**Exploring the relationship between perceived mental health and perceived physical health: a simple linear regression model based on BRFSS self-report data**

Bill Yuanchen Liu

Brown School, Washington University in St. Louis

Applied Linear Modelling

Professor Joe Steensma

Sept 27, 2018

**Introduction**

The current paradigm in our understanding of health is most notably defined by the World Health Organization in 1948, as “*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*.”1 This was described as ambitious at the time, hailing from a previous understanding of health as the absence of primarily physical illness or disease.2 As the field of mental health and social well-being grew in more recent years, it is increasingly evident that there are important determining factors outside of physical testing that contributes to disease. Furthermore, these factors are often multi-layered, interrelated, and influenced by each other. It is therefore important to see this “physical, mental, and social well-being” as a multifactorial, complex, self-influencing model.2

The interaction between the “body” and the “mind” is difficult to study mainly due to its broad concept and underlying mechanisms. However, there are large population-based studies that can control for confounding variables and establish causation. A population-based prospective cohort study conducted by Patten found that physical illness is a risk factor for later development of new depressive episodes. This is echoed by similar studies in other populations.3 Analyses of types of illnesses within the classification of physical illnesses found similar and statistically significant risks.3,4 Goldberg highlighted three pathways in which physical disease can lead to depression.4 There is a proportional relationship between different pains experienced by a patient and the prevalence of depression.5 Also, the perceived risk of disability of physical illnesses places a mental burden on patients.6 This effect is observed in more elderly patients. Finally, physical changes like a change in the allostatic load may directly affect an individual’s ability to experience and cope with pain and depression.7

Similarly, studies have shown that having a poor mental health can also lead to an increased risk in having illnesses that is traditionally viewed as having a more physical cause. Using the example of depression, it can precede chronic physical illnesses like coronary heart disease, colorectal cancer, or multiple sclerosis.8 Schizophrenia can also be associated with morbidity risk in heart disease and respiratory disease.4,9

The common moderating factor that affects both physical and mental health outcomes is lifestyle. If an ecological model were to be constructed, lifestyle can influence the state of both physical and mental health. Example of lifestyle factors include exercise, which improves physical and mental well-bring; diet, and other factors like smoking.9,10 Therefore, it is important to realize that, with a shift in the understanding of health, the factors that affect an individual’s quality of life is also redefined according to new research in mental health.

An important concept in health-related quality of life (QOL) assessment is perceived health. Revuelta *et al.* suggested that perceived mental health can independently predict mortality and morbidity in end-stage renal disease patients.11 The Center for Disease Control (CDC) identified “Healthy Days Measures” to measure self-rated general health in order to assess limitations and disparities in public perception.12 These are validated across multiple studies in a range of populations.12,13 Notably, these measures are useful as a global indicator of disability in the general population.12

Considering the interaction of mental health and physical health mentioned above, it is important to investigate the correlation between perceived mental health and perceived physical health due to a lack of research in this direct area. Having knowledge in this area will pave way for comorbidity studies between physical diseases and mental diseases, helping to build a more holistic healthcare. Given the acknowledgement of association between the two factors, the primary goal of this study is to determine the extent of correlation between the number of days a person feels unwell in mental health and physical health. What is the extent of correlation explained by a Simple Linear Regression (SLR) model used to predict the number of days unwell from mental health and from physical health? The null hypothesis (H0) is there is no linear model that can predict bad mental health days based on bad physical health days. The alternative hypothesis (Ha) is that there is a linear model that can predict bad mental health days based on bad physical health days.

**Methods**

Secondary data was collected from the 2017 Behavioral Risk Factor Surveillance System (BRFSS) Questionnaire accessed from https://www.cdc.gov/brfss/questionnaires/pdf-ques/2017\_BRFSS\_Pub\_Ques\_508\_tagged.pdf. Specific sampling methods, data processing, and data weighting procedures are described in its guidebook.14,15 Data was taken from the core questionnaire administered to all 53 states and territories. Data collection methods consist of landline and cellphone interviews using approved procedures highlighted in the overview. Demographic adjustments and geographic stratification were conducted as described.14 To clean the data for analysis, none, not sure, refused, and missing values for both mental health days and physical health days were eliminated. Only those with a value between 1 and 30 days were used according to the study design.

IBM SPSS version 25 was used to conduct all statistical analyses. As part of generating primary descriptive statistics, the mean, 95% confidence interval (CI), standard deviation, minimum, and maximum were generated using the explore function. To test normality, Kolmogorov-Smirnov test, Shapiro-Wilk test, Q-Q plot, and box plot of the variables were generated. A scatter plot with Pearson’s r and best fit line were generated.

A primary SLR was done using mental health days as the independent variable and physical health days as the dependent variable. Estimates, CI, Durbin-Watson, model fit, case-wise diagnostics, and descriptives were generated. Two plots, ZPRED over ZRESID, SRESID over ZPRED, were calculated, as well as standardized residuals, Cook’s distance, standardized DfBetas, and standardized DfFit.

The model was assessed using diagnostics to determine the assumptions of SLR. Influential values were assessed using Cook’s d, ZDFfit, ZDFbeta, and ZResid. The model will be redone if changes were made to the dataset.

**Results**

450642 data points were downloaded from the 2017 BRFSS Questionnaire. Filtering for mental health days (MENTHLTH) and physical health days (PHYSHLTH) between 1 and 30, 79668 cases were selected for analysis (Table 1). For MENTHLTH, the mean is 12.85 days (95%CI [12.77, 12.92]), variance is 118.566, standard deviation is 10.889, minimum is 1, and maximum is 30. For PHYSHLTH, the mean is 12.80 days (95%CI [12.72, 12.88]), variance is 127.598, standard deviation is 11.296, minimum is 1, and maximum is 30 (Table 2).

For tests of normality, the KS-test statistic for MENTHLTH is 0.184 (Sig. 0.001), and PHYSHLTH 0.198 (Sig. 0.001) (Table 3). A visual inspection of the histogram, Q-Q plot, and box plot shows no normality for both variables (Appendix B). A scatter plot of PHYSHLTH over MENTHLTH shows a best fit line with the equation y = 7.05 + 0.45x. The R2 linear is 0.186.

A significant SLR model was constructed to predict PHYSHLTH using MENTHLTH [F(1, 79666) = 18231.918, p<0.001] (Table 4). The linear model explains the data significantly better than the mean. Assuming we are looking for a p-value of 0.05, the critical value for an F statistic with 1 and 79666 degrees of freedom is 5.024. The F statistic is larger than the critical F; the model is significantly better than the mean at explaining PHYSHLTH based on MENTHLTH. In this case the model explained 18.6% of the variability of PHYSHLTH. The slope (b1= 0.448) is significantly different from zero (t = 135.026, p<0.001). The true slope of the model falls between 0.441 and 0.454 (95%CI [0.441, 0.454]). The variable, MENTHLTH, is a significant contributor to the model. For every additional day of mental health days unwell reported, the days of physical health days unwell reported increases by 0.448 (Table 4). This relationship is positively linear.

The regression equation is Y = 7.048 + 0.448xMENTHLTH.

Checking for assumptions, independence of observations is assumed through study design. Independence of errors is shown through Durbin-Watson (1.946). Linearity is checked in scatter plot, no visible curve is observed (Graph 1). Homoscedasticity assumption is passed through Steensma-Squint test with no visible heteroscedasticity observed (Graph 2). The assumption for normally distributed errors were violated, showing a general bimodal distribution (Graph 3).

Checking for influential values using Cook’s d, Standardized Residuals, Standardized DFBetas, Standardized DFFits, and Leverage Value returned no outstanding observations. No observation was found to have undue influence on the model or the sample (Table 5). Since no observations was removed, the regression was not redone.

**Discussion**

This analysis was successful in creating a statistically significant linear model of the correlation from MENTHLTH and PHYSHLTH data collected from the BRFSS survey. Several data manipulation methods were used to filter research-specific data. Through this process, many data points were eliminated, and this may pose a problem for the generalizability of the resulting model. Considering our research question, which compares the days unwell in MENTHLTH and PHYSHLTH, it makes sense to eliminate answers that are missing, don’t know, or refused. Those with zero number of days unwell was also eliminated, which consisted 60% (276509 entries) of all values.15 These were eliminated to ensure that the correlation was focused on those people that feels unwell, but in the process we reduced our ability to inference to those that felt fine for the duration of the survey question.

Neither of the variables were distributed normally. This is confirmed through the KS-test as well as visual inspection of the histogram, Q-Q plot, and box plot. It is not distributed normally. This suggests some concern regarding the design of the questions, since the BRFSS Questionnaire surveys a large, weighted, national sample.

A significant SLR model was constructed to predict PHYSHLTH using MENTHLTH. For every additional day an individual feels unwell in mental health, he/she may feel 0.448 days unwell physically. This not only agrees with literature in establishing a correlation between perceived mental health and perceived physical health, we can predict using a SLR based on the extent of correlation found. MENTHLTH explains 18.6% of the variability in PHYSHLTH.

To ensure the validity of our model, assumptions were checked using diagnostic tests. Durbin-Watson test suggested that our independence of errors is normal. No visible curve relation is observed in scatter plot (Graph 1). The assumption of homoscedasticity is also done visually, with no visible heteroscedasticity trend observed (Graph 2). However, plotting the regression standard residual, we see a general bimodal distribution (Graph 3). This suggests that there is some relationship unaccounted for that bisects our samples into two different groups. This indicates a great need for future research to see why there is a difference in low number of days unwell and high number of days unwell. For the sake of comparison, individuals who answered zero days of unwell for either MENTHLTH or PHYSHLTH can be added. This will form a more complete picture of those who were surveyed.

After calculating for influential values, no outstanding observations were found, whether it is sorted ascending or descending. This suggests that the sample’s distribution is not significantly affected by any one outliers.

Another consideration for this study is the large population size surveyed by BRFSS. Through protocol outlined online, the data is geographically and demographically adjusted. We can assume that it gives a fair representation of the United States population, but to make further investigations this must be confirmed. With large datasets, we can also do sub-group analyses without the fear of reducing group representation. This creates exciting opportunities for further research, especially looking at trends across geographic regions or demographic cohorts.

The nature of our variable is numerical, quantified in days. However, it is only a falsely continuously variable, as it is limited between 1 and 30. Not only was zero not included, the large sample size ensured that every increment of 1 day had a large number of data. This can be seen in the scatter plot in Graph 1. This contributed to the statistical significance of our model, but also poses limitations. In future research, the research question may be modified to include a larger range of days unwell. However, in doing so, it is important to consider patient recall bias and other problems.

There is inherent limitation on our study based on the definition of “healthy days”. The definition limits this study to be a cross-sectional representation of the population. The survey does not capture any information about the context, existence, cause, or duration of days unwell. Each of these factors may be of interest to the researcher. However, to analyze these, more specific study design needs to be tested.

Another limitation identified with this study is the reporting bias of self-report surveys. It may be easier for individuals to remember 1 or 2 sick days, or if they are sick for all 30 days, but it is harder to distinguish those numbers in the middle. For example, the difference between 17 to 18 may not be as well recalled as 1 to 2, or 29 to 30. This recall bias may contribute to the bimodal distribution of residuals, and calls for adjustments in survey design. Inherent with all self-report studies, it is important to account for recall bias.

**Conclusion**

In short, this study found a statistically significant linear model to predict perceived days unwell for physical health from mental health. However, it is not generalizable to other populations because it does not meet all assumptions. The model can still help us understand the interaction between perceived mental health and perceived physical health, and provides valuable foundation for further research in this area. The extent to which these two variables correlate to each other calls for considerations for physicians when making decisions on physical health interventions for patients. It shows the importance of having a holistic view in patient health recovery that emphasize the interaction between physical and mental health of the individual.

**Works Cited**

1. World Health Organization. Constitution of WHO: principles. WHO. http://www.who.int/about/mission/en/. Accessed September 28, 2018.

2. Huber M, Knottnerus JA, Green L, et al. How should we define health? *BMJ*. 2011;343:d4163. doi:10.1136/bmj.d4163

3. Patten SB. Long-term medical conditions and major depression in a Canadian population study at waves 1 and 2. *J Affect Disord*. 2001;63(1-3):35-41.

4. Goldberg D. The detection and treatment of depression in the physically ill. *World Psychiatry*. 2010;9(1):16-20. doi:10.1002/j.2051-5545.2010.tb00256.x

5. Dworkin SF, Korff MV, LeResche L. Multiple Pains and Psychiatric Disturbance: An Epidemiologic Investigation. *Arch Gen Psychiatry*. 1990;47(3):239-244. doi:10.1001/archpsyc.1990.01810150039007

6. Prince MJ, Harwood RH, Blizard RA, Thomas A, Mann AH. Impairment, disability and handicap as risk factors for depression in old age. The Gospel Oak Project .5. *Psychol Med*. 1997;27(2):311-321. doi:10.1017/S0033291796004473

7. Ormel J, Kempen GI, Penninx BW, Brilman EI, Beekman AT, van Sonderen E. Chronic medical conditions and mental health in older people: disability and psychosocial resources mediate specific mental health effects. *Psychol Med*. 1997;27(5):1065-1077.

8. Collingwood J. The Relationship Between Mental & Physical Health. Psych Central. https://psychcentral.com/lib/the-relationship-between-mental-and-physical-health/. Published May 17, 2016. Accessed September 28, 2018.

9. Mental Health Foundation. Physical health and mental health. Mental Health Foundation. https://www.mentalhealth.org.uk/a-to-z/p/physical-health-and-mental-health. Published August 7, 2015. Accessed September 28, 2018.

10. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Curr Opin Psychiatry*. 2005;18(2):189.

11. López Revuelta K, López G, J F, de Álvaro Moreno F, Alonso J. Perceived mental health at the start of dialysis as a predictor of morbidity and mortality in patients with end-stage renal disease (CALVIDIA Study). *Nephrol Dial Transplant*. 2004;19(9):2347-2353. doi:10.1093/ndt/gfh392

12. Moriarty DG, Zack MM, Kobau R. The Centers for Disease Control and Prevention’s Healthy Days Measures – Population tracking of perceived physical and mental health over time. *Health Qual Life Outcomes*. 2003;1(1):37. doi:10.1186/1477-7525-1-37

13. Ounpuu S, Krueger P, Vermeulen M, Chambers L. Using the U.S. behavior risk factor surveillance system’s health related quality of life survey tool in a Canadian city. *Can J Public Health Ott*. 2000;91(1):67-72.

14. Behavioral Risk Factor Surveillance System. OVERVIEW: BRFSS 2017. July 2018. https://www.cdc.gov/brfss/annual\_data/2017/pdf/overview-2017-508.pdf. Accessed September 27, 2018.

15. Behavioral Risk Factor Surveillance System. LLCP 2017 Codebook Report Overall version data weighted with \_LLCPWT Behavioral Risk Factor Surveillance System. September 2018. https://www.cdc.gov/brfss/annual\_data/2017/pdf/codebook17\_llcp-508.pdf. Accessed September 27, 2018.

**Appendix A**

**Table 1 Case Processing Summary**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case Processing Summary** | | | | | | |
|  | Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| NUMBER OF DAYS MENTAL HEALTH NOT GOOD | 79668 | 100.0% | 0 | 0.0% | 79668 | 100.0% |
| NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD | 79668 | 100.0% | 0 | 0.0% | 79668 | 100.0% |

**Table 2 Descriptives**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptives** | | | | |
|  | | | Statistic | Std. Error |
| NUMBER OF DAYS MENTAL HEALTH NOT GOOD | Mean | | 12.85 | .039 |
| 95% Confidence Interval for Mean | Lower Bound | 12.77 |  |
| Upper Bound | 12.92 |  |
| 5% Trimmed Mean | | 12.55 |  |
| Median | | 10.00 |  |
| Variance | | 118.566 |  |
| Std. Deviation | | 10.889 |  |
| Minimum | | 1 |  |
| Maximum | | 30 |  |
| Range | | 29 |  |
| Interquartile Range | | 19 |  |
| Skewness | | .571 | .009 |
| Kurtosis | | -1.246 | .017 |
| NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD | Mean | | 12.80 | .040 |
| 95% Confidence Interval for Mean | Lower Bound | 12.72 |  |
| Upper Bound | 12.88 |  |
| 5% Trimmed Mean | | 12.50 |  |
| Median | | 7.00 |  |
| Variance | | 127.598 |  |
| Std. Deviation | | 11.296 |  |
| Minimum | | 1 |  |
| Maximum | | 30 |  |
| Range | | 29 |  |
| Interquartile Range | | 22 |  |
| Skewness | | .553 | .009 |
| Kurtosis | | -1.342 | .017 |

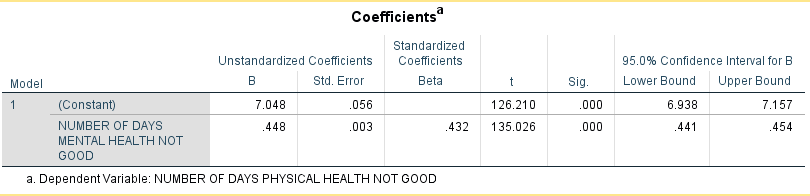
**Table 3 Tests of Normality**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tests of Normality** | | | |
|  | Kolmogorov-Smirnova | | |
| Statistic | df | Sig. |
| NUMBER OF DAYS MENTAL HEALTH NOT GOOD | .184 | 79668 | .000 |
| NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD | .198 | 79668 | .000 |
| a. Lilliefors Significance Correction | | | |

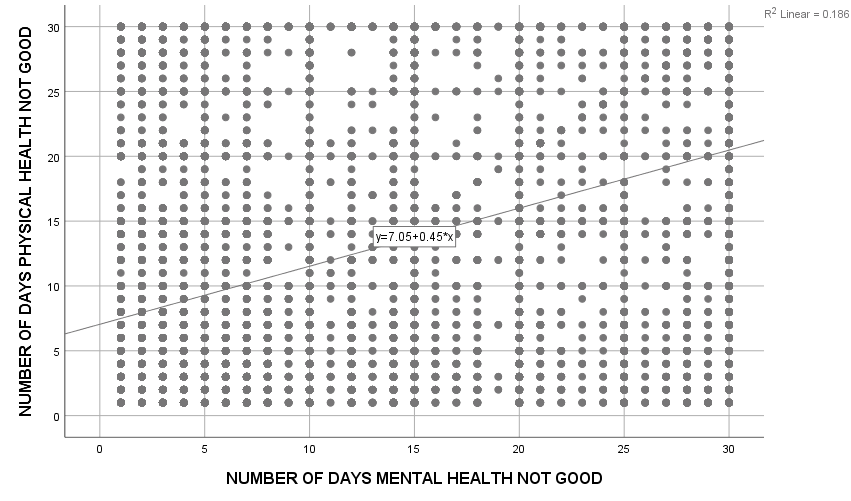
**Table 4 ANOVA and Model Summary for SLR**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| 1 | .432a | .186 | .186 | 10.190 | 1.946 |
| a. Predictors: (Constant), NUMBER OF DAYS MENTAL HEALTH NOT GOOD | | | | | |
| b. Dependent Variable: NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD | | | | | |

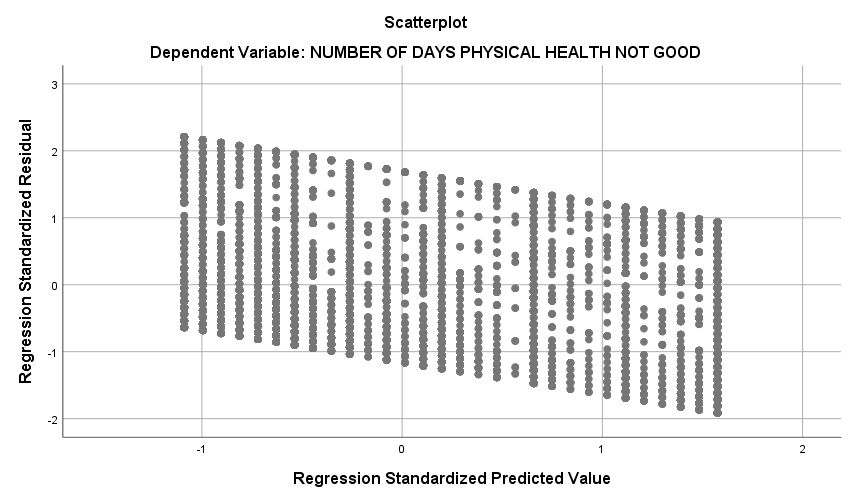
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 1893130.859 | 1 | 1893130.859 | 18231.918 | .000b |
| Residual | 8272205.057 | 79666 | 103.836 |  |  |
| Total | 10165335.916 | 79667 |  |  |  |
| a. Dependent Variable: NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD | | | | | | |
| b. Predictors: (Constant), NUMBER OF DAYS MENTAL HEALTH NOT GOOD | | | | | | |



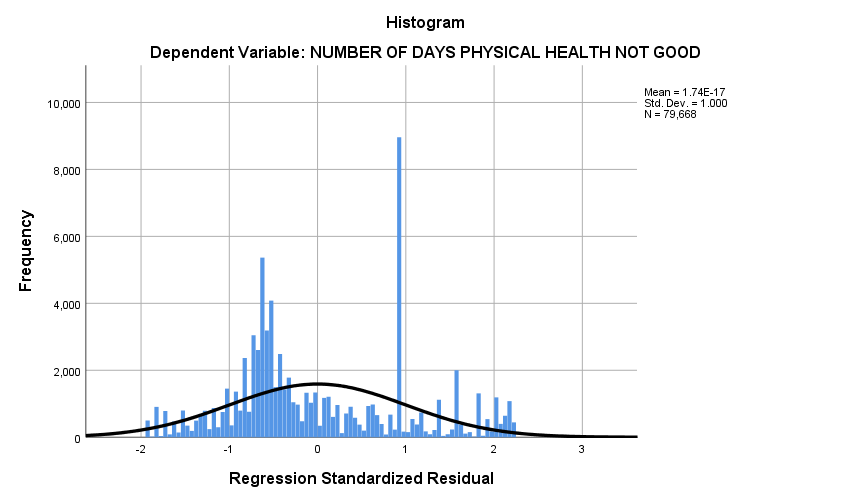
**Graph 1 Scatter plot of MENTHLTH over PHYSHLTH**



**Graph 2 Scatter plot of Residuals ZRESID over predicted ZPRED values**



**Graph 3 Histogram of Regression Residual Frequency**



**Table 5 Table of Top 10 Influential Values**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Top 10 values | Cook’s d | Standardized Residuals | Standardized DFBetas0 | Standardized DFBetas1 | Standardized DFFits |
| 1. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 2. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 3. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 4. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 5. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 6. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 7. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 8. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 9. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |
| 10. | 0.00008 | 2.20851 | 0.01155 | 0.00521 | 0.01156 |