Challenges in Measuring Mental Health Trends

Alden Cheng

June 28, 2023

Abstract

Public awareness of mental health issues has grown in recent years, and there is a common perception that mental health in the population is worsening, concerns that are supported by descriptive evidence. However, changes in public attitudes towards mental health make the interpretation of these trends challenging: low response rates to mental health surveys make their results sensitive to changes in sample selection bias, and even diagnosis rates in comprehensive data sets such as the Medicare data are a function of individuals' willingness to seek professional help. In this paper, I address these measurement issues using a comprehensive data set on all nursing home residents, so it does not suffer from sample selection bias. Similar to Medicare data, it contains information on whether each resident was diagnosed with depression in the recent past, but in addition, it contains a rich set of psychosocial measures, which provides us with a detailed picture of the resident's underlying mental health. I find that while depression diagnoses at admission increased from 19 to 24 percent for residents admitted to a nursing home in California between 2000 and 2010, underlying mental health of these residents (based on observed psychosocial behavior as well as machine-learning predictions using hundreds of covariates) was roughly constant over the same period. These results illustrate the perils of inferring mental health trends from survey evidence or diagnosis trends alone.

1 Introduction

The confluence of several factors has led to increased awareness of mental health issues among the general public, policymakers, as well as academic researchers in the US. For example, Case and Deaton (2015) document a rise in deaths due to suicides and self-poisoning among middle-aged white men, several papers have studied the effect of social media on mental health (Allcott et al. 2020;

I am very grateful to David Autor, David Card, and Frank Schilbach for their valuable comments.

Braghieri, Levy, and Makarin 2022), and 90% of the public think that there is a mental health crisis according to the CNN/KFF Mental Health in America survey in 2022. In addition, numerous pieces of descriptive statistics support this concern that mental health in the US has been worsening in recent years. According to national surveys conducted by Mental Health America, 20.78 percent of adult respondents to a national survey conducted by Mental Health America (MHA) in 2019–2020 reported experiencing a mental illness, up from 18.19 percent in 2015; in addition, 16.39 percent of youth respondents report suffering from at least one major depressive episode in 2019–2020, up from 8.66 percent in 2015. Similar trends are found based on other measures of mental health or based on other surveys (MHA 2015, 2022; Kaiser Family Foundation 2022; Healthy Minds Study).

At the same time, stigma around mental health conditions have generally been decreasing (Pescosolido et al. 2021), which may potentially complicate the interpretation of these trends for several reasons. First, respondents to surveys may be more willing to admit suffering from mental conditions which are less stigmatized. Second, response rates to these surveys are typically quite low (e.g. the response rate for the Healthy Minds Study in 2018–2019 was only 16 percent). It is unclear whether non-response bias can be adequately adjusted for based on observables, and more importantly, characteristics of non-respondents may also change over time due to changes in societal attitudes. Third, increased awareness, reduced stigma and/or the increased availability of mental health resources may increase the number of individuals seeking professional help and thus numbers of diagnoses. Fourth, suicides are typically relatively rare events, and there is often a certain amount of discretion in whether certain deaths should be classified suicides (e.g. fatal single-person car accidents).

The first three reasons listed above are related to sample selection bias. As an example of how serious these measurement issues are, Ibrahim, Kelley, and Adams (2013) find in a review of studies on the prevalence of depression under undergraduate college students that estimates range anywhere from 10% and 84.5%, and moreover, the estimates at the two ends of this extreme are relatively precisely estimated. Even large-scale reputable surveys often have low response rates: the Healthy Minds Study has a 14% response rate in 2020, so while the proportion of respondents reporting major depression was 21%, the true proportion including non-respondents may range anywhere from 3–89%. In this paper, I study this issue of sample selection bias by considering a portion of the US population for whom we have comprehensive data on, specifically, Minimum Data Set 2.0 (MDS) data on the universe of nursing home residents. To the extent that nursing home admissions are not determined by factors related to depression diagnoses, this provides a solution to the issue of sample selection.

Observed trends in depression among nursing home residents from the MDS still may not provide an accurate picture, given that diagnoses patterns may change along with societal attitudes towards depression. Therefore, I measure underlying mental health, using a rich set of health and demographic variables available for each resident. Not only does this reveal trends in underlying mental health, it also identifies changes in diagnoses patterns, i.e., for a given level of underlying mental health, how does the probability of diagnosis differ over time?

I measure underlying mental health in two ways. First, the MDS contains a section on "Indicators of Depression, Anxiety, and Sad Mood". Importantly, the items contained in this section of the survey are relatively objective (as opposed to the subjectivity of depression diagnosis). For example, one item asks if the resident "made negative statements" and provided concrete examples (e.g. "Nothing matters; Would rather be dead, ..., Let me die"), and another item asks if the resident presented "crying, tearfulness".

Second, I make use of the rich set of resident characteristics in the data to obtain a machine-learning prediction of the probability of depression diagnosis — higher predicted probability indicates poorer underlying mental health. Of course, diagnoses patterns change over time so the dependent variable in these predictions is a moving target. To mitigate this, when predicting the probability of depression diagnosis for a resident who is admitted in a particular year, I make the prediction based on a model estimated with the other years in the data, which is similar to a cross-validation approach.

I find that while depression diagnoses at admission increased from 19 to 24 percent for residents admitted to a nursing home in California between 2000 and 2010, underlying mental health of these residents stayed relatively constant over the period. These results are robust to different measures of underlying mental health, such as observed psychosocial measures (which I also sometimes refer to as "mood variables", or machine-learning predictions using hundreds of covariates. These results illustrate the perils of inferring mental health trends from survey evidence or diagnosis trends alone.

I describe my main data source in section 2, and discuss the empirical framework for my analysis in section 3. Section 4 presents results of my analysis, and section 5 concludes.

2 Data

The main data source for this project is the Minimum Data Set 2.0 (MDS), which contains data on the universe of residents admitted to a nursing home that receives federal funding.[1] For each resident,

[1]This accounts for roughly 96% of all nursing homes in the US.

a comprehensive assessment is filled in at entry, followed by quarterly assessments or whenever there is a significant change of status, and finally when the resident is discharged. I mainly use the assessment form filled in at admission, since I want to isolate trends in depression diagnoses from nursing homes' effects on residents' mental health.

The data contains a rich set of health and demographic variables for each resident at entry. Importantly for our purposes, it contains dozens of variables on residents' mood, behavior, and psychosocial wellbeing, information on which is rarely available in healthcare data sets. I use these variables as proxies or predictions of residents' baseline mental health. There is also an item in the assessment form for whether the resident suffers from depression, and the coding for this "require a physician-documented diagnosis in the clinical record". Hence, this variable represents depression diagnosis at the time of or prior to the resident's admission to the nursing home, and I will study how this varies over time, holding other resident characteristics constant.

Table 1 shows summary statistics for my sample. We observe that residents tend to be quite old, about 60 percent are female, one-third are married, and most residents are white. The depression score variable summarizes 16 questions related to residents' mood and behavior patterns. This variable is defined as the sum of each of the answers to the 16 questions (which are associated with worse mental health, and can take values one or two depending on frequency), so it can take any value between 0 and 32, and we observe that the average depression score is quite low.

Comparing the subpopulations that are diagnosed with depression versus not, we observe that diagnosed individuals tend to be slightly younger at admission, and are more likely to be white. In addition, the average depression score for individuals diagnosed with depression is roughly two times that of inviduals who are not diagnosed, which suggests that the mood and behavior variables do contain some predictive power for depression diagnosis.

3 Framework for Studying Changes in Diagnosis Patterns

Suppose there is some "true" underlying index of depression Y_i^*. In my first approach for measuring residents' underlying health, I will simply assume that it is a function of residents' responses to the questions in the section of the MDS, titled "Indicators of Depression, Anxiety, and Sad Mood". Given that these survey questions tend to ask about very specific symptoms and include examples (e.g., residents making certain statements, or exhibiting particular behaviors such as crying), it seems reasonable to assume that these variables are measured relatively consistently over time and are generally

Table 1: Summary Statistics for Residents

	(1) All	(2) Depressed	(3) Not Depressed
Age	77.605	75.635	78.152
	(12.908)	(13.219)	(12.767)
Female	0.611	0.661	0.597
	(0.488)	(0.474)	(0.491)
Married	0.338	0.327	0.342
	(0.473)	(0.469)	(0.474)
White	0.736	0.793	0.720
	(0.441)	(0.405)	(0.449)
Black	0.069	0.052	0.074
	(0.253)	(0.222)	(0.261)
Hispanic	0.109	0.097	0.113
	(0.312)	(0.296)	(0.316)
At Least Bachelor's Degree	0.135	0.139	0.134
	(0.342)	(0.345)	(0.341)
Depression Score	1.390	2.176	1.172
	(2.262)	(2.653)	(2.089)
Number of Residents	717,163	155,837	561,326

Notes: This table contains summary statistics for residents who had their first stays in a nursing home in California between 2000 and 2010.

less affected by social attitudes towards mental health.

In the simplest specification, I take the sum of residents' scores for each question in this section (where higher scores denote worse mental health) as the measure of underlying mental health and call this the "depression score", although I also consider other definitions later. Comparing the time series of average depression score of admitted residents to the time series of average diagnoses rates for the same residents yields insights into changes in diagnoses patterns: if the average depression score remains constant or falls over time whereas diagnoses rates are increasing over time, this suggests that holding mental health constant, individuals are being diagnosed with depression at a higher rate over time.

However, we may worry that measurement of the "depression score" may also be affeected changes in stigma towards mental health. For example, when mental health is more stigmitized, nurses may pay less attention to behavior associated with depression, so that these symptoms are less likely to be reported in the survey. Moreover, since the depression score and depression diagnoses are measured on the same scale, it is difficult to compare differences in the magnitudes of these trends. So, for my second measure of underlying mental health, I take a more data-dependent approach, by defining mental health as the probability of being depressed, obtained from a machine learning model with a rich set of predictors.

To elaborate, suppose that the underlying mental health Y_i^* is related to observed resident characteristics X_i and unobservables u_i by the following equation:

$$Y_i^* = f(X_i) + u_i,$$

where lower values of Y_i^* denotes better mental health. I will make a couple of observations about this setup. First, this is not a causal relationship, i.e., X_i and u_i are variables that predict the depression index. Second, I assume here that the relationship is time-invariant, an assumption that I will relax later on. Hence, I exclude from X_i variables that are clearly outcomes of depression diagnoses, e.g., the use of antidepressants, considering that diagnoses patterns may change over time. In relation to the data, we can think of X_i as corresponding to health and demographic variables that would be recorded consistently over time.

As we noted earlier, diagnoses patterns may not be time-invariant. So, suppose that potential diagnosis for individual i at time $t \in \mathcal{T}$ is given by the function:

$$D_i(t) = \mathbb{I}\left[Y_i^* + \mu_{t(i)} > 0\right] = \mathbb{I}\left[f(X_i) + \mu_t + u_i > 0\right],$$

with u_i is distributed i.i.d. with mean zero and variance σ_u^2. If μ_t is increasing in t, then this means that probability of diagnosis is increasing over time (holding mental health constant). For now, I am only using the assessment at admission for each individual, so for each individual, we only observe actual diagnosis $D_i = D_i(t(i))$ at the time of admission $t(i)$.

Hence, to study possible trends in μ_t, we would like to compare individuals across time to see whether conditional on a rich set of observables X_i, probability of diagnosis $Pr(D_i = 1|X_i, t)$ is increasing in t. Specifically, we can estimate:

$$Pr(D_i = 1|X_i, t) = Pr\left(f(X_i) + \mu_t + u_i > 0|X_i, t\right),$$

and examine how much the estimates of μ_t vary across t. In order for this estimation to recover the time effects μ_t, I make the following unconfoundedness assumption.

A1) (Unconfoundedness)

$$\{D_i(t)\}_{t\in\mathcal{T}} \perp \{\mathbb{I}[t(i)=t]\}_{t\in\mathcal{T}}|X_i.$$

In words, this means that conditional on X_i, knowing which year the resident is admitted in $t(i)$ tells us nothing about her potential outcomes (whether or not they would be diagnosed if admitted in every possible year). An example of how this may be violated is if characteristics (not included in X_i) of residents admitted over different years are very different, and these characteristics are also associated with underlying mental health. The richness the data on resident characteristics X_i makes this scenario less likely.

The exercise above assumes that the relationship between resident characteristics and depression diagnosis remains constant (other than the time trend). Here, I describe an alternative approach that allows $f(X_i)$ to differ over time at the cost of requiring that the distribution of resident characteristics X_i is invariant over time.

For each $t = 1, ..., T$:

1. Estimate $Pr(D_i = 1|X_i, t)$ using period t data, which gives us an estimate $\hat{f}_t(X_i)$.
2. Using the model from year t, I predict $\hat{p}_i^{(t)}$ for individuals admitted in all other years $t' \neq t$. (The superscript (t) denotes predictions based on the model estimated using year t data.)
3. Plot the average value of $\hat{p}_i^{(t)}$ (which I denote by $\bar{\hat{p}}_{t'}^{(t)}$) for each $t' \neq t$, as a function of the year t'.

The question this exercise seeks to answer is: what is the probability that individuals in year t' would be diagnosed as having depression, based on the diagnosis criteria that is prevalent in year t. The procedure above produces a time series plot of $\bar{\hat{p}}_{t'}^{(t)}$ for each t, and assuming the distribution of X_i is time-invariant, if the threshold for depression diagnosis is getting lower over time, the curve for $\bar{\hat{p}}_{t'}^{(t)}$ should be shifted upwards for prediction models fitted on more recent data.

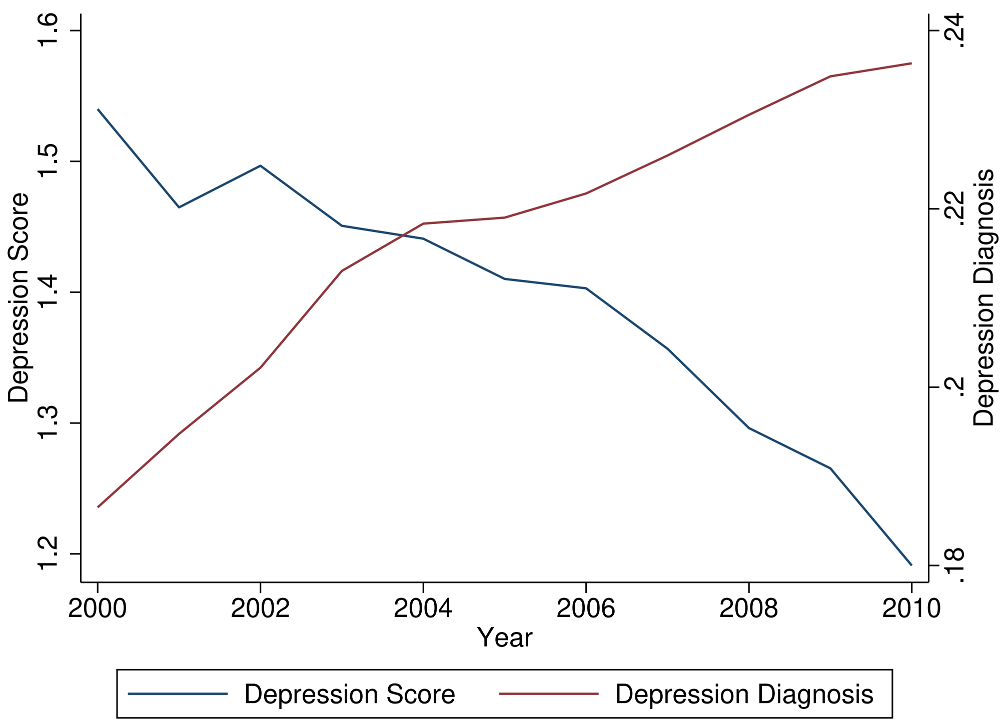
For the data-dependent measures of mental health, I explore using different sets of resident characteristics X_i is predictors of depression diagnosis. First, I consider only variables the mood variables recorded during the first two weeks after admission (X_i^m), since we may worry that some of the other variables (e.g., demographic variables) may not be associated with true mental health, but are instead associated with the probability of seeking treatment. The downside of this approach is the possibility that the way in which nurses record the mood variables may also change over time (e.g., because of changes in public perception towards mental health issues). To account for this possibility, I alternatively use only non-mood variables X_i^{nm} (i.e., demographic and health variables) for my predictions. Finally, I also obtain predictions based on all resident charateristics $X_i = (X_i^{m\prime}, X_i^{nm\prime})'$.

4 Results

4.1 Main Results

Figure 1 shows that according to the first measure of mental health — the "depression score" — underlying mental health of residents at admission seems to be improving over time, and yet rates of depression diagnosis for the same group of residents are also increasing over time. This suggests that holding mental health constant, individuals are being diagnosed with depression at a higher rate over time.

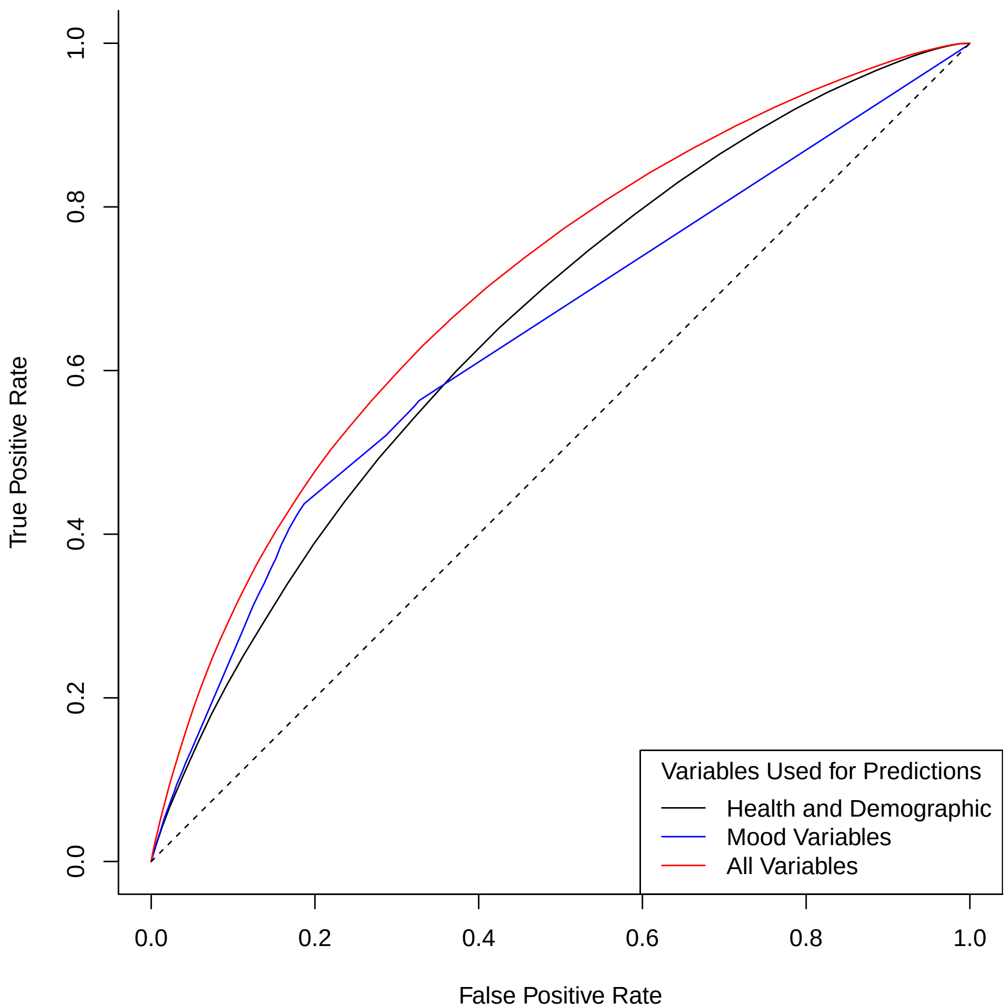
Figure 1: Trends in Depression Diagnoses and Depression Score Over Time



Before discussing the results that use more resident characteristics to control for underlying mental health Y_i^*, I first provide evidence that these variables are informative predictors of mental health. To do so, I obtain cross-validated predictions of depression for the models using either only health and demographic variables, only mood variables, or all variables to predict depression, and plot the receiver operating characteristic curves (ROC curves) for them.

The ROC curve plots the true positive and false positive rates if we use different thresholds to classify depression using a model's predictions. ROC curves for completely uninformative models (i.e., that randomly classifies residents as depressed with different probabilities) will lie on the 45-degree line, so models with ROC curves to the northwest of the diagonal and with greater areas under the curves (AUCs) are more informative. The results in Figure 2 show that the ROC curves for our models predicting depression are all above the 45-degree line, so they are informative to a certain degree. Interestingly, we also see that predictions made based on only health and demographic variables, or only mood variables do not perform much worse than predictions based on all variables.

Figure 2: ROC Curves for Prediction Models of Depression



Now, I consider instead data-dependent measures of underlying mental health. Specifically, I estimate double lasso regressions of depression diagnosis on the time trends and resident characteristics, with the time trends playing the role of the "treatment variable". I also include nursing home fixed effects in some specifications allow for the possibility that some nursing homes tend to admit residents of poorer mental health.

Columns 1 and 2 of Table 2 show that the estimated linear time trend is positive and statistically significant regardless of whether we include nursing home fixed effects, which is consistent with depression diagnosis threshold becoming lower over time. Columns 3 and 4 estimate time trends non-parametrically by include year fixed effects, and we observe that these year effects are steadily increasing over time, providing further support for the notion that the threshold for depression diagnosis is becoming lower over time.

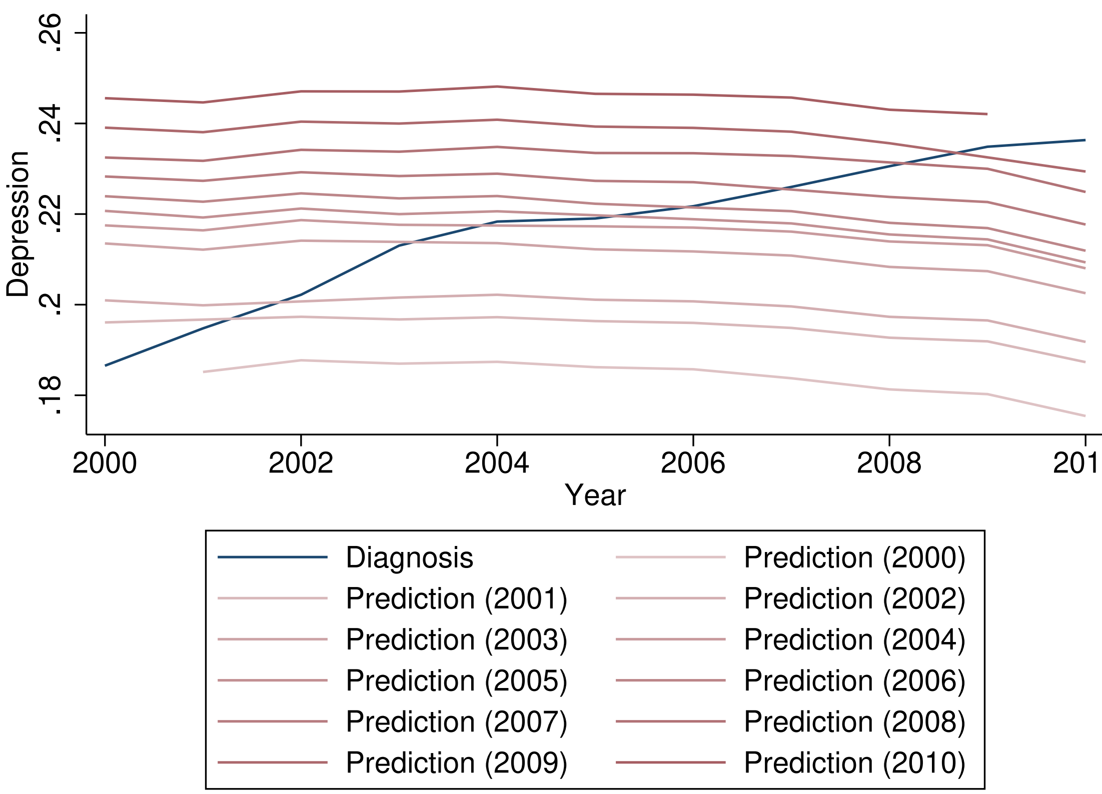
Table 2: Estimated Trends in Depression Diagnoses

	Depression Diagnosis (Percent)			
	(1)	(2)	(3)	(4)
Linear Time Trend	0.5196	0.4701		
	(0.0245)	(0.2306)		
Year: 2002			0.7488	0.8064
			(0.2234)	(0.2233)
Year: 2003			1.8625	1.8923
			(0.2256)	(0.2257)
Year: 2004			2.4246	2.4499
			(0.2247)	(0.2248)
Year: 2005			2.5438	2.6081
			(0.2237)	(0.2239)
Year: 2006			2.7807	2.8783
			(0.2232)	(0.2236)
Year: 2007			3.1793	3.3041
			(0.224)	(0.2244)
Year: 2008			3.6608	3.7413
			(0.2245)	(0.2249)
Year: 2009			4.0183	4.0863
			(0.2276)	(0.2279)
Year: 2010			4.476	4.547
			(0.2437)	(0.2439)
Controls for Resident Characteristics	X	X	X	X
Nursing Home Fixed Effects		X		X
N	717,169	717,169	717,169	717,169

Notes: This table shows estimated time trends from a double lasso regression of depression diagnosis on resident characteristics, where the time trend plays the role of the treatment variable. Time trends are either parametrized as a linear trend, or as year fixed effects (with 2000-2001 as the omitted category). Standard errors are clustered at the nursing home level.

Next, I provide show results from the approach that allows for the relationship between resident characteristics and diagnosis to vary over time (above and beyond what can be captured by a time trend). Figure 3 shows that the threshold for depression diagnosis seem to be getting lower over time, since the lines corresponding to predictions (based on all variables) using data from more recent years are shifted upwards. Moreover, these lines are relatively flat, which is consistent with the assumption that the distribution of X_i is time-invariant. Strikingly, we observe that the actual fraction of residents diagnosed with depression at admission is increasing over time. So, based on observed diagnoses alone, we might reach the wrong conclusion that mental health of more recent cohorts is getting worse, when in fact it is roughly constant, or even slightly decreasing, as Appendix Figures 5, 6, and 7 show. Appendix Figure 4 shows that these results remain qualitatively unchanged if we used predictions of depression based on only health and demographic variables, or mood variables alone.

Figure 3: Trends in Depression Diagnoses and Depression Index Based on Predictions from Various Years



A potential concern with the estimates above is that the "controls" X_i are endogenous. For example, suppose a resident originally had mental health Y_i^0, and if she is diagnosed with depression, this leads to lifestyle improvements (i.e., changes in X_i), which in turn improves mental health, so that mental health at admission Y_i^* is better than mental health at the time of diagnosis Y_i^0. Nonetheless, I show in Appendix Section 5 in a stylized model that this is likely to bias our estimate of the time trends downwards (in which case the true rise in diagnosis rate conditional on mental health is even larger than we estimated).

4.2 Heterogeneity in Diagnosis Patterns

Having shown that conditional on mental health, the probability of being diagnosed with depression increased over the period of 2000 and 2010, I next study whether the probabiliy of diagnosis also varies by other resident characteristics. To do so, I regress a dummy for depression on various resident characteristics, measures of the resident's underlying mental health, as well as nursing home and year fixed effects. As the measure of resident's underlying health, I use either the "depression score" (which is the sum of the mood variables), the components that make the score (i.e., the mood variables), as well as the predicted probability of suffering from depression (where the prediction omits the effect of the characteristic of interest in the heterogeneity analysis).

The results in Table 3 shows that conditional on underlying mental health, black and hispanic residents are less likely to be diagnosed with depression, whereas residents with at least a Bachelor's degree are more likely to be diagnosed, consistent with an existing literature documenting treatment gaps for disadvantaged groups (Finkelstein et al. 2012; Cowan and Hao 2021). Similarly, older residents are less likely to be diagnosed with depression, which may be due to older cohorts perceiving greater stigma against mental health (Bharadwaj, Pai, and Suziedelyte 2015). These results are robust to different controls for underlying mental health, as well as time and year fixed effects.

This interpretation requires that conditional on the depression score, these characteristics are not associated with mental health. However, in order for this type of "omitted variables bias" to completely explain the heterogneeity in probability of diagnosis by resident characteristics, minority race would have to be positively associated with underlying mental health (conditional on depression score) and education has to be negatively associated, which seems unlikely.

5 Conclusion

In this paper, I study trends in mental health and depression diagnosis among residents admitted to nursing homes in California between 2000 and 2010. The results show that while the rate of

Table 3: Heterogeneity in Diagnosis Rates

(a) Year Fixed Effects

	Depression Diagnosis								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race: Black	-0.0548***			-0.0545***			-0.0702***		
	(0.00293)			(0.00305)			(0.00285)		
Race: Hispanic	-0.0278***			-0.0280***			-0.0282***		
	(0.00265)			(0.00271)			(0.00247)		
Bachelor's Degree or Higher Education		0.0152***			0.0146***			0.0101***	
		(0.00160)			(0.00161)			(0.00143)	
Age			-0.00203***			-0.00191***			-0.00352***
			(0.000192)			(0.000163)			(0.000173)
Controls for Mental Health Condition	Depression Score			Mood Variables			Predicted Depression Probability		
Year Fixed Effects	X	X	X	X	X	X	X	X	X
N	717,169	717,169	717,163	717,169	717,169	717,163	717,169	717,169	717,163
R-squared	0.037	0.036	0.039	0.062	0.060	0.064	0.092	0.095	0.094

Notes: Standard errors are clustered at the nursing home level.

(b) Year and Nursing Home Fixed Effects

	Depression Diagnosis								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race: Black	-0.0478***			-0.0458***			-0.0590***		
	(0.00231)			(0.00223)			(0.00220)		
Race: Hispanic	-0.0249***			-0.0268***			-0.0226***		
	(0.00200)			(0.00196)			(0.00183)		
Bachelor's Degree or Higher Education		0.0131***			0.0130***			0.00777***	
		(0.00128)			(0.00127)			(0.00118)	
Age			-0.00258***			-0.00239***			-0.00385***
			(0.000129)			(0.000127)			(0.000120)
Controls for Mental Health Condition	Depression Score			Mood Variables			Predicted Depression Probability		
Year Fixed Effects	X	X	X	X	X	X	X	X	X
Nursing Home Fixed Effects	X	X	X	X	X	X	X	X	X
N	717,164	717,164	717,158	717,164	717,164	717,158	717,164	717,164	717,158
R-squared	0.051	0.051	0.056	0.078	0.078	0.082	0.106	0.108	0.107

Notes: Standard errors are clustered at the nursing home level.

depression diagnosis increased over this period, other measures of mental health based on observed resident behavior and machine-learning predictions using a rich set of resident characteristics remain relatively constant over the same period. This suggests that diagnosis patterns for depression changed over this period: holding mental health constant, individuals from later cohorts are more likely to be diagnosed with depression. This may either be due to changes in how physicians diagnose mental health conditions, or more likely (based on anecdotal evidence), that individuals become more willing to seek mental help over time (perhaps due to a reduction in social stigma towards mental health). In addition, I find that conditional on my measures of mental health, minority and older residents are less likely to be diagnosed with depression, whereas highly educated residents are more likely to be diagnosed, consistent with the previous literature on treatment gaps.

However, it should be noted that my finding that mental health seemed relatively constant for nursing home residents over 2000–2010 should not be generalized to wider segments of the population since the mental health risk factors they face may be very different. For example, the rise of social media is potentially much more consequential for the mental health of adolescents compared to the nursing home population.

The results in this paper caution against interpreting survey evidence or diagnosis rates alone as evidence of worsening mental health trends over time. Strong conclusions about mental health trends can only be reached if one has data that does not suffer from sample selection bias, and that has a rich set of predictors of mental health (not merely diagnosis of clinical outcomes alone). Many commonly used data sets such as mental health surveys or diagnosis rates from administrative data do not satisfy both of these criteria, due to either low response rates (for the former), or a paucity of direct mental health measures (in the latter case). Hence, even as policymakers around in the US and other countries strive to tackle the issue mental health (Ridley, Rao, Schilbach, and Patel 2020), greater efforts must be devoted towards data collection in order for policymakers to have an accurate picture of mental health in their populations.

References

[1] Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110(3): 629–676.

[2] Bharadwaj, Prashant, Mallesh M. Pai, and Agne Suziedelyte. 2017. "Mental Health Stigma." *Economics Letters* 159: 57–60.

[3] Braghieri, Luca, Ro'ee Levy, and Alexey Makarin. 2022. "Social Media and Mental Health." *American Economic Review* 112(11): 3660–3693.

[4] Case, Anne, and Angus Deaton. 2015. "Rising Morbidity and Mortality in Midlife among White Non-Hispanic Americans in the 21st Century." *Proceedings of the National Academy of Sciences* 112(49): 15078–15083.

[5] Cowan, Benjamin W., and Zhuang Hao. 2021. "Medicaid Expansion and the Mental Health of College Students." *Health Economics* 30(6): 1306–1327.

[6] Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3): 1057–1106.

[7] Healthy Minds Network. 2019. *The Healthy Minds Study: 2018–2019 Data Report*. Ann Arbor, MI: University of Michigan.

[8] Healthy Minds Network. 2021. *The Healthy Minds Study: Fall 2020 National Data Report*. Ann Arbor, MI: University of Michigan.

[9] Ibrahim, Ahmed K., Shona J. Kelly, Clive E. Adams, and Cris Glazebrook. 2013. "A Systematic Review of Studies of Depression Prevalence in University Students." *Journal of Psychiatric Research* 47(3): 391–400.

[10] KFF and CNN. 2022. *KFF/CNN Mental Health in America Survey*. October 2022.

[11] Mental Health America. 2015. *The State of Mental Health in America 2015*. Alexandria, VA: Mental Health America.

[12] Mental Health America. 2022. *The State of Mental Health in America 2022*. Alexandria, VA: Mental Health America.

[13] Pescosolido, Bernice A., Andrew Halpern-Manners, Liying Luo, and Brea Perry. 2021. "Trends in Public Stigma of Mental Illness in the US, 1996–2018." *JAMA Network Open* 4(12): e2140202.

[14] Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science* 370(6522): eaay0214.

Appendix

A1 Stylized Model of Endogenous Controls

Consider the following toy example, where for simplicity, we assume diagnosis D_0 is given by $D_0 = \beta_{X_0} X_0 + \delta t + u_i$, where u_i is uncorrelated with X_0 and t, t is uncorrelated with X_0, and $\delta > 0$. Here, X_0 represents healthy behaviors prior to needing nursing home care (so we have $\beta_{X_0} < 0$) and we demeaned all variables (so that D_0 actually takes values in $\{-\bar{D}_0, 1 - \bar{D_0}\}$). Moreover, suppose that depression diagnosis leads to better behavior, so that when the resident is admitted to the nursing home, we observe $X_1 = X_0 + \alpha D_0$, where $\alpha > 0$. Now, we estimate the following

$$D_0 = \tilde{\beta}_{X_1} X_1 + \tilde{\delta} t + \tilde{u},\ \mathbb{E}[(X_1, t)' \tilde{u}] = 0.$$

Now, we want to figure out how the estimated time trend $\tilde{\delta}$ relates to the true time trend. In the equation above, we can view X_0 as the "omitted variable". Next, we residualize t of X_1. To derive a convenient expression, first note that:

$$\begin{aligned} X_1 &= X_0 + \alpha D_0 \\ &= X_0 + \alpha(\beta_{X_0} X_0 + \delta t + u_i) \\ &= (1 + \alpha \beta_{X_0}) X_0 + (\alpha\delta) t + \alpha u_i. \end{aligned}$$

Since u_i is uncorrelated with t, this implies that $\alpha\delta$ is the regression coefficient from regressing X_1 on t. Now, considering the reverse regression, we find that the coefficient γ here is given by:

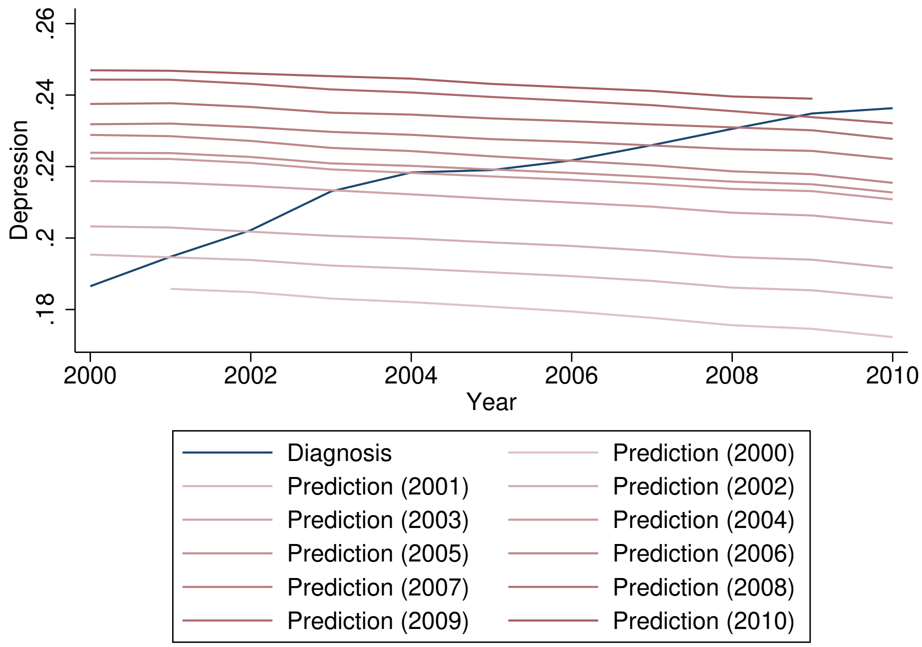
$$\gamma = \frac{\rho^2_{X_1,W}}{\alpha\delta},$$

where $\rho_{X_1,W}$ is the correlation between X_1 and W. So, by the Frisch-Waugh theorem, we obtain:

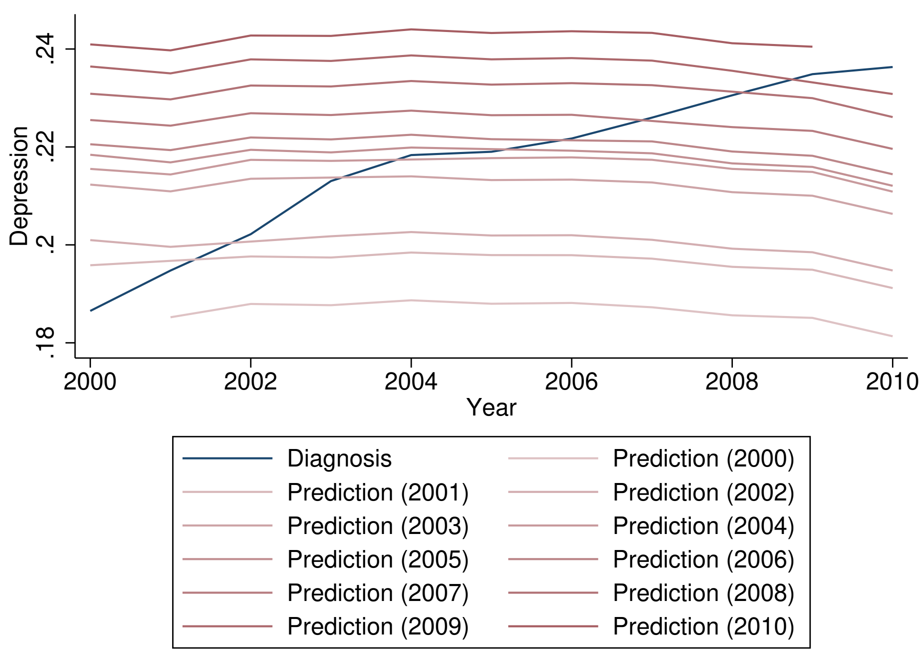
$$\begin{aligned} \tilde{\delta} &= \frac{Cov\left(D_0, W - \frac{\rho^2_{X_1,W}}{\alpha\delta} X_1\right)}{Var(W)} \\ &= \delta - \frac{\rho^2_{X_1,W}}{\delta}, \end{aligned}$$

where we also used the fact that δ is the regression coefficient from regressing D_0 and t, since t is independent of X_0.

Figure 4: Trends in Depression Diagnoses and Depression Index Based on Predictions from Various Years

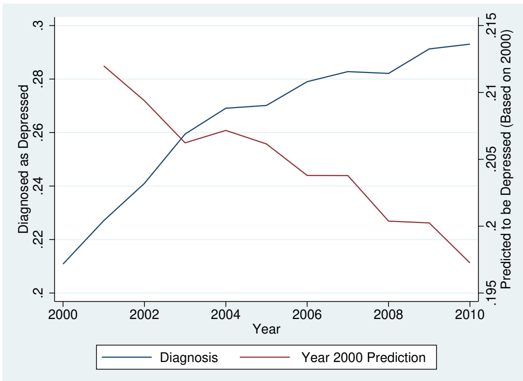


(a) Predictions Based on All Non-Mood Variables

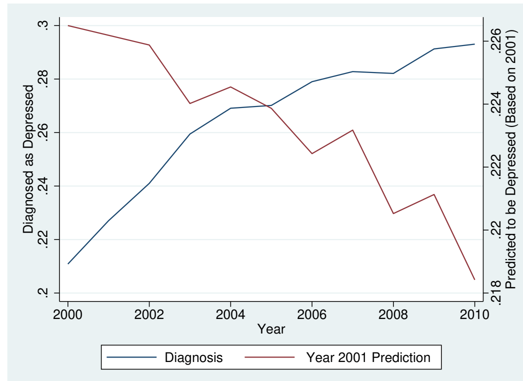


(b) Predictions Based on All Mood Variables

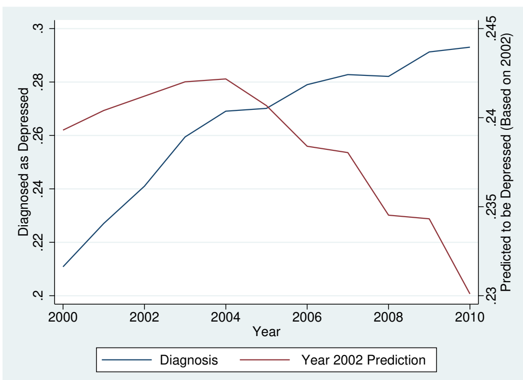
Figure 5: Trends in Depression Diagnoses and Depression Index Based on Predictions from Various Years (Predictions Based on All Variables)



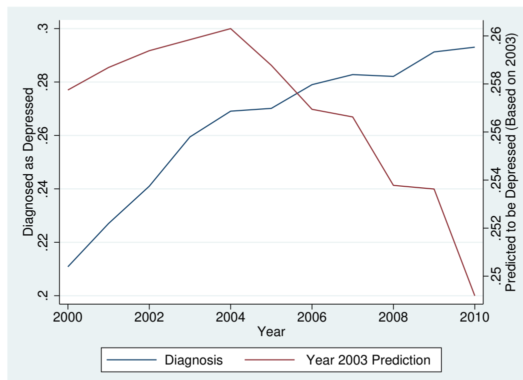
(a) 2000



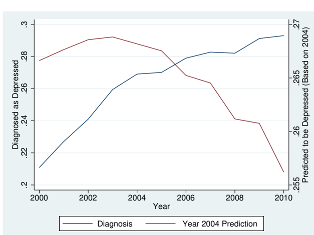
(b) 2001



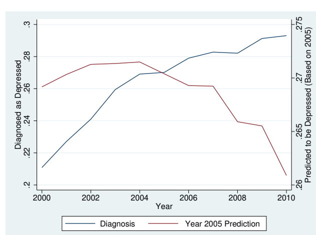
(c) 2002



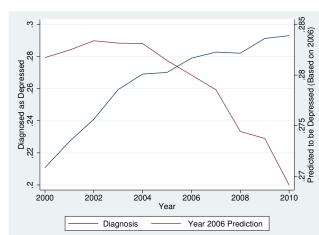
(d) 2003



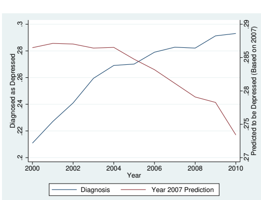
(e) 2004



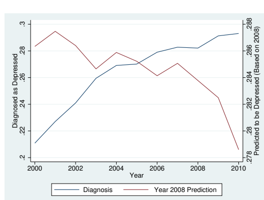
(f) 2005



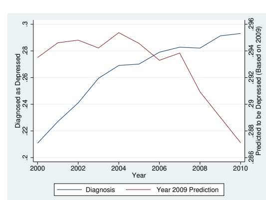
(g) 2006



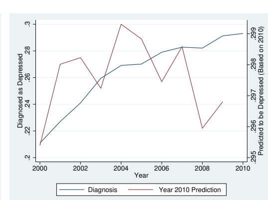
(h) 2007



(i) 2008

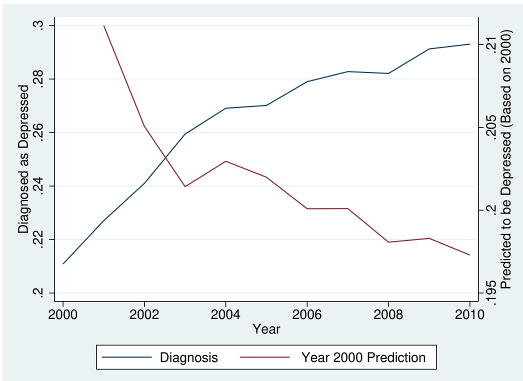


(j) 2009

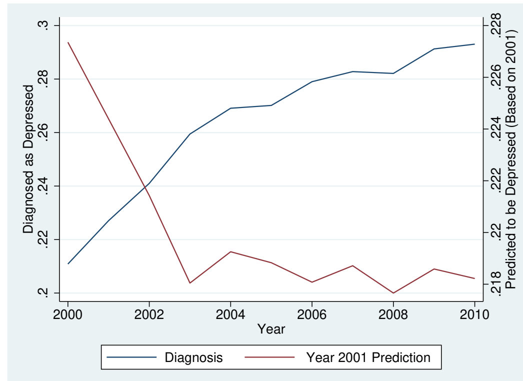


(k) 2010

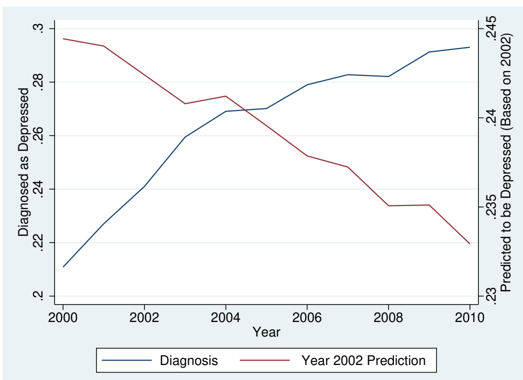
Figure 6: Trends in Depression Diagnoses and Depression Index Based on Predictions from Various Years (Predictions Based All Non-Mood Variables)



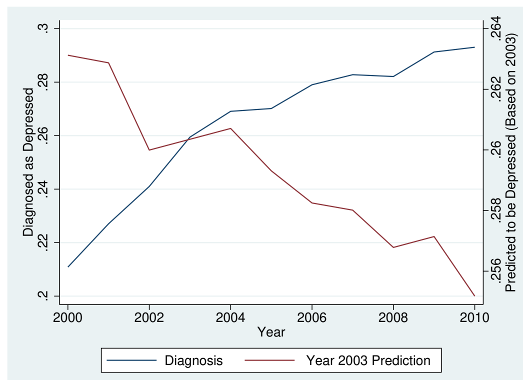
(a) 2000



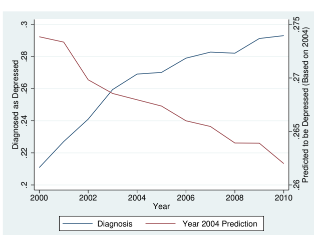
(b) 2001



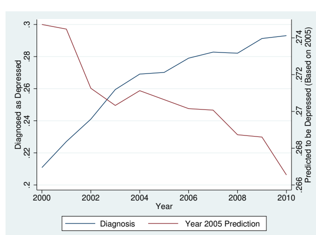
(c) 2002



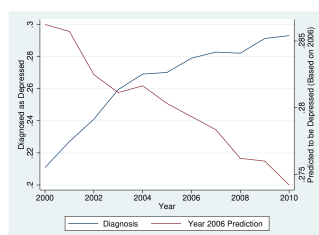
(d) 2003



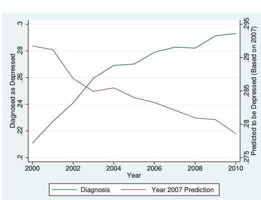
(e) 2004



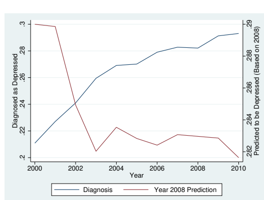
(f) 2005



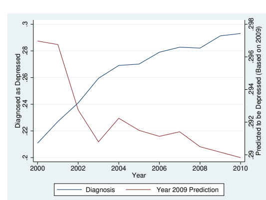
(g) 2006



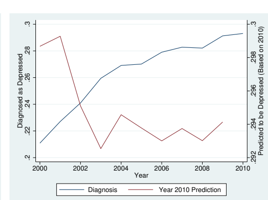
(h) 2007



(i) 2008

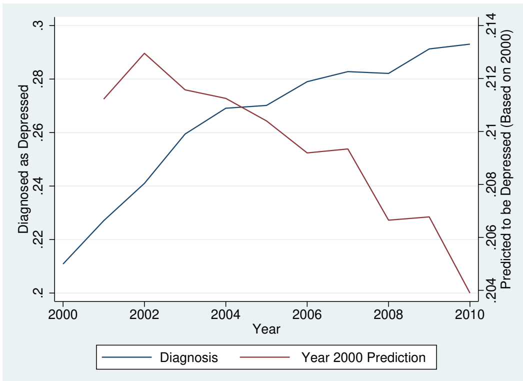


(j) 2009

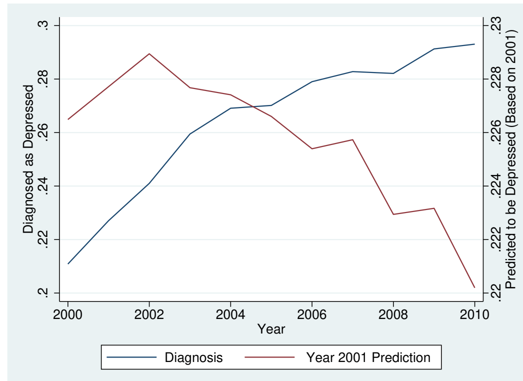


(k) 2010

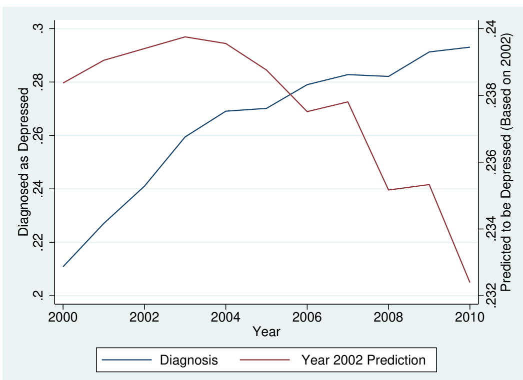
Figure 7: Trends in Depression Diagnoses and Depression Index Based on Predictions from Various Years (Predictions Based on Mood Variables)



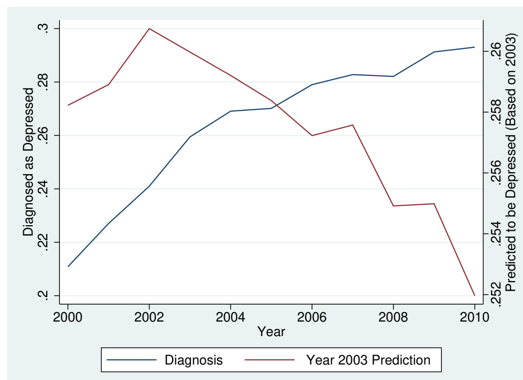
(a) 2000



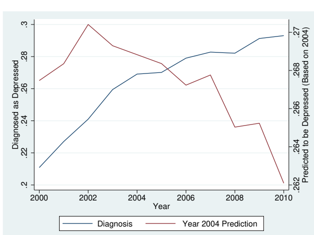
(b) 2001



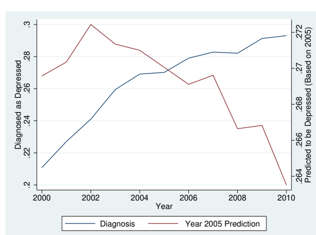
(c) 2002



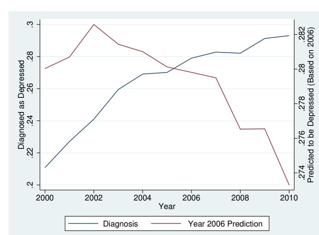
(d) 2003



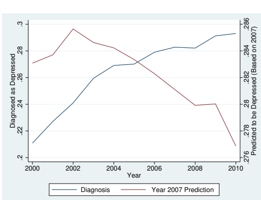
(e) 2004



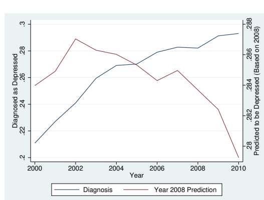
(f) 2005



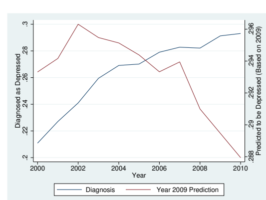
(g) 2006



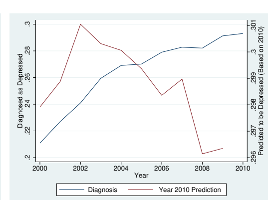
(h) 2007



(i) 2008



(j) 2009



(k) 2010