DIP: Natural history model for major depression with incidence and prevalence

Melike Yildirim [a], Bradley N Gaynes [b], Pinar Keskinocak [a,c], Brian W Pence [d], Julie Swann [e,*]

[a] School of Industrial and Systems Engineering and Center for Health and Humanitarian Systems, Georgia Institute of Technology, Atlanta, Georgia, USA
[b] Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[c] Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[d] Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[e] Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina, USA

ARTICLE INFO

Keywords:
Major depression
Natural history
Prevalence
Incidence
Recall bias
Lifetime prevalence

ABSTRACT

Background: Major depression is a treatable disease, and untreated depression can lead to serious health complications. Therefore, prevention, early identification, and treatment efforts are essential. Natural history models can be utilized to make informed decisions about interventions and treatments of major depression.
Methods: We propose a natural history model of major depression. We use steady-state analysis to study the discrete-time Markov chain model. For this purpose, we solved the system of linear equations and tested the parameter and transition probabilities empirically.
Results: We showed that bias in parameters might collectively cause a significant mismatch in a model. If incidence is correct, then lifetime prevalence is 33.2% for females and 20.5% for males, which is higher than reported values. If prevalence is correct, then incidence is .0008 for females and .00065 for males, which is lower than reported values. The model can achieve feasibility if incidence is at low levels and recall bias of the lifetime prevalence is quantified to be 31.9% for females and 16.3% for males.
Limitations: This model is limited to major depression, and patients who have other types of depression are assumed healthy. We assume that transition probabilities (except incidence rates) are correct.
Conclusion: We constructed a preliminary model for the natural history of major depression. We determined the lifetime prevalences are underestimated and the average incidence rates may be underestimated for males. We conclude that recall bias needs to be accounted for in modeling or burden estimates, where the recall bias should increase with age.

1. Introduction

Major depression is a common mental illness, which affects roughly 17.3 million adults in the United States. It is more prevalent among women (10.2%) than men (6.2%) (Kessler et al., 2010). Major depression (or simply depression) is also a leading cause of disability (Kessler et al., 1999). The annual direct and indirect costs of depression, which are mostly caused by decreased productivity and increased healthcare utilization, were estimated at $210.5 billion in 2010 (Greenberg et al., 2015). Although major depression is a widespread disease, only 33% to 50% (Harman et al., 2006; Kessler et al., 2003; Pincus et al., 1998) of patients are diagnosed or receive adequate treatment in primary care settings. If depression is not detected and treated, it can cause functional impairment and contribute to poor health outcomes. Therefore, prevention and treatments are essential. Models have been used to analyze various components of challenges surrounding major depression.

Several types of modeling approaches can be used to study major depression (Ali Afzali et al., 2012; Kolovos et al., 2017). Natural history models, which describe how diseases develop and progress over time can be used to compare health outcomes (e.g., suicide (Gong et al., 2019), or quality/disability-adjusted life-years (Julien et al., 2020)) under different scenarios including pharmacological treatment and/or psychological treatment (Kolovos et al., 2017; Yan et al., 2019), screening effectiveness (Jiao et al., 2017) or prediction of future outcomes (Lin et al., 2019). Models may be represented with differential equations or by Markovian ones (stochastic, where future states depend

* Corresponding author. Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina, USA.
E-mail address: jlswann@ncsu.edu (J. Swann).

https://doi.org/10.1016/j.jad.2021.09.079
Received 9 December 2020; Received in revised form 11 August 2021; Accepted 26 September 2021
Available online 5 October 2021

on the present state), and they may be analyzed mathematically or computationally. Natural history models, which describe how diseases develop and progress over time, can help inform decisions about interventions and treatments.

This paper studies Markov models of major depression through the use of steady-state analysis, which is a mathematical tool to understand effects over time. Fundamentally, the analysis can provide the distribution of a population among a set of states after a period of time. Bush and Zaremba (1971) note that steady-state analysis should be performed for any setting when a stochastic model is used. Steady-state analysis was described for the context of diseases by Bush and Zaremba (1971), and used for various diseases by others (Barendregt et al., 2003; Goenka et al., 2021; Kruijshaar et al., 2002; Kruijshaar et al., 2003). Examples of uses include characterizing natural history models, parameterizing unavailable data in a Markov model, or checking the internal consistency of parameters (Faissol et al., 2009; Gopalappa et al., 2018).

A natural history model for major depression has been developed for different study cohorts including 40-year-old primary care patients (Valenstein et al., 2001) or adults with newly diagnosed major depression (Ross et al., 2019)). Generally, studies are built with Markov models and input parameters are validated based on fits to historical data (e.g., prevalence or incidence separately).

For disease models to be broadly useful, good estimates should be available for parameters such as incidence, prevalence, and many others, which can come from personal health records (Lin et al., 2016, 2018) or surveys (Eaton et al., 2007). In survey data, incidence is calculated as the proportion of people who newly develop a condition during a particular time period. Annual prevalence represents the rate of the study population who had major depression within the 12 months before the study. In comparison, lifetime prevalence is estimated from the percentage of the people who had experienced at least one episode of depression at the time of the interview. Annual estimates of incidence are between 2.3 to 15.9 per 1000 people for major depression (Eaton et al., 1997; Mattisson et al., 2005; Murphy et al., 2002), and lifetime prevalence is usually reported as ranging from 10% to 20% (Bland, 1992; Kessler et al., 2003; Kessler et al., 2010).

Several articles have identified challenges associated with incidence and lifetime prevalence values (Kessler et al., 2003; Kessler et al., 2010; Patten, 2009; Patten et al., 2012; Takayanagi et al., 2014). For example, for the population studied by Eaton et al. (1997), applying the lowest value for annual incidence of 2.3 from age X to Y would result in a lifetime prevalence of 50%, which is much higher than the maximum values reported in Kessler at al. (2003). Several hypotheses have been suggested as causes such as recall bias and birth cohort effects (Patten et al., 2010; Patten et al., 2012).

One potential cause of the mismatch between incidence and lifetime prevalence is recall bias on episodes of depression (Patten et al., 2012; Takayanagi et al., 2014; Wells and Horwood, 2004). Recall bias is observed when study participants are less likely to recall their events or experiences from the past; one measure of it is the true value relative to the reported value. In the literature, the recall bias of lifetime prevalence of major depression is estimated as 35-291% (Andrews et al., 1999; Foley et al., 1998; Giuffra and Risch, 1994; Knauper and Wittchen, 1994; Kruijshaar et al., 2005; Patten, 2003, 2009; Takayanagi et al., 2014). From the Netherlands Mental Health Survey and Incidence Study (NEMESIS) the calculated recall bias is reported as 38% (Kruijshaar et al., 2005).

We build a simplified natural history model of major depression with incidence and prevalence (DIP) that mimics the disease progression without interventions. We utilize a Markov chain model with model inputs derived from published data, and we mathematically show how parameters can be inconsistent. The structure of our model and the equilibrium analysis (Bush and Zaremba, 1971) allow us to quantify a set of parameters that is feasible under a particular scenario of a multi-year analysis. We obtain estimates of parameters that are reasonable over long periods of time which can be used in longer models. We demonstrate an approach that can be generalized for other problem settings where calibration is needed. Our findings show how parameters need to be adjusted for more complex natural history models or simulations of major depression, which are used to study screening, treatment, or other interventions. Our model can also be a framework upon which deeper models can be built including ones accounting for age, prior history and uncertainty.

2. Methods

2.1. Population-level natural history model

We outline a simplified natural history model of major depression with incidence and prevalence (DIP), which consists of four health states: healthy, depression, remission, and death (Fig. 1). People who have never had a major depressive disorder in their lifetime are defined as in the "healthy" state. The depression state contains people who have met the diagnostic criteria for major depression in the last 12 months, while the remission state contains patients with a history of depression who have not satisfied the diagnostic criteria in the past 12 months. The death state includes deaths from all causes, including depression-related and not depression-related. From each state, there exists a transition to the death state, which is also a reservoir.

We utilize a discrete-time Markov chain model with a sequence of stochastic and state-to-state transitions. Patients are allowed to transition between states at the end of each year. Our study population represents adults age 18 or older, which is consistent with published data (e.g., prevalence and incidence) in the literature. We develop identical but separate models for males and females. Lifetime prevalence is the sum of the proportion of the population in the depression (in the last 12-months) and remission (past depression) states. A newly introduced population is assigned initial states based on reported prevalence at age 18 (arcs 7-10 in Fig. 1)

We use a closed system, i.e., with a constant size population. A new individual enters the system from the reservoir when a person dies (Bush and Zaremba, 1971). As in Bush and Zaremba (1971) and other studies with steady state analysis, our model is finite, irreducible, and aperiodic, so the model will converge. Steady state analysis provides a way of determining the long-run outcome (e.g., lifetime prevalence) of applying a particular transition probability matrix (e.g., annual incidence).

We use a stationary Markov model, where the transition probabilities are identical for every Markov cycle. Steady state is an interesting property of Markov chains, where if one simulates a process over an extended time, the state distribution will converge or reach an equilibrium (Gopalappa et al., 2018). A unique equilibrium or steady state exists regardless of the initial distribution as a result of the features of the Markov model (irreducible and aperiodic; refer to Ross (2010) for additional information). Steady-state analysis is used to obtain that equilibrium, and the process can be used to test parameters and transition probabilities empirically.

2.1.1. Population-level model parameterization

We derive the initial model parameters from highly cited studies, including the nationally representative U.S. cohort studies of the National Comorbidity Survey Replication (NCS-R) (Kessler et al., 2010) and Epidemiologic Catchment Area (ECA) (Eaton et al., 2007). The rates are generally consistent with others; for example, the prevalence values of depression from NCS-R are similar to the reported percentages from the National Institute of Mental Health (National Institute of Mental Health n.d). We use annual transition probabilities specific to males and females throughout the model, except from depression to remission (see Table 1 for initial inputs). These parameters are reported in the literature based on weighted averages of survey population (Eaton et al., 2007; Kessler et al., 2010).

We compute the weighted average of the age-specific death rate

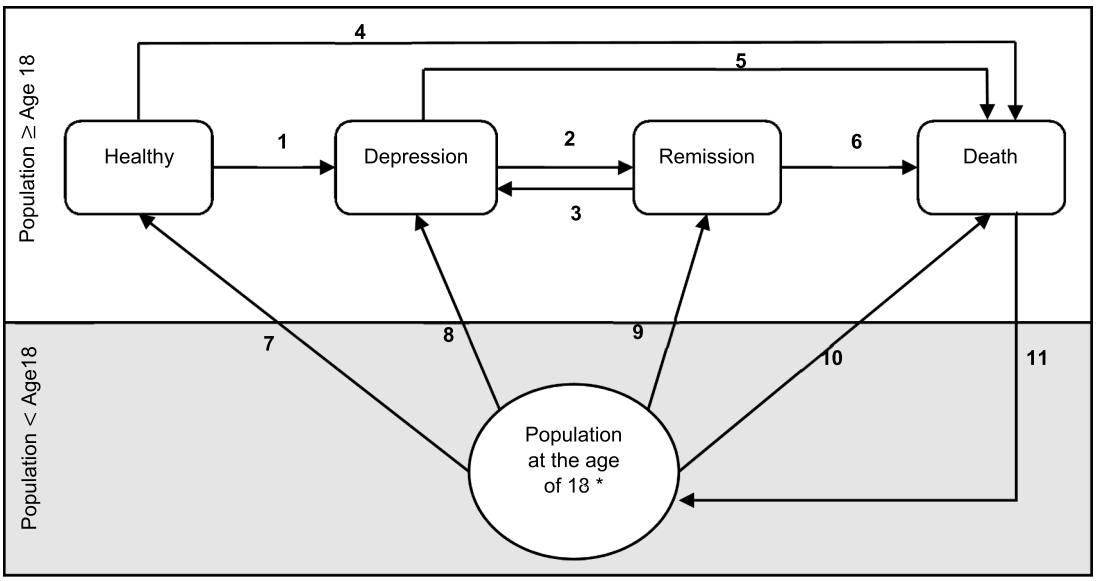

Fig. 1. Depiction of the Markov model for the natural history of major depression.
Notes: Boxes represent health states; arrows represent allowed transitions between states * This is a dummy state.

Table 1
Transition probabilities of major depression: Markov model.

Definition	Transitions	Parameter mean [Female, Male]	Reference
Incidence	1	[0.0039, 0.0021]	(Eaton et al., 2007)
Achieving remission	2	0.45	(Brodaty et al., 1993; Thase et al., 2005; Whiteford et al., 2013) [a,b]
Recurrence of depression	3	[0.324, 0.281]	(Kessler et al., 1994)[c]
Mortality for general population (healthy or in remission)	4, 6	[0.016, 0.017]	(Arias et al., 2017)[d]
Mortality for depressive patients	5	[0.025, 0.0269]	(Arias et al., 2017; Cuijpers et al., 2014) [e,f]
Prevalence at the age of 18 (Healthy, Dep., Rem., Death)	7, 8, 9, 10	[(0.763, 0.136, 0.101, 0.000305), (0.848, 0.072, 0.079, 0.000751)]	(Arias et al., 2017; Kessler et al., 2010)
Reentrance (refer text for more information)	11	1	

[a] Full remission rates for untreated patients are 0.37 (Whiteford et al., 2013) and for treated patients 0.47 (Thase et al., 2005). It is assumed that 65% of the patients are diagnosed (Simon & VonKorff, 1995) and, one-third of the diagnosed patients received treatment (Waitzfelder et al., 2018). The full remission rate is, on average, 0.39.

[b] Partial remission with the existence of residual symptoms is seen for 24% of the population in 3.8 years follow-up (Brodaty et al., 1993), which corresponds to the rate of 6.32% in a year.

[c] Probability for the population in the age range 15-54.

[d] Weighted average.

[e] Denotes that this source is a primary source.

[f] Denotes that this source used secondarily for an adjustment (multiplier) to the primary source.

(Arias et al., 2017) based on the proportion of each age in the standard U.S. population (Arias et al., 2017). Depressive patients are more likely to have one or more other comorbid conditions; accordingly, they have an elevated risk of death than healthy patients. Therefore, their lifetime may be shorter than patients without depression. In our analysis, we use the relative risk of mortality in depressed patients as 1.58 compared to the non-depressed population (Cuijpers et al., 2014).

In the main results, we focus on the general steady-state, that is, where the system would be stable. The steady-state results hold on average and in the long term. Without the consideration of birth cohort effects on major depression rates, the assumption of stable disease will be valid. In section 2.2., we parameterize the model with age-specific rates for incidence, lifetime prevalence, or recall bias.

To examine the model with parameters obtained from the literature, we build two sets of linear equations for the Markov model, with one equation for each gender; see Appendix 1 in the supplementary appendix for the equations. We conduct our steady-state calculations using R (version 3.6.2) (Team, 2013).

The equations, which are built on parameters such as annual incidence, recurrence and recovery rates are infeasible with reported prevalence (steady-state distribution) data which is P_{female} (healthy, depression, remission, death) = (0.755, 0.102, 0.127, 0.016) and P_{male} (healthy, depression, remission, death) = (0.832, 0.062, 0.089, 0.017) (Arias et al., 2017; Kessler et al., 2010; Kessler et al., 1997). Therefore, we further calibrate the parameters to build a model.

2.1.2. *Calibration*

As documented (Eaton et al., 1997), an infeasible system of equations for depression can be caused by different reasons, e.g., resulting in a prevalence rate that is too high, an incidence that is too low, or both. In the following sections, we hypothesize each of these cases for evaluation.

In hypothesis 1, we assume that the reported lifetime prevalence of each state obtained from the literature is correct, along with the initial prevalence at age 18. Therefore, one or more inaccurate transition probabilities for incidence may lead to infeasibility with lifetime prevalence values. Thus, we calibrate incidence rates and determine the steady-state distribution of the Markov model for average of lifetime and annual prevalence, comparing the results to reported values from the literature. This hypothesis corresponds to inaccurate reported incidence values.

In hypothesis 2, we assume that reported transition probabilities (including for incidence) are correct, as is the initial prevalence at the age of 18 (Kessler et al., 2010; Kruijshaar et al., 2005). We focus on finding fitted values for lifetime prevalence for each model state and compare them to the evidence from the literature. We calculate the

steady-state distribution of the Markov model with average values. This hypothesis is consistent with inaccurate lifetime prevalence, such as from recall bias.

In hypothesis 3, because of the high discordance between calculated and reported rates of incidence and lifetime prevalence, we evaluate the case that both sets of parameters are incorrect. We perform the calibration by examining scenarios where the incidence is equal to the mean, lower, and upper bound obtained from the Baltimore ECA study (Eaton et al., 2007) and calculating the steady-state distribution the range of prevalence values. We calculate the prevalence (annual or lifetime) values to fit within bounds of prevalence from the literature (Patten, 2003; Patten et al., 2010). We denote the difference between reported and calculated prevalence as recall bias. Note that the analysis could have been performed in the same way beginning with the range of lifetime prevalence values, assuming recall bias, and calculating the corresponding range of incidence values.

2.1.3. Sensitivity analysis

We evaluated various levels (low, medium, and high) of incidence and prevalence rates given in previous sections. Specifically, we extend lifetime, annual prevalence, and recall bias calculations for 20% higher than the upper bound and 20% lower than the lower bound from the point estimates of remission and recurrence rates reported in Table 1.

2.2. Age-specific parameters and transient analysis of natural history model

It is important to verify if the population-level model results hold for different age groups. Due to the existence of birth cohorts with different transition probabilities, we use transient analysis to capture the distribution of the population across states after a defined period of time. This is not a full steady state but is another mathematical tool to understand the implications of a set of parameters on the system.

Lifetime prevalence relates to the cumulative impact of incidence from many prior years. Therefore, a short period of time is insufficient to observe the accumulated effect (Kruijshaar et al., 2002). We performed additional scenario-based analysis, with age-specific values of incidence, prevalence, and mortality in DIP model, as shown in Table S1 (see the supplementary appendix). The analysis is valid for short to medium periods of time.

We calculated lifetime prevalence from age 18 to the ending ages of 20, 34, 49, 64, 79, and 90. The ages 20, 34, 49, and 64 were selected based on reported prevalence from Kessler et al. (2010), and the older ages were added to better analyze the group 65+. Our initial average mortality risk ratio is 1.58 for depressed patients (Cuijpers et al., 2014). Age-specific analysis has the potential to show much larger gaps between feasible and reported values. To explore an additional scenario, we also extended the age-specific analysis with a mortality risk ratio up to 4.0.

3. Results

This section collected our findings under five categories; hypotheses I, II, III, sensitivity analysis on remission and recurrence, and age-specific analysis of lifetime prevalence. For sensitivity analysis, we will use the best-fitting model from the previous hypothesis testing.

Note that the feasible pairs of incidence and lifetime prevalence in steady state can be calculated for a given initial prevalence. Feasible pairs of incidence and lifetime prevalence, given initial conditions, recurrence, remission, and mortality are shown in Fig. 2.

3.1. Hypothesis I: prevalence rates are correct

Assuming the stated prevalence values in each state, we calculate the annual incidence rates from the steady-state distribution in the model as .0008 and .00065 for females, and males, respectively. Compared to the reported average incidence rates (Table 1) of .0039 and .0021, the incidence rates from the model calculations were lower (~79% lower for females and ~69% for males). The calculated values are also substantially lower than the lower bounds in the Baltimore ECA study (Eaton et al., 2007) (the lower bounds are .0029 and .0013, for females and males, respectively). Fig. 2 displays the incidence and lifetime prevalence pairs under Hypothesis 1.

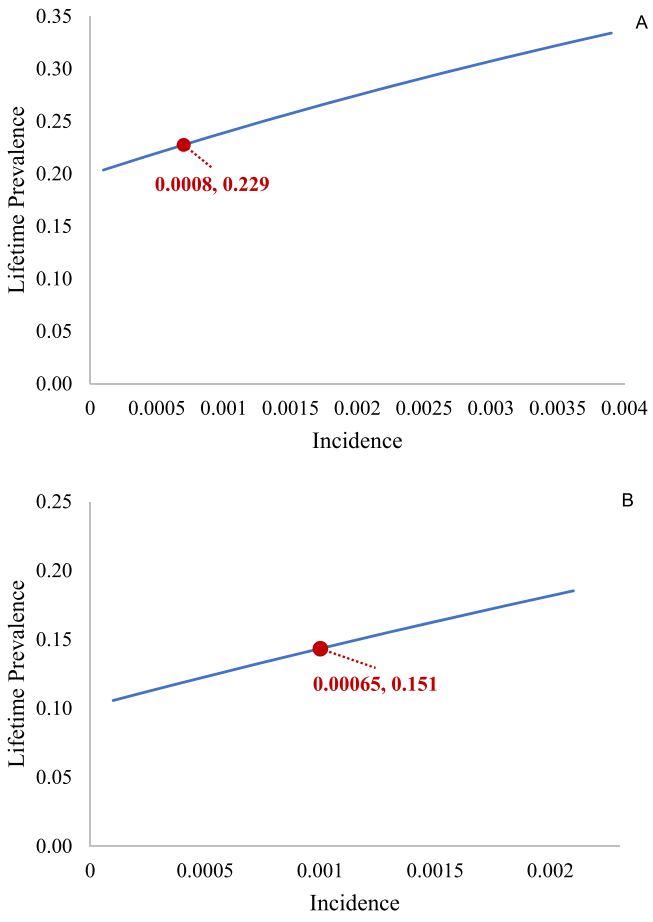

Fig. 2. Feasible pairs of incidence and lifetime prevalence, with initial conditions, recurrence, remission, and mortality.
Note: Labeled pair shows values resulting from prevalence rates from the literature and calculated incidence; A shows Female, and B shows Male.

3.2. Hypothesis II: transition probabilities for incidence are correct

Given incidence rates, we calculate the lifetime prevalence of depression from the model as 33.2% for females and 20.5% for males. However, the reported prevalence from national surveys was 22.9% and 15.1%, respectively (Kessler et al., 2010). The calculated lifetime prevalence rates with known incidence are 45% higher for females and 35.9% higher for males than reported values.

Using the Markov model and the transition probabilities, we plotted the prevalence to compare the expected values with the reported values. Fig. 3 shows that the transition probabilities from the literature lead to higher depression and remission prevalence than is stated in the literature. Additionally, we observed that fewer individuals had never experienced an episode of depression in their lifetime (65.1% calculated vs. 75.5% reported for females and 77.7% calculated vs. 83.2% reported for males) in the model results compared to the reported values.

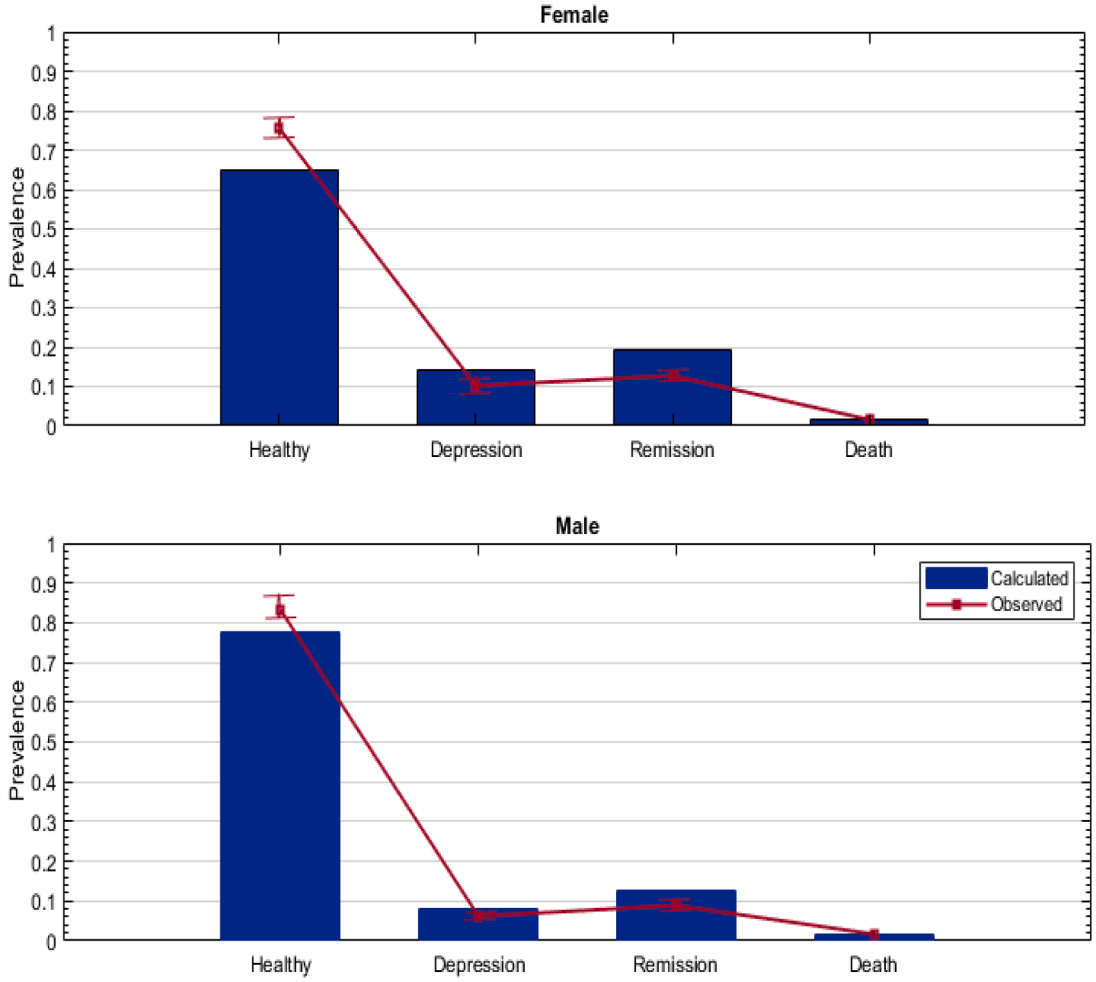

Fig. 3. Comparison of calculated prevalence from the Markov model and observed prevalence from the literature for each health state assuming that incidence is correct, where the observed proportion of healthy people is higher than the steady state analysis.
* Observed prevalence rates are obtained from Kessler et al. (1) and calculated as 95% CI.

3.3. *Hypothesis III: prevalence and incidence rates are biased*

Using several scenarios for incidence values, we obtain calculated values for prevalence from the steady-state model and recall bias necessary to match reported prevalence values.

We quantified that lifetime prevalence in steady state feasibly ranged from 30.2% to 36.4% for females (vs. 17.6 to 24.5% for males) and 12-month prevalence ranged from 12.8% to 15.4% for females (vs. 6.8 to 9.6% for male). Even with the lowest incidence rate, the lifetime prevalence values in steady state are higher (as shown in Fig. S1) than the values in the literature, which are 22.9% and 15.1%, respectively (Kessler et al., 2010), for females and males.

We calculated the recall bias of the lifetime prevalence based on the gap between the estimated lifetime prevalence from the Markov model (as shown in Table 2) and the reported lifetime prevalence in the literature. The gap between calculated and reported lifetime prevalence ranged from 31.9% to 59% for females and 16.3% to 62.3% for males. Additionally, for the lower bound of incidence (which was 25.6% and 38.1% lower than the average incidence, respectively for females and males), we calculated the recall bias for the 12-month prevalence of major depression as 25.5% for females and 9% for males. In comparison, the average recall bias identified in the literature is 60% for female and 63% for males (Kruijshaar et al., 2005).

Table 2
Calibration of the Markov model and calculated prevalence and recall bias.

Gender	Low	Medium	High
	Incidence*		
Female	0.0029	0.0039	0.0051
Male	0.0013	0.0021	0.0033
Calculations from Markov Model			
	Lifetime Prevalence (%)		
Female	30.2	33.2	36.4
Male	17.6	20.5	24.5
	12-month Prevalence (%)		
Female	12.8	14.1	15.4
Male	6.8	8.0	9.6
Calculated Recall Bias			
	Lifetime Prevalence (%)		
Female	31.9	45.0	59.0
Male	16.3	35.8	62.3
	12-month Prevalence (%)		
Female	25.5	38.2	51.0
Male	9.0	29.0	54.8

* 95% CI of the incidence rate in the Baltimore ECA study (Eaton et al., 2007).

3.4. *Sensitivity analysis results*

We reported sensitivity analysis result in Fig. S7 in the supplementary appendix. All recall biases quantified are positive, except for recall of 12-month prevalence in males for two cases. The two cases with recall

bias equal to zero or negative (calculated prevalence is higher than reported prevalence) both have incidence equal to the lower bound; in one the remission rate is 20% higher than the point estimate, and in the other the recurrence rate is 20% lower than the point estimate.

3.5. Age-specific analysis of lifetime prevalence

We performed age-specific analysis, including additional parameter settings of higher mortality for people who are depressed or with increased incidence for the younger population. In Table S2, we report the calculated lifetime prevalence values using mean incidence rates. Additionally, based on the upper and lower bounds (95% CI) of the incidence rate (Table S1), we calculate the range, as shown by the bars included on Figs. S2 and S3.

Figs. S2 and S3, which also display the reported lifetime prevalence (Kessler et al., 2010), show that the calculated lifetime prevalence starts to drop after the age of 49, as the reported does. However, the difference between reported and calculated prevalence is increasing with age, except for age 90.

Table S3 shows the recall bias (the difference between reported and calculated prevalence) that would be necessary for the system to be feasible for low, medium, and high values of the incidence rate. For some ages (e.g., < 65), there are recall bias values that could be possible, especially for low and medium incidence. On the other hand, for the highest ages.s, recall values above 100 are not possible, suggesting that recall bias alone does not explain the infeasibility in the system based on incidence, lifetime prevalence, and recall bias.

Furthermore, we observed that recall bias still exists even if the mortality risk of people with depression was much higher than others, e.g., a 4-fold increase (Figs. S2 and S3 in the supplementary appendix) and if there is a significant (e.g., 6-fold) increase in incidence among younger birth cohorts (Fig.s S4 and S5 in the supplementary appendix).

4. Discussion

In hypothesis I, we found that assuming prevalence values from the literature are correct resulted in an incidence rate that is lower than reported in the literature (Eaton et al., 2007), so this hypothesis seems unlikely to be true. In hypothesis II, we found that assuming incidence values from the literature are correct resulted in lifetime prevalence values that are higher than are reported in Kessler et al. (2010), so this hypothesis also seems unlikely to be true. In hypothesis III, we reported recall bias estimates for incidence rates at their lowest and highest reported values. Even with the lower bound of incidence, we find that the lifetime prevalence in steady state is higher than that reported in practice. This finding suggests that either incidence is over-reported (which seems unlikely given that it is reflecting a specific snapshot in time) or that lifetime prevalence as reported may be underestimated. We find that the model can be in steady state with the lowest reported values of incidence and with reasonable recall bias (32% for females, 16% for males).

One of the possible explanations of high lifetime prevalence among younger populations is a "cohort effect". E.g., younger generations may have an elevated risk of depression. However, our incidence adjustment based on cohort effects (Twenge et al., 2019) showed that birth cohorts alone do not explain the gap between incidence and resulting lifetime prevalence for older generations (in the supplementary appendix Fig. S4 S5).

On the other hand, we may see fewer patients with depression in older age groups because of the elevated risk of mortality of patients with current depression. We account for increased mortality in patients with depression in our model, and we conclude this alone does not explain the results. Our age-specific analysis further supports this conclusion.

Cohort effects and increased mortality of patients with current depression are insufficient to explain the gap between reported and calculated lifetime prevalence (Appendices 2 and 3) (Patten et al., 2010). There may be other factors that contribute to the pattern, such as the changes in diagnostic criteria. Conditions that are diagnosed (or denoted) as depression now, were called other names prior to the past half-century (e.g., "anxiety" or "melancholy") (Horwitz, 2010). If so, then survey instruments may need to account for this when questioning patients from previous birth cohorts.

Our age-specific analysis showed that the recall bias would need to increase until the age of 79 for the feasibility of other parameters (Table S3). To be more conservative on the calculation of bias, we used annual and lifetime prevalence from NCS-R (Kessler et al., 2010), which reported relatively higher rates than the ECA study (Eaton et al., 2007) from which incidence rates are obtained.

Recall problems may increase with age because of high risk for cognitive decline. Mental health problems may also fade in the face of physical ailments that increase with age (Bor, 2015). As well as age, the number of previous episodes and time since the last episode may affect reporting. It has been reported that 10% of patients did not report their depression episodes at onset (Patten et al., 2012), and recall bias can exist when the recall period is as short as one week (Zanni, 2007).

In our results, we calculated relatively low recall bias rates for males would be necessary to match incidence and lifetime prevalence in the steady state. This does not seem true because some studies show that females have a better memory than males (Lundervold et al., 2014). In the age-specific analysis, we observed that men have higher recall bias in older ages ($\geq$79 years old) than females (Table S3 and Fig. S5 in the supplementary appendix). However, men may hide psychological problems, and they may be reluctant to seek help for their conditions (Lee and Owens, 2002; O'Brien et al., 2005). This indicates that the incidence rate may be underestimated for males, which would increase the recall bias. Additionally, another concern is measurement bias, which may lead to underestimating the disease burden among men. In general, tools and questions that are used in surveys detect the symptoms the same way for men and women (Smith et al., 2018).

One recent study showed a relatively higher lifetime prevalence (14.7% for males and 26.1% for females) of major depression than previous studies, using The National Epidemiologic Survey on Alcohol and Related Conditions III (Hasin et al., 2018). However, concerns about the underreporting rates still exist even with cumulative estimates (Wells and Horwood, 2004). Cumulative evaluations from multiple interviews may also underestimate the true lifetime prevalence of major depression because of the patients who did not recall their lifetime event in all interviews (Takayanagi et al., 2014). Therefore, the lifetime prevalence values estimated from simulation studies were higher than the general population surveys, either one-time retrospective or cumulative evaluations (Kruijshaar et al., 2005) estimated the lifetime prevalence of 20% for men and 30% for women from a microsimulation study.

Our model provides a preliminary framework for the natural history of major depression. Additionally, we quantified the calibration for incidence and lifetime prevalence. Our findings mathematically prove and support the arguments around the potential discordance between incidence and lifetime prevalence rates, which can be solved in part by adjusting for recall bias.

Developed Markov models can also be used by other researchers to estimate recall bias or other adjustments needed to ensure the feasibility of parameters for models. Furthermore, our models are available for others, and we provide steady-state calculations as an R script (Appendix 4 in the supplementary appendix).

4.1. Limitations

This study has several limitations related to the model structure, data inputs, and calibration process. First, we have included a relatively small number of states rather than characterizing based on the severity of depression (low, moderate, and high) and recovery procedure (partial

remission, full remission). Additionally, our model is limited to major depression, and patients who have other types of depression are assumed healthy. Second, the stationary (time-homogeneous) assumption of the model may not entirely match with practice; however, it is useful to show the stability of disease distribution and quantify adjustments (Bush and Zaremba, 1971). Besides, we support the analysis by providing an age-specific analysis of lifetime prevalence with time-dependent parameters (e.g., incidence, prevalence, and mortality). Furthermore, in all the calibration process steps, we assume that transition probabilities (except incidence rates, lifetime and annual prevalence at the age of 18) are correct, while we obtained the data from various studies from different years.

5. Conclusions

Average incidence estimates have often seemed unrealistically high (Eaton et al., 1997; Patten, 2008) because they would lead to excessively high lifetime prevalence (33.2% for females and 20.5% for males); however, lifetime prevalence is probably much higher (Kruijshaar et al., 2005; Takayanagi et al., 2014) than reported in general population surveys.

In the literature, studies reported recall bias of 38% or more (Andrews et al., 1999; Kruijshaar et al., 2005). Our rates are consistent with this rate while additionally suggesting that the average incidence rates may be underestimated for males.

Our conclusions apply to future models or estimates of burden. For all populations, we recommend using incidence rates that are low to medium from reported values, and lifetime prevalence values that are medium to high for reported values. We conclude that recall bias needs to be accounted for in modeling or burden estimates, where the recall bias should increase with age. We concur that incidence for males is likely under-reported, and additional refinements may be considered for survey instruments to capture gender or age-based effects. We recommend that modelers continue to use steady state analysis, especially as reported values are updated over time, to ensure that parameters used within models are consistent across multiple types of environments.

Author statement

Funding

This research was supported in part by the Center for Health and Humanitarian Systems, the William W. George Endowment at Georgia Tech, the following Georgia Tech benefactors: Andrea Laliberte, Claudia L. and J. Paul Raines, and Richard E. “Rick” and Charlene Zalesky, and the Allison Distinguished Professorship at NC State University.

Declaration of Competing Interest

None reported.

Acknowledgements

The authors are grateful to the Editors and anonymous reviewers for their thoughtful, constructive comments. We gratefully acknowledge Hannah Smalley, Ph.D., for her comments and suggestions on the first draft of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2021.09.079.

References

Ali Afzali, H.H., Karnon, J., Gray, J., 2012. A critical review of model-based economic studies of depression. PharmacoEconomics 30, 461–482.
Andrews, G., Anstey, K., Brodaty, H., Issakidis, C., Luscombe, G., 1999. Recall of depressive episode 25 years previously. Psychol. Med. 29, 787–791.
Arias, E., Heron, M., Xu, J., 2014. United States life tables. Natl. Vital Stat. Rep. 66, 1–64.
Barendregt, J.J., Van Oortmarssen, G.J., Vos, T., Murray, C.J.L., 2003. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. Popul. Health Metr. 1, 4. -4.
Bland, R.C., 1992. Psychiatric disorders in America: the epidemiologic catchment area study. J. Psychiatry Neurosci. 34–36.
Bor, J., 2015. Among the elderly, many mental illnesses go undiagnosed. Health Affairs 34, 727–731.
Bush, J.W., Zaremba, J., 1971. Estimating health program outcomes using a Markov equilibrium analysis of disease development. Am. J. Public Health 61, 2362–2375.
Cuijpers, P., Vogelzangs, N., Twisk, J., Kleiboer, A., Li, J., Penninx, B.W., 2014. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am. J. Psychiatry 171, 453–462.
Eaton, W.W., Anthony, J.C., Gallo, J., Cai, G., Tien, A., Romanoski, A., Lyketsos, C., Chen, L.S., 1997. Natural history of diagnostic interview schedule/DSM-IV major depression. The baltimore epidemiologic catchment area follow-up. Arch. Gen. Psychiatry 54, 993–999.
Eaton, W.W., Kalaydjian, A., Scharfstein, D.O., Mezuk, B., Ding, Y., 2007. Prevalence and incidence of depressive disorder: the Baltimore ECA follow-up, 1981-2004. Acta. Psychiatr. Scand. 116, 182–188.
Faissol, D.M., Griffin, P.M., Swann, J.L., 2009. Bias in Markov models of disease. Math. Biosci. 220, 143–156.
Foley, D.L., Neale, M.C., Kendler, K.S., 1998. Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. Psychol. Med. 28, 857–870.
Giuffra, L.A., Risch, N., 1994. Diminished recall and the cohort effect of major depression: a simulation study. Psychol. Med. 24, 375–383.
Goenka, A., Liu, L., Nguyen, M.-H., 2021. SIR economic epidemiological models with disease induced mortality. J. Maths. Econ. 93, 102476.
Gong, J., Simon, G.E., Liu, S., 2019. Machine learning discovery of longitudinal patterns of depression and suicidal ideation. PLOS ONE 14, e0222665.
Gopalappa, C., Guo, J., Meckoni, P., Munkhbat, B., Pretorius, C., Lauer, J., Ilbawi, A., Bertram, M., 2018. A two-step markov processes approach for parameterization of cancer state-transition models for low- and middle-income countries. Med. Decis. Making 38, 520–530.
Greenberg, P.E., Fournier, A.A., Sisitsky, T., Pike, C.T., Kessler, R.C., 2015. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J. Clin. Psychiatry 76, 155–162.
Harman, J.S., Veazie, P.J., Lyness, J.M., 2006. Primary care physician office visits for depression by older Americans. J. Gen. Intern. Med. 21, 926–930.
Hasin, D.S., Sarvet, A.L., Meyers, J.L., Saha, T.D., Ruan, W.J., Stohl, M., Grant, B.F., 2018. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 75, 336–346.
Horwitz, A.V., 2010. How an age of anxiety became an age of depression. The Milbank quarterly 88, 112–138.
Jiao, B., Rosen, Z., Bellanger, M., Belkin, G., Muennig, P., 2017. The cost-effectiveness of PHQ screening and collaborative care for depression in New York City. PLOS ONE 12, e0184210.
Julien, J., Ayer, T., Bethea, E.D., Tapper, E.B., Chhatwal, J., 2020. Projected prevalence and mortality associated with alcohol-related liver disease in the USA, 2019–40: a modelling study. The Lancet Public Health 5, e316–e323.
Kessler, R.C., Barber, C., Birnbaum, H.G., Frank, R.G., Greenberg, P.E., Rose, R.M., Simon, G.E., Wang, P., 1999. Depression in the workplace: effects on short-term disability. Health Aff. (Millwood) 18, 163–171.
Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., National Comorbidity Survey, R., 2003. The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R). JAMA 289, 3095–3105.
Kessler, R.C., Birnbaum, H., Bromet, E., Hwang, I., Sampson, N., Shahly, V., 2010. Age differences in major depression: results from the national comorbidity survey replication (NCS-R). Psychol. Med. 40, 225–237.
Kessler, R.C., Zhao, S., Blazer, D.G., Swartz, M., 1997. Prevalence, correlates, and course of minor depression and major depression in the national comorbidity survey. J Affect. Disord. 45, 19–30.
Knauper, B., Wittchen, H.U., 1994. Diagnosing major depression in the elderly: evidence for response bias in standardized diagnostic interviews? J. Psychiatr. Res. 28, 147–164.
Kolovos, S., Bosmans, J.E., Riper, H., Chevreul, K., Coupé, V.M.H., van Tulder, M.W., 2017. Model-based economic evaluation of treatments for depression: a systematic literature review. PharmacoEconomics - Open 1, 149–165.
Kruijshaar, M.E., Barendregt, J., Vos, T., de Graaf, R., Spijker, J., Andrews, G., 2005. Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. Eur. J. Epidemiol. 20, 103–111.
Kruijshaar, M.E., Barendregt, J.J., Hoeymans, N., 2002. The use of models in the estimation of disease epidemiology. Bull. World Health Organ. 80, 622–628.
Kruijshaar, M.E., Barendregt, J.J., Van De Poll-Franse, L.V., 2003. Estimating the prevalence of breast cancer using a disease model: data problems and trends. Popul. Health Metr. 1, 5.
Lee, C., Owens, R.G., 2002. Issues for a psychology of men’s health. J. Health Psychol. 7, 209–217.

Lin, Y., Huang, S., Simon, G.E., Liu, S., 2016. Analysis of depression trajectory patterns using collaborative learning. Maths. Biosci. 282, 191–203.
Lin, Y., Huang, S., Simon, G.E., Liu, S., 2018. Data-based decision rules to personalize depression follow-up. Scientific Reports 8, 5064.
Lin, Y., Huang, S., Simon, G.E., Liu, S., 2019. Cost-effectiveness analysis of prognostic-based depression monitoring. IISE Trans. Healthcare Syst. Eng. 9, 41–54.
Lundervold, A.J., Wollschlager, D., Wehling, E., 2014. Age and sex related changes in episodic memory function in middle aged and older adults. Scand. J. Psychol. 55, 225–232.
Mattisson, C., Bogren, M., Nettelbladt, P., Munk-Jorgensen, P., Bhugra, D., 2005. First incidence depression in the lundby study: a comparison of the two time periods 1947-1972 and 1972-1997. J. Affect Disord. 87, 151–160.
Murphy, J.M., Nierenberg, A.A., Laird, N.M., Monson, R.R., Sobol, A.M., Leighton, A.H., 2002. Incidence of major depression: prediction from subthreshold categories in the stirling county study. J. Affect Disord. 68, 251–259.
National Institute of Mental Health, Major Depression.
O'Brien, R., Hunt, K., Hart, G., 2005. It's caveman stuff, but that is to a certain extent how guys still operate': men's accounts of masculinity and help seeking. Soc. Sci. Med. 61, 503–516.
Patten, S.B., 2003. Recall bias and major depression lifetime prevalence. Soc. Psychiatry Psychiatr. Epidemiol. 38, 290–296.
Patten, S.B., 2008. Major depression prevalence is very high, but the syndrome is a poor proxy for community populations' clinical treatment needs. Can. J. Psychiatry 53, 411–419.
Patten, S.B., 2009. Accumulation of major depressive episodes over time in a prospective study indicates that retrospectively assessed lifetime prevalence estimates are too low. BMC Psychiatry 9, 19.
Patten, S.B., Gordon-Brown, L., Meadows, G., 2010. Simulation studies of age-specific lifetime major depression prevalence. BMC Psychiatry 10, 85.
Patten, S.B., Williams, J.V., Lavorato, D.H., Bulloch, A.G., D'Arcy, C., Streiner, D.L., 2012. Recall of recent and more remote depressive episodes in a prospective cohort study. Soc. Psychiatry Psychiatr. Epidemiol. 47, 691–696.
Pincus, H.A., Tanielian, T.L., Marcus, S.C., Olfson, M., Zarin, D.A., Thompson, J., Magno Zito, J., 1998. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. JAMA 279, 526–531.
Ross, E.L., Vijan, S., Miller, E.M., Valenstein, M., Zivin, K., 2019. The cost-effectiveness of cognitive behavioral therapy versus second-generation antidepressants for initial treatment of major depressive disorder in the United States: a decision analytic model. Ann. Intern. Med. 171, 785–795.
Ross, S.M., 2010. CHAPTER 4 - Markov Chains. In: Ross, S.M. (Ed.), Introduction to Probability Models, Tenth Edition. Academic Press, Boston, pp. 191–290.
Smith, D.T., Mouzon, D.M., Elliott, M., 2018. Reviewing the assumptions about men's mental health: an exploration of the gender binary. Am. J. Men's Health 12, 78–89.
Takayanagi, Y., Spira, A.P., Roth, K.B., Gallo, J.J., Eaton, W.W., Mojtabai, R., 2014. Accuracy of reports of lifetime mental and physical disorders: results from the baltimore epidemiological catchment area study. JAMA Psychiatry 71, 273–280.
Team, R.C., 2013. R: A language and environment for statistical computing.
Twenge, J.M., Cooper, A.B., Joiner, T.E., Duffy, M.E., Binau, S.G., 2019. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset. J. Abnorm. Psychol. 128, 185–199.
Wells, J.E., Horwood, L.J., 2004. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol. Med. 34, 1001–1011.
Yan, C., Rittenbach, K., Souri, S., Silverstone, P.H., 2019. Cost-effectiveness analysis of a randomized study of depression treatment options in primary care suggests stepped-care treatment may have economic benefits. BMC Psychiatry 19, 240.
Zanni, G.R., 2007. Patient diaries: charting the course. Consult. Pharm. 22, 472–476, 479-482.