

Evaluating the Influence of Socioeconomic Factors on Drug Abuse Through An Uniform AI-Driven Scale

Wei-Chen Lee¹

Received June 15, 2025

Accepted March 27, 2026

Electronic access April 30, 2026

The global drug epidemic continues to escalate into a severely serious public health concern. It's an epidemic heavily influenced by different social aspects, including wealth (including, but not limited, to poverty), education, peer pressure, mental health, and other social relationships (family, isolation, etc.). Despite these five main factors having a wide recognition of having a significant relationship to drug abuse, public funding faces enormous challenges as policies struggle to find a focus. In response, this study incorporates an AI-driven scale to evaluate the "qualities" of the five different factors in regard to its urgency (Ur), reach (Rc), and recurrence (Re). The three measures collectively concern over the severity, scale, and persistence of socioeconomic factors and its relationship with drug abuse. Distinguishing the three dimensions allows for more precise and accurate evaluation, as well as for more interpretation space when the results are put to practical use. Through the three evaluation measures, legislators of different scales and perspectives are offered a better insight of what really is most important at the considered time and place. The ranking of the factors based on the data in the US from 1990 to 2020 through evaluation measures is as the following: Urgency: Mental Health (9.33) > Wealth (9.00) = Peer Pressure (9.00) > Other Social Relationships (8.89) > Education (7.56); Reach: Mental Health (9.00) > Peer Pressure (8.89) > Education (8.67) > Wealth (8.44) > Other social relationships (7.67); Recurrence: Wealth (8.56) > Peer Pressure (8.22) = Other Social Relationships (8.22) > Mental Health (8.00) > Education (6.67). As a conclusion, through the evaluations given by an AI-driven scale, by using urgency, reach, and recurrence of a factor's relationship to drug abuse, legislators could find better focus in regard to their concerns. With a more comprehensive view on the drug epidemic as a whole, we hope to see a better future with public funding solving the problems it is intended to.

Keywords: LLM, AI, drug abuse, socioeconomic, substance abuse, substance use, social factors, public health, drug epidemic, education, peer pressure, family, wealth, poverty, mental health, isolation, drug prevention

Introduction

The global drug epidemic, in recent years, has escalated into one of the most significant public health concerns. In search of effective solutions, the question of where we should focus our efforts sustains as current solutions present inconsistencies in finding success with the lack or instability of public funding¹. Current researches have identified key socioeconomic factors that have significant relationship regarding its influence² with drug abuse. These socioeconomic factors include wealth, education, peer influence, mental health, and other social relationships such as families.

As we develop an increasingly complete understanding of these bidirectional relations, a comprehensive overview on the relative importance can be conducted in the purpose of finding an aim for public funding. By evaluating the relative importance among each of the five factors, future decisions can be made with a clearer focus, and a more precise solution can be

developed in search to solve the drug epidemic.

However, current studies tend to operate within its own framework, utilizing distinct measures, definitions, and data analysis methods, making a comprehensive overview of the relative importance difficult, if not impossible, using traditional statistical tools. That is, as traditional statistical tools, such as odds ratios, Hedges' g, and regression models, rely on consistent effect sizes and comparable datasets³, conditions rarely meet in large scale, socially complex issues like drug abuse.

Portugal in 2001⁴ introduced a series of drug policies aimed to decriminalize the use and possession of all illicit drugs. Policymakers at the time faced competing claims from research, where some emphasized structural poverty, and other highlighted mental health disorders. Since these studies all operate on varying sampling frames, outcome definitions, and statistical measures, with traditional methods the policymaker failed to compare the magnitude of each factor.

To address this challenge, this study proposes a different approach: to use an AI-driven uniform rating (ranking) scale

¹ VIS@betterworld lab Experimental Education Institution, Taiwan, Republic of China (ROC)

that evaluates three parameters of each factor's influence on drug abuse: urgency (Ur), reach (Rc), and recurrence (Re). Unlike traditional models, this AI-driven scale utilizes natural language processing⁵ and adaptive weighting⁶ to synthesize qualitative and quantitative evidence across studies, providing a dynamic and comparable ranking of social determinants.

The scale would almost directly allow for comparisons across varying frames, definitions, and measures. An example of this scale is practiced through gathering data from 1990 to 2020 done in primarily Western countries especially in the US. Through this AI-driven scale, the challenges of conventional effect size analysis can be solved, and insights from prior studies are utilized to rank the relative importance among five socioeconomic factors. The tool is aimed to ultimately offer practical, data-informed results for policymakers, educators, and public health officials. In the fight against drug abuse, a strategic focus is not optional, it's essential.

Literature Review

Introduction

Human behavior is inherently social, which is a concept that can be traced back to classical philosophy. The famous Greek philosopher Aristotle described humans as social animals. Yet, the severity of the impact in relation among individuals on drug abuse were never realized until recent years⁷. Just as rates among individuals with Substance Use Disorder (SUD) differ among communities², a variety of factors among human to human relationships have a severe impact on SUD⁸. Scientific evidence suggests that this is a result of the oxytocinergic system⁹ interfering with both medicinal and social interactions. Besides the brain that needs to be carefully examined, the influence of social relationships on the problem of drug abuse should also be sincerely described, in detail of course.

Existing literature, as the following paragraphs will describe, highlights key factors and research results related to factors such as education, poverty, peer pressure, mental health, and other social relationships (including but not limited to family and isolation).

However, while these factors have been examined individually, a comprehensive synthesis of how these factors compare in importance remains underdeveloped. (The literary review is conducted with the assistance of OpenAI ChatGPT model 4o, with the prompt: “[p]lease go through the articles I’ve provided through study titles. Find sentences that fit these given labels: education, poverty, family, community, isolation, peer pressure and mental health. Give me a list of sentences that fit into the labels, and provide proper citations for the sentences. Make sure each and every output is verified and correctly sourced.”)

Other Social Relationships: Family and Isolation

The family is often considered the smallest unit of society, but the weight “family” holds is more than what a small society holds. The dynamics between parents, children, and drugs are often overlooked, but importantly, they are easily misrecognized¹⁰. Parental substance use has a significant influence on the growth and development of children, specifically, it alters the oxytocin signaling, affecting their vulnerability to addiction. The dynamic extends beyond the parent-child relationship¹¹, as it reaches into the “family” structure. Studies show that families affected by opioid addiction have heightened risks in intergenerational substance use patterns¹². The phenomenon of substance abuse is thereby described, but not examined. Questions such as “how has drug abuse penetrated into the structure of families?” and “how does substance abuse affect the relationships among individuals within a family, and what consequences are then led to?” are still unanswered. That is, the impact of the upbringing of individuals on their future drug abuse and the possible solution to drug abuse due to familial impacts are still hanging.

The lack of human interaction for an individual is deadly. Social isolation and opioid use disorder are bidirectionally linked¹³. This means that they create a cycle in which loneliness leads to drug abuse, drug abuse leads to loneliness¹⁴, and so on. Isolation is, therefore, not only a result of addiction, but also a force that drives continuation¹⁵. The phenomenon makes solving the drug epidemic also a quest to solve for social isolation, meaning for more job seeking opportunities, care units, and social building networks. The drug epidemic is not a problem itself, but a result of various structural flaws, which is what makes finding a solution so difficult.

Peer Pressure

As the saying goes, “schools are the first society (outside of the family) that children meet”; the influence of peers, classmates, and friends is also crucial in uncovering drug abuse as a social phenomenon¹⁶. The relationship can be a hindrance, as friends who are unaware of addiction for a user, can create a barrier in seeking help¹⁵. Peer groups can also reinforce substance use¹⁴. But peer pressure and its impact on drug users is different from the family, as a family often serves for motivation, and peers often serve as reinforcement. Although the relationship varies across cases, the trend can still be described, especially when considering the environment and consequences of poverty.

Education

Shifting the perspective of viewing school as a social institution to a place of education, its relationship with drug abuse also changes. Generally, the hypothesis that low education

attainment leads to higher risk of drug use is supported, but to which degree the link of education is to drug abuse is still rather unclear¹⁷. Furthermore, as drug prevention education since its implementation of the 19th century¹⁸ has now developed into a heavily regarded intervention in school systems¹⁹, its efficiency is still constantly debated. While some studies have demonstrated its success in preventing (delaying) substance use, others questioned its long-term effectiveness, especially when they lack sustained engagement. Thus, the role of education in drug abuse prevention must be more complex than previously assumed²⁰.

Poverty

Poverty also builds strong structural correlation with drug abuse. Among individuals with unstable employment, financial stress can increase susceptibility to substance use, and access to health care and treatment resources are also hindered because of the lack of accumulation²¹. As similar to what researchers have found in previous topics: there is a correlation, but the understanding stops there. This makes it difficult to be practical, to come up with systematic solutions, and to from-the-stem eliminate substance abuse as a whole.

Mental Health

In recent years, the topic of mental health is no longer invisible, and its relationship to drug abuse must be discussed beyond biology. In fact, the relationship is causal, reciprocal, and deeply entangled²². Disorders such as depression, anxiety, and PTSD all have shown to increase susceptibility to substance use, often as a form of self-medication. Unfortunately, oftentimes, what began as a form of medication for mental health develops into a form of dependency on the substance. Mental illness also weakens social bonds, decreases motivation to seek outside assistance, and interferes with decision making processes, all of which contributes to the initiation and maintenance of drug use¹⁴. Conversely, substance use alters the brain in exacerbating symptoms of mental illness²³, forming an endless cycle of mutual reinforcement. That is, the relationship between substance use and mental illness is cyclical and expansive.

Measurements

In this study, we later introduce measurements Urgency, Reach, and Recurrence as evaluation scores. Their specific definitions and grading methods will be detailed in section Measures. Urgency was emphasized in the health impact pyramid²⁴ in prioritizing interventions. Reach is concerned in sources like the WHO who evaluates health problems primarily based on population affected. And lastly, recurrence,

which additional drug-biology related researches all emphasize relapse cycle and chronic vulnerability²⁵, highlighting the importance of persistence in drug abuse analytics.

Beside the three however, there were some other options for evaluation that were eliminated. For example, economic cost, policy feasibility, and statistical significance (p-values) were all once considered. Economic cost was vulnerable due to difficulty in standardizing across the considered factors. Furthermore, it requires modeling assumptions that invalidate research results. Policy feasibility is high context-dependent, and would have introduced interpretive bias beyond quantitative analysis. Statistical significance is highly sensitive to sample size, but more importantly, it does not reflect practical importance therefore is against the objective of this study.

Gap

Today, what is known is often repeated. Drug abuse has strong connections to social, psychological, and economical conditions. Specifically, on each factor of peer pressure, education, poverty, mental health, and other social relationships there have been a large basis of studies that indicate the harmfulness of the relationships. However, though studies have sometimes explored the connection between each factor, there has yet to be a comprehensive review to map an overview of the complex connection web.

Methods

Overview

Motivated by acknowledging the difficulties of public funding in finding a solution to the worsening drug epidemic, this study aims to identify a focus, thereby ranking, the relative importance of different socioeconomic factors and their influence on drug abuse. This study especially considers five socioeconomic factors (see literature review) including wealth, education, peer pressure, other social relationships, and mental health. This study introduces a uniform scale to measure the relative relationship between socioeconomic factors and their influence on drug abuse (see Methods Justification for further elaboration).

Data Collection

Data of previous studies on related topics will first be conducted through literature search and mass search engines. The comprehensive literature search will first be conducted using online academic medical databases including the following: CDC (Centers for Disease Control and Prevention), NIDA (National Institute on Drug Abuse), WHO (World Health Organization), SAMHSA (Substance Abuse and Mental Health

Services Administration), NIH (National Institutes of Health), PubMed, KFF (Kaiser Family Foundation), OECD Health Statistics.

The focus of data collection will be placed on peer-reviewed articles published between 1990 to 2025. Since the 1990s, countries particularly in Northern America and Europe saw the rise of prescription opioid use which later developed into the opioid epidemic. On a side note, other historical drug trends like the first opioid wave from the 1870s to 1910s, or the post WWII heroin trend all show similar trends. However, this study targets the opioid epidemic as it is the most socially and culturally relevant example of most modern drug epidemics. This study attempts to evaluate modern factors and its influence on modern drug epidemics, therefore using studies since the start of trends similar to today's epidemic would be appropriate. The criteria of inclusion is of the following: studies that report quantitative data on drug abuse outcomes and cause; studies that examine socioeconomic, psychological, and/or environmental factors; and studies that include a clear definition of its data and analysis. Exclusion criteria include qualitative-only studies and studies with insufficient (incomplete, irrelevant, arbitrary, etc.) data reporting.

Every study must surround the topic of drug abuse, specifically mentioning texts including “drug abuse”, “substance abuse”, or “illicit use of medication”. Each research also must include either one of the following topics, and studies mainly the factor's relationship to drug abuse. First, wealth, which the text must mention specific keywords multiple times including “wealth”, “poverty”, “economic despair”, and “economic difficulties”. Second, education, which the text must mention specific keywords multiple times including “high school students”, “middle school students”, “elementary school students”, “anti-drug education (courses)”, “education level”, and “education opportunities”. Third, other social relationships, which the text must mention specific keywords multiple times including “relationship issues/difficulties”, “family”, and “isolation”. Fourth, peer pressure, which the text must mention specific keywords multiple times including “peer influence”, “peer pressure”, and “peer”. Fifth, mental health, which the text must mention specific keywords multiple times including “mental health”, “mental ill(ness)”, “mental disorder”, “personality disorder”, “trauma”, “trauma-induced stress”, “PTSD”, “depression”, “anxiety” and other mental disorders defined in DSM-5-TR.

These five (wealth, education, other social relationships, peer pressure, and mental health) will together be referred to as “study factors” or “factors” for short.

Note that search terms “illicit medicine use”, “drug abuse”, and “substance abuse” are all required in this study. Illicit medicine use is defined as non-prescribed, unlawful, and/or unintended consumption of substances that in other cases may be recognized as medication. Substance abuse and drug abuse,

on the other hand, are umbrella terms used to describe harmful use of both legal and illegal substances. These terms are contextually distinct, and to include all terms in data collection is important as to include medication misuse regardless of their varying social perception, access pathways, regulations, and intervention strategies.

Methods Justification

In consideration of the limitations of traditional effect size analysis, an AI-driven uniform weighting scale will be used. The scale is an attempt to evaluate the quality of quantitative statistics. Typically, any attempt to assess relative relationships between quantitative statistics will be done by experts of the field. However, as the nature of the socioeconomic factors, bias caused by any experts' personal experience would heavily influence the result. With AI-based LLM, the bias can be significantly reduced in consideration of the nature of AI being developed from a plethora of documented studies. AI could reduce the bias caused by personal experience and emotion. Later, in Prompting, this study attempts to guide AI-based LLMs into a comparatively unbiased qualitative assessment of the previously collected statistics.

Measures

Each study factor with its relevant statistics will be evaluated based on the following three categories: urgency (Ur), reach (Rc), and recurrence (Re). These three will be collectively referred to as the evaluation measures. Together, the measurements offer a comprehensive view for evaluating, not only of the magnitude of the concern, but also its temporal persistence and population-level impact, making them especially relevant for analyzing complex, multifactorial phenomena such as drug abuse. These were chosen, not only because of their relevance to public health and drug use analysis, but also for their ability to capture distinct aspects of “influence” that traditional effect size analysis couldn't. The three dimensions allow for an analysis beyond binary associations or isolated statistical effects.

First, Urgency (Ur) is the evaluation that is based on the degree of harm, danger, and public health implied by the data. This also includes whether direct behaviors are affected and the vagueness of the relationship between the factor and drug abuse. With 1 being the least “urgent” and 10 being the most. Second, Reach (Rc) captures the amount or proportion of a population affected by the factor. With 1 being the most small scale influence and 10 being the most large scale, possibly global or systematic. Third, Recurrence (Re) refers to the consistent nature of the relationship between the factor and drug abuse. That is, this includes whether the influence is persistent, cyclic, or one-time. With 1 being one-time and rarely consistent and with 10 reflecting a systemic, chronic,

and cyclic recurrence pattern. It is important to clarify that a low score in recurrence for any factor cannot be equivalent or drawn to higher effectiveness of intervention programs without further research. A low score in recurrence can only mean that the relationship between the factor and its related drug abuse studies show to be one-time and non-consistent. An example interpretation for the factor “Education” is the relationship with drug abuse.

The purpose of this study is to provide structured dimensions through which prior statistical findings can be interpreted with consistency and transparency. It is not to claim that LLMs can validate or empirically measure these constructs in the traditional psychometric sense. As this study does not attempt to construct a formal latent measurement model, the three measurements: Urgency, Reach, and Recurrence, are intentionally defined as interpretive dimensions derived from public health logic.

Large Language Models (LLMs)

Multiple LLM (Large Language Models) will be incorporated and the averaged score given by each LLM model will be the measure for the final evaluation metric. Given that LLMs fall into different categories of types based on how the model is trained, accessed, used or enhanced, this study incorporates different types for a more comprehensive and complete overview on the topic. For future researchers, as AI technology evolves, it is recommended to use the most up-to-date models. However, it is important to note that only models of similar nature or “type” will be suited for replicating the methods described herein.

The following LLMs will be used (listed as: Company – Model – type of LLM, followed by a justification and explanation paragraph):

First, OpenAI – ChatGPT model o4 – Closed-Weight is one of the most advanced and widely used language models available to the general public. Ever since the chatbot (model 3.5) was launched in November of 2022, the AI language model marked the beginning of the times of AI. Its ability to produce high accuracy factual responses is significant to this study as it is capable of engaging in conversations and providing explanations for its reasoning. Second, Google – Gemini 2.5 Flash – Closed-Weight is an AI-based LLM developed by Google. It is designed for a high-level understanding and reasoning across sources of text, images, and other forms of information. It is notable for its integration with Google’s massive data infrastructure, enabling it to provide strong responses based on both real-time and contextual knowledge. Its inclusion in this study plays a role in providing qualitative responses from quantitative data in a comprehensive manner. Third, Mistral AI – Le Chat – Open-Weight is a conversational interface built on Mistral’s open-weight Large Language Models. The mod-

els especially place emphasis on a transparent AI architecture, allowing any user to train, download, inspect, modify, and fine-tune the model. Le Chat, therefore almost entirely different from ChatGPT and Gemini, allows interpretations not influenced by proprietary datasets. In the context of this study, it would allow for interpretations based solely upon the 48 selected research papers. Later in the “data analysis” section, the effects and changes of so will be discussed, and other biases will be noted.

Besides the general characteristics of these LLMs, these models were all chosen for their outstanding performance in structural reasoning, instruction-following, and domain-general language understanding. Specifically, ChatGPT was chosen for its evaluative judgement, Gemini was for its contextual reasoning, and Le Chat was for its transparency oriented model. In general these models all satisfy the qualities of being strong in interpretability, reasoning, and evaluation, making them great candidates for this study.

Prompting

With the bases of the evaluation scale established and the selected LLMs defined, the following guidelines will serve as outlines for prompting a response for each LLM. That is, because of the foundational difference in the cornerstones of each model, a slightly different approach in prompting has to be taken. For example, a different prompt must be created between ChatGPT and Mistral AI as one relies on proprietary knowledge bases and the other relies on user-specific-weights. Therefore, standardized prompting guidelines are necessary to minimize variation in outputs and ensure that all results are assessed under comparable conditions, while accounting for the differences among LLMs.

The standardized guidelines must be followed for an accurate replication of this study. Any LLM response that violates any of the guidelines is invalid. The standardized guidelines are as follows.

First principle is on Pre-Prompt Acknowledgement. All prompts should be given to an LLM that has not been previously operated by any user whatsoever: the account of the LLM must be brand-new. The model must have no prior user interaction, saved memory or context, or exposure to any prompts. A minimal contextual framework should be given once at the start of each LLM’s evaluation cycle. Meaning, it shall serve only the purpose to clarify task expectations. The pre-prompt acknowledgement should be less than 100 English words. The contextual framework given should be as simple as possible, allowing the LLM enough information to understand without leading it towards any direction. The contextual message must aim to avoid introducing bias and must not suggest any expected ranking or outcome.

The second principle is Uniform Objective and Process

Framing. The LLM must be asked to evaluate the relationship between a factor and drug abuse. The prompt must mention the words “drug abuse”. The prompt should be less than 200 English words, not including the statistics given to the LLM. The LLM should be asked only to evaluate one of the five study factors at a time, and to assign a score to one of the three evaluation measures at a time. All LLM should be informed to use solely the statistics provided as bases of reasoning and assessment. The prompt includes a clear, structured request to assign a numerical score ranging from 1 to 10, explaining 1 being the least, and 10 being the most. For either one of the evaluation measures, a minimal explanation is required to inform the LLM of the definition of the measure. The following words are required to mention for each measure: Urgency (Ur): harm and severity; Reach (Rc): population scope; Recurrence (Re): chronicity or frequency. For either one of the study factors, the complete name and title of the factor should be clearly stated. The complete names are as follows: Wealth (including, but not limited to, poverty); Education level; Peer Pressure; Other Social Relationships (including, but not limited to, family relationships and social isolation); Mental Health. Finally, the LLM should be asked to provide a two sentence explanation for the numerical assessment. Additionally, the explanation should be asked to be as concise and specific as possible.

Third, please review Architecture-Specific Adjustments. Architecture-Specific Adjustments are required to account for inherent differences in LLM models that may systematically affect the results of this research. As previously explained, LLM models vary in architecture, training, and context-processing mechanisms. The adjustments are a strategy to reduce these inherent differences, while still counting on the structural judgement and reasoning these LLMs share. For Closed-Weight (e.g., ChatGPT, Gemini) models, the prompt should instruct the LLM to ignore internal training knowledge and rely solely on the given data. And additionally, the prompt should emphasize on the conciseness for the explanation. For Open-Weight (e.g., Le Chat) models, the prompt should provide full, embedded data inline in the prompt. Besides, the LLM should be asked to evaluate only on what is shown and given. For SALM (None that is incorporated in this specific study, but can be used in future studies.) models, the prompt should clearly request no web retrieval. And again, the LLM should be asked to evaluate only on what is shown and given.

Fourth, see the following on Response Formatting Consistency. The response should be as the following format to ensure consistency and comparability: (Isn’t required to mention in the prompt) the response should be in English language, name and title of the study factor, evaluation measure the LLM is assessing, a numerical score for the evaluation measure, and lastly a 2 sentence explanation of the assessment. The total re-

sponse must not exceed 100 (English) words.

Fifth, for Meta-Awareness Prompting, after 3 prompts, gathering for all 3 evaluation measures for 1 study factor, the LLM should be asked to self-assess the confidence of its assessments. The LLM should be asked to state any insufficiency, conflicts, or inconclusive output explicitly to ensure minimal uncertainty and ambiguity. The following is the exact replica of the meta-awareness prompt that will be integrated in this study. Minimal changes will be allowed, as long as the purpose of the meta-awareness prompt stays consistent.

The guideline framework is designed to ensure that evaluations performed by LLMs are isolated from previous influence, standardized across models, and interpretable across repeated runs. The evaluations should allow for valid cross-model comparison and transparent documentation on how and why each assessment is scored. The following is an example prompt constructed according to the standardized guidelines mentioned above. Sentences corresponding to specific guidelines are annotated using (number, letter) that match the sections outlined above.

[Minimal Contextual Framework]

“You are participating in a research study evaluating the relationship between socioeconomic factors and drug abuse. You will be shown data from peer-reviewed studies published between 1990 and 2025. Your task is to assign a 1–10 score for one specific evaluation measure, based only on the data provided in each prompt. Please do not use any outside knowledge.”

[Prompt 1, to closed-weight models]

“Evaluate the following data related to wealth (including, but not limited to, poverty) and its influence on drug abuse. Assign a score from 1 to 10, with 1 being the least and 10 the most, for its Urgency, defined as the harm and severity implied by the relationship. Use only the statistics below to make your assessment. After scoring, provide a two-sentence explanation of your assessment, based strictly on the data given below.

Statistical Data Excerpt: ...

Please respond in the following format:

Factor: Wealth (including, but not limited to, poverty)

Evaluation Measure: Urgency

Score (1–10): [your score]

Explanation (2 sentences): [your explanation]”

After feeding the LLM with a prompt that completely follows the above guidelines, the researcher should follow the protocol as described below to ensure consistency and avoid any possible bias or technical difficulties. **Result Averaging and Consistency Protocol:** first, after executing the same prompt 3 times for all 3 evaluation measures of 1 study factor, there should exist 3 numerical scores and 2 sentence explanations for each of the scores. Second, each prompt should then be executed 2 more times (3 in total) with minor prompt variations in wording, order, and phrasing. Third, final Ur, Rc, Re values per factor are averaged. Fourth, outputs should be compared internally and if any outlier appears the whole cycle should be redone.

Evaluation

The final score assigned to each factor for a given evaluation measure will be derived from the combined average of across all LLMs used. Specifically, after executing each prompt thrice per model (according to the Results Averaging and Consistency Protocol), the average score would be calculated. These individual LLM averages will then be again, averaged across multiple LLMs to produce a final cross-model score for that factor's evaluation measure. As an example, in the case of the "Urgency (Ur)" evaluation for the factor "Wealth", the final score would be calculated by averaging the three-run results within each model (e.g., ChatGPT, Gemini, Le Chat, and Perplexity AI), and then computing the overall average across all four models.

Interpretation

The final, averaged scores for all three evaluation measures of each (Urgency, Reach, and Recurrence) factor will be interpreted and discussed both independently and collectively to assess the relative importance and influence of the socioeconomic factor on drug abuse.

Each measure reflects a distinct interpretation of public health relevance: Urgency – the interpretation of the score is as the relative degree of immediate harm and severity of the factor's influence on drug abuse. The higher the score in Urgency, the more severe and detrimental the influence is. The score of Reach implies the relative population scope of the influence. The higher the score is, the more likely the relationship between the factor and drug abuse is to be a national or even international. Recurrence reflects the patterns of association, particularly in concern of whether the relationship is one-time or persistent. That is, the high scores in Recurrence reflect systemic or chronic patterns of association. Factors that demonstrate high scores in all three measures will be considered as more influential and significant in the appropriate context of the given data collection.

An important note to take is that the three dimensions are not intended to be collapsed into a single composite score unless the researcher provides a clear rationale for doing so. Instead, the interpretation should be done through multidimensional profiles, allowing for a more thorough analysis and interpretation based on the perspective of the interpreter.

In this study in particular, sources were collected primarily from Western countries, especially in the US. The purpose of so is to operate on a more comprehensive and complete database for a general interpretation of nationally common socioeconomic factors. Therefore, the interpretation of this study will prioritize understanding national patterns within the US context. Additionally, differences in score consistency across LLMs should be noted. Self-reported uncertainty from the models (see Standardized Guidelines) will be used to further guide interpretation.

The results gathered through this study must be qualitative. It is important, in the results, to cite the given score, explanation, and uncertainties to ensure validity of the drawn conclusions.

Data Analysis

Source Collection

First, 50 studies were found on the previous described large medical database, including CDC (Centers for Disease Control and Prevention), NIDA (National Institute on Drug Abuse), WHO (World Health Organization), SAMHSA (Substance Abuse and Mental Health Services Administration), NIH (National Institutes of Health), PubMed, KFF (Kaiser Family Foundation), and OECD Health Statistics.

The specific search strategy used in this study consists of the following search terms: "drug abuse" or "substance abuse" or "substance use disorder" or "illicit drug use", which is followed by Boolean operator AND to connect with core socioeconomic factors including search terms: "wealth" or "poverty" or "economic hardship" or "economic despair"; "education" or "school" or "education level"; "peer pressure" or "peer influence" or "friends"; "family relationship" or "social isolation" or "social relationship"; "mental health" or "mental disorder" or "depression" or "anxiety" or "PTSD" or "stress". Reference lists of relevant articles were also manually reviewed to identify all eligible sources.

Data Extraction

Then, data were drawn from the studies to form a database of different factors and its relationship to drug abuse. Please see Appendix for a complete chart of the collected data.

The data were drawn through the AI model Consensus under the prompt: "Please read through the following articles

and extract any statistical evidence in the format of ‘statistic, value, unit, description’.” Consensus is a LLM that specializes in reading and extracting information from researches given, qualifying its use to extract quantitative data through the given research papers. Consensus (no version description, May 2025) is an AI-powered search engine designed specifically for academic research. Its main function is to extract and summarize findings from scientific articles, literature, and research. Its database utilizes literature from Semantic Scholar, PubMed, OpenAlex, etc., and consists of over 200M+ papers as of 2025.

Additionally, all statistics and data drawn using Consensus were manually reviewed to ensure validity and reliability of the data. Each statistic was compared against the original cited source. The purpose of using Consensus was to improve efficiency in drawing statistics from cited sources under the sacrifice of possibly losing minor statistics.

For each factor, a corresponding selection of data will be drawn from the complete chart of collected data. See the following for the correspondence: Education: data number 1–30; Mental Health: data number 31–100; Peer Pressure: data number 101–120; Wealth: data number 121–146; Relationship: data number 147–232 (all include both the start and end, e.g. data number 1–30 includes both data number 1 and data number 30).

Prompting & Evaluation Collection

Following the standardized prompting guidelines, a separate prompt will be written for each of the chosen LLMs. For each LLM, three prompts will be created, and the same three will be applied to all 3 measures in all 5 factors. A total of 135 (3 LLMs × 5 factors × 3 measures × 3 evaluations per measure) pieces of evaluation should be collected.

The pre-prompt contextual framework would be identical throughout all of the run evaluations. The contextual framework is as follows: “You are participating in a research study evaluating the relationship between socioeconomic factors and drug abuse. You will be shown data from peer-reviewed studies published between 1990 and 2025. Your task is to assign a 1–10 score for one specific evaluation measure, based solely on the data provided in each prompt. Please do not use any outside knowledge.”

Besides the pre-prompt contextual framework, a main prompt will be created according to the standardized prompting guidelines. According to the Standardized Prompting Guidelines, each prompt will be slightly different in consideration of the variety among the LLMs. Besides, a slight variation and change of word choice or sentence structure is made between each prompt as according to the Results Averaging and Consistency Protocol. For the exact prompts used, see Appendix.

Results

The collected numerical evaluations are as below. See Appendix for the evaluation of both the numerical results and explanation. The explanations will later be cited in the Results section.

Table 1 Education

Education	Ur	Reach	Recurrence
ChatGPT – 1	8	9	7
ChatGPT – 2	8	9	7
ChatGPT – 3	8	9	7
Gemini – 1	8	9	7
Gemini – 2	7	9	7
Gemini – 3	8	9	7
Le Chat – 1	7	8	6
Le Chat – 2	7	8	6
Le Chat – 3	7	8	6
Average	7.56	8.67	6.67

Table 2 Mental Health

Mental Health	Ur	Rc	Re
ChatGPT – 1	10	9	9
ChatGPT – 2	10	9	9
ChatGPT – 3	10	10	9
Gemini – 1	9	8	7
Gemini – 2	9	9	7
Gemini – 3	9	9	7
Le Chat – 1	9	9	8
Le Chat – 2	9	9	8
Le Chat – 3	9	9	8
Average	9.33	9.00	8.00

The final, averaged results, standard deviations, and confidence intervals (95%) are shown in the following tables. The highest scores of each column are highlighted in green, and the lowest are in red.

Generally, as the charts of SD and corresponding 95% confidence intervals can tell, the results didn’t vary much, with the biggest SD being 1.0. This indicates a small, limited dispersion among the scores generated by LLMs. The narrow confidence intervals reflect stable, consistent central tendencies across repeated tests and model architectures.

Table 3 Peer Pressure

Peer Pressure	Ur	Rc	Re
ChatGPT – 1	9	9	9
ChatGPT – 2	9	9	9
ChatGPT – 3	9	9	9
Gemini – 1	9	7	8
Gemini – 2	9	7	8
Gemini – 3	9	8	7
Le Chat – 1	9	9	9
Le Chat – 2	9	9	9
Le Chat – 3	9	9	9
Average	9.00	8.44	8.56

Table 4 Wealth

Wealth	Ur	Rc	Re
ChatGPT – 1	9	9	8
ChatGPT – 2	9	9	9
ChatGPT – 3	9	9	9
Gemini – 1	9	9	8
Gemini – 2	9	9	8
Gemini – 3	9	9	8
Le Chat – 1	9	9	8
Le Chat – 2	9	8	8
Le Chat – 3	9	9	8
Average	9.00	8.89	8.22

Table 5 Relationship

Relationship	Ur	Rc	Re
ChatGPT – 1	9	8	9
ChatGPT – 2	9	8	9
ChatGPT – 3	9	8	9
Gemini – 1	9	6	7
Gemini – 2	9	6	8
Gemini – 3	9	8	8
Le Chat – 1	9	9	8
Le Chat – 2	9	8	8
Le Chat – 3	8	8	8
Average	8.89	7.67	8.22

Table 6 Final Averaged Results

Final Averaged	Urgency	Reach	Recurrence
Education	7.56	8.67	6.67
Mental Health	9.33	9.00	8.00
Peer Pressure	9.00	8.44	8.56
Wealth	9.00	8.89	8.22
Other Social Relationships	8.89	7.67	8.22
(lowest)	7.56	7.67	6.67

Table 7 Standard Deviation

Standard Deviation	Urgency	Reach	Recurrence
Education	0.53	0.50	0.50
Mental Health	0.50	0.50	0.87
Peer Pressure	0.00	0.88	0.73
Wealth	0.00	0.33	0.44
Other Social Relationships	0.33	1.00	0.67

Table 8 Confidence Intervals 95%

Confidence Intervals 95%	Urgency	Reach	Recurrence
Education	7.15–7.96	8.28–9.05	6.28–7.05
Mental Health	8.95–9.72	8.62–9.38	7.33–8.67
Peer Pressure	9.00–9.00	7.77–9.12	8.00–9.11
Wealth	9.00–9.00	8.63–9.15	7.88–8.56
Other Social Relationships	8.63–9.15	6.90–8.44	7.71–8.73

interpreted separately. Starting with the measure “Urgency”, the following bar graph is created for better visual understanding.

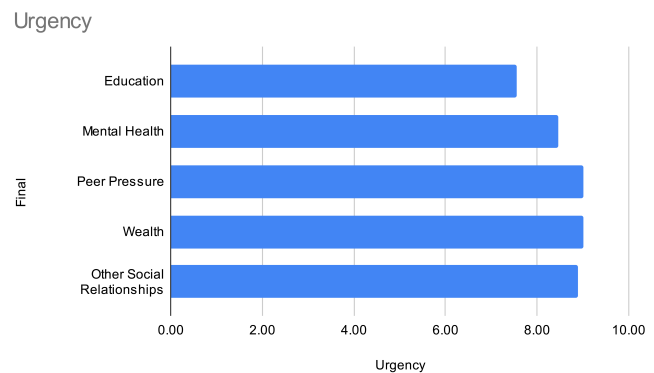


Fig. 1 Urgency scores by factor

Based on the evaluation, Mental Health shows the highest in Urgency with a score of 9.33. The results suggest a stronger need for emphasis in policies regarding mental health, espe-

Discussion & Limitations

Results Discussion

To further analyze the results, following the Interpretation section under Methods, the three evaluation measures should be

cially in concern with immediate harm and damages caused by a drug epidemic. On the contrary, Education scores the lowest in Urgency with 7.56. However, it must also be noted that one of the LLMs, ChatGPT in the Meta-Awareness Prompt responded that particularly for Education on Urgency, the data “[most] were static without longitudinal depth beyond this window” and that “the data focus on education as a contextual backdrop but do not directly measure how educational variables modulate the substance use outcomes.” Future data points added to the analysis would improve accuracy and precision to the study.

Meanwhile, Wealth (9.00), Peer Pressure (9.00), and Other Social Relationships (8.89) all have relatively similar scores in regard to Urgency. The representation means for a moderately significant relationship between drug abuse and the three factors, in which case, a legislator must be cautious and consider context to solve for the immediate harm and severity caused by the relationship.

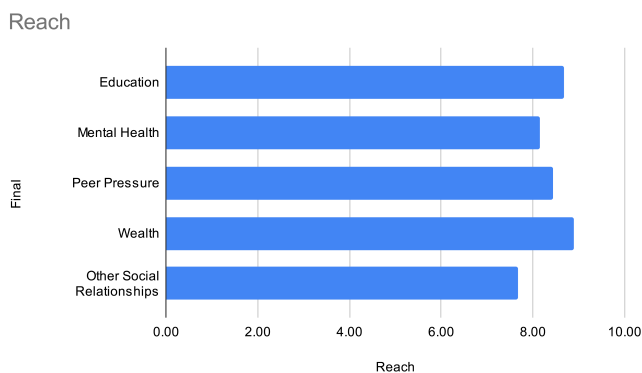


Fig. 2 Reach scores by factor

The “Reach” of these statistics imply the population scope implied by the relationship between the factor and drug abuse. With this in mind, Mental Health (9.00) and Peer Pressure (8.89) each rank number one and two in regards to the factor. These are the two most widely impactful factors when determining a trend of drug abuse.

A key insufficiency, however, is that the lowest scoring factor “Other Social Relationships” were recorded by Gemini in its Meta-Awareness Prompt to have data that “does not explicitly define or provide clear statistical data directly linking”. It states that “the presented statistics is implied rather than explicitly quantified in every instance”. An improved and renewed dataset is required for a more precise analysis.

The reach factor is especially helpful for legislators of smaller scale, meaning in consideration of a wide-spread drug epidemic, Mental Health and Peer Pressure, or the following, Education, should be most sincerely viewed. On the

other hand, Other Social Relationships, covering family and other relationships scored the least. Such results can be interpreted that the relationships implied are very group-specific, not widespread, and certainly should be considered the last in regards to national matters.

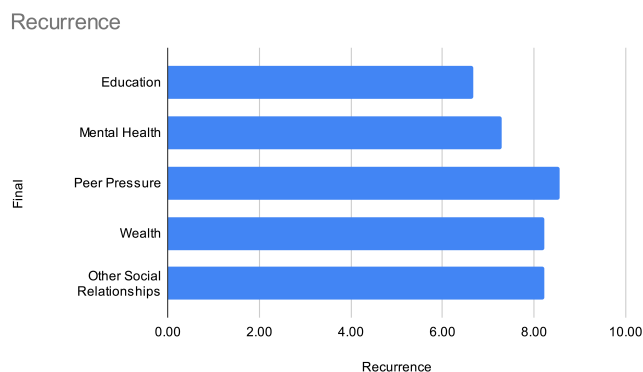


Fig. 3 Recurrence scores by factor

It is obvious that Education (6.67) scores much lower than any other factor. This goes to show that the factor of education and its relationship with drug abuse present low chronicity and frequency rates. It’s important to draw that a low recurrence score cannot be drawn directly to the effectiveness of intervention programs. On the other hand, Wealth (8.56) presents the high score in recurrence. As the relationship suggests a more chronic relationship between Wealth and drug abuse, it is a direction that urgently needs to be focused on solving.

On another note, Le Chat in its Meta-Awareness Prompt response noted that “[t]he data on family and peer influence is compelling, but the interplay between these factors and other socioeconomic variables is not entirely clear, leading to some uncertainty in the assessment.” As this study concerns a factor over which each factor overlaps, some scores may be overlooked especially in a close race like in scores for Recurrence. Recognizing this blindspot and making further judgments are therefore required for these scores to be put into real world context, as much as the scores are a strong indicator.

In 2023, California passed a series of laws (namely SB43, SB326, and AB531) targeting the factors of mental health and infrastructure. The series of laws aligns with this research’s result, in which both poverty and mental health are in reach and urgency. However, the subsequent policy reversal reintroduces criminal penalties, suggesting a divergence between enforcement-driven policies and evidence-based prioritization.

Meta-Awareness

Looking back at the methods of this study, LLMs are heavily relied on to evaluate the qualities of quantitative data. Therefore, it is crucial to recognize the meta-awareness prompting as well as the results, as it is the LLMs' own reflections on the strengths and limitations of its evaluations. ChatGPT reported a "high level of confidence" for its evaluations, citing the "clarity and granularity of the data set" including "multiple overlapping indicators across time... and usage types." The limitations, on the other hand, ChatGPT acknowledged insufficiencies such as a "lack of trend continuity", "missing denominators or sample descriptions," and "no direct linkage to educational intervention impact (the meta-awareness is asked here after evaluating the factor Education)". Gemini, similarly, expressed high confidence but noted a need for "interpretive leaps" due to the implied rather than implicit connection between the factor and data. Le Chat raised concerns over absent temporal context, demographic detail, and methodological transparency. Future researchers should aim to use LLM in such matters that resolves the mentioned limitations to the best of technology availability.

As a conclusion, meta-awareness is exercised by explicitly identifying areas of ambiguity. The self-assessments underline the necessity of treating the evaluations as context-sensitive and probabilistic rather than definitive. The conclusions of this research remain robust within the scope of available data, but the precision and generalizability should be understood as limited by gaps and assumptions that were identified by the LLMs themselves.

General Limitations

As explained in Data Collection, the sources of this study were found through large databases that mainly consist of studies in the English language. Furthermore, a majority of the sources collected were based in the US. The results of this study are limited in regard to their generalizability, and future studies must consider incorporating studies and statistics from different language and geographical areas.

One of the major issues in compiling quantitative data and assigning it a qualitative score is that whether the quantity of data entries are equal or not, biases would appear in either case. In this study, where the quantity of entries vary from subject to subject, the Saliency bias where larger input blocks appear more "important" is structurally unavoidable. As a counter, we independently evaluate each factor and prompt such that the evaluation is made under the instruction to use only provided statistics without cross-factor comparison. Additionally, variability metrics along with repeated prompt cycles and cross model averaging were used to reduce and acknowledge the bias.

Circling back to the topic of human expert biases. As mentioned previously, the motivation of the AI evaluation method was to overcome the gap in quantitative methods, as well as to eliminate human bias when the topic is viewed qualitatively. However, LLMs still inherit systematic biases from human-fed training data, but that is qualitatively different from individualized bias of human experts. It is most important to acknowledge that this study does not claim to eliminate bias, but rather, to reduce the influence of human's normative framing through standardized prompting, repeated tests, and cross-model/architecture data gathering. Furthermore, this study utilizes a meta-awareness prompt to record any biases or inaccuracies that the LLM self-reflected on.

On a further note, we recognize that recent mechanistic interpretability work is using sparse autoencoders (SAEs) to steer internal features of LLMs. Generally, these SAE-based steering require access to model internals, which by the time of this study is yet to be available to us and to the general public. Knowing that the technology exists, it is a matter of time in which our limited ability to implement mechanistic bias suppression could be overcome and improved in future studies.

Future Directions

As future AI research technology advances, uses of SAE and other mechanisms to steer internal features will become key to eliminating biases resonated to human-fed training materials. Future researchers could experiment on whether selectively steering internal models as the one suggested in this study will alter evaluation outcomes in structured scoring frameworks.

However, even with improved mechanisms and technology, on the topic of qualitative, objectively viewing statistics of varying measures will remain complex. The broader question on whether language-based systems, operating through tokenization without reference to form can suffice as qualitative judgments will guide not just studies in the topic of drugs, but all future scientific evaluations.

Acknowledgements

Thank you Dr. B, Mr. McLean, and Mr. Collin for all your support, suggestions, and critique throughout this study. Thank you (in alphabetical order) Francis L., Jimmy H., Steven W., and the rest of ENG4U-B2 during the 2025 Spring semester for excellent guidance throughout the struggles and challenges that were met during the conducting of this paper.

References

- 1 *National Institute on Drug Abuse*, <https://nida.nih.gov/research-topics/covid-19-substance-use>.

-
- 2 D. Daley, M. Smith and D. Scharf, *Forgotten but not gone: The impact of the opioid epidemic and other substance use disorders on families and children*, <https://doi.org/10.15367/com.v20i2-3.189>.
 - 3 C. Ialongo, *Understanding the effect size and its measures*.
 - 4 X. Rêgo, M. Oliveira, C. Lameira and O. Cruz, *20 years of Portuguese drug policy—developments, challenges and the quest for human rights*.
 - 5 Y. Chen, H. Wang, K. Yu and R. Zhou, *Artificial intelligence methods in natural language processing: A comprehensive review*.
 - 6 Y. Li, Y. Gao, J. Jin, J. Nan, Y. Meng, M. Wang and C. Chen, *Adaptive weights-based relaxed broad learning system for imbalanced classification*.
 - 7 *Loneliness and Social Support Networks: Findings from the KFF Survey of Racism, Discrimination and Health - Findings - 10412 — KFF*.
 - 8 M. Kariisa, *Vital Signs: Drug Overdose Deaths, by Selected Sociodemographic and Social Determinants of Health Characteristics — 25 States and the District of Columbia, 2019–2020*, <https://doi.org/10.15585/mmwr.mm7129e2>.
 - 9 C. Jones, I. Barrera, S. Brothers, R. Ring and C. Wahlestedt, *Oxytocin and social functioning*, PMID: 28867943; PMCID: PMC5573563.
 - 10 R. A. Cree, *Health Care, Family, and Community Factors Associated with Mental, Behavioral, and Developmental Disorders and Poverty among Children Aged 2–8 Years — United States, 2016*, <https://doi.org/10.15585/mmwr.mm6750a1>, www.cdc.gov/mmwr/volumes/67/wr/mm6750a1.htm#suggestedcitation.
 - 11 S. Gabel, *Family Variables in Substance-Misusing Male Adolescents: The Importance of Maternal Disorder*, <https://doi.org/10.3109/00952999809001699>.
 - 12 M. Rizk, R. Herzog, S. Dugad and I. Galynker, *Suicide risk and addiction: The impact of alcohol and opioid use disorders*.
 - 13 N. Christie, *The role of social isolation in opioid addiction*.
 - 14 P. Albertín and L. Íñiguez, *Using drugs: The meaning of opiate substances and their consumption from the consumer perspective*.
 - 15 J. Ravitz and I. Hayes, *Inside the secret lives of functioning heroin addicts*.
 - 16 S. Bashirian, *Application of the Theory of Planned Behavior to Predict Drug Abuse Related Behaviors among Adolescents*.
 - 17 P. Chatterji, *Illicit Drug Use and Educational Attainment*, <https://doi.org/10.1002/hec.1085>.
 - 18 J. Neiderhiser, D. Reiss, P. Lichtenstein, E. Spotts, J. Ganiban and A. Pike, *Four factors for the initiation of substance use by young adulthood: A 10-year follow-up twin and sibling study of marital conflict, monitoring, siblings, and peers*.
 - 19 K. Fothergill, M. Ensminger, K. Green, R. Crum, J. Robertson and H. Juon, *The impact of early school behavior and educational achievement on adult drug use disorders: A prospective study*.
 - 20 T. Rakhshani, S. Kamranpoor, A. Kamyab, A. Yari and A. Khani Jeihooni, *The effect of an educational intervention in prevention of drug abuse in students*.
 - 21 A. Shafi, M. McDonough, B. Tofighi and D. Fiellin, *Synthetic opioids: A review and clinical update*.
 - 22 A. T. Dinwiddie, *Reported Non-Substance-Related Mental Health Disorders among Persons Who Died of Drug Overdose — United States, 2022*, <https://doi.org/10.15585/mmwr.mm7334a3>, www.cdc.gov/mmwr/volumes/73/wr/mm7334a3.htm.
 - 23 M. Czeisler, *Mental Health, Substance Use, and Suicidal Ideation during the COVID-19 Pandemic*, <https://doi.org/10.15585/mmwr.mm6932a1>, www.cdc.gov/mmwr/volumes/69/wr/mm6932a1.htm.
 - 24 F. Frieden, *A framework for public health action: the health impact pyramid*, <https://pubmed.ncbi.nlm.nih.gov/20167880/>.
 - 25 N. Volkow, M. Michaelides and R. Baler, *The Neuroscience of Drug Reward and Addiction*.