FOUNDATION OF ANALYTICS
Analysis of Somatization Dataset

Troy Dixon
The University of Tulsa
Email: tdd5232utulsa.edu

Oisin Spillane
The University of Tulsa
Email: oms5064@utulsa.edu

Silas Hughes
The University of Tulsa
Email: sch7390@utulsa.edu

ABSTRACT

This study investigates the somatization symptom dimension using data derived from the Symptom Checklist-90 (SCL-90), and the dataset given, aiming to explore its predictive patterns and clinical implications. The primary objectives were to identify key predictors of somatization and evaluate the relationships between somatization, anxiety, and depression across varying age groups, including younger patients in particular. Utilizing Excel and R for data analysis, we produced histograms, graphs, and summaries of average and co-occurring symptoms. Additionally, logistic regression and standard deviation were employed, with k-fold crossvalidation ensuring robust model validation. Evaluation metrics, including accuracy and sensitivity, provided a comprehensive assessment of model performance. Our findings revealed that certain SCL-90 items significantly contribute to predicting somatization severity, along with depression and anxiety playing a role. This research highlights the value of data-driven approaches, supported by visualization and statistical methods, in advancing the understanding and management of psychological disorders

KEYWORDS: Time, Critical Thought, Defensiveness, Pain, Sensitivity, Somatization, Ideation

INTRODUCTION

SCL-90 is the questionnaire presented to the outpatient respondents, but the intent of measurement is not described by the data set if there was any intent of the survey communicated to the respondents. It is likely that the survey was an optional survey taken after psychiatric in-patient treatment or for similar mental conditions. The history of the SCL-90 questionnaire is noted in the literature review below. Somatization as the physical manifestation of strong mental conditions often involves holding tension in the body ordinarily compared to mental constructs.

We began by categorizing survey responses into two primary groups: physical pain (e.g., dizziness and headaches) and negative ideation (e.g., guilt, self-blame). Each survey question was assigned to one of these categories based on the type of response it generated. A key focus of our study was the question C_4, "Are you bothered by faintness or dizziness?".

We hypothesized that dizziness might serve as a strong indicator of somatization, linking physical symptoms to underlying mental stress. Our analysis also considered response biases, informed by a comprehensive review of relevant literature. Specifically, we focused on headrelated pain—headaches and dizziness—and their potential connections to feelings of guilt and self-blame.

To visualize this, we created charts to analyze patterns of blame, pain, public fear, and compared these relationships across different age groups within the dizziness dataset. In addition, we tested the hypothesis that fear of surveillance technology, or concern about others knowing, might be a significant factor influencing responses. However, the T-test results were statistically insignificant, suggesting that other, unidentified variables may have been at play. This led us to explore the influence of mental constructions and speculative reasoning on responses and biases.

Furthermore, we investigated the relationship between psychological tension, unwanted thoughts, surveillance concerns, and contemplation. Lastly, we explored whether feelings of soreness or being blocked were related to the survey results.

To analyze these relationships, we used feature selection and model training techniques to prepare and process the data. We carefully selected features that were most relevant to our research questions.

The following is what we argue we aren't able to validate: We applied predictive models tailored to the unique challenges of this study, justifying their use and explaining how they contributed to addressing our primary research objectives.

LITERATURE REVIEW Literature Review of Topic 1

From this literature review we discuss order bias in surveys, particularly how the sequence of questions can influence the responses. Order bias occurs when earlier questions affect how respondents answer later ones. This is important because, in some cases, the survey may be more "painful" to take for certain respondents, depending on the order of the questions.

You are also referencing a study that explores question order effects and how the visual layout of the survey (such as item-by-item vs. grid formats) can affect respondents' answers. Your point is that understanding these effects is crucial when designing surveys, especially for mental health assessments, where both psychological and physical factors need to be considered.

In summary, you are emphasizing the importance of careful survey design to avoid biases—such as order bias—so that the data collected is both accurate and reliable. Order bias link – painful to take the survey for some but not for all? Question order effect refers to the phenomenon that previous questions may affect the cognitive response process and respondents' answers.

The findings highlight the importance of considering both psychological and physical factors in mental health assessments. The study also underscores the potential biases introduced by the order and context of survey questions, emphasizing the need for careful survey design to ensure accurate and reliable data.

Literature Review of Topic 2

Maladaptive Strategies: This likely refers to ineffective or harmful coping strategies that individuals use to deal with stress or psychological issues. These strategies can exacerbate mental health problems rather than help resolve them.

The SCL-90 Validity:

Convergent Validity: The nine dimensions of the SCL-90 (a psychological symptom inventory) were found to correlate with similar measures from other tests, indicating that the SCL-90 is valid in measuring what it intends to (mental health symptoms).”

Discriminant Validity: However, the SCL-90 also correlated with measures that were not related to what it was supposed to be measuring. This suggests that the test may not always effectively distinguish between different psychological constructs, indicating a lower level of discriminant validity.

Development of a Framework: The researchers plan to create a system or framework that categorizes different psychological conditions and mental health issues. This would help in organizing and understanding various aspects of mental health based on the results of tests like the SCL-90.

Study Reference: The study you referenced, titled "Discriminant and Convergent Validity of the SCL-90 in Psychiatric Inpatients," was conducted by W. David Dinning and Ronald G. Evans at the Central Louisiana State Hospital. This study explores the validity of the SCL-90 in psychiatric inpatients, specifically looking at how well the test measures mental health symptoms and whether it accurately distinguishes between different mental health conditions.

Literature Review of Topic 3

Galaxy Cluster Scale large scale variable clusters as analyzing a statistics model or guide to framework. (which includes the concept of bias in arrangement across unique people). Discriminate and convergent delivery built a scale to measure its concept as a model. Clustering and stratifying sample classifications. If you put the wrong topics together the sensitivity of bias increases the traditional shared meaning theory which produces ideation from discovery or potential ratio like how much.

This is the negative idea about the survey when being asked of sensitivity itself. The context of the survey is too vague to ascertain true validity and suffers from a discriminate validity. When testing a hypothesis about a specific property of galaxy clusters (e.g., the relationship between galaxy density and mass), both convergent and discriminant validity are critical for determining whether your findings hold true across multiple measures and are not being influenced by extraneous variables. Like the “on the stage” sensitivity.

Literature Review of Topic 4

These are different groups administered the same survey. In analyzing the psychological symptoms of students in an undergraduate program for elementary mathematics teaching, a model was built to assess the validity of the survey.

Our hypothesis raised ethical concerns regarding the potential impact of ideation and pain, particularly considering the age of the participants (particularly the minor age respondents).

The study also examined maladaptive strategies, focusing on how mental disorders might influence treatment approaches or medical interventions. A key limitation of the study is the difference between the SCL-90 and the revised SCL-90-R. The results indicated that females in Turkey were more sensitive to psychological symptoms than males, and that students in regular education programs were more sensitive than those in evening or night classes. The study also explored the Maladaptive Psychological Model of Intervention (MPMI), which investigates ethical considerations and mental constructs in survey responses. It pointed out a construction of measurement that would perhaps filter bias through a framework of correlation? As the SCL90 could influence how participants answer due to its structure. This is one major limitation of the MPMI model as applied to the SCL-90 questionnaire.

Literature Review Summary Tables

Citation	Purpose of Study	Data Analytics Methods	Key Findings	Relevance to Project	Difference from Other Studies
Garralda (1996).	Investigation into somatization in children.	Use of ICD and DSM systems, and the Children's Somatization Inventory.	Strong links with family health problems and somatization in children	Important for looking at the different age ranges, to identify trends.	Specific Data on children, which is relevant to our paper looking at varying age groups.

Citation	Purpose of Study	Data Analytics Methods	Key Findings	Relevance to Project	Difference from Other Studies
Goldberg & Bridges (1988).	Presentation of somatic symptoms – Psychiatric illness in a Primary care setting.	Survey of 15 family practices, with use of the DSM-III criteria.	About 19% of patients with existing diseases fulfilled the criteria for somatization .	Use of DSM-III, which is a survey similar in design to the SCL-90.	Co-occuring symptoms used for somatization prediction.

Citation	Purpose of Study	Data Analytics Methods	Key Findings	Relevance to Project	Difference from Other Studies

Asmundson & Katz (2009),	To examine how anxiety sensitivity contributes to heightened pain sensitivity	Psychometric assessments along with mediation and regression analysis to	Anxiety sensitivity as a strong predictor of somatic symptoms	Provides a psychological framework for understanding the cognitive – affective	
	and its link to somatization.	assess predictive factors of somatic complaints		mechanisms behind somatization and it's maintenance.	

Citation	**Purpose of Study**	**Data Analytics Methods**	**Key Findings**	**Relevance to Project**	**Difference from Other Studies**
Hardt (2000).	Use of the SCL-90 to analyse and treat patient group.	A total of 3683 patients completed SCL-90-R as part of evaluation.	Somatization patients linked with depression (49%), anxiety (43%), and chronic pain (40%).	Use of SCL90 to view trends in patients and link to our own dataset.	More concrete data on anxiety & depression as predictors.

Citation	**Purpose of Study**	**Data Analytics Methods**	**Key Findings**	**Relevance to Project**	**Difference from Other Studies**

Haug et al. (2004).	To examine the connection between anxiety, depression, and functional somatic symptoms in a large community sample.	Large population survey in Norway – 50,377 analysed.	The association of anxiety, depression, and functional somatic symptoms was equally strong in men and women (mean number of somatic symptoms men/women in anxiety: 4.5/5.9, in depression: 4.6/5.9, in comorbid	Links into our research on correlation between anxiety, depression and somatic symptoms.	Exceptionall y large dataset used.
			anxiety and depression: 6.1/7.6, and in no anxiety or depression: 2.6/3.6) and in all age groups.		

Overall, the existing literature points to relevance of somatization symptoms in younger patients being prevalent, along with links between somatization and anxiety/depression. We tried to build on this by examining mental stress and pain responses from the 90-question survey.

DATA, MATERIALS AND RESEARCH METHODS

Survey for psychology is a 0-5 template**:** *Not at all, a little bit, moderately, quite a bit, extremely* For the pain to ideation we broke into two categories for the outpatient survey. None of the respondents selected zero. We calculated the mean averages of each question among the groups and opinions of self that are measurable and analyzed them.

We began by categorizing survey responses into two primary groups: physical pain (e.g., dizziness and headaches) and then negative ideation (e.g., guilt, self-blame). Each survey question was assigned to one of these categories based on the type of response it generated. A key focus of our study was the question, "Are you bothered by faintness or dizziness?" We hypothesized that dizziness might serve as a strong indicator of somatization, linking physical symptoms to underlying mental stress.

Categorization for measurements:

Time Related (bothered by): feeling blue, worried, lonely, blocked, self blame, critical, hopeless future, guilt.

Physical Related (bothered by): decisions, low energy, slowed, irritable, nervous, easily hurt feelings, concentration, tense, difficulty, effort, guilt.

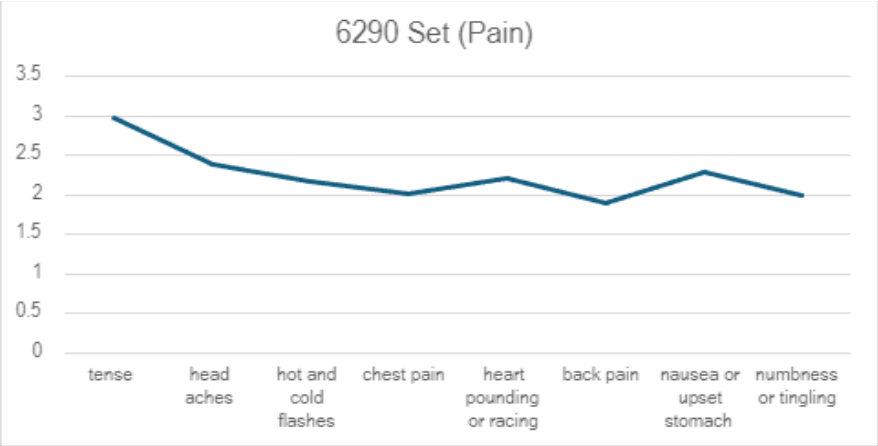
To analyze these relationships, we used feature selection and model training techniques to prepare and process the data. We carefully selected features that were most relevant to our research questions like anxiety, depression and age comparisons to infer the meaning of the mean answers.

Specifically, we focused on head-related pain—headaches and dizziness—and their potential connections to feelings of guilt and self-blame. To visualize this, we created charts to analyze patterns of blame, pain, public fear, and also compared these relationships across different age groups within the dizziness dataset.

However, the T-test results were inconclusive, suggesting that other, unidentified variables may have been at play. In the survey the mean average of “others are not sympathetic” scores higher than other questions and as “often” but is an indicator of response bias in that a portion of the interviewees are bothered by the survey itself. This led us to explore the influence of mental constructions and speculative reasoning on the responses, as well as the biases introduced by question order and the overall survey design.

Data and Materials

Somatization study with ‘Outpatients’



2 Moderate and 3 Often

This survey group is reporting pain as more than moderate on average. The initial dataset and statistics show that of 6290 Set participants the average for:

- tense 2.97 nearly often.
- headaches was 2.4 a little more than moderately.
- nausea or upset stomach 2.29 a little more than moderately.

- heart pounding or racing 2.216 a little more than moderately.
- hot and cold flashes 2.174 a little more than moderately.
- chest pain was 2.02 moderately.
- numbness or tingling 2.0 a little more than moderately.
- back pain was 1.9 almost moderately.

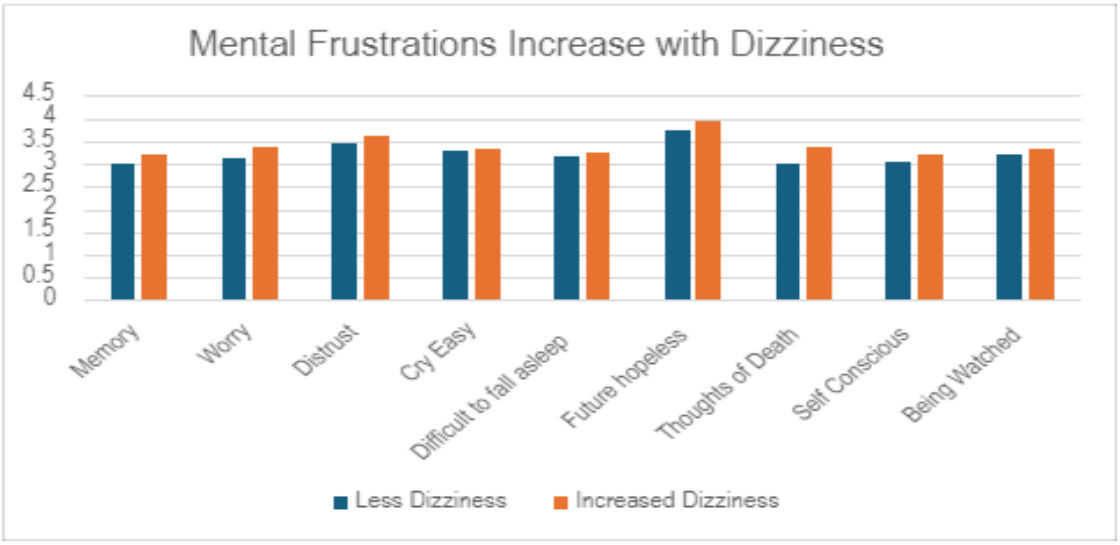
The initial dataset and statistics reveal that non-physical factors, such as thinking or contemplating ideas, had an average mean score of 2.41. This indicates that participants were bothered by these factors between a “moderate” amount of time and “quite a bit”. This is higher than the average mean score for physical pain 2.28, indicating that participants were bothered by physical pain slightly more than a ‘moderate amount of the time’, but less than they contemplate ideas by .13 (the difference between 2.41 and 2.28) The difference between 2 and 3 when 2 is moderate and 3 is often.

This indicates that the 6,290 respondents reported more about “negative thinking” than physical ailments overall. Will this ratio stay the same under particular combinations of variables in/for ideation? A continuing research project.

Tense	2.97
Head aches	2.4
Nausea or upset stomach	2.29
Hot and cold flashes	2.174
Chest pain	2.02
Numbness or tingling	2
Back pain	1.9

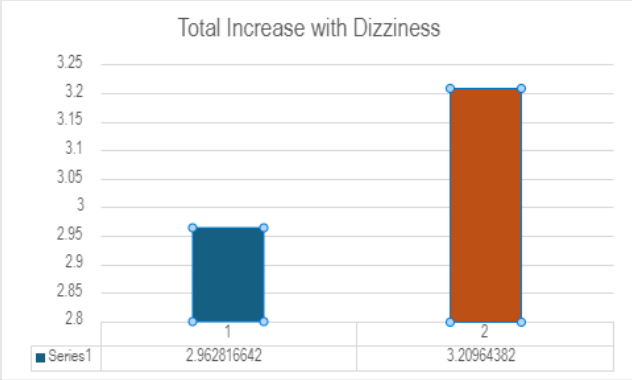
Dizziness and Contemplation Subsets

Of the 6290 initial data set this subset involves 2 Sets of 1350 which only include those who considered being dizzy as a *3 – Often to 4 & 5 extremely*.



The graph shows as dizziness increases so do implications of negative contemplation-time and other pains.

The accumulation of the differences suggest significance dizziness implies more mental stress as noted. If they report more dizziness they are generally more sensitive to every question.



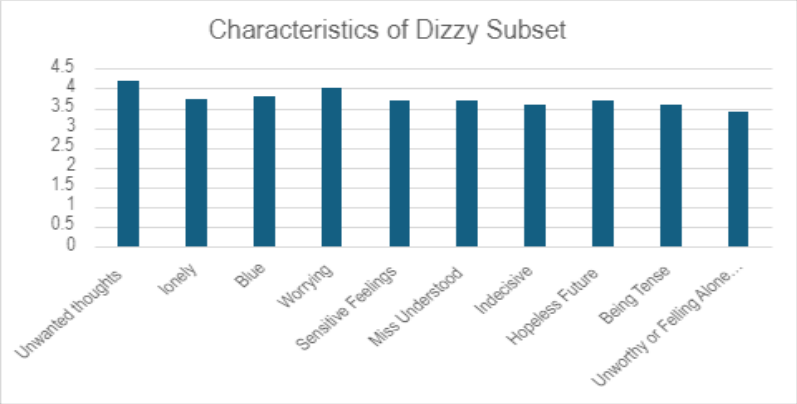
The correlation does not necessarily mean causation and is proportionally small. But it also rules in favor of our focus which is somatization which is energy in thinking.

As dizziness increases notice significant mental comparisons:

- Unwanted thoughts, words or ideas that won't leave your mind (4.2 to 4.4)
- Low energy and irritability both increase (3.8 and 4.1)
- Loss of sexual interest also increased from (3.3 to 4.25)

- Easily have your feelings hurt from (3.7 to 3.9)
- Headaches increased from roughly (3.1 to 3.4)
- Nervousness increased from (3.6 to 3.9)
- Unworthy or feeling alone while in public is increased to (3.5)
- Heart pounding and nausea or upset stomach increased (3.1 & 3.2 to 3.5 & 3.6)
- Other relevant elements like loneliness increase slightly with higher reports of dizziness

Of the 6290 initial data set this subset involves 2 Sets of 1350 which only include those who considered being dizzy as a *3 – Often to 4 & 5 extremely*



We will observe and showcase a potential example or correlation of the physical manifestation of ideas from the survey present themselves. The first group will include individuals who report dizziness levels averaging from 3 to 5, where 3 indicates ‘quite a bit,’ 4 indicates ‘extreme,’ and 5 indicates ‘most distressing.’ The highest correlated rankings for this group align more closely with the dizziness factor as follows:

- Unwanted thoughts, words, or ideas that won’t leave your mind (4.2)
- Lonely, feeling blue or worrying are also (3.75, 3.8 and 4.0)
- Easily have their feelings hurt or feel that people don’t understand both at (3.7)
- Feeling hopeless for the future is (3.7)
- Indecisive and trouble concentrating are both (3.6)
- *Being tense* and feeling that things are too much effort are both (3.6)
- Unworthy or feeling *alone while in public* is (3.4)

The second group will exclude those who reported a 3 and will only include the extreme and most distressing cases (4 and 5) to see which variables increase with the extreme.

The second subset of the same category only includes extreme dizziness ratings of rating 4 to 5 most distressing instead of the average of ratings 3.5 including moderate cases. All mental ideations increased with higher dizziness reports as assumed.

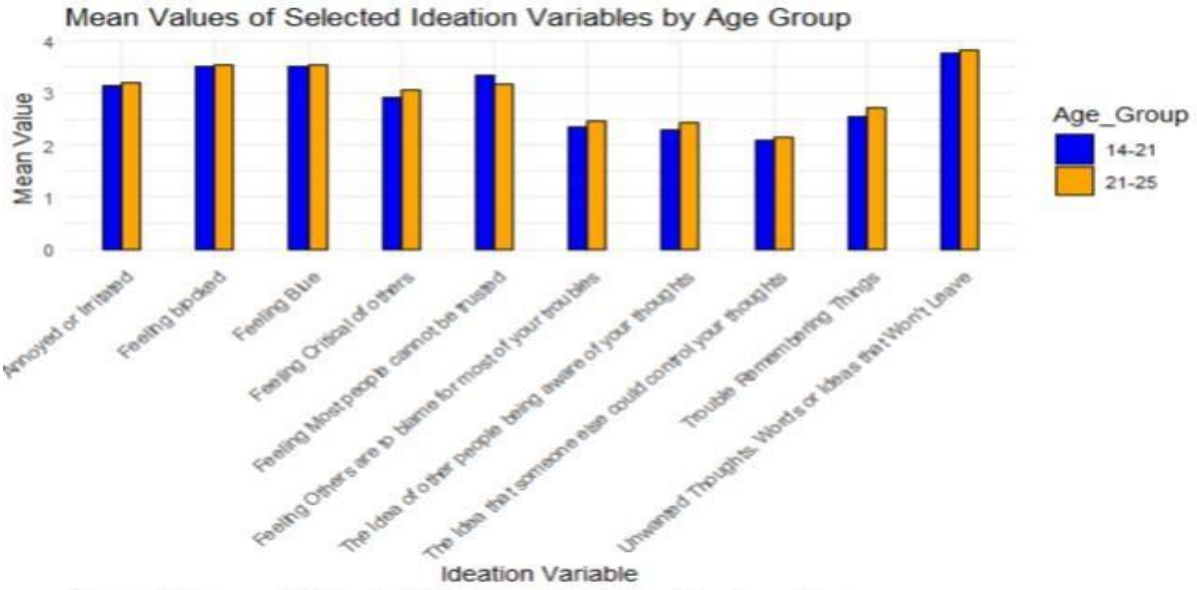
The conclusion of this section was to show the increase in physical ailments or previously related questions as the persistence of dizziness increased.

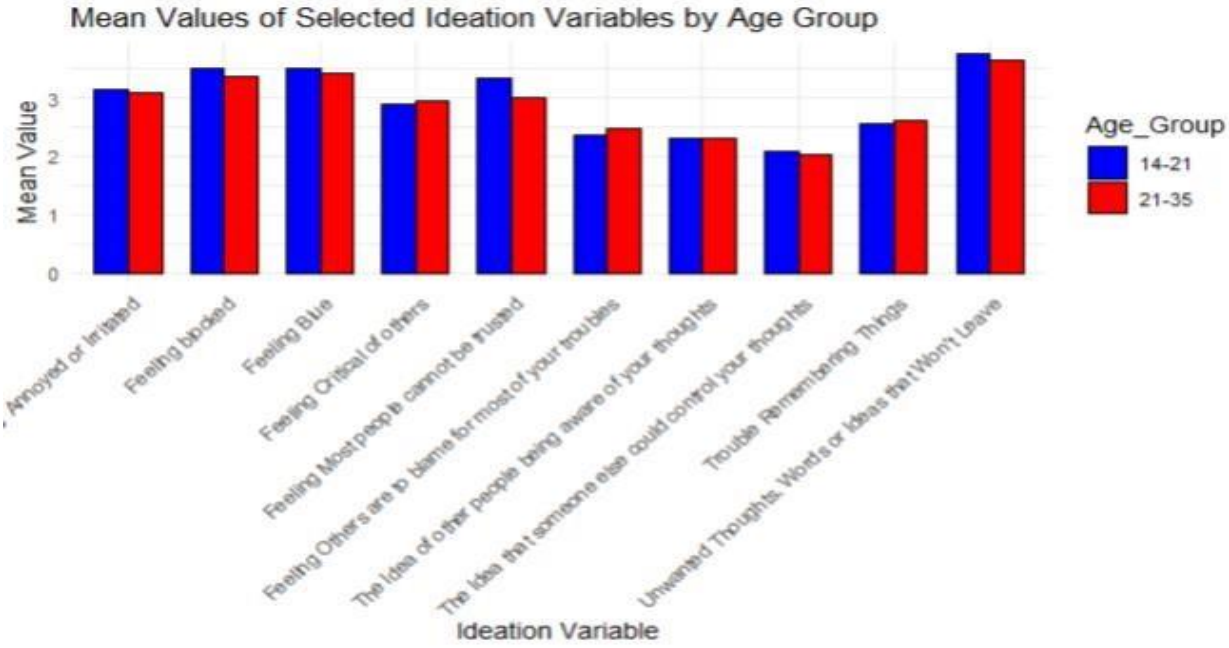
Age Section:

The following chart shows a sensitivity for the 21-25 group in the following questions:

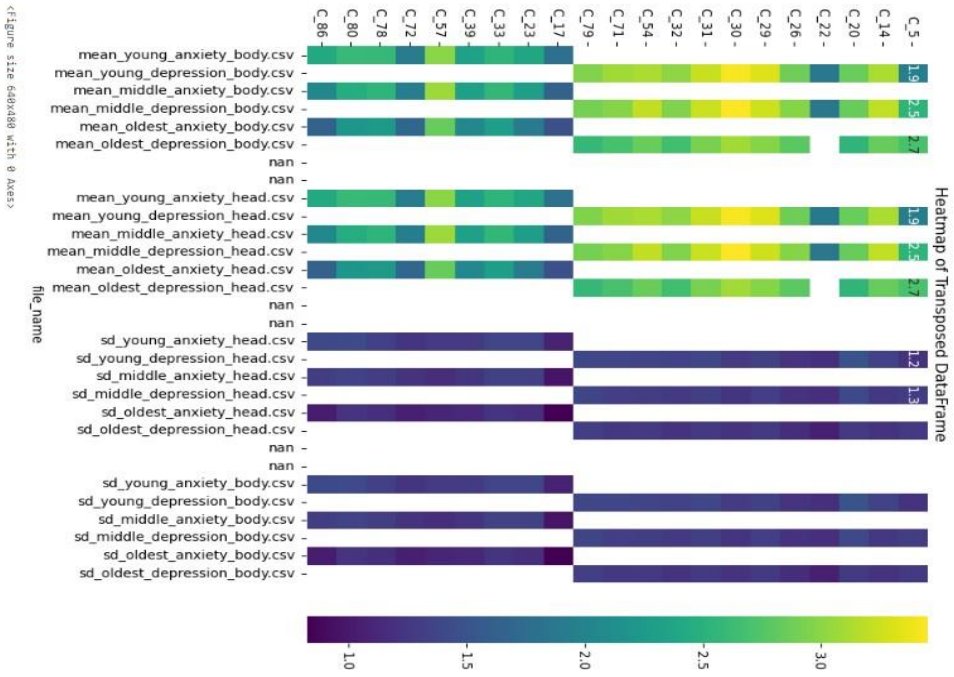
- Feeling blocked and feeling blue as there is a discrepancy between the categories.
- Others being aware of your thoughts a discrepancy between categories • Others could control your thoughts is a discrepancy between categories
- Unwanted thoughts there is a discrepancy between the categories.

In this comparison we've listed a slightly higher sensitivity of 21-25 for the previous questions.





Depression and anxiety as related to pain in the head or body.



The following questions are related to the pains associated with both anxiety and depression for the head and body as categorized above and generated in a heat map. The yellow indicates a higher rating among associated questions. Note that C_57 is tension or feeling keyed up. C_33 and C_80 reveal that they are associated with ”fear” and ”feeling that things are unreal” the continuing results indicate a psychological induction.

C_29 Feeling lonely
C_31 Worrying too much about things
C_14 Feeling low in energy or slowed down
C_32 Feeling no interest in things
C_54 Feeling hopeless about the future
C_71 Feeling everything is an effort
C_57 Feeling tense or keyed up

Dataset Description and Exploratory Analysis

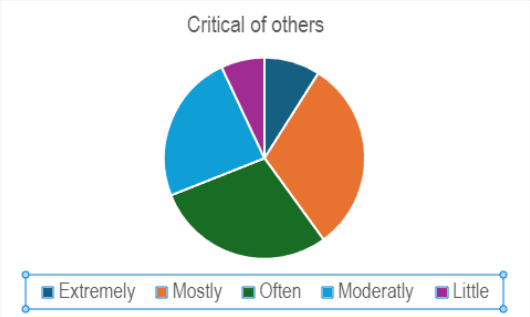
The data set was obtained as a voluntary outpatient survey generally described 6290 patients across 375 doctors. We transformed data by age groups and identified features of age groups for questions. We arranged the data by age groups between ideation and pain.

We created a subset of the most critical of the group called the 314 critical blaming group.

The Blame Element Subsets

As an example of ideation compared to physical pain or ailments, we consider two questions from the SCL-90 survey regarding blaming and the time category.

Among 375 doctors and 6,290 participants, the variation in how participants perceive the survey could significantly affect the sample, particularly regarding feelings of guilt or blame. The survey pie charts show results from all 6,290 participants on questions of blame and criticism of others in general, such as “Believe that others are to blame for their problems” and “Feeling critical of others.”



Critical of others subset from SCL-90 or introduce (The 314 Set)

These percentages are the highest among those who selected a 5 for being highly critical of others, subset totaling 314 out of 6,290 respondents.

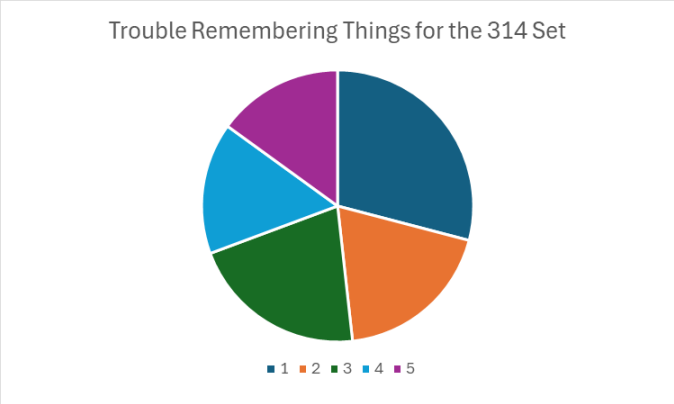
The set of the extreme 314 filtered as overly critical group of 6290 set examining the energy of a negative critical subset (stress energy evaluation)

The 314 set includes 5 ratings for the questions: ‘extremely critical of others’ and “Feeling others are to blame for most of their troubles” filtered set includes 4.5% of the total (314 of 6290 total participants)

The dataset of individuals who blame others for their problems reveals the following statistics from the survey as significantly greater than other questions:

- 77% Rating (4-5) Often to always report loneliness
- 82% Rating (4-5) Often to always felt that others do not understand them or are unsympathetic toward them.
- 45% Rating (4-5) Often to always feel that things take too much effort or are too difficult.
- 48% Rating (5) Extremely or always indicate that they feel “Worthless”.
- 76% Rating (1) a little bit of the group report they “hear uncontrollable voices in their head.”
- 41% Rating (1) a little bit indicate that “having to repeat the same actions such as touching, counting, washing” and with Rating (4,5) often or extreme at 26%.
- 65% Rating (1) a little bit, of the group report that they are “afraid to go out of their house alone.
- 60% Rating (1) a little bit of the same group say they “feel afraid they will faint in public.”

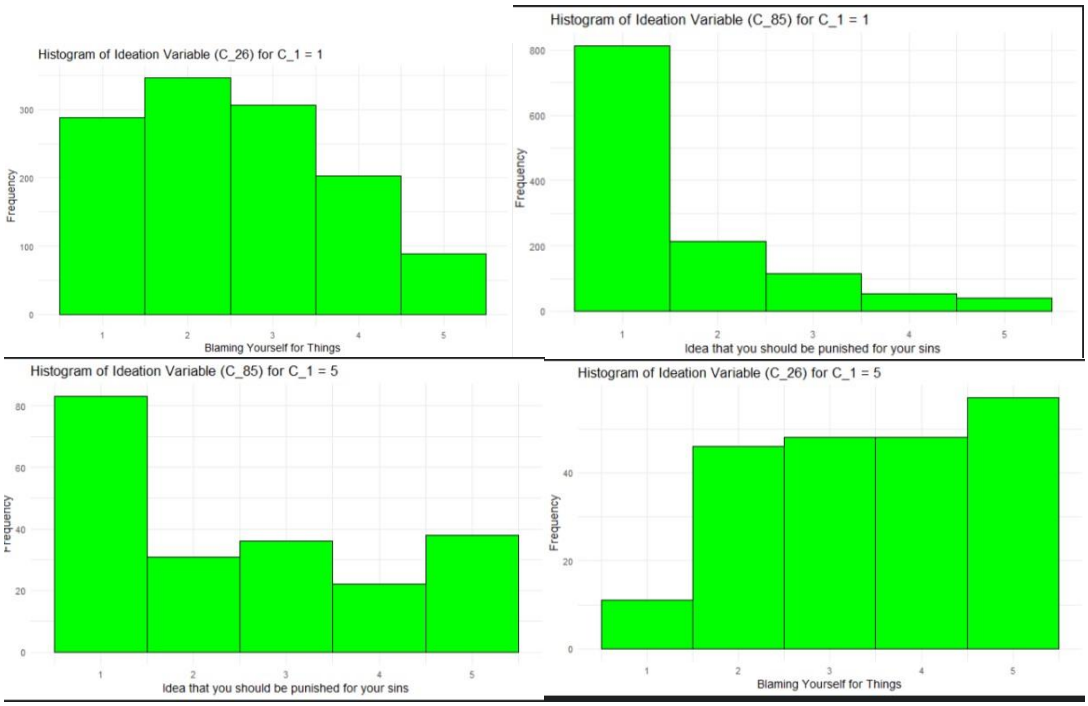
The 314 Critical Negative Set (high blame group)

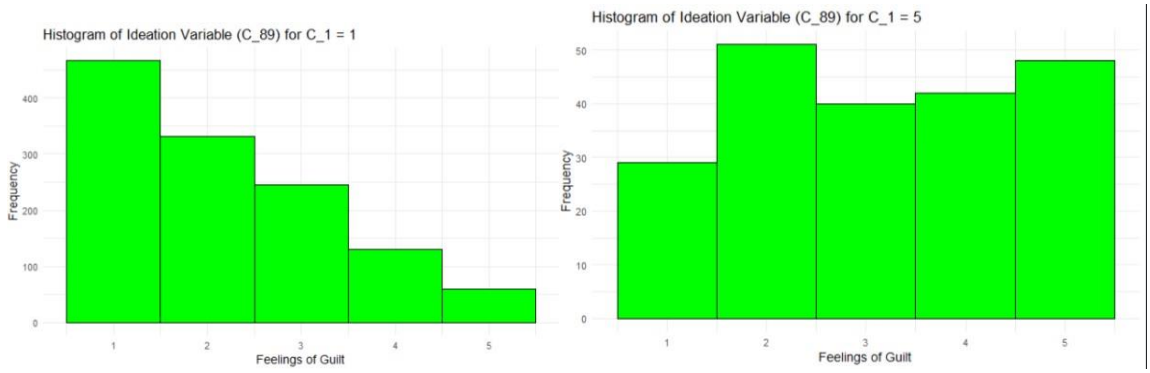


The data indicates that those who are more critical of others regarding their problems are 46% afraid in open spaces and roughly 18% unafraid in open spaces or on the streets.

In the same group, feeling weak or hot and cold spells were evenly distributed in amount among the group in levels of a little bit, moderate, often to extreme.

The following graphs exemplify our argument C_1 meaning low headaches and C_5 meaning high headaches





As the evidence indicates somatic activity related to the contemplation of guilt and blame for either self or others.

We also attempted to frame a theory as an impulse study for tension and ideation after recognizing the mean average of tension being highest from a previous pain group.

To continue suggested research, we introduced three subcategories within the time category: measurable opinions which is the time factor for the reasonable duration of the ideation. Dynamic ideations vary in mental emotion, a third category are questions only related to others, and a final group focused on being in public as a further continuing suggestion for analysis.

Results

Findings:

Of the entire dataset 6290 the following were ***listed as 3*** or more which is more *often* (bothered by): f*eeling blue, worried, lonely, blocked, self blame, irritable, critical of others, hopeless future*

Of the entire dataset 6290 the following were ***listed as 2.8*** much more than *moderately* (bothered by): d*ecisions, low energy, slowed, nervous, easily hurt feelings, concentration, tense*

Discussion and Conclusions

Like algorithms affecting children consuming social media and the design of 1-to-many and many-to-1 groupings for user experience or advertising, the study has the same ethical questions that are relative to the structured contexts through the ordering of questions or the context of the administration. People are simply different in how they perceive and think. One author felt being observed impacted their confidence.

This indicates the side effects concerns of psychotropic drugs and the ethics of business. The authors of this paper are not psychologists or doctors. Due to somatization concept we conclude order and response bias while attempting to measure idea time verses pain.

The survey is a shield to validate psychiatrists and obstructs ordinary thought with the patient. People are simply different. The survey seems to be Inducive theory. How many of the participants liked the survey or if they felt another person may read it. Like, who do you feel will

read your responses? Or, if they remember within the moment of the survey how it would be relational to anything.

Additionally, we investigated the relationship between psychological tension which scores highest on physical pain and unwanted thoughts, surveillance concerns, and contemplation.

Somatization is holding tension in the body ordinarily in mental constructs related to the head, mental “block”, “mind going blank”, and “trouble remembering things”.

Lastly, we explored whether feelings of soreness or being trapped (blocked) were related to each other through anxiety as a mental authoritative obstruction. We felt it induces the condition as an authoritative pressure.

Limitations and Future Work

The original SCL 90 is not conclusive or inconclusive in this paper. We don’t know if the data sets are accurate from the information we have. Can we determine if the survey context involves doctor-patient relationships with varying diagnoses? Are the illnesses or conditions all the same, such as being classified as “bipolar,” “anxiety,” or “depression”? The results are speculatively biased, or we just want to see how to build a framework model. The limitation is that the data set is potentially generic.

From the data set provided the dynamic of heterogeneous is noted. -375 different doctors and 6290 patients. Determining the context of the survey’s administration would affect the bias. For instance, medical condition. The context of the data indicates the word “outpatient”. Are they outpatients from a recent inpatient experience?

Or is this a survey that is taken of those with current diagnosis for somatization at a later visit? The assumption is that it is a varying collection. The ordering or coupling of the question may create bias or affect the intention of the interviewee or their relationship with the administrator.

We also considered whether the question should reflect the overall bias of the survey or have a more direct meaning, such as relating to an accident, injury, crime, or malpractice indicating potential subjective meaning across different unknown scenarios to limited data details like gender.

This approach addresses the survey’s statistical validity concerning the order or administration of questions. The relationships between meanings would be more sensitive in a physically aggravated population. This can also portray a mental stereotype in communication or misinterpretation behind the meaning of the survey’s intent; a voluntary survey among 375 various Doctors. We considered ordering bias in the event of a non-homogeneous survey.

The context in which the survey is delivered can influence participants’ intentions when answering. Additionally, somatization may exaggerate participants’ opinions or objectives, introducing bias varying in their responses; response bias.

Invalidity in a diagnosis doesn’t mean the data is irrelevant, “inaccurate” or unimportant. The correlation could be an essential pivot.

References

Asmundson, G. J., & Katz, J. (2009). Understanding the co-occurrence of anxiety disorders and chronic pain: state-of-the-art. *Depression and anxiety*, *26*(10), 888-901.

Bass, C., & Benjamin, S. (1993). The management of chronic somatisation. *The British Journal of Psychiatry*, *162*(4), 472-480.

Garralda, M. E. (1996). Somatisation in children. *Journal of child Psychology and Psychiatry*, *37*(1), 13-33.

Goldberg, D. P., & Bridges, K. (1988). Somatic presentations of psychiatric illness in primary care setting. *Journal of psychosomatic research*, *32*(2), 137-144.

Hardt, J., Gerbershagen, H. U., & Franke, P. (2000). The symptom check-list, SCL-90-R: its use and characteristics in chronic pain patients. *European Journal of Pain*, *4*(2), 137-148.