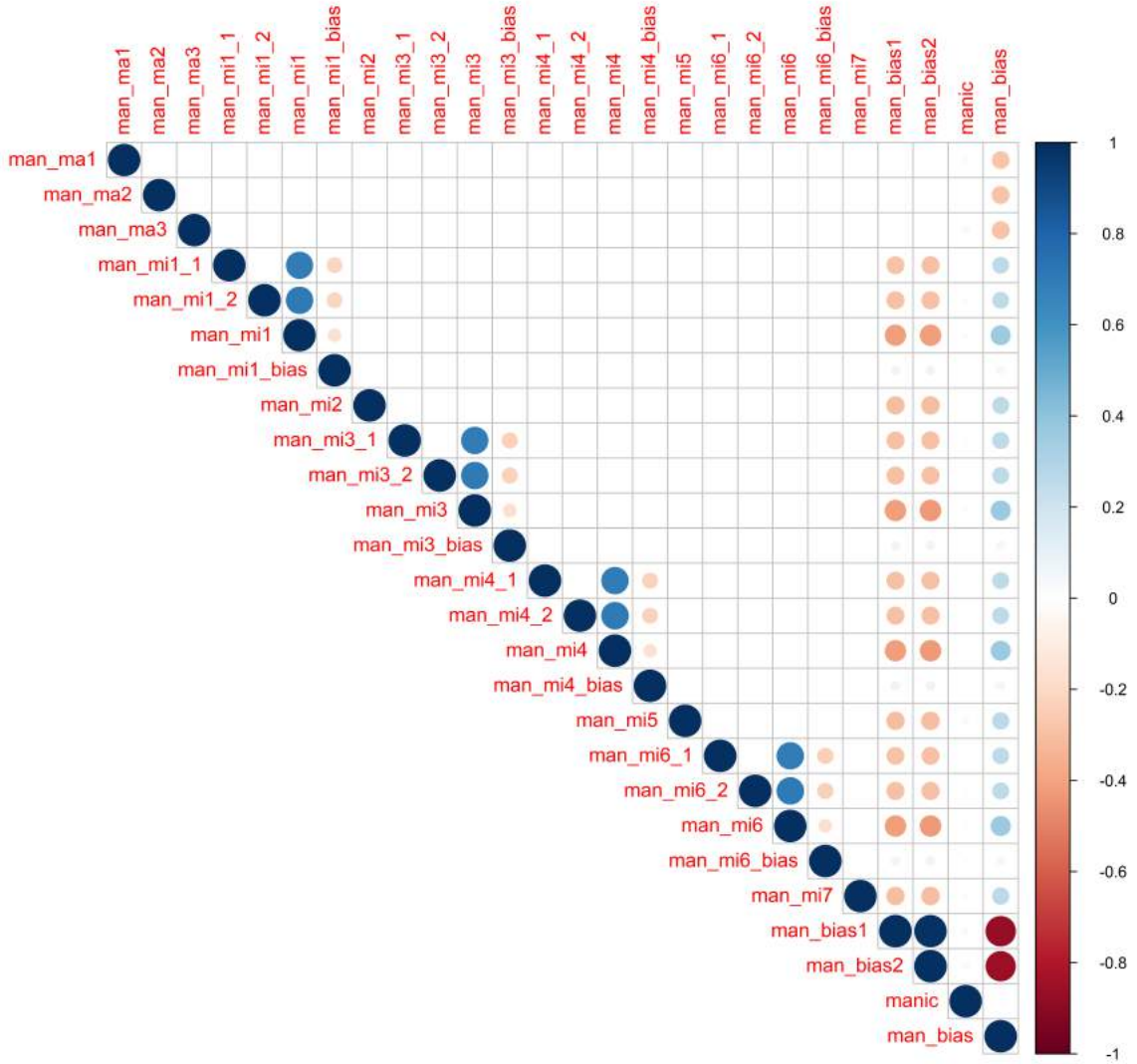

mde_mi4
mde_mi6
mde_mi8
dys_ma
dys_mi1_1
dys_mi1_2
dys_mi1
dys_mi1_bias
dys_mi4
dys_mi6
dys_mi
dys_mi_bias
dys
dys_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

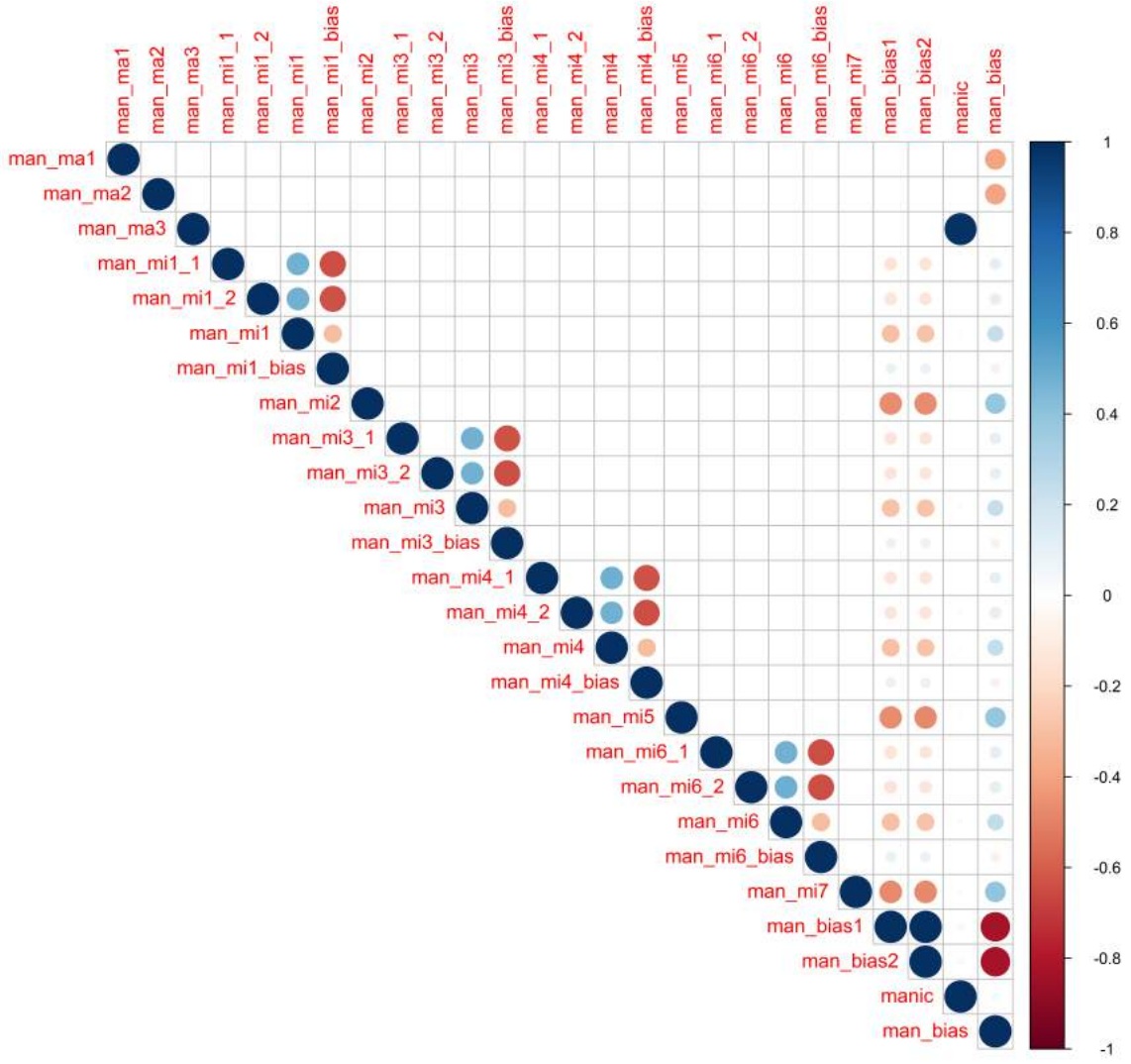


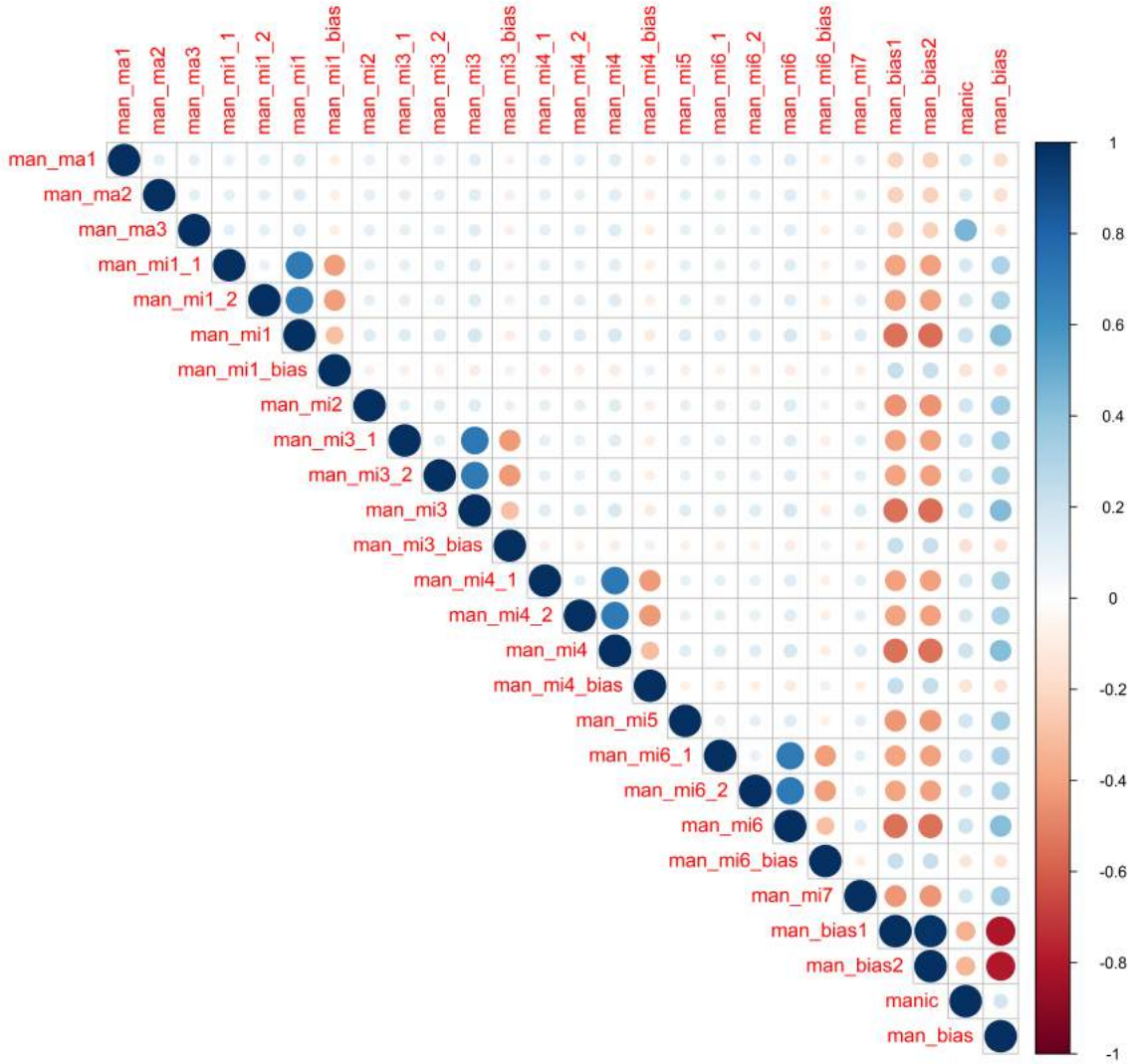


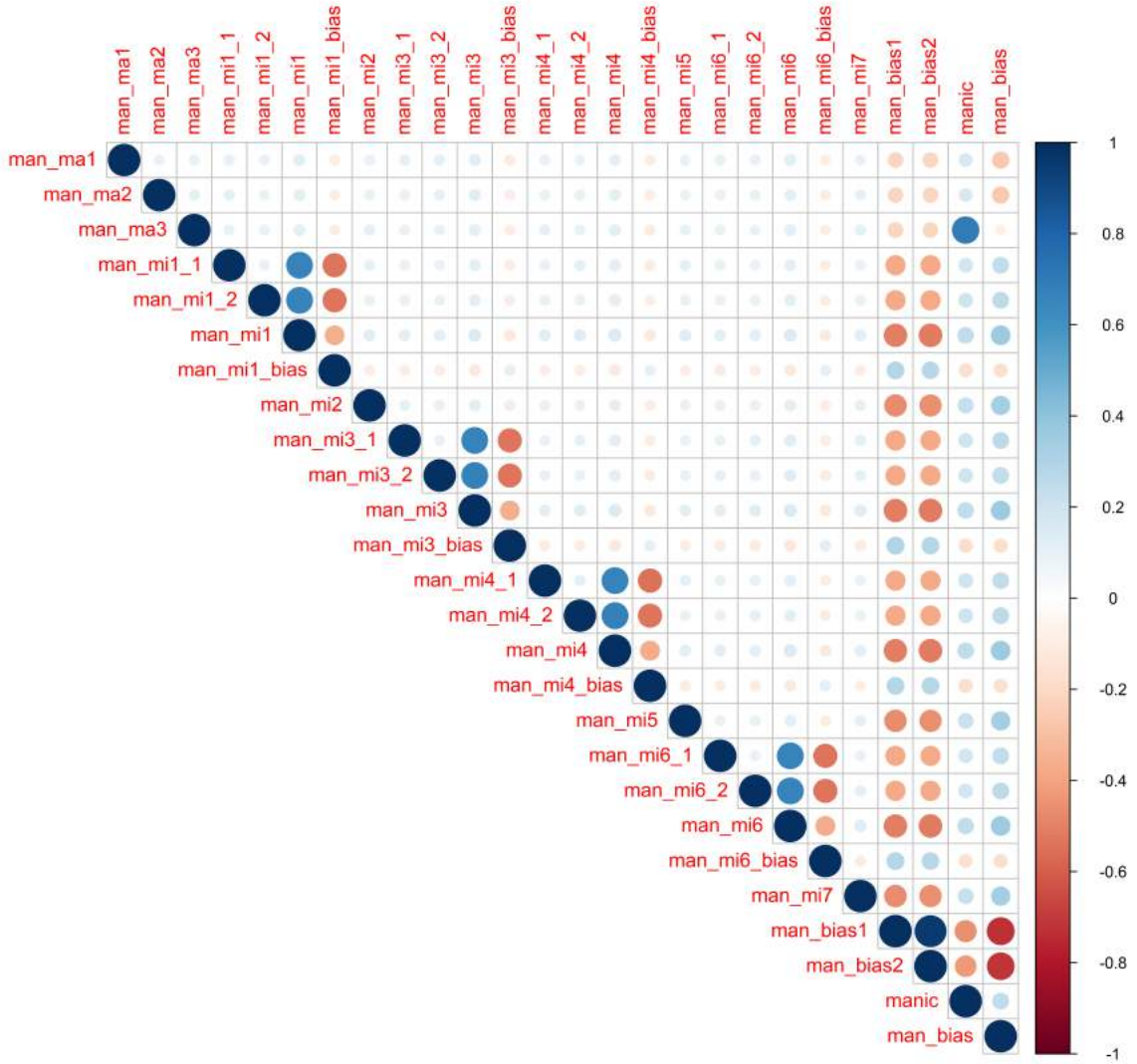

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

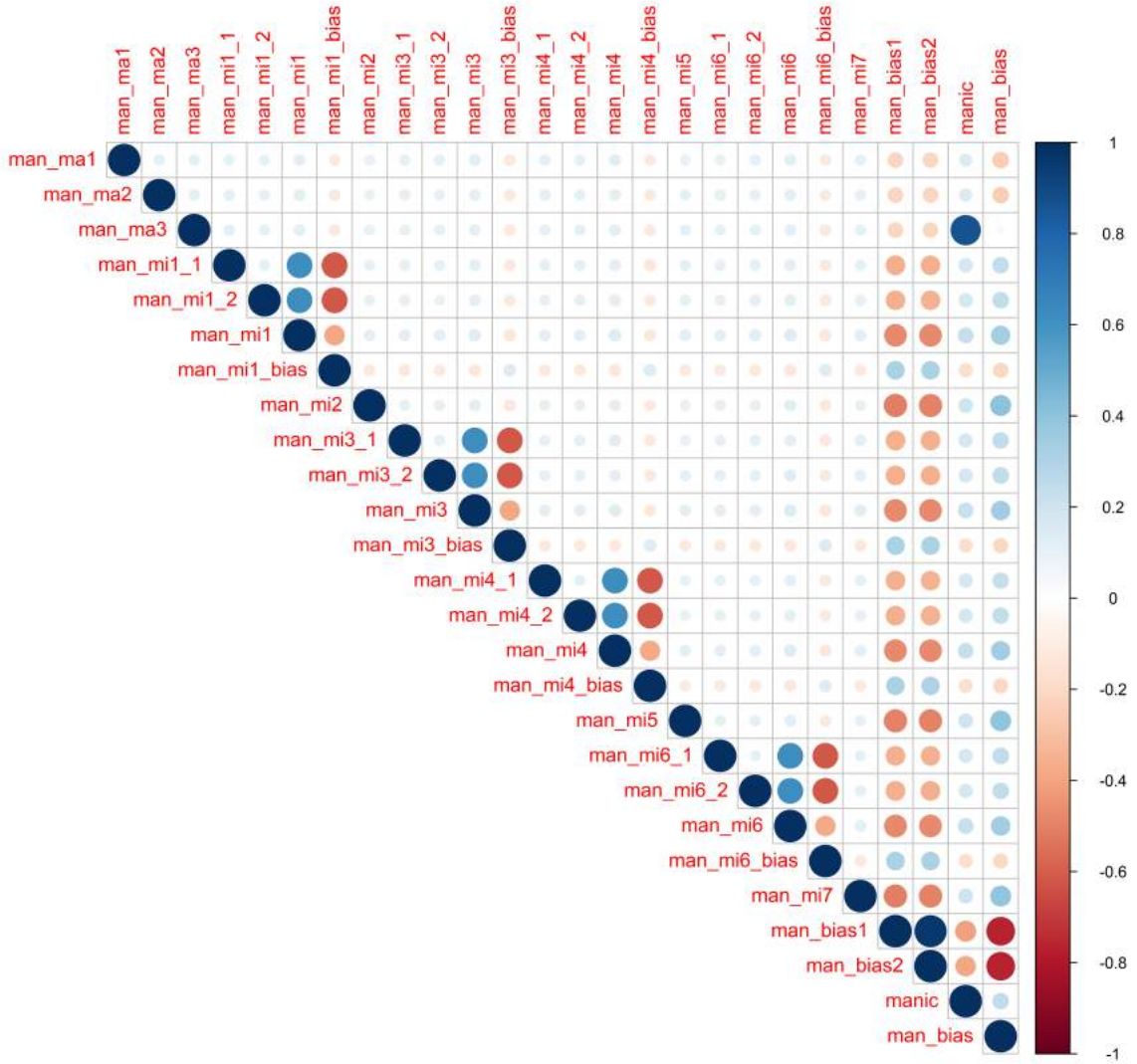


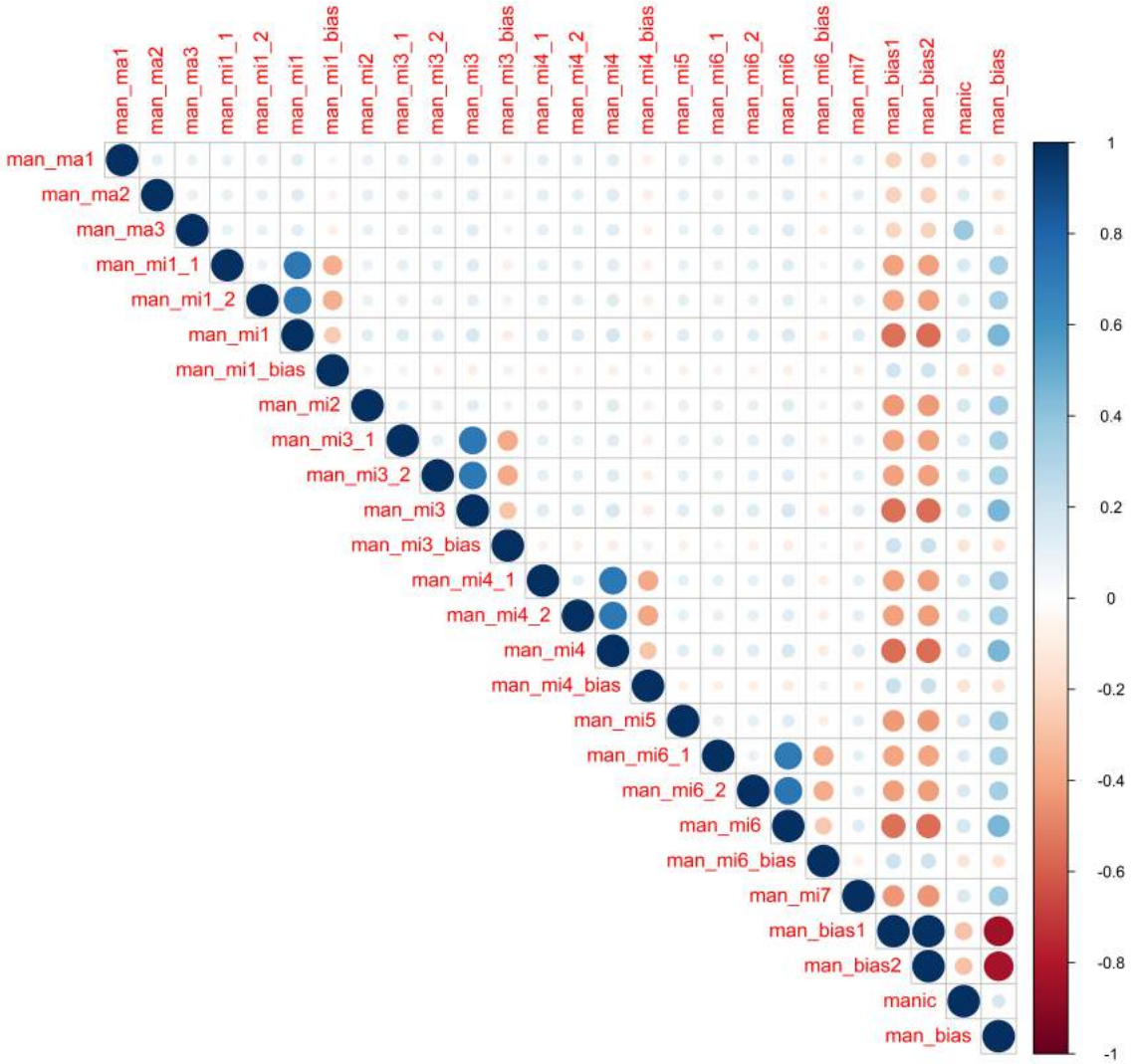

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

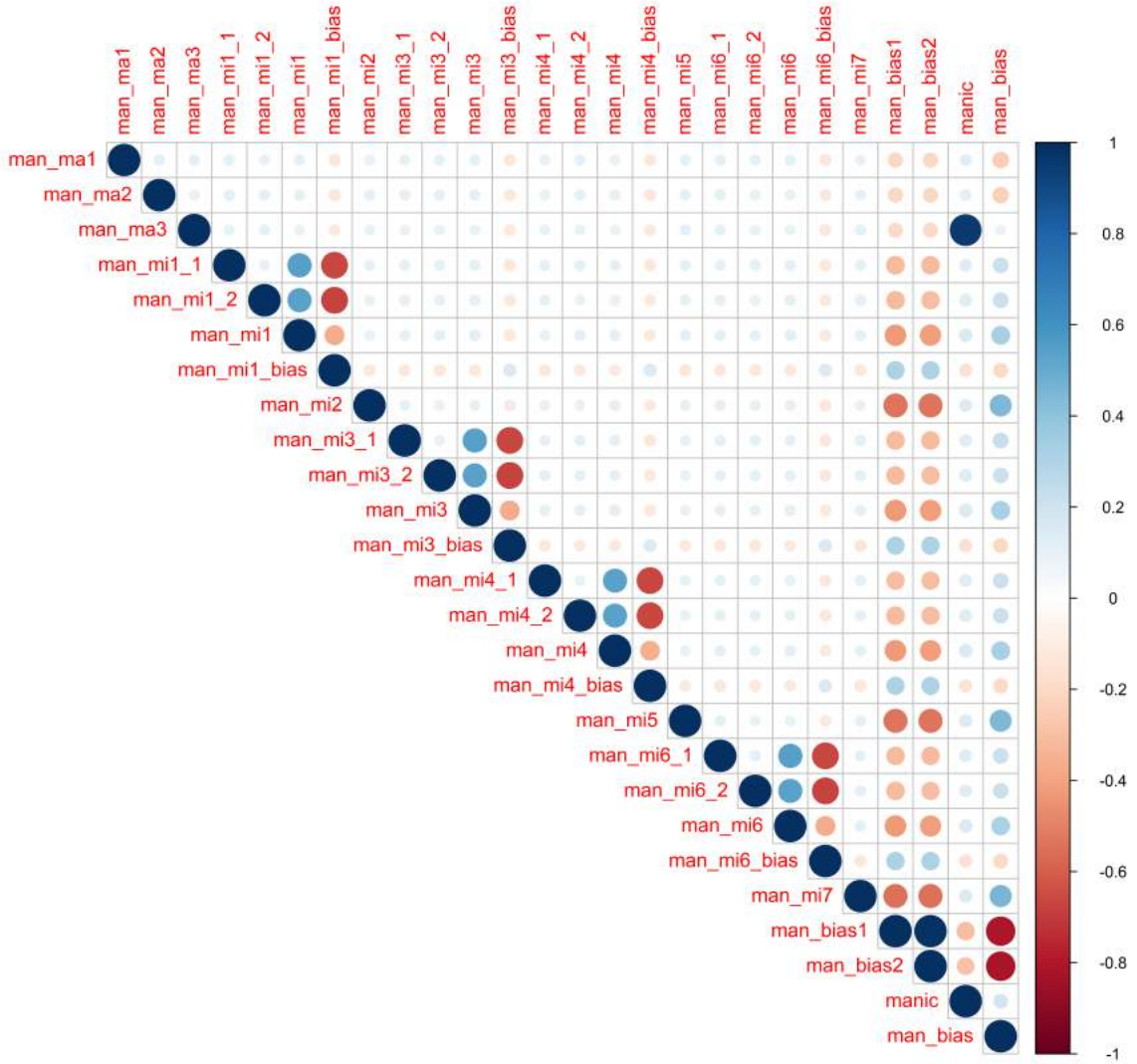

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

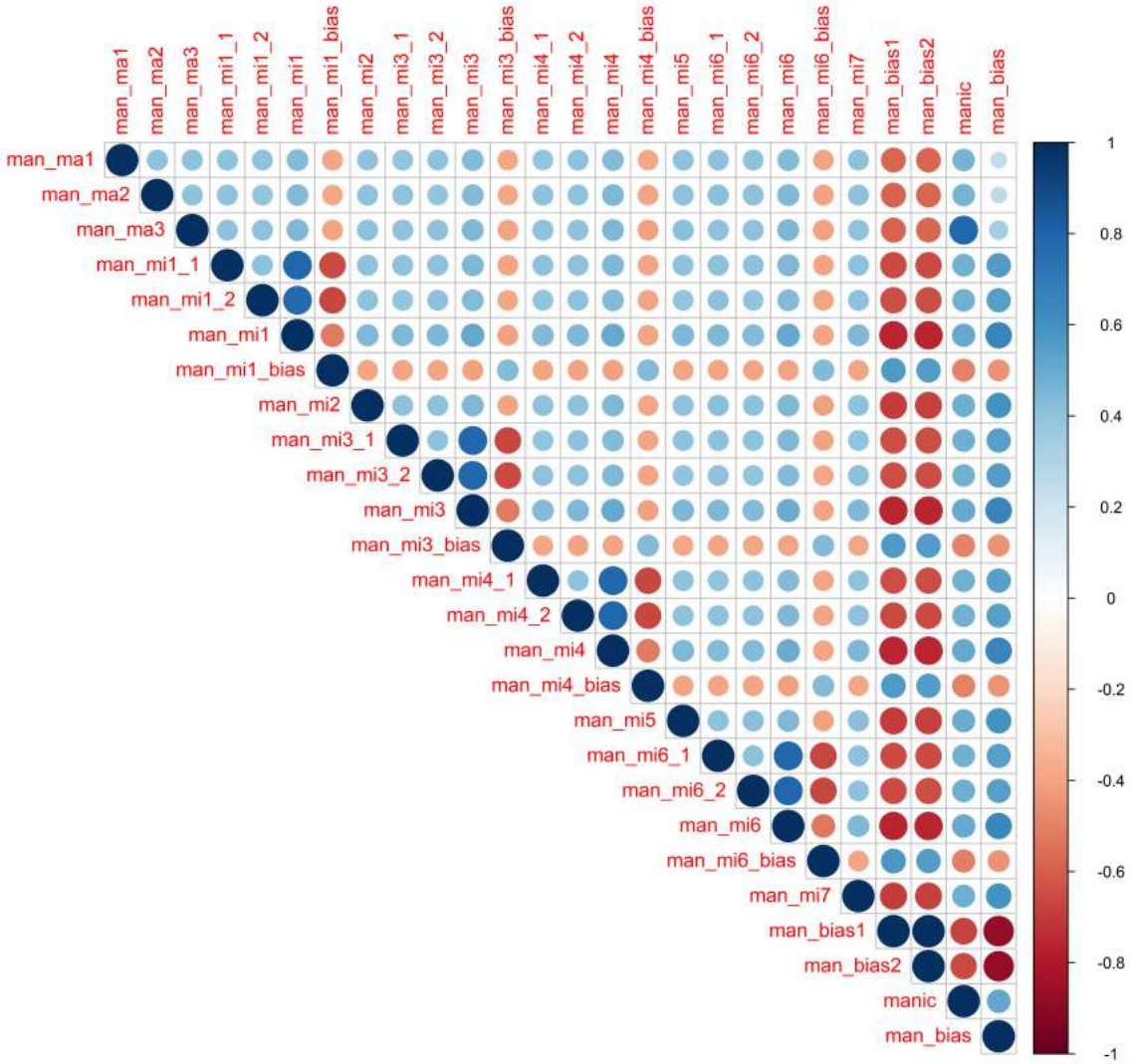


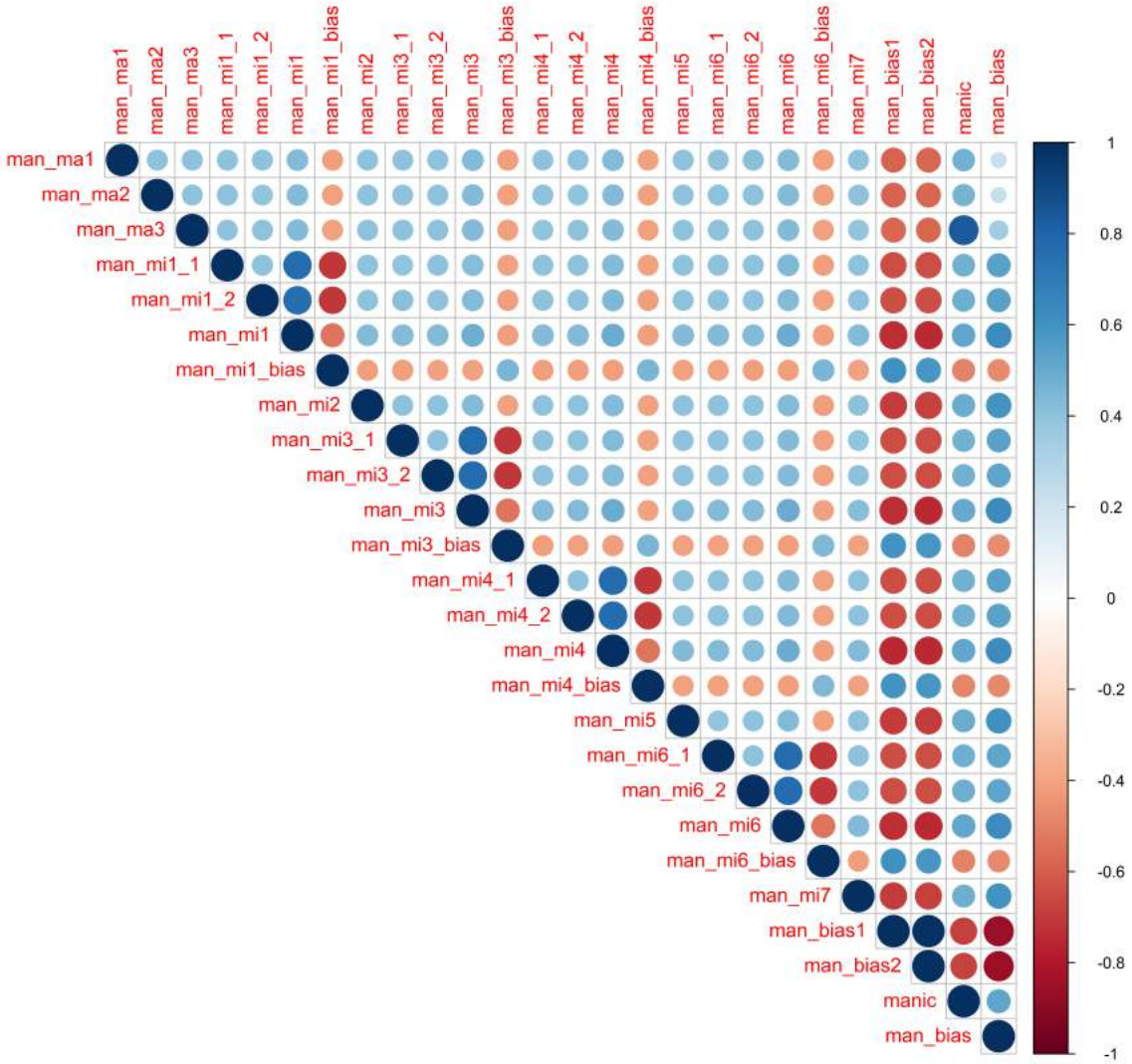

man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

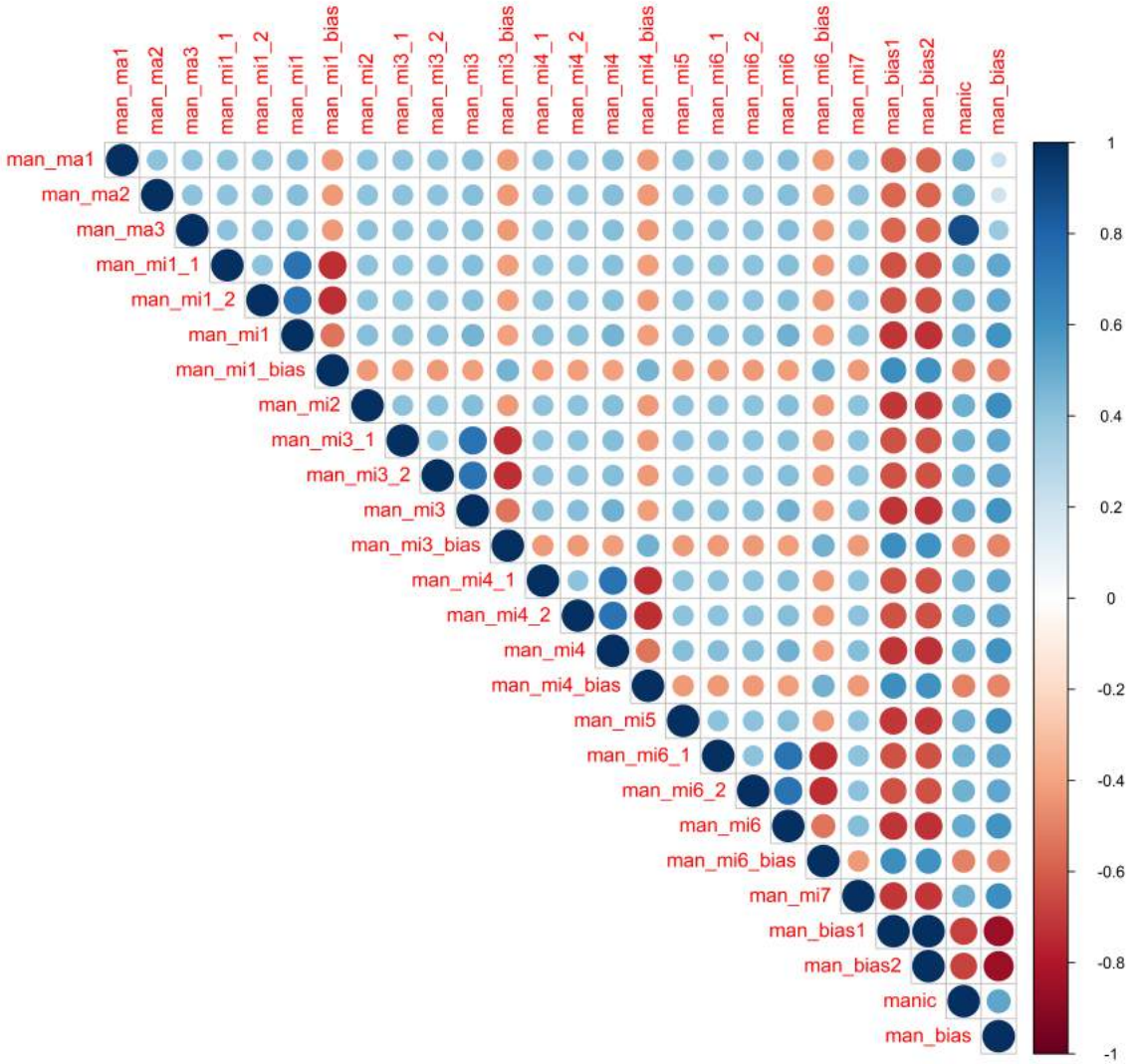



man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1




man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


man_ma1
man_ma2
man_ma3
man_mi1_1
man_mi1_2
man_mi1
man_mi1_bias
man_mi2
man_mi3_1
man_mi3_2
man_mi3
man_mi3_bias
man_mi4_1
man_mi4_2
man_mi4
man_mi4_bias
man_mi5
man_mi6_1
man_mi6_2
man_mi6
man_mi6_bias
man_mi7
man_bias1
man_bias2
manic
man_bias
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1