BMJ Open

# 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,[6] Yen-Po Cheng,[6] Yi-Chun Lai,[7] Wei-Chih Chen[8,9,10]

**To cite:** Chao Y-S, Lin K-F, Wu C-J, *et al*. Simulation study to demonstrate biases created by diagnostic criteria of mental illnesses: major depressive episodes, dysthymia, and manic episodes. *BMJ Open* 2020;**10**:e037022. doi:10.1136/bmjopen-2020-037022

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-037022).

Received 26 February 2020
Revised 24 June 2020
Accepted 15 August 2020



For numbered affiliations see end of article.

**Correspondence to**
Dr Wei-Chih Chen;
wiji.chen@gmail.com

## 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 relationships between the diagnoses and the input symptoms, as well as the magnitudes of biases created by diagnostic criteria and introduced into the diagnoses of mental illnesses with large disease burdens (major depressive episodes, dysthymic disorder, and manic episodes).

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

**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 0 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 and 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.

## Strengths and limitations of this study

- 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.
- Biases due to data censoring or categorisation were created by the diagnostic criteria and introduced into the intermediate variables and the three diagnoses of mental illnesses in simulated populations.
- The diagnostic criteria modified the importance of the input symptoms; certain input symptoms or bias variables were weighted more than expected in simulated populations.
- 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.
- 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.[1–3] 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 meet 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.[4 5] 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.[6 7] Recent studies on composite measures have found composite diagnostic criteria problematic because biases can be introduced while aggregating information from input variables.[7] The biases emerge while the sums of input variables are censored or while input variables are transformed inadequately.[7 8] 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.[7] For example, categorising continuous variables considers individuals in the same group homogenous and disregards the heterogeneity between individuals in the same categories.[7] Such practices induce biases and decrease measurement precision.[7 8]

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.[7] Frailty is diagnosed based on several symptoms and characterised by weakness and vulnerability to adverse health events.[7] While using one of the most widely used diagnostic criteria, the Biological Syndrome Model scores, to diagnose frailty,[9] biases alone can explain more than 71% of the variances of the frailty diagnosis.[7] The biases introduced by data censoring and data categorisation can better explain the frailty diagnosis than the input symptoms.[7]

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 online supplemental file 1. 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-online supplemental file 1. 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.[10] The symptoms for the diagnosis of manic episodes correlated to one another. Third, the

**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 precisely by patients | |
| | **Parameters of input symptoms of the same diagnosis for each simulation** | |
| 1 | Population sizes | 100 000 |
| 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 |

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 and we needed to examine various combinations of these epidemiological measures. There were studies on the prevalence of mental illnesses,[11 12] 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.[7 13–19] 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 and the diagnostic criteria in the DSM-IV-TR were strictly followed. However, these assumptions did not hold in the real world.[20] 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 tables 2–4. 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 0 and 1 for the absence and presence of the symptoms, criteria or the diagnoses, respectively. For example, two symptoms, 'insomnia' and 'hypersomnia', were extracted from one of the minor criteria for the diagnosis of major depressive episodes.[1] Two other symptoms, '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.[7] In other words, 'poor appetite or overeating' equalling the sum of two input variables, 'poor appetite' and 'overeating,' and a bias variable to achieve censoring of the sum of both variables.[7] The sum of two binomial variables could be 0, 1 and 2 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.[7] 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, categorisation, 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 confirm the diagnosis of dysthymic disorder among those with a multiplicative product of 1. Individuals could be assigned with 0 or 1 for whether they met both criteria, while the sum of major and minor criteria were 0, 1 or 2 for the individuals. Linearly, a bias variable with values of −1 or 0 was created and those meeting both the major or minor criteria were assigned with −1.[7] For categorisation of continuous variables, bias variables were required to remove the variations between the subjects in the same categories.[7] Other equations to generate the intermediate variables and the diagnoses were listed and explained in tables 2–4.

### 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).[7] A visual presentation of how bias variables were generated was published.[7] 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. All bias variables were described in tables 2–4.

### 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 hypothesised 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.[21 22] The intermediate and diagnosis variables were derived according to the equations in tables 2–4. For each combination of prevalence rates and between-variable correlations, the populations were simulated for 100 times to obtain the mean values and 95% CIs of derived prevalence rates, as

**Table 2** The input symptoms, intermediate variables and bias variables for the diagnosis of major depressive episodes based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision

| 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)<br>2. Categorising 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 2 weeks. | | | | | | |
| | | Depressed mood for more than 2 weeks. | mde_ma1 | | | | | |
| | | Loss of interest or pleasure in daily activities for more than 2 weeks. | mde_ma2 | | | | | |
| **Minor criteria (at least 5 of the symptoms including the two in major criteria)** | mde_mi | | | | | | | |

Continued

**Table 2** Continued

| 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 |
|---|---|---|---|---|---|---|---|---|
| | | 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 | | | |

Continued

**Table 2** Continued

| 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 |
|---|---|---|---|---|---|---|---|---|
| | | | | 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 categorisation (choosing three domains in minor criteria)** | mde_bias1 | | | | | | | Bias introduced to categorise the sum of the number of confirmed symptoms in the minor criteria |
| **Information due to categorisation (choosing four domains in minor criteria)** | mde_bias2 | | | | | | | Bias introduced to categorise the sum of the number of confirmed symptoms in the minor criteria |

Continued

Table 2 Continued

| 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 |
|---|---|---|---|---|---|---|---|---|
| **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 categorisation |

well as the adjusted R squared and p values derived by approximating the diagnosis variables.

### Diagnosis approximation

Due to the existence of the biases, the input symptoms were not likely to fully explain the diagnoses.[7] Therefore, the diagnoses were approximated by the input, bias and intermediate variables individually and collectively.[7 13 15 17] The approximation was conducted using forward-stepwise linear regressions.[7 13 15 17 23] The interpretability of the diagnoses by the input symptoms and bias variables was assessed via adjusted R square ranging from 0 to 1: 0 suggested that the input symptoms were unrelated to the diagnosis, and 1 suggested that the input symptoms perfectly explained the diagnosis.[15 16 24–27]

All statistical analyses were conducted within the R environment (V.3.4.1)[28] and RStudio (V.1.0.153).[29] 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 online supplemental file 1. The derived correlations between the input symptoms were close to assumed levels in the online supplemental file 1. 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 0 and 1 (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 the assumed correlations and prevalence rates of the input symptoms in figure 2.

### Prevalence of mental illnesses

The derived prevalence rates of three diagnoses were plotted against the assumed prevalence rates and

**Table 3** The input symptoms, intermediate variables and bias variables for the diagnosis of dysthymic disorder based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision

| 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 categorisation |
| **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 two items)** | | | dys_mi | | | dys_mi=dys_mi1+dys_mi2+ dys_mi3+dys_mi4+dys_mi5+ dys_mi6+dys_mi_bias | | Categorising 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 | | | |

Continued

Table 3 Continued

| 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 |
|---|---|---|---|---|---|---|---|---|
| | | 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 | | | |

Continued

**Table 3** Continued

| 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 |
|---|---|---|---|---|---|---|---|---|
| | | | | 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 categorising 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 categorisation (dys_mi_bias) |

*The input symptoms used for the diagnosis of both major depressive episodes and dysthymic disorder.

**Table 4** The input symptoms, intermediate variables and bias variables for the diagnosis of manic episodes based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision

| 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),<br>2. multiplication for the condition of presenting irritable mood (… x man_ma3), and<br>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;<br>4. the bias variables introduced by categorising 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**) | | | | | | | | |

Continued

Table 4 Continued

| Classification of symptoms | Criterion variable | Major or minor criteria (domains) | Domain variables | Symptoms | Symptom variables | Equations | Approximation | Mechanisms related to introducing biases |
|---|---|---|---|---|---|---|---|---|
| | | A distinct period of abnormally and persistently elevated, expansive or irritable mood, lasting at least 1 week (or any duration if hospitalisation 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 (three or more of the following symptoms have persisted; four 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 | | | |

Continued

**Table 4** Continued

| Classification of symptoms | Criterion variable | Major or minor criteria (domains) | Domain variables | Symptoms | Symptom variables | Equations | Approximation | Mechanisms related to introducing biases |
|---|---|---|---|---|---|---|---|---|
| | | Decreased need for sleep (eg, 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 | | | |

Continued

Table 4 Continued

| Classification of symptoms | Criterion variable | Major or minor criteria (domains) | Domain variables | Symptoms | Symptom variables | Equations | Approximation | Mechanisms related to introducing biases |
|---|---|---|---|---|---|---|---|---|
| | | Distractibility (ie, 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 (eg, engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments) | man_mi7 | | | | | |

Continued

Table 4 Continued

| Classification of symptoms | Criterion variable | Major or minor criteria (domains) | Domain variables | Symptoms | Symptom variables | Equations | Approximation | Mechanisms related to introducing biases |
|---|---|---|---|---|---|---|---|---|
| Information of diagnosis due to categorisation (choosing at least three symptoms) | man_bias1 | | | | | | | Bias introduced by categorising the number of input symptoms confirmed in the minor criteria |
| Information of diagnosis due to categorisation (choosing at least four symptoms) | man_bias2 | | | | | | | Bias introduced by categorising 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 categorisation, man_bias1 and man_bias2 |

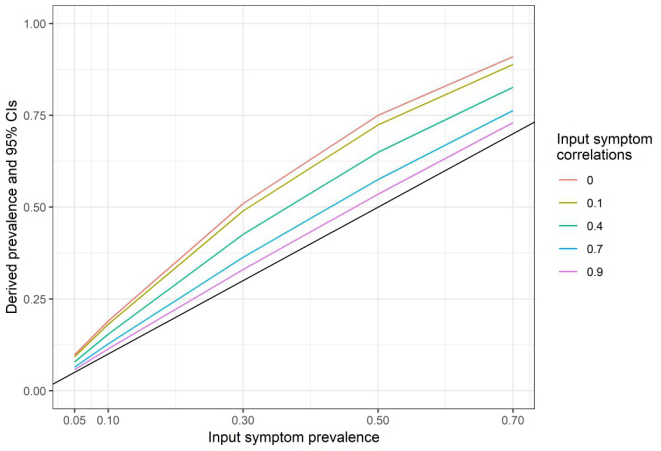


**Figure 1** The prevalence rates of an intermediate variable for the diagnosis of major depressive episodes by assumed input symptom prevalence and correlations. The intermediate variable is 'significant unintentional weight loss or gain' and the input symptoms are 'significant unintentional weight 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 the prevalence rates of the intermediate variable larger than those of the input symptoms. CI, confidence interval.

correlations of the input symptoms in figures 2–4 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 0 correlations and 0.7 prevalence rates than the

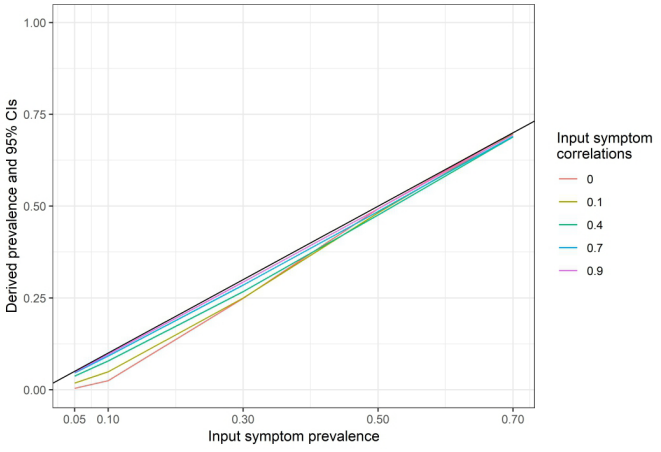


**Figure 2** The prevalence rates of dysthymic disorder by assumed input symptom prevalence and correlations. Dysthymic disorder is diagnosed when both the major (depressed mood most of the day for more days than not, for at least 2 years) and minor criteria (at least two of the six items) are met. The black line represents the situation where the prevalence rates of the input symptoms are the same as those of dysthymic disorder. Lines below the black lines have dysthymic disorder prevalence rates lower than those of the input symptoms. CI, confidence interval.

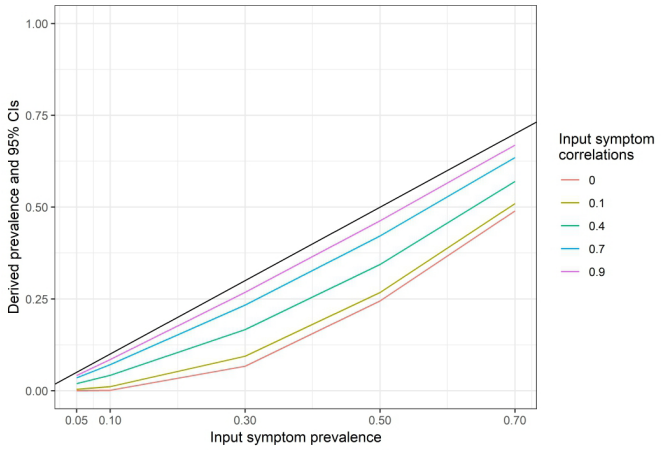


**Figure 3** The prevalence rates of major depressive episodes by assumed input symptom prevalence and correlations. Major depressive episodes are diagnosed when both major and minor criteria are confirmed. The black line represents the situation where the prevalence rates of the input symptoms are the same as that of major depressive episodes. Lines below the black lines have the prevalence rates of major depressive disorder lower than those of the input symptoms. CI, confidence interval.

prevalence rate (0.679) assuming 0.1 correlations and 0.7 prevalence rates of the input symptoms. When compared across figures 2–4, 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).

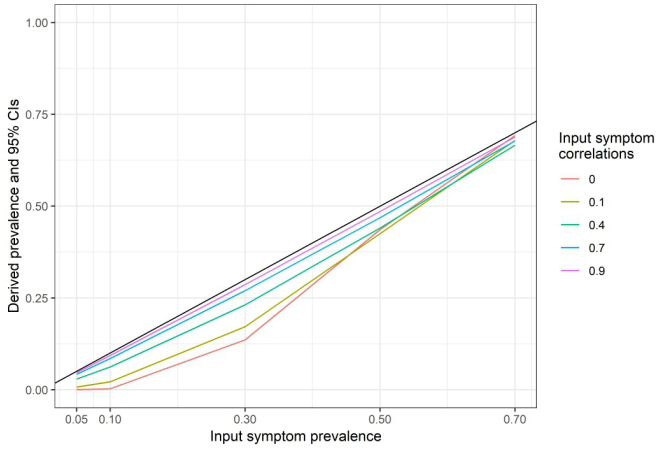


**Figure 4** The prevalence rates of manic episodes by assumed input symptom prevalence and correlations. Manic episodes are diagnosed when the symptoms present as described in the diagnostic manual. The black line represents the situation where the prevalence rates of manic episodes are the same as those of the input symptoms. Lines below the black lines have prevalence rates of manic episodes lower than those of the input symptoms. CI, confidence interval.

**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) |

CI, confidence interval.

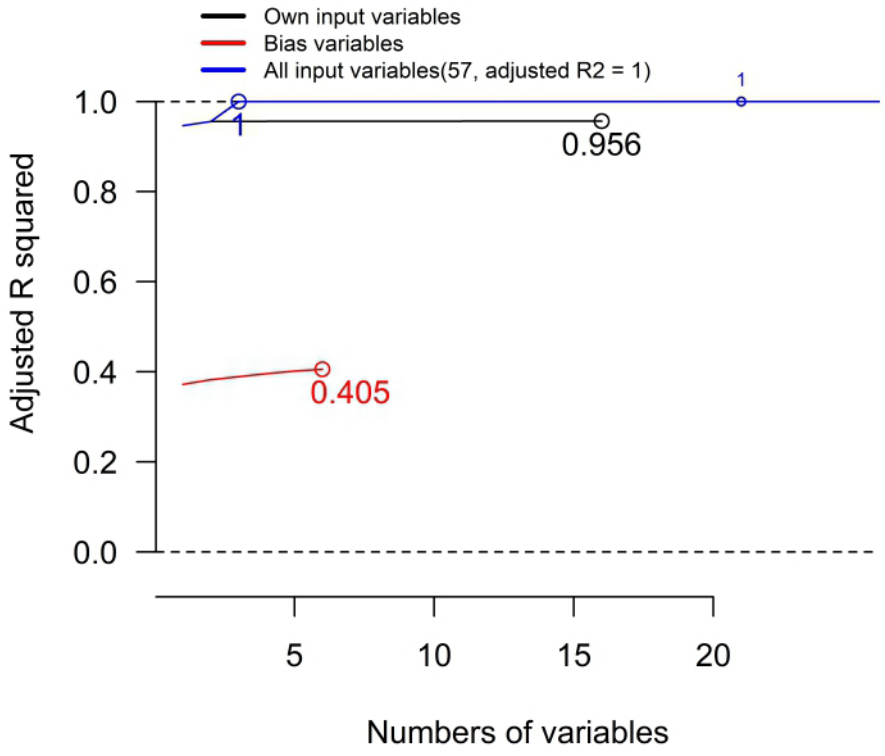


**Figure 5** The approximation of the diagnosis of dysthymic disorder by input symptoms, bias variables and both, measured by R-squared. The diagnosis of dysthymic disorder is approximated by all variable, including input symptoms and bias variables, using forward-stepwise regression. The selection of the variables was determined by adjusted R-squared. Circles are the maximal adjusted R-squared achieved by the regression with input symptoms, bias variables, or both of them. 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 in this figure.

### Associations between the diagnoses and 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 0.956 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 tables 6 and 7, respectively.

For the diagnosis of major depressive episodes, the first or second items in the major criteria (variable names: mde_ma1 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 and 0.974, depending on the assumed prevalence rates and between-variable correlations. Based on a high correlation with the diagnoses, certain input variables or symptoms were more important than others, 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 categorisation 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 categorisation of the confirmed input symptoms

**Table 6** The individual input symptoms that best explained the diagnoses based on adjusted R-squared: major depressive episodes, dysthymic disorder and manic episodes by assumed input symptom prevalence and correlations

| 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 |

Continued

Table 6 Continued

| Assumed correlations between input symptoms | Assumed prevalence of input symptoms | Major depressive episodes | Dysthymic disorder | Manic episodes |
|---|---|---|---|---|
| **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) |

See table 2 to 4 for variable definitions. Adjusted R-squared is derived from linear regressions using individual input symptoms as predictor with 95% confidence intervals (CIs) derived from 100 simulations for each combination of assumed input symptom prevalence and correlations.
CI, confidence interval.

in the minor criteria (dys_mi_bias) were the leading bias variables. For the diagnosis of manic episodes, the bias due to the categorisation 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 categorisation 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 0 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

Table 7 The individual bias variables that best explained the diagnoses based on adjusted R-squared: major depressive episodes, dysthymic disorder and manic episodes by assumed input symptom prevalence and correlations

| 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 |

Continued

Table 7 Continued

| Assumed correlations between input symptoms | Assumed prevalence of input symptoms | Major depressive episodes | Dysthymic disorder | Manic episodes |
|---|---|---|---|---|
| **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) |

See table 2 to 4 for variable definitions. Adjusted R-squared is derived from linear regressions using individual bias variables as predictor with 95% confidence intervals (CIs) derived from 100 simulations for each combination of assumed input symptom prevalence and correlations.
CI, confidence interval.

individually better than the bias variables of the other two diagnoses.

### Approximating the diagnoses with input symptoms

When the diagnoses were approximated by all of 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 together 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 0 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 0 between-variable correlations and 0.7 prevalence rates for the input symptoms (adjusted R-squared=0). With increasing assumed

**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) |

Adjusted R-squared is the maximal values from the forward-stepwise linear regressions using all input symptoms as candidate predictors with 95% confidence intervals (CIs) derived from 100 simulations for each combination of assumed input symptom prevalence and correlations.
CI, confidence interval.

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 with 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. The proportions of the diagnosis variance explained by bias variables also varied widely from 0 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.[7] There are three major mechanisms of introducing biases: censoring, data categorisation[8] and multiplication of input symptoms.[7] These mechanisms introduce information or biases that cannot be fully explained by the input symptoms.[7] 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 with each other. The design of diagnostic criteria transform input symptoms into various diagnosis prevalence rates with implicit upper limits (ie, no more prevalent than the input symptoms), unacknowledged differential weights on the input symptoms (ie, certain input symptoms better explaining the diagnoses) and the introduction of biases (ie, due to censoring, data categorisation or multiplication).

We are the first to notice that the prevalence rates of the three diagnoses are lower than those of the input symptoms if input symptoms are 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.[7] 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 $[C(2,6)+C(3,6)+C(4,6)+C(5,6)+C(6,6)] \times (0.7)^6 = 37 \times 0.117 = 4.35 > 100\%$). 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.[7] 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

**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) |

Adjusted R-squared is the maximal values from the forward-stepwise linear regressions using all bias variables as candidate predictors with 95% confidence intervals (CIs) derived from 100 simulations for each combination of assumed input symptom prevalence and correlations.
CI, confidence interval.

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. This 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 emphasised 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 the prevalence of input symptoms are similar), unacknowledged weights to input symptoms (certain input symptoms explaining the diagnoses much better) and biases that are 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, given their input symptoms of 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,[20 30] diagnoses are not well made particularly by non-psychiatrists,[31] and there are two diagnostic systems (the DSM and the International Classification of Disease) that require efforts to harmonise.[32] Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved for interpreteability, clinical use without introducing biases, and better connection to clinical decisions. Certain measures and biomarkers have been proven useful to identify mental illnesses.[33 34] We are developing methods that better detect symptom-based conditions and applying syndrome mining techniques[35] to search for neglected mental illnesses.

## 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.

**Author affiliations**
[1]Independent researcher, Montreal, Quebec, Canada
[2]Psychiatry, National Taiwan University Hospital Yun-Lin Branch, Douliou, Taiwan, Taiwan
[3]Département d'informatique, UQAM, Montreal, Quebec, Canada
[4]Internal Medicine, Taipei Hospital, Xinzhuang, Taipei, Taiwan
[5]Pathology, Changhua Christian Healthcare System, Changhua, Taiwan, Taiwan
[6]Surgery, Changhua Christian Healthcare System, Changhua, Taiwan, Taiwan
[7]Chest Medicine, National Yang Ming University Hospital, Ilan, Taiwan, Taiwan
[8]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[9]Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei, Taiwan, Taiwan
[10]Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan

**Contributors** Y-SC conceptualised and designed this study, managed and analysed data and drafted the manuscript. K-FL assisted in the interpretation of the diagnostic criteria. C-JW assisted in data management and computation. H-CW, H-TH, L-CT, Y-PC, Y-CL and W-CC participated in the design of this study. All authors reviewed and approved the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** Y-SC is currently employed by the Canadian Agency for Drugs and Technologies in Health. The other authors declare that there is no conflict of interest.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as online supplemental information. No real-world data used—all analysis are based on simulations reproducible with the files in the online supplemental materials.

**ORCID iD**

Yi-Sheng Chao http://orcid.org/0000-0002-4019-3586

## REFERENCES

1 American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. In: *Text revision (DSM-IV-TR®)*. Fourth ed. 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 Chao Y-S, Wu C-J, Wu H-C, HC W, *et al*. Composite diagnostic criteria are problematic for linking potentially distinct populations: the case of frailty. *Sci Rep* 2020;10:2601.
4 Feighner JP, Robins E, Guze SB, *et al*. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 1972;26:57–63.
5 Kendler KS, Muñoz RA, Murphy G. The development of the Feighner criteria: a historical perspective. *Am J Psychiatry* 2010;167:134–42.
6 Chao Y-S, Wu C-J. PP46 when composite measures or indices fail: data processing lessons. *Int J Technol Assess Health Care* 2018;34:83.
7 Chao Y-S, Wu H-C, Wu C-J, *et al*. Index or illusion: the case of frailty indices in the health and retirement study. *PLoS One* 2018;13:e0197859.
8 Barnwell-Ménard J-L, 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;34:936–49.
9 Cigolle CT, Ofstedal MB, Tian Z, *et al*. Comparing models of frailty: the health and retirement study. *J Am Geriatr Soc* 2009;57:830–9.
10 Brown TA, Chorpita BF, Korotitsch W, *et al*. Psychometric properties of the depression anxiety stress scales (DASS) in clinical samples. *Behav Res Ther* 1997;35:79–89.
11 Lim GY, Tam WW, Lu Y, *et al*. Prevalence of depression in the community from 30 countries between 1994 and 2014. *Sci Rep* 2018;8:2861.
12 Smith DJ, Nicholl BI, Cullen B, *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:e75362.
13 Chao Y-S, C-J W. PD26 principal component approximation: Canadian health measures survey. *International Journal of Technology Assessment in Health Care* 2019;34:138–9.
14 Chao Y-S, Wu C-J, Wu H-C, *et al*. Trend analysis for national surveys: application to all variables from the Canadian health measures survey cycle 1 to 4. *PLoS One* 2018;13:e0200127.
15 Chao Y-S, Wu H-C, Wu C-J, *et al*. Principal component approximation and interpretation in health survey and Biobank data. *Frontiers in Digital Humanities* 2018;5.
16 Chao Y-S, Wu H-C, Wu C-J, *et al*. Stages of biological development across age: an analysis of Canadian health measure survey 2007–2011. *Front Public Health* 2018;5:2296–565.
17 Chao Y-S, C-J W. PD25 principal component approximation: medical expenditure panel survey. *International Journal of Technology Assessment in Health Care* 2019;34:138.
18 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 Inf Sci Syst* 2014;2:5.
19 Chao Y-S, Wu H-T, Scutari M, *et al*. A network perspective on patient experiences and health status: the medical expenditure panel survey 2004 to 2011. *BMC Health Serv Res* 2017;17:1472–6963.
20 Bonnin JE. *Treating without diagnosis: psychoanalysis in medical settings in Argentina*, 2015.
21 Leisch F, Weingessel A, Hornik K. *On the generation of correlated artificial binary data*, 1998.
22 Leisch F, Weingessel A, Hornik K. bindata: generation of artificial binary data, 2012. Available: http://CRANR-projectorg/package=bindataRpackageversion09-19
23 Lumley T, Lumley MT, 'leaps' P. Regression subset selection Thomas Lumley based on Fortran code by alan Miller, 2013. Available: https://cran.r-project.org/package=leaps [Accessed 18 Mar 2018].
24 Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. Second ed. Springer New York, 2009.
25 James G, Witten D, Hastie T, *et al*. *An introduction to statistical learning: with applications in R*. New York: Springer, 2013.
26 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:e0183997.
27 Chao YS, Wu H-tieng, Wu C-J. Feasibility of classifying life stages and searching for the determinants: results from the medical expenditure panel survey 1996–2011. *Front Public Health* 2017;5:2296–565.
28 R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2016.
29 RStudio Team. *RStudio: integrated development for R. Boston*. MA: RStudio, Inc., 2016.
30 Demyttenaere K, Bonnewyn A, Bruffaerts R, *et al*. Clinical factors influencing the prescription of antidepressants and benzodiazepines: results from the European study of the epidemiology of mental disorders (ESEMeD). *J Affect Disord* 2008;110:84–93.
31 Margolis RL. Nonpsychiatrist house staff frequently misdiagnose psychiatric disorders in general Hospital inpatients. *Psychosomatics* 1994;35:485–91.
32 First MB. Harmonisation of ICD-11 and DSM-V: opportunities and challenges. *Br J Psychiatry* 2009;195:382–90.
33 Husain SF, Tang T-B, Yu R, *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.
34 Ho CSH, Zhang MWB, Ho RCM, CSH H, Ho R. Optical topography in psychiatry: a CHIP off the old block or a new look beyond the mind-brain frontiers? *Front Psychiatry* 2016;7:74.
35 Chao Y-S, Wu C-J, Wu H-C, HC W, *et al*. Using syndrome mining with the health and retirement study to identify the deadliest and least deadly frailty syndromes. *Sci Rep* 2020;10:5357.