

Predicting Depression Using Health Care Data

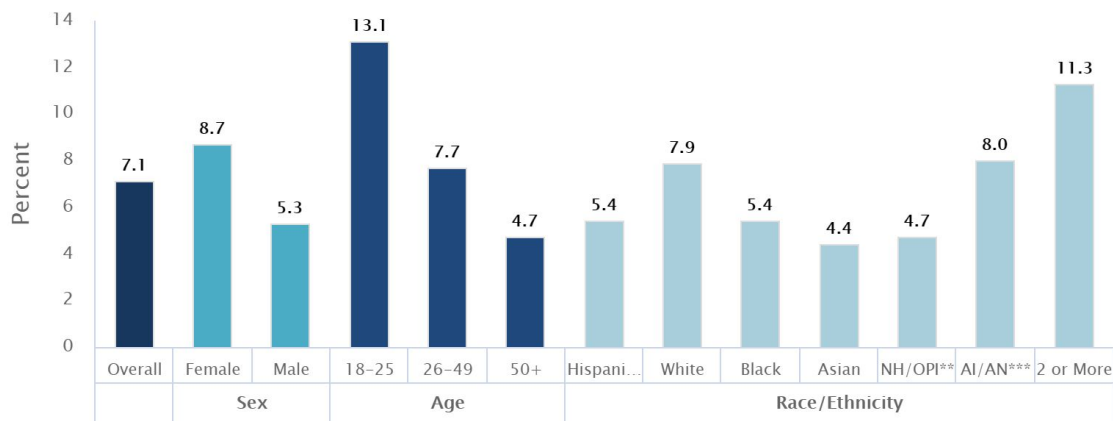
By Vivienne DiFrancesco

Depression

- 264 million people globally have depression - WHO
- 7.1% of US adults had a major depressive episode in 2017 with young adults being the most affected - NIMH

Past Year Prevalence of Major Depressive Episode Among U.S. Adults (2017)

Data Courtesy of SAMHSA



The Problem

- Physicians have become the front line for handling mental health disorders.
- The majority of physician visits are driven by mental disorders and few people follow up with a mental health professional.
- 90% of those who commit suicide have a mental health disorder and 40% had visited their doctor in the last month

The Problem

- Study published in JAMA - Patients who receive more holistic care with doctors for mental health reduce healthcare costs and improve patient outcomes.
- This can be unfeasible for many reasons
 - Buy-in, logistics, training
- Machine learning could fill the gap to predict patients who are depressed.

Using machine learning to identify individuals with depression could connect patients with the help they need more quickly and easily while reducing healthcare costs and burden on physicians.

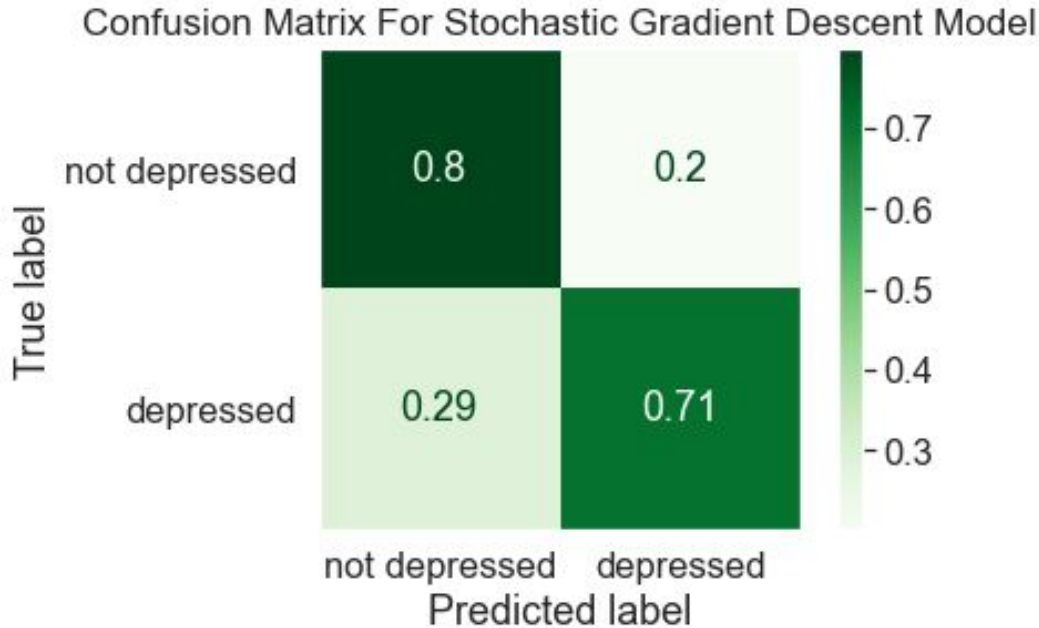
The Data

- CDC NHANES data
- 2005 to 2018 comprising 36259 entries total
- Tried to use data that was consistent across years and could reasonably be found in a patient's medical file
- Created labels of who was depressed and not depressed based on the “depression screener” in the data

Methods

- Classification of depressed or not depressed
- Roughly followed OSEMiN method
 - Obtain, scrub, explore, model, interpret
- Modeling was done from simple going to complex
- Wanted to use as few features as necessary
- Imbalanced data means accuracy was not a good metric
- Recall weighted more heavily to reduce false negatives

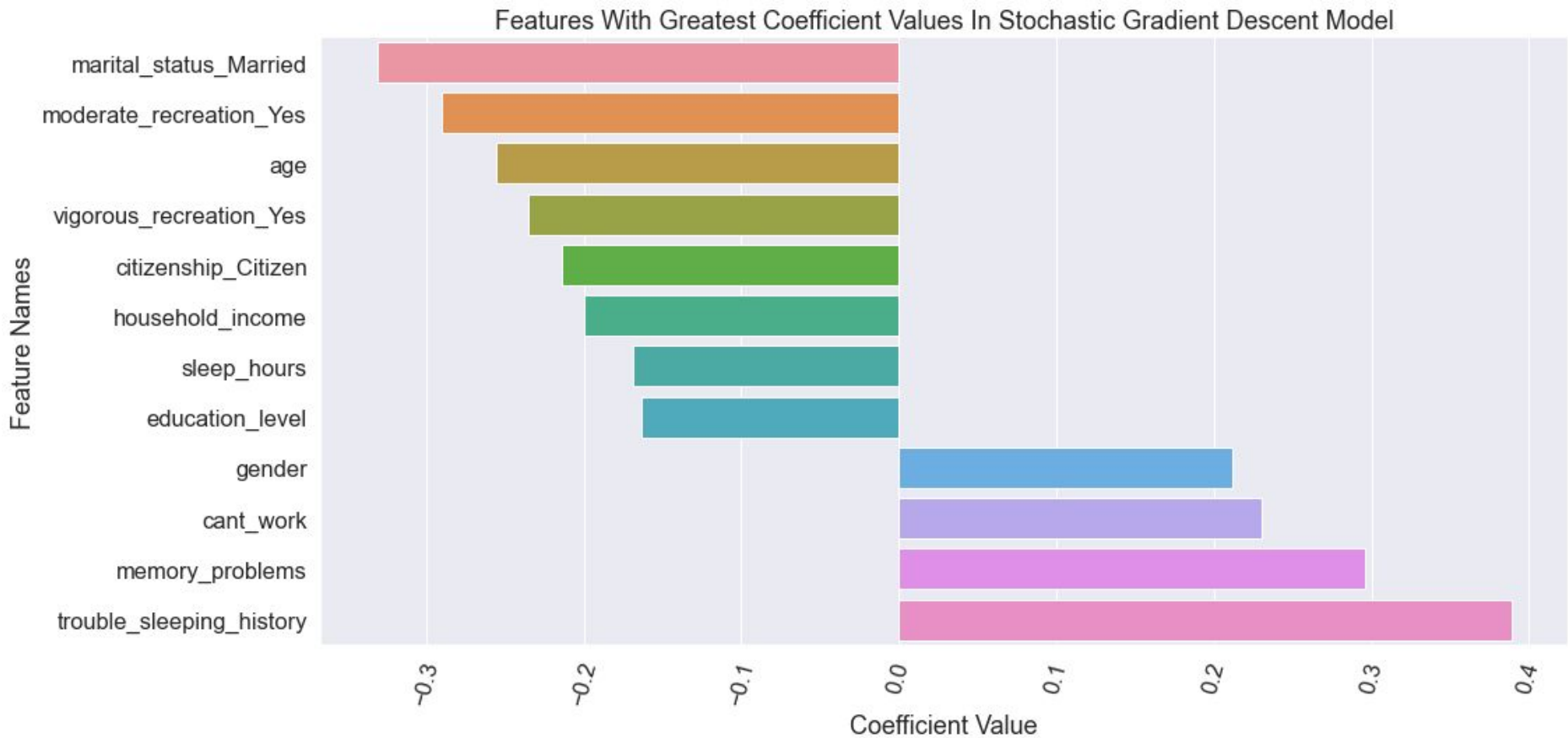
Best Model Results



Logistic regression is a model that calculates probabilities of entries being in one or the other class then uses a 50% threshold to make the prediction.

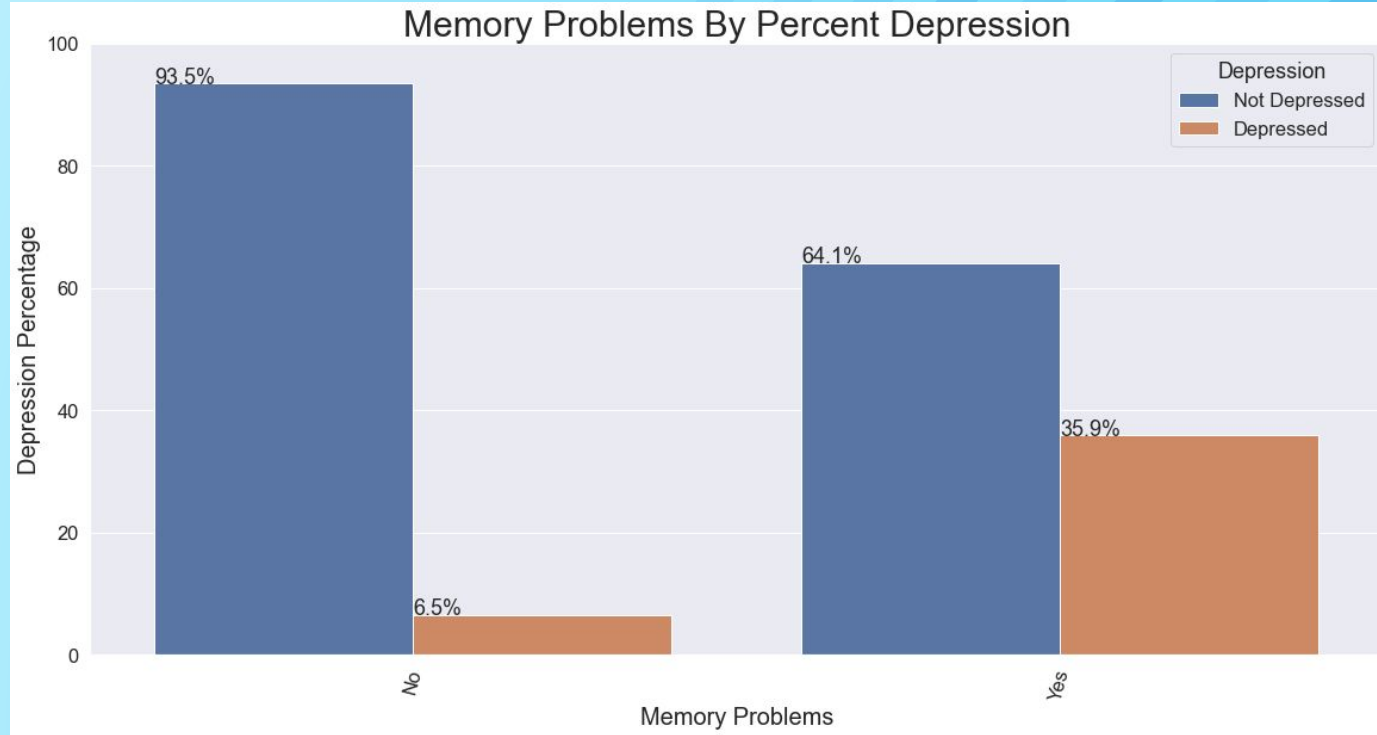
Gradient descent was used to maximize on a logistic regression.

Best Model Results



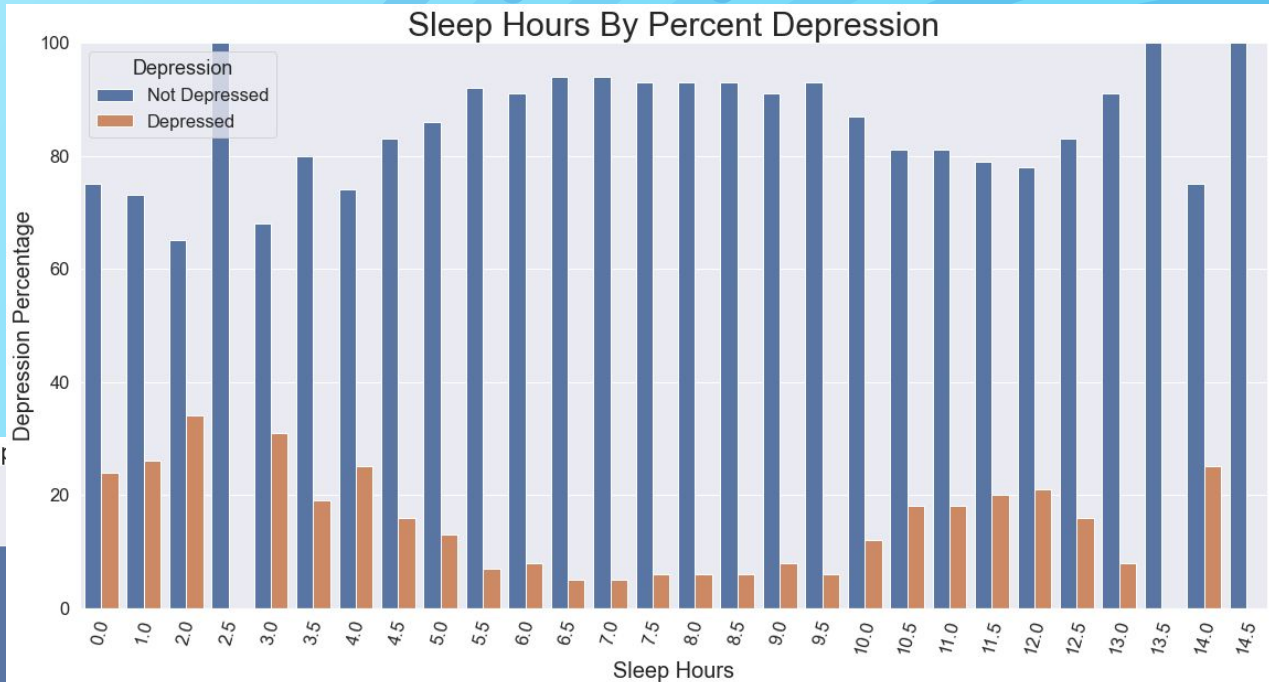
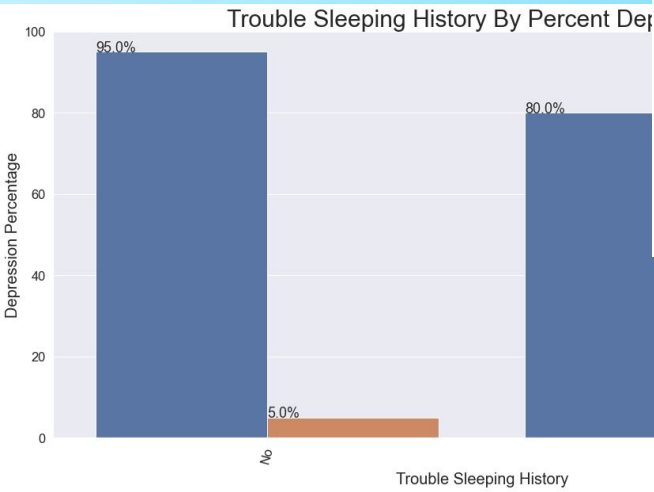
Recommendation 1:

Watch for patients with memory problems



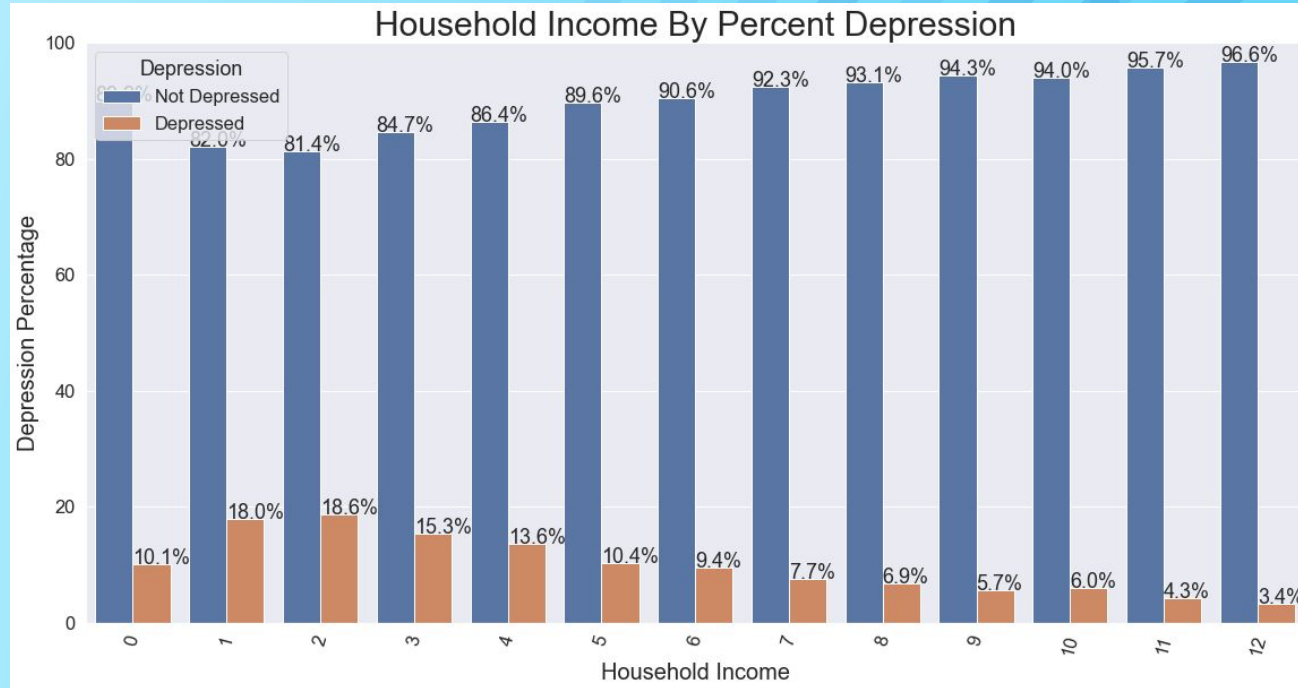
Recommendation 2:

Watch for patients who have trouble sleeping and sleep too much or too little



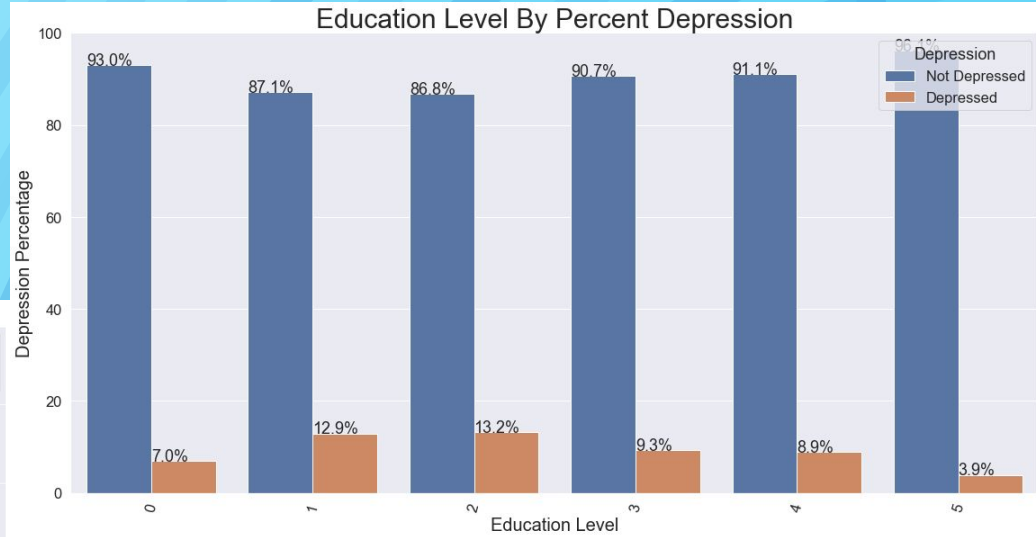
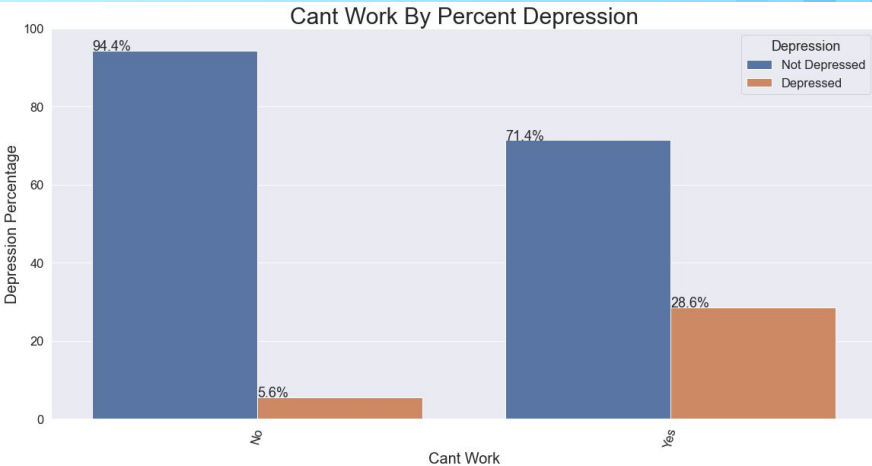
Recommendation 3:

Watch for patients who can't work, have low household income, and/or low education



Recommendation 3:

Watch for patients who can't work, have low household income, and/or low education



Recommendations

Recap

1. Watch for patients with memory problems
2. Watch for patients with sleep troubles
3. Watch for patients who can't work, have low income, or low education

Future Work

- Try different models - perhaps neural networks
- Add more data
- Tuning and testing of parameters



Thank you for your time

Any questions?

Appendix

[Click here](#) to be directed to a dashboard to explore the data in this project.

PHQ-9

Take the PHQ-9 online
here:
<https://www.mdcalc.com/phq-9-patient-health-questionnaire-9>

PATIENT HEALTH QUESTIONNAIRE - 9								
Over the <u>last 2 weeks</u> , how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day				
1. Little interest or pleasure in doing things	0	1	2	3				
2. Feeling down, depressed, or hopeless	0	1	2	3				
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3				
4. Feeling tired or having little energy	0	1	2	3				
5. Poor appetite or overeating	0	1	2	3				
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3				
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3				
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3				
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3				
<p style="text-align: right;"><i>FOR OFFICE CODING</i></p> <p style="text-align: right;"> <u>0</u> + <u> </u> + <u> </u> + <u> </u> =Total Score: <u> </u> </p>								
<p>If you checked off <u>any</u> problems, how <u>difficult</u> have these problems made it for you to do your work, take care of things at home, or get along with other people?</p> <table style="width: 100%; text-align: center;"> <tr> <td>Not difficult at all <input type="checkbox"/></td> <td>Somewhat difficult <input type="checkbox"/></td> <td>Very difficult <input type="checkbox"/></td> <td>Extremely difficult <input type="checkbox"/></td> </tr> </table>					Not difficult at all <input type="checkbox"/>	Somewhat difficult <input type="checkbox"/>	Very difficult <input type="checkbox"/>	Extremely difficult <input type="checkbox"/>
Not difficult at all <input type="checkbox"/>	Somewhat difficult <input type="checkbox"/>	Very difficult <input type="checkbox"/>	Extremely difficult <input type="checkbox"/>					
Copyright © 2010 Pfizer, Inc. All rights reserved.								