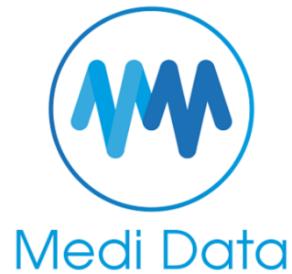


Data3001 Final Report



The Study of the Efficacy and Safety of Cannabis in Relieving the Symptoms of Arthritis

By **Medi-Data**

Heqing Huang (z5248161)

Jingyi Xu (z5249166)

Alexandria Yoo (z5308575)

Yimeng Zhang (z5212791)

Alan Lin (Rongcong Lin) (z5286173)

Medi-Data

Our team aims to study the effectiveness of cannabis as pain relief for arthritis treatment.

Our mission is to apply statistical techniques to medical assessment and ensure the public can access certified instructions for safe medical practice.

Executive Summary

The use of medical cannabis has been legalized in Australia since 2016. Patients seek to use this botanical medicine to treat a wide range of medical conditions. Although anecdotal evidence abounds, medical professionals frequently claim that they need more scientific evidence to feel comfortable prescribing cannabis. For this project, the app, Strainprint™, was used to collect arthritis patient's daily medication status (including symptoms at the time, type of medication, and reaction after taking the medicine). The purpose of the study was to determine the efficacy and safety of marijuana for relieving symptoms of arthritis. The raw data comprised of 58% females and a population where 52% of patients were older than 45. 391 patients with a 12-week medication history were considered in this project. The most frequently reported symptoms were physical pain, insomnia, inflammation, joint stiffness, and anxiety. By comparing the patient's self-reported pre- and post-medication severity, we observed a successful relief in symptoms for all arthritis patients, with an average severity reduction of 3 levels. In addition, the severity of symptoms and the type of medication ingested was found to have had a significant impact on the efficacy of the cannabis drug. There are also several side effects associated with taking cannabis drugs, the most common of which are thirst and sleepiness. Future clinical studies would further consolidate our findings for this project.

Background

Arthritis is a condition that affects the normal function of muscles, bones, and joints and usually causes pain, swelling, stiffness, fatigue, and an overall limitation of a person's mobility (NPSMEDIINEWISE, n.d.). Areas generally afflicted by arthritis include the hands, knees, hips, and neck to name a few. To Arthritis direct report in 2021, Arthritis affects nearly 4 million Australians, affecting people of all ages, ethnicities, and genders. Although it is not an age-related disease, arthritis affects women more often than males and some forms are more common in older individuals than in younger people. While certain symptoms may continue to be present for many years, others may change over time.

Despite the current countermeasures held in place to slow down and minimise the symptoms and occurrence of arthritis, there is yet to be a cure for arthritis. With the number of people diagnosed with the condition growing ever so drastically, the need for further study and research is vital to better address this medical condition.

One main area of interest, as of recently, is the use of medicinal cannabis for the treatment of pain in arthritis patients. Recent studies suggest promising results on the efficacy of cannabis in treating chronic pain for arthritis in animals (Hammell et al. 2015). However, there is still a lack of evidence from human clinical trials. Medical cannabis is cannabis that is specifically used for

medical purposes and for that purpose only. In Australia, cannabis is an illegal drug that has the potential to pose serious health hazards when not taken with caution. For this reason, there is still a lot of apprehension held towards cannabis use without compromising the safety of patients on cannabis. Cannabinoids work in the body by affecting the normal functioning of the central nervous system. Without the proper communication between nerve receptors in areas such as the brain, this may result in detrimental psychological effects such as delusions, hallucinations, and lost sense of self. Through this project, and based on real patient data of Arthritis patients, Medi-Data aims to identify the suitability of cannabis according to the efficacy and safety of the drug. With significant literature review of previous works on similar topics, our team plans to further broaden current studies of medical cannabis use. The importance of our work exists in supporting the aptness of cannabis in relieving the symptoms experienced in arthritis patients. From this project, the final results of the project will be used as evidence that is either for or against the use of cannabis in the medical field.

One of the first objectives our team had set for this task was to study the characteristics of the data provided in the Strainprint_Arthritis_data.xlsx file. This includes, but is not limited to, the proportion of the different age groups involved, the distribution of each gender, the number of individuals categorised under each symptom type as well as the frequency of each class of data, just to name a few. After a thorough initial analysis, we then shifted our attention to our main objective for the project, which was to analyse the safety and efficacy of cannabis as a medicinal product (De Briyne et al. 2021). In addition, we then furthered our study by identifying relationships present within the dataset.

This research paper has employed 2464 patients' data collected from Strainprint™ concerning the efficacy, safety, and product type information to capture the potential impact medical cannabis has based on pre-recorded patient symptoms. As a team, Medi-Data firmly believes that a systemic and well-run assessment has the potential to build and support a positive outlook on cannabis in arthritis patients.

Scope

At the end of this project, Medi-Data aims to deliver a detailed report of the efficacy and safety of cannabis as an effective medicinal based on thorough analysis and prior research on topics of similar context.

Assumptions considered in advance of the project are as follows:

- Patient records in the Strainprint_Arthritis_data.xlsx file have been drawn randomly from the StrainprintTM database
- Patients are recording their symptoms honestly and under the instructions provided in the StrainprintTM app
- Dosage units registered under inhalations, sprays, and fingertips are consistent across all patients
- Patients are not experiencing any other medical conditions that may potentially affect the extent to which they experience their symptoms of arthritis

- No interaction presents between any other medications with the cannabis taken throughout patients' participation in the StrainprintTM app

However, upon observing the given dataset, we can see patients have been relatively inconsistent in their intake of cannabis. To comply with a fair and effective study and with reference to the, Cahill et al (2021) article, our team deduced 12 weeks to be the ideal time frame to consider for this project. This was given the fact that the above research paper indicated patients' pain to not improve significantly over the course of 6 weeks.

To assess the efficacy of medicinal cannabis, our team first divided the problem into smaller sub-categories. This involved a careful consideration of the ingestion method, symptoms, strain type and dosage of cannabis received by each arthritis patient. For the ingestion component, our team examined the ingestion method that had the highest weighted efficacy for each strain type. This was completed via a set criterion of comparing the average weighted efficacy of ingestion method for each strain type. For the symptom's component, we first deduced the symptoms that benefited the most from cannabis intake. The criteria for this also being based on the proportion of positive weighted efficacy scores. We then established the most useful strain type for each symptom, with such being based on the highest average weighted efficacy of a strain type for each symptom. To determine whether the post-medication score was significantly lower than the pre-medication score, we made use of the Wilcoxon hypothesis test and set as the critical point.

Assuming the severity of symptoms to coincide with pre-medication scores, we further determined the ingestion methods best suited to patients in accordance with their symptom severity. We started off by classifying pre-medication scores into the following three categories, with each describing a different degree of severity: 'minor', 'moderate' and 'severe'. We observed patterns present within each severity subgroup with products of type 'high THC', 'high CBD' and 'Balanced' and used this information to denote the preferred cannabis chemotypes as well as ingestion method for each severity subgroup. This was in addition, supported by the use of 95% confidence intervals of weighted efficacies for the top three ingestion methods. Moreover, the top three most effective cannabis strain type was further used to aid a rigorous assessment of the efficacy. Dosage was also considered in our study of efficacy through scatter plot visualisations of weighted efficacy vs. dosage measure and by superimposing fitted lines constructed via additive regression modelling.

To assess the safety, the top five side effects based on the whole dataset were considered. In addition to this, our team determined the top three positive, negative, and neutral side effects according to the sub-categories of age, gender as well as for the complete dataset. We then established our own safety scores to correspond mathematically to the positive emotion score minus the negative emotion scores (treating neutral emotions to a 0 score). A safety score was obtained for each patient record and therefore used as a key criterion for assessing the safety of medicinal cannabis.

Consistency of results irrespective of time was further studied as a part of additional research. This was conducted via an initial grouping of data according to age. We then filter out patients whose records are less than 3, divide the whole data set into 3 periods of times of taking cannabis equally. This was to ensure no uneven data was generated. Using line plot and paired t-test to check the tendency and difference in average weighted efficacy and safety scores between each

period, considering 5% significance level and corresponding degrees of freedom, we were able to compare each t-stats result with critical value and having results whether the duration between period is stable with cannabis treatment.

The success of this project will be assessed based on three aspects. In the research analysis and communication aspect, validated assessment tools will be used to judge both efficacy and safety. Secondly, statistical inferences, in addition to tables and figures, will further enhance our findings in an easy-to-follow visual format. In the discussions section, results from the analysis will be equitably compared with previous studies and results. Lastly, for the conclusion, insights from the study will be provided to further benefit future clinical trials and medical practice.

Plan

Approach

Stage 1: Data Cleaning and literature research:

By conducting this stage, we aim to have a clear view about data composing and decide the direction of team's research. Medi-data has done several things on this stage:

- Conducting exploratory analysis on the dataset to give the team a brief understanding on the problem.
- Formatting data to have a correct and easy to analysis form
- Removing redundant value and extracting a 12-week dataset which includes records of patient whose treatment period has equal to 12 weeks
- Searching feasible method for further analysis
- Researching literature about cannabinoids (THC, CBD) and other useful papers

Stage 2: Data analysis and modelling

- Classify patient characteristics for a detailed view of the population composition
- Based on the patient characteristic, use mean methods to conduct an analysis for the efficacy and safety of cannabis
- Figure out a method to calculate the safety score, which counts the difference between positive emotive and negative emotive.
- Sorting the side effects to gain a rank of side effects
- Build linear models and decision trees to analyse
- Use t-test and line plot to analyze sustainability of cannabis

Stage 3: Visualisation:

- Use Power BI to generate visualisation
- Find suitable graphs to represents each measurement
- Remove some blank data to give clearer view

Stage 4: Finalising the findings

- Determine population characteristics
- Based on the set measurements of scale, draw conclusions on the efficacy and safety of cannabis

- Summarise the factors affecting the weighted efficacy and safety of cannabis
- Provide recommendations regarding the safe and effective use of cannabis

Comparison on initial plan

After Medi-Data approaching problems, difference between current plan and original plan exists:

- Additional analyses were added to support further study, like sustainability
- Method to the problem were detailed.
- Adding model part to the project.

Apart from these minor changes, team's plan was remained consistent and pivot through the project.

Unexpected Issues

1) Measurements of safety:

In the data analysis period, Medi-Data has encountered the problem of defining safety and measuring it, our team has a thought of using POQ-SF method to measure the safety score at the first try, however, figure out this method may not be suitable for our project. After consideration and exploration on the data, our team found the emotive is used to count the number of side effects, therefore, Medi-Data has decided a particular way of calculating the safety score, which is to calculate the difference between positive emotive and negative emotive to produce a safety score for each record.

2) Different units on cannabis consumption:

When conducting Dosage vs. Effectiveness Analysis, Medi-Data has found the dosage units are different from mg/ml to drops, inhalation and so on. To avoid unit effects, we plotted only the most effective mode of intake for each symptom with respect to dose. (e.g., the unit of dosage is the same for each method of ingestion)

Findings

1. Demographics

In the original dataset, 2464 patients participated in filling out their reports of cannabis use in which 391 patients who provided 12 weeks' worth of records were selected as the analytical sample. The age of patients within this sample ranged from 20 to 77 years old. The proportion of women in the sample was 58.8%, evidently higher than male participants. In terms of age distribution, there are obvious features of aging, where more than 50% of patients were 45 years old and over (Table 1.1).

Table 1.1 | The characteristics of the arthritis patient cohort.

Patient Characteristics	n(%)
Age	(Mean age: 45.7 years)
18 - 24	6(1.5)
25 - 34	65(16.6)
35 - 44	116(29.7)
45 - 54	108(27.6)
55+	96(24.6)
Biological Sex	
Female	230(58.8)
Male	158(40.4)
Unknown	3(0.8)

In addition, we observed patients to have also provided their symptoms of arthritis prior to having had cannabis. Pain from arthritis was the most significant reaction, experienced by approximately 95 percent of patients, and was equally common in both males and females (1:1.02) (Table 1.2). It was also worth noting that from the results of our statistical analysis, women were found to be more sensitive to the discomforts of arthritis than men. The most prominent symptom being that of joint stiffness, where the ratio of gender between male and female is 1:1.18 respectively.

Table 1.2 | The information of each symptom in gender.

Mental Disorders	Joint Stiffness
250 (63.94%)	223 (57.0%)
Female : Male	Female : Male
1.16 : 1	1.18 : 1
Inflammation	Insomnia
241 (61.6%)	249 (63.7%)
Female : Male	Female : Male
1.05 : 1	1.09 : 1
Pain	
369 (94.4%)	
Female : Male	
1.02 : 1	

Instructions: n (%) under the symptom refers to the number of patients who had the condition, each patients can record multi-symptoms. The ratio between Female and Male is the comparison between the percentage of the symptom in female and in male.

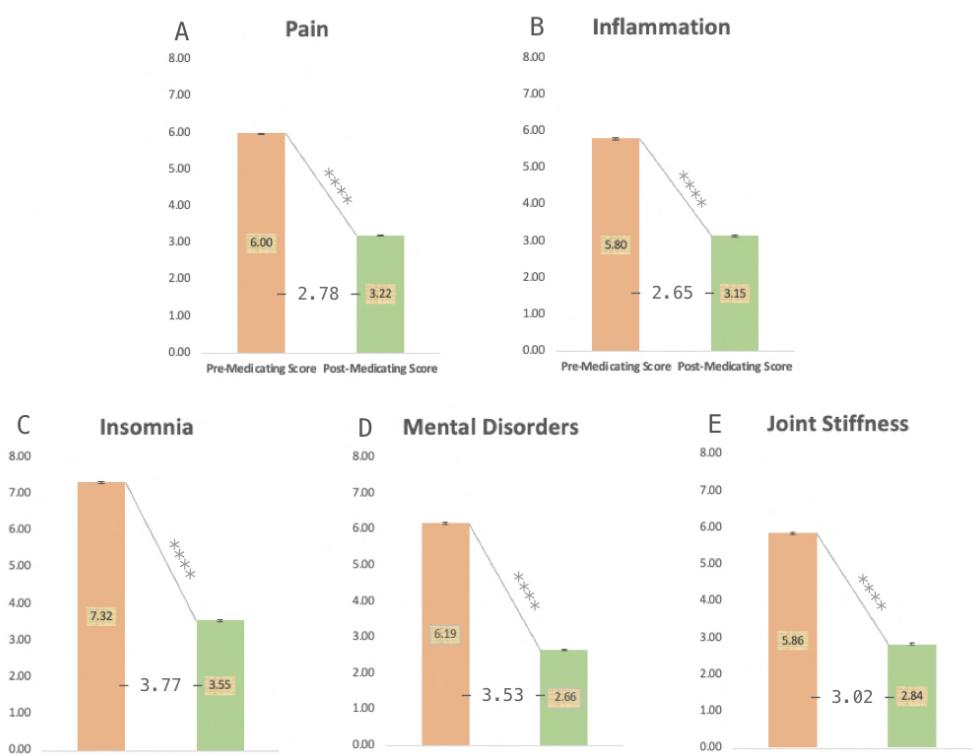
2. Efficacy Results

In pharmacology, efficacy is about the maximum response achieved with a drug in research settings. To assess the efficacy of cannabis-based medicines in the treatment of arthritis-related symptoms comprehensively. Strainprint™ introduces a robust efficacy assessment tool called Weighted Efficacy, the detailed explanation is provided in the Appendix. The analysis consists of four parts including a Wilcoxon test between pre-medicating score and post-medicating score across symptoms, the variation of weighted efficacy in gender and age groups across symptoms, the variation of weighted efficacy in ingestion methods across symptoms, and a simulation of doses on weighted efficacy under the generalized additive model.

2.1. Aim1: Overall change in medicating score

There was a significant reduction in weighted efficacy scores for all patients with symptoms, meaning that the symptoms had been alleviated. The most significant reduction in scores was for insomnia, followed closely by mental disorders.

Figure 2.1 | The comparison between pre- and post- medicating score for arthritis-related symptoms (and Wilcoxon Test)



(****P < 0.0001)

Instruction: Pre-medicating Score and Post-medicating Score are patients' self-reported severity of their symptoms before and after taking the cannabis-based medicines(CBM). The number labeled on the bar refers to the average score, and the tiny symbol positioned on the top of the bar refers to standard error. Wilcoxon Test is used to determine whether a significant change is happened after taking CBM.

2.2.1 Pain

As depicted in Figure 2.1 (panel A), results revealed a significant reduction in symptoms of pain following cannabis use (Wilcoxon test $P < 0.0001$). Based on patient-provided responses to cannabis treatment, 36.8% of them felt “comfortable” with their daily functioning, 15.4% reported “less aware of pain”, and 13.2% reported “free of pain”. Furthermore, cannabis also made them have

some improvements in mental conditions, nearly half of patients reported being relaxed and 22% felt happy.

2.2.2 Inflammation

As shown in Figure 2.1 (panel B), patients reported a significant reduction in inflammation after using cannabis (Wilcoxon Test $P < 0.0001$). Besides, more than half of the post-ingestion records indicated that patient had a “relaxed” feeling, and nearly 35% responded with a “comfortable” reaction, and 26% of them feel “pain free” or “less aware of pain”.

2.2.3 Insomnia

Figure 2.1 (panel C) shows that there exists a significant reduction in symptoms of insomnia following cannabis use (Wilcoxon test $P < 0.0001$), which indicates majority of patients with insomnia have improved sleep problems. Moreover, 27.62% of insomniacs reported that “sleepy” was a reaction after taking the medicine.

2.2.4 Mental Disorders

Arthritis patients also experience significant reduction in psychiatric symptoms after using cannabis (see Figure 2.1, panel D). Further, after treatment with cannabis medication, no patients reported a “depression” reaction and 98.7% of patients’ “anxiety” disappeared, with a further 22% of them feeling ‘happy’.

2.2.5 Joint Stiffness

As depicted in Figure 2.1 (panel E), results indicated there was a significant decrease in symptoms of joint stiffness (Wilcoxon test $P < 0.0001$). In addition, nearly half of the patients felt “relaxed” after the treatment, and more than one-third of them noticed an improvement in their day-to-day functions. Moreover, 30% of them feel “pain free” or “less aware of pain”.

2.2. Aim2: effects of gender and age group on weighted efficacy

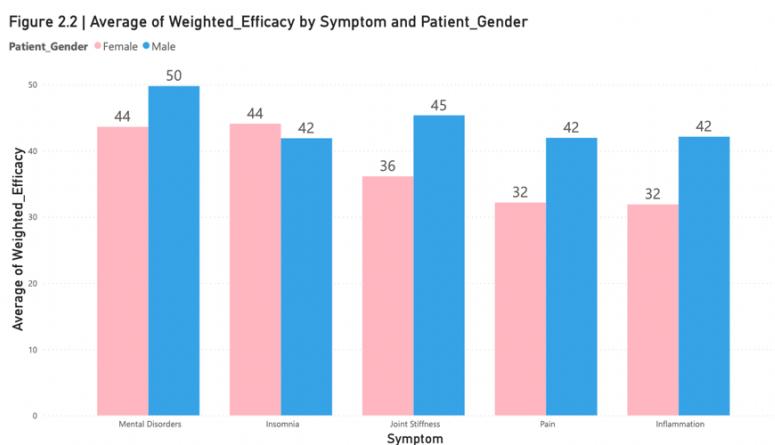
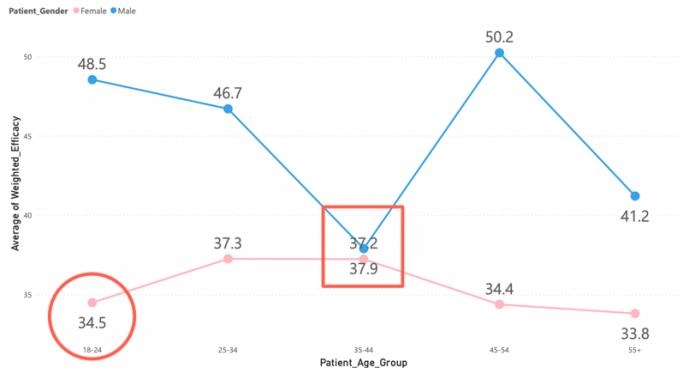


Figure 2.2 and Figure 2.3 show the average of weighted efficacy through different symptom of male and female. Male experience better in most symptom than females except Insomnia. The weighted efficacy of females was ranged from 33.8 to 37.2 (Mean = 35.44, SD = 1.67), and there was no significant difference in the efficacy of cannabis for different age groups. The weighted efficacy distribution for males ranged from 37.9 to 50.2 (Mean = 44.9, SD = 5.17), with men in the 35-44 year age group benefiting the least and men in the 45-54 year age group benefiting the most.

Figure 2.3 | Average of Weighted_Efficacy by Patient_Age_Group and Patient_Gender

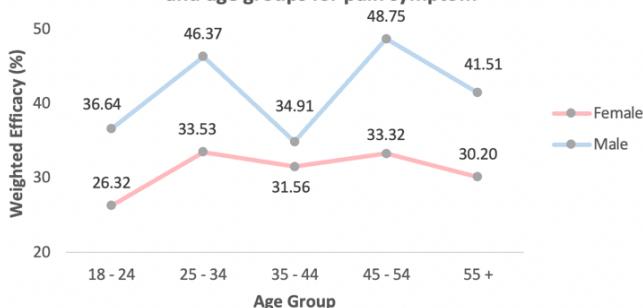


The following is our analysis of the effectiveness of the drug in male and female patients with various symptoms.

2.3.1 Pain

In the men population, age had a significantly stronger effect on cannabis in reducing pain symptoms than in women. More specifically, except for young women aged 18-24 who received a weighted score of less than 30, women patients of all ages had a weighted score of around 32. While male patients in both the 25-34 and 45-54 age groups perceived a significant greater reduction in pain symptoms compared to other age groups (see Figure 2.4).

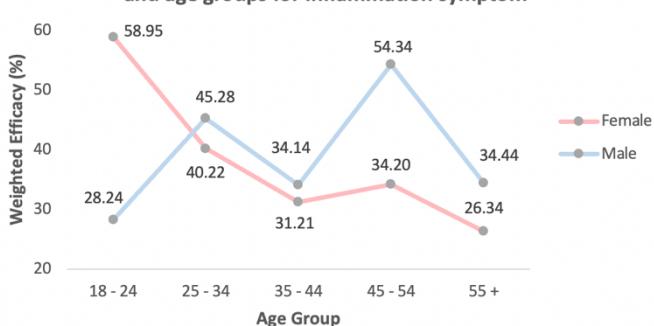
Figure 2.4: The change of weighted efficacy in gender and age groups for pain symptom



2.3.2 Inflammation

The weighted efficacy of cannabis medicines on inflammation symptom was highly associated with the patient's age and gender. The efficacy of cannabis in female patients decreases with age, from almost 60% in young age to 26% in elder age (Figure 2.5). In terms of the effectiveness in the male cohort, patients aged 45 to 54 perceived a greater reduction in inflammation compared to other age groups.

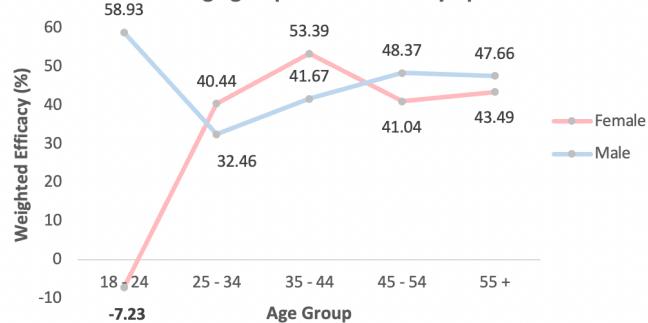
Figure 2.5: The change of weighted efficacy in gender and age groups for inflammation symptom



2.3.3 Insomnia

As shown in Figure 2.6, use of cannabis for insomnia in young women aged 18-24 years can instead worsen symptoms (-7.23 weighted efficacy). In contrast, young males in the same age group were highly satisfied with the cannabis treatment, with an average weighted efficacy of 59%, indicating a significant improvement in insomnia.

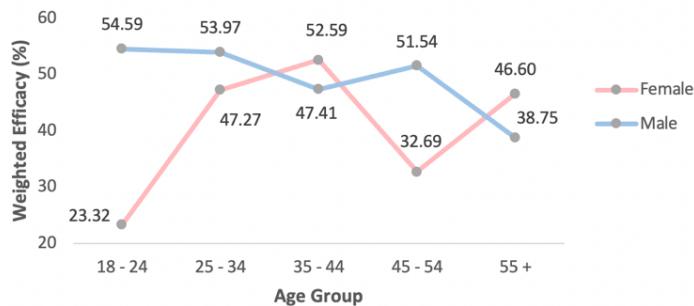
Figure 2.6: The change of weighted efficacy in gender and age groups for insomnia symptom



2.3.4. Mental Disorders

There was no significant difference in the efficacy of cannabis in treating mental disorders in men aged 18-55 years, while men over 55 years benefiting less than other age groups. Weighted efficacy scores for women fluctuated more between ages. Women aged 18-24 benefited least, with an efficacy score of 23%, but females aged 35-44 with mental disorders benefited more than twice as much (Figure 2.7).

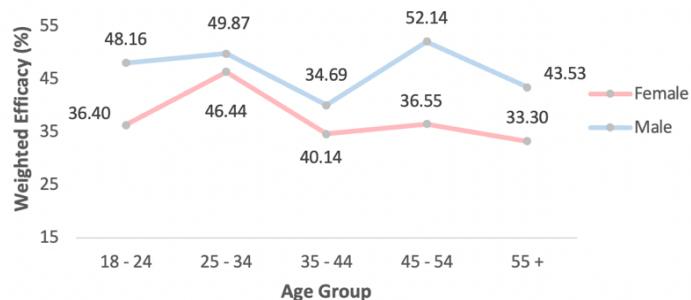
Figure 2.7: The change of weighted efficacy in gender and age groups for mental disorders



2.3.5 Joint Stiffness

As depicted in Figure 2.8, men have a higher weighted efficacy score than women across all age groups. The most remarkable gap occurred in the age group 45 to 54-year-olds, where men had an average weighted efficacy score of 52.14% compared to 36.55% for women.

Figure 2.8: The change of weighted efficacy in gender and age groups for joint stiffness



2.3. Aim3: effects of ingestion method and THC&CBD on weighted efficacy

- The severity of the patient's symptoms is classified according to the premedication score: 1-3 for mild, 4-6 for moderate and 7-10 for severe.
- The ingestion methods are classified into three categories: oral, Smoke, and external use. The specific classification is as follows:
 - Oral: edible, oral, pill, tincture, oil
 - Smoke: dab bubbler, dab portable, smoke, vape, concentrate
 - External use: transdermal, topical, spray

2.4.1 Pain

For all levels of pain, the 'smoke' intake approach performed better than the other approaches.

In terms of cannabis chemistry type, high CBD medicines dominated patient choice for mild and moderate pain with (95% confidence interval on weighted efficacy [CI]: 33.48 to 37.26) and (95% confidence interval on weighted efficacy [CI]: 37.90 to 39.86) respectively, but high concentrations of THC gave better relief for patients with severe pain with (95% confidence interval on weighted efficacy [CI]: 45.80 to 46.92) (Table 3.1).

Table 3.1 | Information on preferred cannabis chemotypes for the optimal ingestion method on different pre-medicating scores to treat a pain symptom.

Pre-medicating Score	Ingestion Method	Cannabis Chemotype	95% CI
[1 - 3] Mild (3234 observations)	Smoke	High THC	[28.14, 31.46]
		High CBD	[33.48, 37.26]
		Balanced	[25.50, 34.78]
[4 - 6] Moderate (14895 observations)	Smoke	High THC	[32.91, 34.01]
		High CBD	[37.90, 39.86]
		Balanced	[28.75, 31.91]
[7 - 10] Severe (12737 observations)	Smoke	High THC	[45.80, 46.92]
		High CBD	[43.51, 46.21]
		Balanced	[36.85, 39.85]

Instructions: 95%CI refers to such chemotype medicines has 95% confidence will achieve the weighted efficacy within this interval under the optimal ingestion method and the severity of the symptom. High THC: the content of thc in cannabinoid $\geq 60\%$, High CBD: the content of cbd in cannabinoid $\geq 60\%$, Balanced: the content of thc in cannabinoid $< 60\%$ and $> 40\%$.

2.4.2 Inflammation

The result show that “oral” intake method is more suitable for patients with mild inflammation, while the “smoke” intake method is more effective in treating patients with moderate to severe inflammation (grade 4 and above) (Table 3.2). Furthermore, higher CBD chemotypes provided better results in treatment for all levels of inflammation.

Table 3.2 | Information on preferred cannabis chemotypes for the optimal ingestion method on different pre-medicating scores to treat an inflammation symptom.

Pre-medicating Score	Ingestion Method	Cannabis Chemotype	95% CI
[1 - 3] Mild (969 observations)	Oral	High THC	[28.36, 48.18]
		High CBD	[29.22, 34.48]
		Balanced	[26.58, 45.50]
[4 - 6] Moderate (3484 observations)	Smoke	High THC	[33.90, 36.30]
		High CBD	[35.74, 39.38]
		Balanced	[23.97, 28.83]
[7 - 10] Severe (2669 observations)	Smoke	High THC	[43.20, 45.66]
		High CBD	[42.82, 48.02]
		Balanced	[30.70, 35.78]

Instructions: 95%CI refers to such chemotype medicines has 95% confidence will achieve the weighted efficacy within this interval under the optimal ingestion method and the severity of the symptom. High THC: the content of thc in cannabinoid $\geq 60\%$, High CBD: the content of cbd in cannabinoid $\geq 60\%$, Balanced: the content of thc in cannabinoid $< 60\%$ and $> 40\%$.

2.4.3 Insomnia

Patients with mild insomnia are not advised to take the medicine, as 43% of observations show that consuming the cannabis leads to more severe insomnia. Moderate patients who ingest cannabis by "smoking" it may be the best option for symptom relief, but severe patients do best with oral medication for insomnia (see Table 3.3).

Based on these patients' experiences, THC-dominated drugs are more effective for patients with moderate and severe insomnia.

Table 3.3 | Information on preferred cannabis chemotypes for the optimal ingestion method on different pre-medicating scores to treat an insomnia symptom.

Pre-medicating Score	Ingestion Method	Cannabis Chemotype	95% CI
[1 - 3] Mild (107 observations)		No optimal ingestion method and any chemotype medicines since 47 (43.93%) observations indicates the symptom goes worse after taking cannabis based medicines	
[4 - 6] Moderate (1108 observations)	Smoke	High THC High CBD Balanced	[22.49, 28.13] [5.22, 29.36] [17.62, 47.30]
[7 - 10] Severe (3076 observations)	Oral	High THC High CBD Balanced	[56.15, 60.09] [43.62, 52.78] [54.08, 58.76]

Instructions: 95%CI refers to such chemotype medicines has 95% confidence will achieve the weighted efficacy within this interval under the optimal ingestion method and the severity of the symptom. High THC: the content of the in cannabinoid $\geq 60\%$, High CBD: the content of cbd in cannabinoid $\geq 60\%$, Balanced: the content of thc in cannabinoid $< 60\%$ and $> 40\%$.

2.4.4 Mental Disorders

For the treatment of different levels of mental disorders, although the best intake method is consistent, being "smoke", but the recommended chemical types of medicines are different. As the Table 3.4 shows, THC-dominant medicine is suitable for mild symptoms, high CBD medicines are more effective for moderate symptoms, and medicines with a balanced THC to CBD ratio are most effective in relieving severe disorders.

Table 3.4 | Information on preferred cannabis chemotypes for the optimal ingestion method on different pre-medicating scores to treat mental disorders.

Pre-medicating Score	Ingestion Method	Cannabis Chemotype	95% CI
[1 - 3] Mild (925 observations)	Smoke	High THC High CBD Balanced	[34.03, 39.95] [32.32, 49.02] [25.32, 54.90]
[4 - 6] Moderate (3497 observations)	Smoke	High THC High CBD Balanced	[44.43, 46.47] [56.70, 61.08] [29.82, 36.04]
[7 - 10] Severe (3586 observations)	Smoke	High THC High CBD Balanced	[49.29, 51.33] [55.61, 61.29] [55.07, 66.95]

Instructions: 95%CI refers to such chemotype medicines has 95% confidence will achieve the weighted efficacy within this interval under the optimal ingestion method and the severity of the symptom. High THC: the content of the in cannabinoid $\geq 60\%$, High CBD: the content of cbd in cannabinoid $\geq 60\%$, Balanced: the content of thc in cannabinoid $< 60\%$ and $> 40\%$.

2.4.5 Joint Stiffness

In patients with joint stiffness, “smoke” performs better than other intake methods in all degrees of symptom. In terms of cannabis compounds, products contain similar amounts of CBD and THC work better for patients with mild symptoms (95% confidence interval on weighted efficacy [CI]: 31.97 to 50.31). For patients with moderate to severe symptoms, high CBD levels are more recommended, and joint stiffness of greater severity responded better to this chemical type of product with (95% confidence interval on weighted efficacy [CI]: 41.61 to 44.59) and (95% confidence interval on weighted efficacy [CI]: 51.32 to 56.54) (Table 3.5).

Table 3.5 | Information on preferred cannabis chemotypes for the optimal ingestion method on different pre-medicating scores to treat a joint stiffness symptom.

Pre-medicating Score	Ingestion Method	Cannabis Chemotype	95% CI
[1 - 3] Mild (713 observations)	Smoke	High THC High CBD Balanced	[29.26, 34.76] [29.29, 37.47] [31.97, 50.31]
[4 - 6] Moderate (3566 observations)	Smoke	High THC High CBD Balanced	[41.01, 43.31] [41.61, 44.59] [29.83, 34.80]
[7 - 10] Severe (2622 observations)	Smoke	High THC High CBD Balanced	[44.75, 47.43] [51.32, 56.54] [35.55, 40.37]

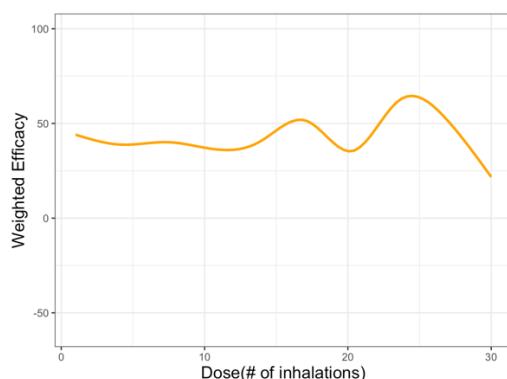
Instructions: 95%CI refers to such chemotype medicines has 95% confidence will achieve the weighted efficacy within this interval under the optimal ingestion method and the severity of the symptom. High THC: the content of thc in cannabinoid >= 60%, High CBD: the content of cbd in cannabinoid >= 60%, Balanced: the content of thc in cannabinoid < 60% and > 40%.

2.4. Aim4: effects of dose on change in weighted efficacy

Scatter plots of 'weighted efficacy' and 'dosage measure' were plotted and fitted lines were superimposed using a generalized additive model method. The relationship between the dose of the optimal intake method and the effectiveness of cannabis for each symptom was analyzed according to the graphs.

2.5.1 Pain

Figure 4.1: Pain patients use smoke for ingestion



For patients with various levels of pain, 24-25 inhalations were considered the most effective dose to “smoke” cannabis, with a weighted efficacy greater than 50% (Figure 4.1). In comparison, a dose of 30 inhalations was considerably less effective than any other dose.

2.5.2 Inflammation

Figure 4.2 - A

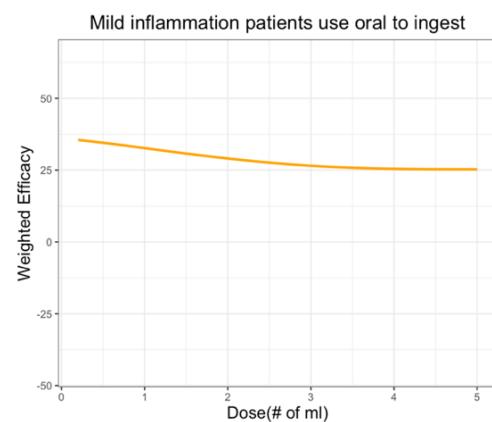


Figure 4.2 - B



Dosage was not a significant factor affecting cannabis effectiveness, either for mild inflammation patients using oral cannabis or for moderate to severe patients using smoked cannabis, but higher doses can instead reduce medicine effectiveness (e.g. 5 ml oral product, 30 ml smoke product) (Figure 4.2).

2.5.3 Insomnia

Figure 4.3 - A

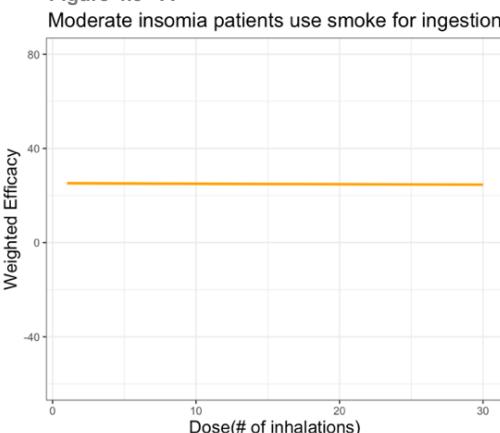
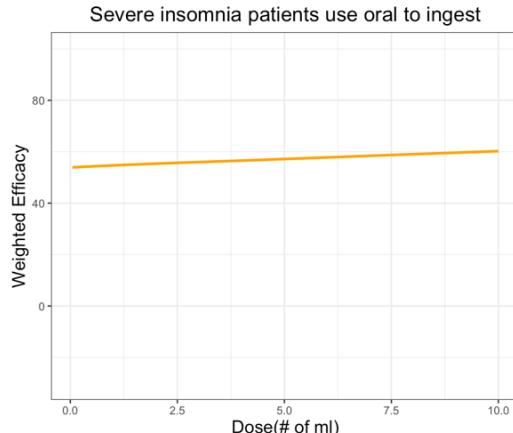


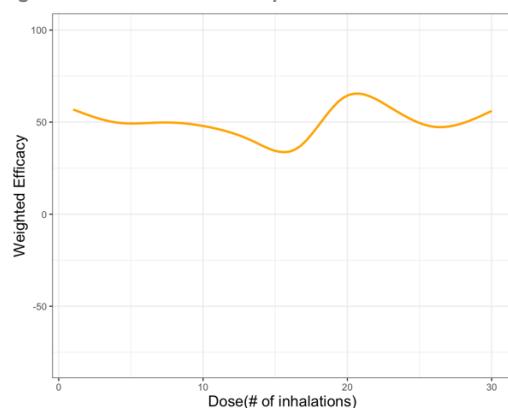
Figure 4.3 - B



For moderate insomnia patients using “smoke” ingested cannabis, the change in dose had no effect on the effectiveness of the medicine, i.e. 30 inhalations had the same effect as 1 inhalation. Similarly, for patients with severe insomnia using oral cannabis, the effect of dose variation on efficacy was still insignificant (Figure 4.3).

2.5.4 Mental Disorders

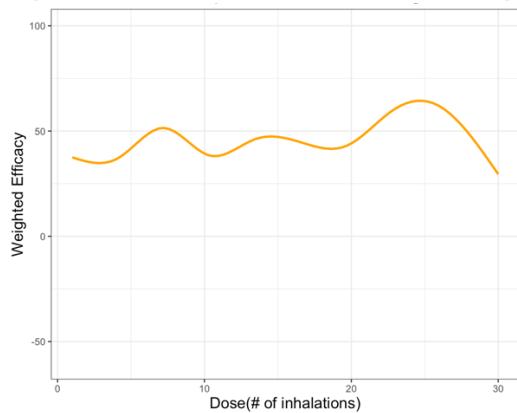
Figure 4.4: Mental disorders patients use smoke for ingestion



Ingestion of different doses of cannabis through 'smoke' has different effects on the alleviation of mental disorders. 16 inhalations were considered less effective than any of the other doses; 20 inhalations are most effective in reducing anxiety and depression (Figure 4.4).

2.5.5 Joint Stiffness

Figure 4.5: Joint stiffness patients use smoke for ingestion

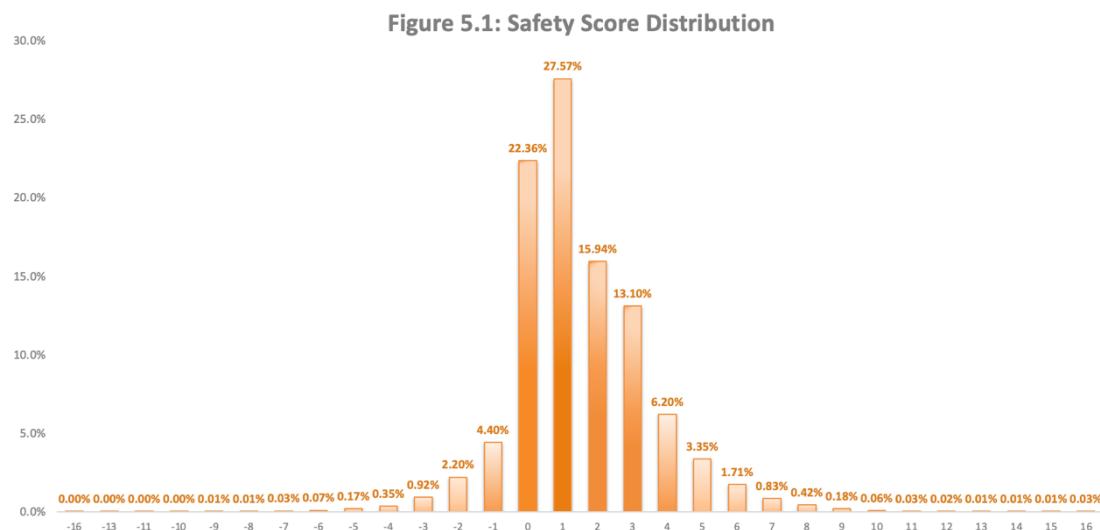


For patients with joint stiffness who ingested cannabis using "smoke", medicine effectiveness tended to increase with higher doses. Patients with 25 inhalations reported weighted efficacy values above 50% (Figure 4.5).

Cannabis-based medicines do alleviate arthritis-related symptoms supported by the analysis of medication efficacy. Firstly, the severity of symptoms can be reduced by 3 levels on average after taking such medicines. Drugs tend to be more effective in males than females, and young females aged 18 to 24 are less adaptable to the drug than others. Besides, Smoke is relatively more effective than oral and external use in the treatment of symptoms. In terms of cannabis chemotype, High CBD is the primary choice for inflammation and High THC is the optimal type for insomnia. Another practical finding is that the change in dose does not variate the medication efficacy.

3. Safety Results

In pharmacology, safety describes the toxicity of the drug. Are they some moderate side effects that go away after a while or do adverse side effects continue and cause more harm than benefits? To assess the safety of cannabis-based medicines in the treatment of arthritis-related symptoms comprehensively. Medi-data introduces Safety Score criteria based on the features of the dataset, the detailed explanation is provided in the Appendix 2. The study consists of four parts including the most frequent side effects after taking cannabis, the distribution of safety score, the variation of safety score in ingestion methods and times, and the variation of safety score in gender and age groups.



80% of the safety scores were between 0 and 3, with the frequency ranging from 1 (27.57%), 0 (22.36%), 2 (15.94%) and 3 (13.10%) (Figure 5.1), indicating that the positive response of cannabis drugs to most patients is stronger than the negative response.

3.1. Aim1: Top Side-effect

Table 5.2 Most frequent side effects			
Moderate Side Effects	n (%)	Adverse Side Effects	n (%)
Thirsty	8920 (15.6%)	Headache	535 (0.9%)
Sleepy	5214 (9.1%)	Anxious	405 (0.7%)
Hungry	3987 (7.0%)	Nauseous	259 (0.5%)
Red Eyes	2004 (3.5%)	Racing Heart	197 (0.4%)
Couchlocked	1408 (2.5%)	Dizzy	137 (0.2%)

Cannabis-based medications have caused several side effects in patients while treating arthritis and the resulting symptoms. A count of 39 different post-medication reactions provided by patients showed that thirsty was the most common, with a probability of more than 15%. Other moderate side reactions including sleepiness, hunger, red eyes, and couch lock did not exceed 10% (Table 5.2 - A).

The adverse side effects that occur most frequently include headache, anxiety, nausea, rapid heartbeat, and dizziness. Although the probability of any of these reactions is less than 1% (Table 5.2 - B), it is still worth noting because some side effects can be life-threatening if they occur repeatedly or last for a long time (e.g., racing heart).

3.2. Aim2: effects of ingestion methods and taking time on safety score

Table 5.3 Probability of safety score < 0 after taking medicine with different intake methods and time					
	Overall	Morning	Afternoon	Evening	
Smoke	8.6%	8.8%	6.4%	7.1%	11.7%
Oral	7.4%	6.5%	5.1%	6.8%	11.4%
External Use	7.3%	5.2%	5.6%	11.4%	6.7%

Table 5.3 shows that external use of cannabis has the lowest probability of receiving a negative safety score (7.3%) and was relatively safer than smoke use and oral use.

Furthermore, by comparing the cannabis consume on different time of the day, the evidence suggests that evening/overnight use is 5-6% more likely to receive a negative score than morning and afternoon use.

3.3. Aim3: effects of gender and age group on safety score

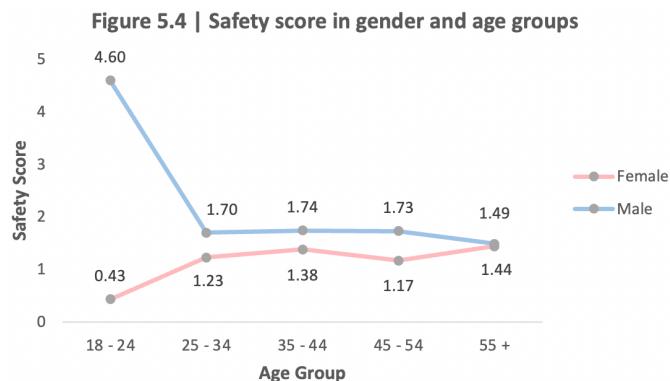


Figure 5.4 illustrates the relationship between average safety scores through age and gender. At any age group, average safety scores are higher for men than women, with a significant difference in the 18-24 age group, where men scored as high as 4.6, while women received only 0.43. However, in the 55+ age group, the difference in safety scores was almost non-existent.

Cannabis-based medicines create more benefits than harm in the treatment of arthritis-related symptoms supported by the safety score analysis. Around 80 percent of records indicate that one or two more positive reactions were raised than the number of side effects after taking medicine. The most common side effects are thirsty and sleepy, some adverse effects also are reported such as a racing heart (0.5%). The administration routes have an impact on medication safety, and the side effects of External use are relatively mild compared with Oral and Smoke. In addition, taking such medicines in the early of the day (morning, afternoon) is safer than late (evening and overnight).

4. Sustainability

The sustainability of medicine can be defined as whether this kind of medicine is able to maintain efficacy and safety over a period. The aim of this topic is to find: For each age group, how is sustainability changed?

4.1 Approach to find sustainability in each age group

4.1.1 Filtering and grouping data

In order to know the sustainability of cannabis, first looking at the number of medications taking for each patient, we decide to separate each patient' medication times into 3 portions. (e.g., If a patient took 12 times of cannabis, we would divide the period into 1-4 times period, 5-8 times period, 9-12 times-period).

The reason and benefit of choosing this method are, the dataset consists different records on different patients, this method can ensure each patient has his own period, this will increase the accuracy of analysis and more reasonable.

After filtering out the three dataset (Period 1, Period 2, Period 3) from the original dataset, we decide to group each dataset into age group and calculate the average weighted efficacy and safety score for each patient in the period.

Table 6.1 is an example output for age group (18-24):

Table 6.1: Average weighted efficacy of 18-24 age group in each period

Patient ID	Period 1	Period 2	Period 3
9192	43.50984	62.78361	64.626667
21236	41.50500	29.30500	49.533898
21252	45.00909	41.10952	66.428571
22866	51.20000	47.13333	43.100000
29433	20.68750	12.73333	6.273333
42914	25.58636	22.26512	32.795349
45391	36.96667	29.66667	22.275000
48122	46.66667	58.46667	33.63333

4.1.2 Line Plot for observing tendency and spread for each patient

By plotting each patient average weighted efficacy against period, we have the following example output plot (Figure 6.2). Where the tendency of average weighted efficacy cannot be captured very well and average safety score has a decreasing tendency along with time in this age group (18-24).

Figure 6.2 – A

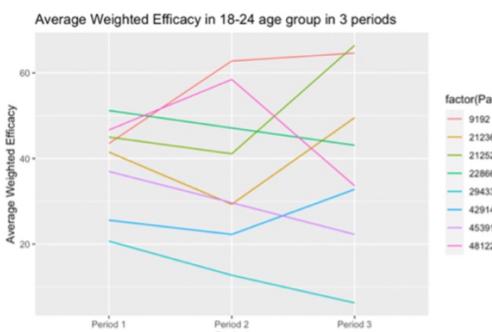
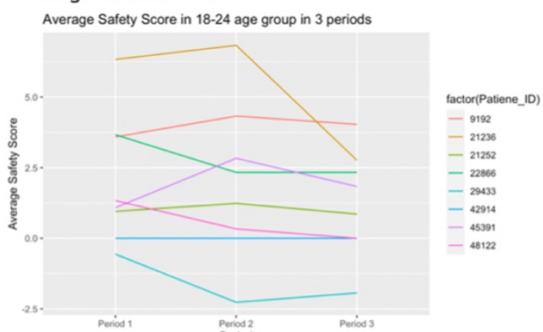


Figure 6.2 – B



paired t-test to check the difference between two data with matching variables. By setting the null hypothesis as:

The average difference in weighted efficacy is 0 from period 1 to period 2

With significance level alpha = 0.05 and degree of freedom = 7, the critical value of t is 1.895, we obtain t-stats = 0.25178.

Since $|0.252| < 1.895$, we cannot reject the null hypothesis, hence we get the result, the average difference in weighted efficacy from period 1 to period 2 is 0.

4.2 Result summary

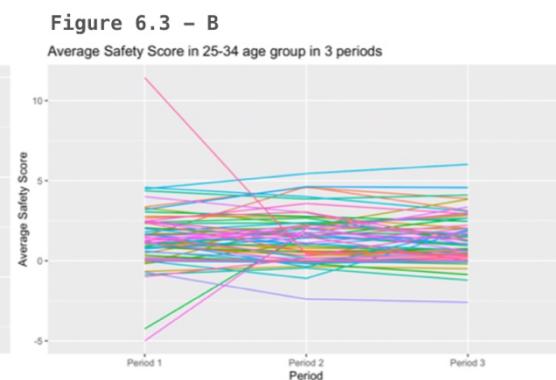
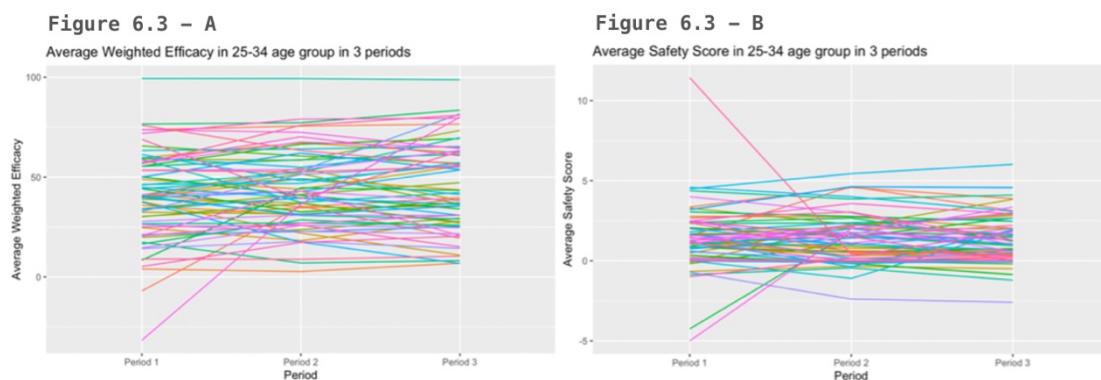
4.2.1 Result Plot

- Age Group (18-24):

Result can be seen in the above example.

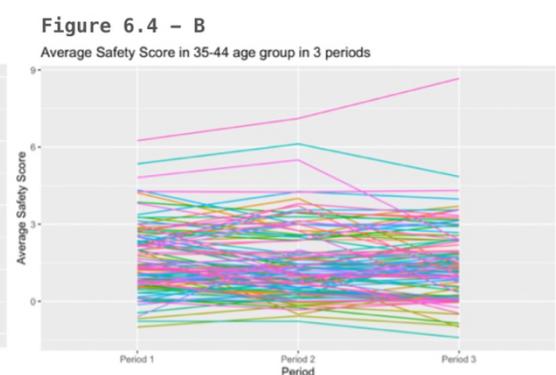
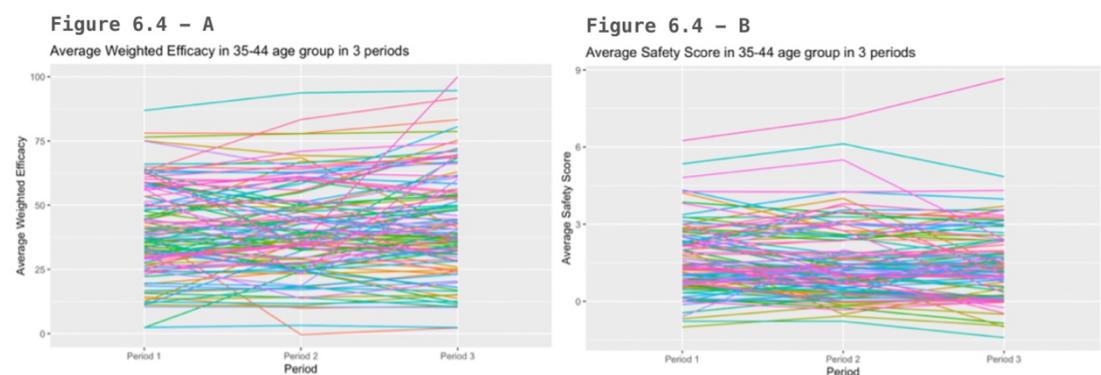
- Age Group (25-34):

From Figure 6.3, there is an increasing tendency in average weighted efficacy across period 1 to period 3 and the majority number is positive. However, the tendency in safety is hard to observe, in long term period, the safety scores seemed to vary a from the previous records.



- Age Group (35-44):

Seen from the Figure 6.4, the difference between each period in weighted efficacy is hard to observe, almost all the average weighted efficacy for each patient is positive. But the safety score has a decreasing trend along with time and having more small number of safety scores.



- Age Group 45-54:

From Figure 6.5, all average weighted efficacy for each patient is positive, and the number vary from each patient a lot, cannot judge the tendency of the average weighted efficacy. For average safety scores, it has a slightly increase along with time, but most of the number is smaller compared to younger patients.

Figure 6.5 – A

Average Weighted Efficacy in 45-54 age group in 3 periods

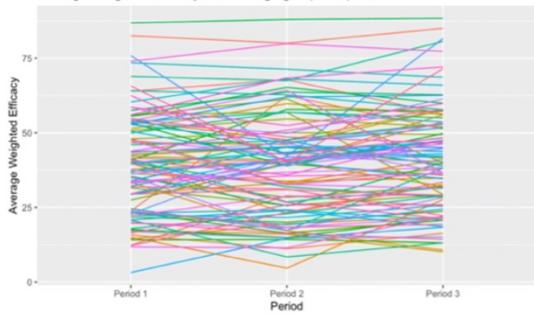
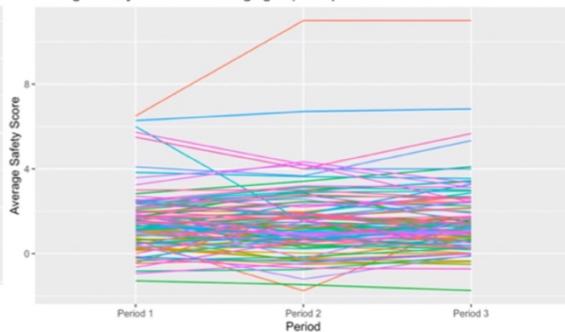


Figure 6.5 – B

Average Safety Score in 45-54 age group in 3 periods



- Age Group 55+:

According to Figure 6.6, the average weighted efficacy varies between patients from period 1 to period 3, the overall tendency is relatively stable. For average safety scores, it has a slightly increasing along with time, however, there are few numbers is negative and some number is close to 0, which indicates older patients has more risks in taking cannabis.

Figure 6.6 – A

Average Weighted Efficacy in 55+ age group in 3 periods

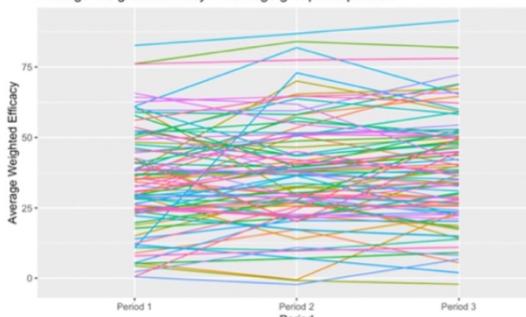
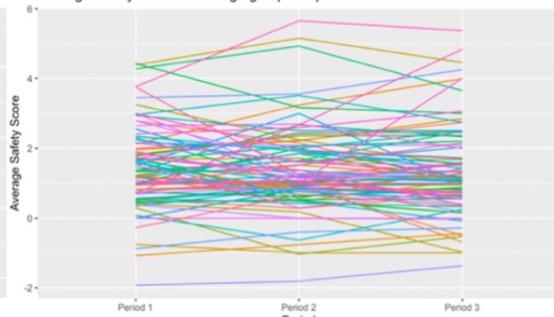


Figure 6.6 – B

Average Safety Score in 55+ age group in 3 periods



4.2.2 Result Paired T-test table:

Table 6.7 Hypothesis result of weighted efficacy

Age Group	Period 1-2		Period 2-3		
	T-stats	Hypothesis	T-stats	Hypothesis	CV
18-24	0.252	NO	-0.329	NO	1.895
25-34	-1.773	YES	-0.773	NO	1.668
35-44	-0.449	NO	-2.93	YES	1.645
45-54	0.286	NO	-2.249	YES	1.645
55+	-2.022	YES	-1.331	NO	1.662

Table 6.8 Hypothesis result of safety score

Age Group	Period 1-2		Period 2-3		CV
	T-stats	Hypothesis	T-stats	Hypothesis	
18-24	0.232	NO	1.442	NO	1.895
25-34	0.007	NO	0.56	NO	1.668
35-44	0.732	NO	1.369	NO	1.645
45-54	0.667	NO	-1.343	NO	1.645
55+	0.692	NO	0.731	NO	1.662

Number under each period represents the t-stats for this age group in this period, and the column next to it represent, whether we reject the null hypothesis, where the last column means the critical value of 5% significance level and correspond degrees of freedom.

Average Weighted Efficacy

- Age Group (18-24):

The null hypothesis has not been rejected for both durations between periods, which can be concluded as cannabis are more sustainable to this age group in weighted efficacy.

- Age Group (25-34):

From period 1 to period 2, the average difference is not equal to 0 since the null hypothesis has been rejected, but during period 2 to period 3, the sustainability is retained. It can be concluded, in this age group, cannabis weighted efficacy will not be stable at the beginning of the treatment.

- Age Group (35-44):

The weighted efficacy remains stable from first period to second period, however, have fluctuation from period 2 to period 3. It is commonly believed medicine will lose its power in long term treatment. Hence, the weighted efficacy maintains stable at short term treatment but has random fluctuation for long term treatment for age group 35-44.

- Age Group (45-54):

Similarly, the test result shows, from period 1 to period 2, the null hypothesis has not been rejected, average weighted efficacy difference between these two periods is 0. But the sustainability cannot maintain from period 2 to period 3 for this age group as well.

- Age Group (55+):

The weighted efficacy varies a lot in period 1 to period 2, one assumption here will be, for older patients, they will need more time to build tolerance to cannabis. From period 2 to period 3, the patient's weighted efficacy is more stable. However, compare the t-stats with other groups, the number is very close to the critical value, this also proved, older patients will have varied weighted efficacy, which means cannabis' weighted efficacy is not very sustainable in this age group.

Average Safety Score

For all age groups, the average safety scores difference between each period are 0 under 5% significance level and corresponding degrees of freedom condition. However, from period 2 to period 3, the t-stats for each age group are closer to critical value, compared with period 1 to period 2. This proved longer term treatment with cannabis, will result in higher possibility to cause side effects.

4.3 Overall Conclusion of sustainability

Cannabis-based medicines have sustainability in most of age group from a medication safety perspective. However, geriatric patients are a vulnerable group that may be suffered from many chronic diseases and the sustainability of such medicines is relatively unstable compared with other groups during treatment. The weighted efficacy variates from each age group, where the age group (25-34) and age group (55+) have more unstable weighted efficacy with cannabis treatment in the early stage, but the age group (35-44) and age group (45-54) is opposite.

5. Modelling

5.1 Weighted Efficacy Prediction Linear Model

Before modeling, setting 70% of the data as train set and 30% of data as test set, for model prediction accuracy.

Multiple independent variables have been added into the linear regression model with weighted efficacy as dependent variable.

There are 3 models in total:

Model 1:

$$\begin{aligned} \text{Weighted Efficacy} = & \text{Patient_Age} + \text{Gender_Female} + \\ & \text{Gender_male} + \text{Time_of_Day_Morning} + \text{Time_of_Day_Evening} + \\ & \text{Time_of_Day_OverNight} + \text{THC} + \text{CBD} + \text{Dosage_Measure} + \\ & \text{Ingestion_Method} \end{aligned}$$

Model 2:

Based on model 1, add additional terms: pre medicating score- post medicating score and post medicating score.

Model 3:

Based on model 2, add additional terms: pre medicating score, pre medicating score square and pre medicating score square*post medicating score.

Model 4:

$$\begin{aligned} \text{Weighted Efficacy} = & \text{CBD} + \text{THC} + \text{Patient Age} + \text{Patient Gender} + \text{Strain Type} + \text{Symptom} + \\ & \text{Dosage Measure} + \text{Ingestion Method} + \text{Pre medicating Score} \end{aligned}$$

Comparing the adjusted R square, MSE and RMSE for each model as follow (Table 6.9)

Table 6.9 Adjusted R square of three MLR

Model	Adjusted R Square	MSE	RMSE
Model 1	0.05042	625.0	25.00
Model 2	0.9559	66.8	8.18
Model 3	0.9817	12.9	3.51
Model 4	0.143	572.3	23.92

From the result, model 3 has the highest Adjusted R square and smallest MSE, RMSE, which means this model explain the highest amount of variability and the most accuracy one.

However, model 3 has multicollinearity among variables, we then choose model 2 as our final MLR model for weighted efficacy

Check the success rate of model 2 for prediction, there is 92.8% probability of successful prediction.

Additionally, by seeing the parameters of ‘pre medicating score – post medicating score’, it shows the higher the difference between pre and post medicating score, the higher weighted efficacy a patient can have via cannabis treatment. One assumption here is, cannabis is more efficient for severe symptom.

However, since adding the ‘post-medicating score’ as regressors, the dependent variable and independent variable are highly related. From model 1, which exclude post-medicating score

from the model, it only obtains a 0.05042 Adjusted R square for model 1 and 0.143 for model 4, therefore, we need to consider other model to proceed.

5.2 Machine Learning – Decision Tree model and Random Forest

The advantages of decision trees are that they are visual, understandable and interpretable. It is also computationally small, fast to classify and can handle high latitude data. The number of data dimensions in this dataset is not high, but there are many categorical variables, so the number of variables increases after one-hot encoding. In addition, the random forest prevents the risk of overfitting compared to the decision tree model. Therefore, we experimented with these two models.

5.2.1 Data preprocessing

Correlation analysis was performed on 6 numerical variables (Figure 6.10), and it was found that there was no high correlation between the variables, so feature merging was not required.

Figure 6.10 | Correlation Matrix



In addition, since Dosage Units is a supplement to the description of Dosage Measure, it should not be input as a training feature, so we exclude it from our model.

5.2.2 Predictors

Numerical predictors: Dosage Measure, Patient Age, Pre-medicating Score, THC, CBD, Weighted Efficacy

Categorical predictors: Patient Gender, Strain Type, Symptom, Ingestion Method

Since the decision tree in sk-learn cannot handle categorical variables, the categorical variables are one-hot encoded before the decision tree algorithm.

- One-hot encoded: convert categorical data to numerical data

One-hot encoding uses N-bit state registers to encode N states, each state has its own independent register bit, and at any time, only one of them is valid.

For example: in the Patient Gender variable, there are 2 categories [Male, Female] and therefore 2 binary variables are needed. A “1” value is placed in the binary variable for the gender and “0” values for the other gender.

Male => 10

Female => 01

5.2.3 Split Data

As when building a linear model, we split the data set by 25% test set and 75% training set to measure the error of a model in predicting quantitative data.

5.2.4 Data Standardization

The purpose of standardization is to process data of different dimensions among multiple features, and scale the data so that each variable has the same range or variance, hence, reduce the impact of scale, feature and distribution differences.

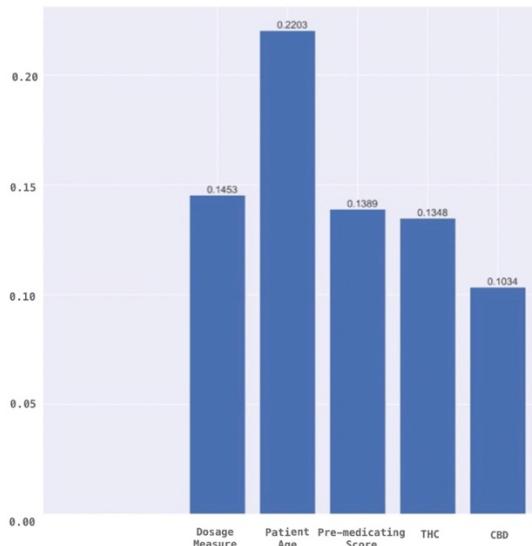
For the decision tree, since each feature is processed separately, it does not involve the adjustment of multiple feature dimensions, and the division of data does not depend on scaling, so the decision tree algorithm does not require feature preprocessing.

5.2.5 Result of the algorithm (Table 6.11)

Table 6.11	MSE	RMSE	R Square
Decision Tree	513.87	22.67	0.21
Random Forest	283.44	16.83	0.56

When the decision tree uses the default parameters, serious overfitting occurs, so we limit the maximum depth = 5, and there is no obvious overfitting afterwards.

Figure 6.12 | Top 5 significant factors in the efficacy of cannabis-based medicines



The number of random forest selection trees is 100, and we use this algorithm to output the importance of each feature when predicting. Then we found the three most important features are patient age, dosage measure, and pre-medicating score (Figure 6.12).

Note: The above figure only contains the variables with the highest feature importance values. The full picture is shown in the appendix.

5.3 Model part conclusion

Since random forest gives the highest R squared value and lowest MSE compared with linear model (MLR model 4) with same regressors , we choose Random Forest as our model to predict and the top 5 most significant factors to impact the efficacy of cannabis-based medicines are patient age, dosage measure, pre-medicating score, the component of THC, and the component of CBD.

Discussion

With increased legal access to medical marijuana and a change in opinion, both socially and medically, patients are looking for cannabis to treat a wide variety of conditions. Referring to the (“Frequently asked questions about medicinal cannabis”, 2022), the cultivation of medical cannabis was licensed by the Australian Federal Government in 2016, and there are now two types of medical cannabis that have been registered with ARTG, namely Nabiximols (Sativex®) under Schedule 8 and Schedule 4 Cannabidiol (Epidyolex®). If a patient's symptoms require a cannabis-based drug, they only need a health practitioner permission to get a prescription. But the price of cannabis drugs fluctuates widely from \$50 - \$1,000 per week, and it is not covered by medical insurance. If its effectiveness is confirmed scientifically in the future, then the affordability of this drug is worth considering by the government.

In this project, strainprint collected a total of 1,214 different medicine use records from 154 brands. Eighty percent of the drugs were high THC, 20 percent were high CBD and 8 percent were balanced.

Medication Efficacy

The results showed that compared with females, male experienced better levels of pain, joint stiffness, inflammation (except in the 18-24 age group) and pain relief after cannabis use than females. For mental disorders (e.g., depression and anxiety) and insomnia symptoms, there were no significant effect differences between two genders. But it is worth noting that the weighted efficacy obtained by females aged 18-24 with these two symptoms was significantly lower than that of others, and the cannabis even had a worsening effect on the insomnia symptoms of female in this age group.

By plotting weighted efficacy against dose, we found no significant differences at any dose, but using the highest dose (e.g. 30 inhalations, 10 ml) instead reduced the drug's effect on relieving pain, inflammation and joint stiffness. Nevertheless, it is important to note that different methods of administration (e.g., smoking, oral, external use) were included together in these analyses and these different methods of administration would affect the potency of the cannabis.

The analysis results based on weighted efficacy also showed that “smoking” was the most effective way of ingestion for most symptoms, followed by “taking by oral”. When smoked, cannabinoids can enter the bloodstream directly through the respiratory system, with a bioavailability of 56%. In comparison, oral ingestion is sublingual, delivering cannabinoids to the bloodstream through the oral mucosa, with a bioavailability of 50-75% (Mckinney, 2018). External application is applied to the skin and the time of action is the slowest.

Medication Safety

Drug safety variates in gender, age, ingestion methods, and taking time based on the analysis of the safety score. In clinical practice, medication safety is subjected to many other factors including the side effects in pregnant and lactating women, the safety to children (pediatrics), the potential risk to old people (geriatrics) as well as drug-drug and food-drug interaction.

Pregnant women should not take the drug without consulting a specialist about its safety as it may

adversely affect the formation of the fetus. Even if the drug is taken just before conception, it can have harmful effects on the fetus (Thamir, 2014). Besides, some drugs may be excreted in breast milk during lactation, which may have negative effects on the infant (Dumont and Black, 2013). A child's vital organs are immature, exposure to certain drugs can lead to toxic side effects as the body often fails to fully metabolize or eliminate the drug. This allows the active substance to remain in the body for a long period of time, increasing levels in the body and possibly producing toxic effects (Yang and So, 2014).

Older people are considered the most vulnerable to drugs because they can suffer from many chronic diseases, such as high blood pressure, diabetes, and high blood cholesterol and lipids in the blood. Possibly requires chronic use of multiple drugs that may compete, what is known as drug-drug interaction (Thamir, 2014). In addition, the physiological functions of many body organs decline with age, especially important organs such as the liver and kidneys. The geriatric population may also suffer from dementia "impaired memory" and so take their medications improperly (Haider et al., 2013)

Although, the benefits are greater than the cost to take cannabis-based medicines to treat arthritis-related symptoms under the patients' self-reported circumstances based on the analysis of the safety score. But such medicines still cannot be defined as "safe" in pharmacology. The reasons include mostly the lack of scientific studies and clinical trials needed to evaluate the safety of these medications in each age group separately and the drug-drug and food-drug interactions. For example, if arthritis patients are prescribed specific arthritis drugs, do this interaction with cannabis-based medicines?

In terms of ingestion method, we have come to the exact opposite conclusion from the efficacy section, "smoke" being the most dangerous form to ingest. This is consistent with studies that now show strong evidence that cannabis causes bronchial inflammation, respiratory symptoms and affects lung function (Lee & Hancox, 2011). Furthermore, dab bubbler/portable is the emerging method of ingestion and is classified by us as "smoke", but little is known about the potential risks of this method. There is also an increased potential for accidents due to the need to use the blowtorch multiple times (Loflin & Earleywine, 2014).

In summary, our most recommended method of intake is "oral", which has a good track record of safety and efficacy. We also recommend that future policies should encourage further research into the health effects of smoking cannabis.

Recommendation

For most tracking sessions, users reported reduced symptoms of pain (87.8%), inflammation (89.5%), insomnia (86.7%), mental disorders (93.7%), joint stiffness (93.1%) after inhaling cannabis. However, this percentage results may be overstated because patients who do not experience relief of symptoms after trying cannabis may discontinue use immediately. Moreover, for different symptoms, only a unified weighted efficacy can be used to measure, so the results also have bias. Therefore, we recommend adding a specific score to each symptom to compare and calculate the effectiveness of cannabis use, such as pain rating change, insomnia rating change, etc.

Although we have made drug recommendations for different symptoms based on the main components of cannabis, the THC and CBD concentration values are entered by the users of the app themselves, so the reliability of some of these data is questionable. Such as: there are huge, unreasonable CBD and THC concentration values (>100), which will affect our judgement. Therefore, we recommend setting fixed values with upper limit for patients to choose and adding an “unknown” option when filling in the two items of CBD and THC.

Finally, we propose to build the predictive model into the backend of the app so that patients can input their personal information and medication information (e.g. duration of medication, THC/CBD level, dose, mode of intake) to see the predicted weighted efficacy as a reference for medication use. Further, a summary report of the results need to be generated. For instance, if the weighted efficacy score is less than 0, the patient is reminded to use the medication with caution or to seek advice from a medical professional.

Peer Review

Review1:

Having read through the report, Medi-Data delivered intriguing insights, covering all symptoms caused by arthritis and the efficacy of cannabis on each symptom as well as its associated risks. The report was well structured and achieved consistency by having the same plots and structure of analysis for all factors, and all plots used in the report were appropriate, making it easier to interpret. They also compared and thoroughly outlined identifiable factors which have the largest efficacies. However, upon observing table 0.1, the data appears highly skewed towards the older demographic (35 - 55+), which hasn't been addressed in the report, which could affect the reliability of the analysis and insights. Furthermore, it is unclear how the safety score was classified and calculated; additional explanations would further clarify this metric and aid understanding. Moreover, whilst there was ample analysis of graphs and tables, Medi-Data seemed to lack deeper insights on how this can be effectively utilised.

Review2:

We liked that the report was well structured and had clear headings – it made it very easy to follow and understand what was being discussed. The visualisations particularly were informative and broke up the information into readable chunks.

We noticed that the title and scope of the report listed safety as a key measure in the study. However, the analysis and exploration of the data was heavily skewed towards the efficacy of the treatment rather than the safety. It would have been good to see some more analysis of the safety aspect of this treatment, as well as some visualisations akin to the many in the efficacy section. Further to this point, there were multiple mentions of a ‘safety score’ for each patient in the scope and plan, however this score is not included anywhere in the findings section.

Suggestion 1. It is not clear how the safety score is classified and calculated.

Response: A detail explanation about safety score has been added in appendix 2.

Suggestion 2. Insufficient analysis on safety aspects.

Response: we added a deeper analysis and visualization on safety part (see Finding – 3. Safety result), which is also analyzed in the sustainability section.

Suggestion 3. Data are skewed towards an older population, which may affect the reliability of the analysis.

Response: It is a description problem and we have rewritten it. The reason behind elder population has arthritis is that they are the most vulnerable group, and the physiological functions of many body organs decline with age.

Reference List

Website Reference

Frequently asked questions about medicinal cannabis. (2020, March 15). Retrieved from <https://www.health.vic.gov.au/drugs-and-poisons/frequently-asked-questions-about-medicinal-cannabis#>

Mckinney, J. (2018, November 20). Consumption methods affect how much cannabis you absorb, HelloMD. <https://www.hellomd.com/articles/consumption-methods-affect-how-much-cannabis-you-absorb/>

Journal Reference

De Briyne, N., Holmes, D., Sandler, I., Stiles, E., Szymanski, D., Moody, S., Neumann, S., et al. (2021). Cannabis, Cannabidiol Oils and Tetrahydrocannabinol—What Do Veterinarians Need to Know? *Animals*, 11(3), 892.

<https://doi.org/10.3390/ani11030892>.

Dumont, T., Black, A., (2013). Pregnancy and isotretinoin therapy. *Canadian Medical Association Journal*, 185(6), 508.

<https://doi.org/10.1503/cmaj.113-2111>.

Haider, B., Schmidt, R., Schweiger, C., Forstner, T., Labek, A., Lampl, C., (2014). Medication Adherence in Patients With Dementia: An Austrian Cohort Study. *Alzheimer Disease & Associated Disorders*, 28(2), 128-133.

<https://doi.org/10.1097/WAD.0000000000000006>.

Hammell, D.C., Zhang, L.P., Ma, F., Abshire, S.M., McILwrath, S.L., Stinchcomb, A.L., Westlund, K.N., (2015). Transdermal cannabidiol reduces inflammation and pain-related behaviours in a rat model of arthritis. *European Journal of Pain*, 20(6), 936-948.

<https://doi.org/10.1002/ejp.818>.

Lee, M.H.S., Hancox, R.J., (2011). Effects of smoking cannabis on lung function, *Expert Review of Respiratory Medicine*, 5(4), 537–547.

<https://doi.org/10.1586/ers.11.40>.

Loflin, M., Earleywine, M., (2014). A new method of cannabis ingestion: The dangers of dabs?, *Addictive Behaviours*, 39(10), 1430-1433.

<https://doi.org/10.1016/j.addbeh.2014.05.013>.

Shaina, P.C., Stephanie, E.L., Patrick, D., Jonathan, E.P., (2021). *Evaluation of Patient Reported Safety and Efficacy of Cannabis From a Survey of Medical Cannabis Patients in Canada*. Cannabis: A Public Health Perspective.

<https://doi.org/10.3389/fpubh.2021.626853>.

Thamir, M.A. (2014). Drug safety: The concept, inception and its importance in patients' health. *Saudi Pharmaceutical Journal*, 24(4), 405-412.
<https://doi.org/10.1016/j.jsps.2014.04.008>.

Yang, M., & So, T. Y., (2014). Revisiting the safety of over-the-counter cough and cold medications in the pediatric population. *Clinical pediatrics*, 53(4), 326–330.
<https://doi.org/10.1177/0009922813507998>.

Appendix 1

The Magic of Weighted Efficacy

Assume we have $X = \text{pre-medicating score}$, $Y = \text{post-medicating score}$, $X, Y \in [1, 10]$

The formula of the change between pre-medicating and post-medicating is:

$$\frac{X - Y}{X}.$$

The formula of weighted efficacy (WE) is:

$$\frac{\frac{X - Y}{X} - \frac{X - Y}{10}}{2}.$$

Weighted efficacy can shrink the emotional error of the given pre- and post-medicating score from patients. If the symptom has some degree of relief (i.e., $(X - Y) > 0$), its' value will be lower than the actual difference between pre- and post-medicating score. However, if the symptom goes worse (i.e., $(X - Y) < 0$), its' value will be larger than the actual difference.

Mathematical Proof:

$$\begin{aligned} & \frac{X - Y}{X} - \frac{\frac{X - Y}{X} - \frac{X - Y}{10}}{2} \\ &= \frac{2(X - Y)}{2X} - \frac{X(\frac{X - Y}{X} - \frac{X - Y}{10})}{2X} \\ &= \frac{2(X - Y)}{2X} - \frac{(X - Y) - \frac{X(X - Y)}{10}}{2X} \\ &= (X - Y) - \frac{X(X - Y)}{10} \\ &= \frac{(10 + X)(X - Y)}{10} \end{aligned}$$

If $(X - Y) > 0$, then the equation will greater than > 0 .

But $(X - Y) < 0$, then the equation will greater than < 0 .

Appendix 2

The Beauty of Safety Score

Motivation of Safety Score:

Medication safety describes the toxicity of drugs in various settings such as gender, age, dose, drug-drug interactions, and drug-food interactions. To assess the toxicity, most of the attention is focused on the severity of the side effects after taking medicines based on the reference to the existing literature. The given dataset is about patients' self-reported taking records and 39 after-medicating reactions. Strainprint™ separated these into 3 categories positive emotive (wanted effects), neutral emotive (moderate side effects), and negative emotive (adverse effects). The question is that many side effects were recorded but the severity of these effects was unknown. For example, some patients got a headache after taking medicines, is it side effects go away after a while or continue to harm the body? This is unknown.

To less the bias of analysis, two penalty coefficients are added in front of neutral emotive and negative emotive, and positive emotive are invited to be the minuend. Zero is considered as the baseline of the Safety Score, if it is less than zero then the drug is relatively unsafe, else if it is greater than zero then the drug is relatively safe.

Explanation of the mechanism:

$$\text{Safety Score} = \frac{\text{num of Positive Emotive}}{- 1 * \text{num of Neutral Emotive} - 2 * \text{num of Negative Emotives}}$$

Penalty Coefficients

Positive Emotive includes Aroused, Comfortable, Creative, Dreamy, Energized, Euphoric, Focused, Giggly, Happy, Less aware of pain, Light, Motived, Pain free, Positive, Refreshed, Relaxed, Talkative, and Upbeat (18 effects).

Neutral Emotive includes Couchloacked, Dry mouth, Foggy, Forgetful, Hungry, Lethargic, No appetite, Red eyes, Restless, Sleepy, Thirsty, Tingly, Tired and Zoned out (14 effects)

Negative Emotive includes Anxious, Depressed, Dizzy, Headache, Nauseous, Paranoid and Racing heart (7 effects)

The penalty coefficient of negative emotive is heavier than neutral emotive since the side effects in the latter are relatively more severe than it in the former.

Limitation of Safety Score:

The penalty coefficients were selected subjectively and were not in a scientific approach since it requires intensive research in comparing the difference of each side effect. To convenience this project analysis, a hybridized safety score mechanism was selected based on the limitation of the given dataset. It includes the benefits and costs of taking medicines and is not purely focused on the harm of taking the drug. In the future, if the severity of each side effect is provided, then the safety score mechanism can be improved.

Appendix 3.1 – Data cleaning coding

SQL

```
create table Patients (
    ePRO_ID integer,
    Patient_ID integer,
    Conditions text,
    Age integer,
    Age_Group text,
    Gender text,
    Session_Date date,
    primary key (ePRO_ID)
);
```

```
copy Patients FROM '/Users/alanlin/Desktop/Data 3001/Project/ePRO_ID.csv' WITH CSV
HEADER;
```

```
delete from Patients where age is null;
```

```
#####
## Step 1 filter out the first session date
## and last session date
#####
alter table Patients
rename column Session_Data TO Session_Date;
```

```
create view Patient_Details as
select Patient_ID, Conditions,
min(Age) as start_age,
max(Age) as end_age,
Gender,
(max(Session_Date) - min(Session_Date) + 1) / 7 AS Weeks
from Patients
group by Patient_ID, Conditions, Gender
order by Patient_ID;
#####
##### ##### ##### ##### #####
```

```
#####
## Step 2 modify the conditions column
#####
create view Patient_Conditions as
select Patient_ID, (string_to_array(Conditions, ',')[1] as condition
```

```

from Patient_Details;
select *
into Patient_Con
from Patient_Conditions
order by Patient_ID;
##### ##### #####
##### #####
## Step 3 generate a new patient_info dataset
#####
create view Patient_Info as
select p_c.Patient_ID, p_c.condition,
p_d.start_age, p_d.end_age, p_d.Gender, p_d.Weeks
from Patient_Con p_c
join Patient_Details p_d
on p_c.Patient_ID = p_d.Patient_ID;

select *
into Patient_Information
from Patient_Info
order by Patient_ID;
#####

#####
## Step 4 export to csv file
#####
alter table Patient_Information
add primary key (Patient_ID);

copy Patient_Information to '/Users/alanlin/Desktop/Data 3001/Project/Patient_Info.csv'
DELIMITER ',' CSV HEADER;
#####

#####
## Step 5 count the patients' session duration
#####
create view data_cleaning as
select p.ePRO_ID, p.patient_ID, p.Conditions,
p.Age, p.Age_Group, p.Gender, p.Session_Date
from buffer_02 b_02
join patients p
on b_02.ePRO_ID = p.ePRO_ID;

create view epro_id_wk12 as

```

```

select p.ePRO_ID, p.Patient_ID, p.Session_Date
from data_cleaning p
join patient_with_12wks p_12
on p_12.Patient_ID = p.Patient_ID;

create view help_01 as
select Patient_ID,
(min(Session_Date) + 83) as end_date
from epro_id_wk12
group by Patient_ID;

create view help_02 as
select e.ePRO_ID, e.Patient_ID, e.Session_Date,
h_01.end_date
from help_01 h_01
full outer join epro_id_wk12 e
on h_01.Patient_ID = e.Patient_ID;

create view help_03 as
select ePRO_ID, Patient_ID,
Session_Date, end_date,
(end_date - Session_Date) + 1 as days
from help_02;

create view help_04 as
select patient_id,
max(case when days <= 0 then days end) as last
from help_03
group by Patient_ID;

create view help_05 as
select h_03.ePRO_ID, h_03.Patient_ID, h_03.Session_Date,
h_03.end_date, h_03.days, h_04.last
from help_04 h_04
full outer join help_03 h_03
on h_03.Patient_ID = h_04.Patient_ID;

select *
into buffer
from help_05
order by Patient_ID, days DESC;

alter table buffer
add happy integer;

```

```

update buffer
set happy = 1
where days >= last;

delete from buffer
where happy is null;

create view epro_id as
select ePRO_ID, Patient_ID
from buffer;

select *
into Subset
from ePRO_id
order by Patient_ID, ePRO_ID;

copy Subset to '/Users/alanlin/Desktop/Data 3001/Project/Subset.csv' DELIMITER ',' CSV
HEADER;

create view help_06 as
select patient_id,
count(*) as records,
(max(session_date) - min(Session_Date)) / 7 as weeks
from buffer
group by patient_id
order by Patient_ID;

create view help_07 as
select *
from help_06
where weeks = 12
order by patient_id;

select *
into Subset_patients
from help_07;

copy Subset_patients to '/Users/alanlin/Desktop/Data 3001/Project/Subset_patients.csv'
DELIMITER ',' CSV HEADER;

create view help_08 as
select s.ePRO_ID, s.patient_id
from Subset s
join Subset_patients p

```

```
on s.Patient_ID = p.Patient_ID;

select *
into Subset_id
from help_08
order by Patient_ID, ePRO_ID;

copy Subset_id to '/Users/alanlin/Desktop/Data 3001/Project/Subset.csv' DELIMITER ',' CSV
HEADER;
##### ###### ##### ##### #####
##### ##### ##### ##### #####
## Step 6 extract all patients_id with >= 12 weeks
##### ##### ##### ##### #####
create view Patient_Wk12 as
select Patient_ID, Weeks
from Patient_Information
where Weeks >= 12
order by Patient_ID;

select *
into Patient_With_12Wks
from Patient_Wk12
order by Patient_ID;

alter table Patient_12Wks
add primary key (Patient_ID);

copy Patient_With_12Wks to '/Users/alanlin/Desktop/Data 3001/Project/Patient_12Wks.csv'
DELIMITER ',' CSV HEADER;
##### ##### ##### ##### #####
```

Appendix 3.2 – Data analysis coding

R-code 1:

```

library("readxl")
library("dplyr")
library("tidyverse")
library(ggplot2)
library(patchwork)

Cannabis <- read_excel("fd.xlsx")
attach(Cannabis)
summary(Cannabis)
head(Cannabis)

hist(`Weighted Efficacy`)

plot('Pre-medicating Score', `Weighted Efficacy`)
hist('Pre-medicating Score')
# Overall 57152 observations
# Overall 391 patients
id_table <- Cannabis%>%count('Patient ID')
min(id_table$n)
max(id_table$n)
mean(id_table$n)
# The frequency range of personal use of the app is 2~895, and the average number of uses is
about 146

##### Count the frequency of each side effect(dummy variables)#####
side_effect <- apply(Cannabis[,24:62],2,sum,na.rm=T)
rev(sort(side_effect))
# The 5 most frequent
# Relaxed:29186, Comfortable:20835, Happy: 11959, Thirsty:8920, Focused:8535

# Positive/Neutral/Negative Emotives
cbd_domi <- Cannabis%>%filter(CBD/THC>20)
cbd_effect <- apply(cbd_domi[,21:23],2,sum,na.rm=T)
rev(sort(cbd_effect))

the_domi <- Cannabis%>%filter(THC/CBD>20)
thc_effect <- apply(the_domi[,21:23],2,sum,na.rm=T)
rev(sort(thc_effect))

# Other side effect for THC or CBD dominant
cbd_domi <- Cannabis%>%filter(CBD/THC>20)

```

```

cbd_effect <- apply(cbd_domi[,24:62],2,sum,na.rm=T)
rev(sort(cbd_effect))

thc_domi <- Cannabis%>%filter(THC/CBD>20)
thc_effect <- apply(thc_domi[,24:62],2,sum,na.rm=T)
rev(sort(thc_effect))

##### Mental Disorders #####
Mental <- Cannabis%>%filter(Symptom=='Anxiety' | Symptom=='Depression')
Mental%>%count() #8008
Mental%>%filter(`Weighted Efficacy` > 0)%>%count()
Mental%>%select(`Patient ID`)%>%distinct()%>%count() # 250
Mental%>%filter(`Patient Gender`=='Male')%>% select(`Patient ID`)%>%distinct()%>%count()
#92
Mental%>%filter(`Patient Gender`=='Female')%>%select(`Patient ID`)%>%distinct()%>%
count() #155
Mental%>%filter(`Patient Gender`=='Unknown')%>%select(`Patient ID`)%>%distinct()%>%
count() #3

ment_id <- Mental%>%count(`Patient ID`)
min(ment_id$n)
max(ment_id$n)
mean(ment_id$n)
sd(ment_id$n)

Mental%>%filter(is.na(Anxious))%>%count() #7900
Mental%>%filter(is.na(Depressed))%>%count() #8008
Mental%>%filter(Happy==1)%>%count() #1776

##### (3) Symptom: Pain #####
Pain <- Cannabis%>%filter(Symptom=='Arm or Leg Pain'|

                           Symptom=='Back Pain - Lower'|

                           Symptom=='Back Pain - Upper'|

                           Symptom=='Joint Pain' |

                           Symptom=='Muscle Pain'|

                           Symptom=='Nerve Pain')

Pain %>% count() #30830
Pain%>%select(`Patient ID`)%>%distinct()%>%count() # 369
Pain%>%filter(`Patient Gender`=='Male')%>% select(`Patient ID`)%>%distinct()%>%count()
#147
Pain%>%filter(`Patient Gender`=='Female')%>%select(`Patient ID`)%>%distinct()%>% count()
#219
Pain%>%filter(`Patient Gender`=='Unknown')%>%select(`Patient ID`)%>%distinct()%>%
count() #3

```

```

mean(Pain$`Patient Age`) #45.83
sd(Pain$`Patient Age`) #11.19
#369 patients (147 males, 219 females, 3 unknown gender; age M=45.83, SD=11.19)
pain_id <- Pain%>%count(`Patient ID`)

# After medication: Less aware of pain=1, Pain free=1
# Pain free patient
NoPain <- Pain %>%filter(`Less aware of pain`==1|`Pain free`==1)
NoPain%>%count() #8780
as.data.frame(table(NoPain$`Strain Type`))
# Most frequently used: Flower: 5931, Oil 1933

# Ingestion method
Pain_Flower <- NoPain%>%filter(`Strain Type`=='Flower')
as.data.frame(table(Pain_Flower$`Ingestion Method`))

Pain_Oil <- NoPain%>%filter(`Strain Type`=='Oil')
as.data.frame(table(Pain_Oil$`Ingestion Method`))

#####
# (4) Symptom:Inflammation #####
Inflammation <- Cannabis%>%filter(Symptom=='Inflammation')
Inflammation%>%count() #7122
Inflammation%>%select(`Patient ID`)%>%distinct()%>%count() # 241
Inflammation%>%filter(`Patient Gender`=='Male')%>% select(`Patient
ID`)%>%distinct()%>%count() #95
Inflammation%>%filter(`Patient Gender`=='Female')%>%select(`Patient
ID`)%>%distinct()%>% count() #145
Inflammation%>%filter(`Patient Gender`=='Unknown')%>%select(`Patient
ID`)%>%distinct()%>% count() #1
mean(Inflammation$`Patient Age`) #46.1
sd(Inflammation$`Patient Age`) #10.15
# 241 patients (95 males, 145 females, 1 unknown gender; age M=46.1, SD=10.15)
infla_id <- Inflammation%>%count(`Patient ID`)

#Inflammation patients with reduced/disappeared symptoms
Inflammation%>%filter(`Less aware of pain`==1|`Pain free`==1|`Comfortable`==1)%>%count()
Inflammation%>%filter(`Less aware of pain`==1)%>%count()#907
Inflammation%>%filter(`Pain free`==1)%>%count() #939
Inflammation%>%filter(`Comfortable`==1)%>%count()#2477

NoInfla <- Inflammation%>%filter(`Pain free`==1)
as.data.frame(table(NoInfla$`Strain Type`))

# Ingestion method

```

```

Infla_Flower <- NoInfla%>%filter(`Strain Type`=='Flower')
as.data.frame(table(Infla_Flower$`Ingestion Method`))

Infla_Oil <- NoInfla%>%filter(`Strain Type`=='Oil')
as.data.frame(table(Infla_Oil$`Ingestion Method`))

#####
# (5) Symptom:Joint Stiffness #####
Stiff <- Cannabis%>%filter(Symptom=='Joint Stiffness')
Stiff%>%count() #6901
Stiff%>%select(`Patient ID`)%>%distinct()%>%count() # 223
Stiff%>%filter(`Patient Gender`=='Male')%>% select(`Patient ID`)%>%distinct()%>%count()
#82
Stiff%>%filter(`Patient Gender`=='Female')%>%select(`Patient ID`)%>%distinct()%>% count()
#141
Stiff%>%filter(`Patient Gender`=='Unknown')%>%select(`Patient ID`)%>%distinct()%>%
count() #0
mean(Stiff$`Patient Age`) #45.74
sd(Stiff$`Patient Age`) #11.52
# 223 patients (82 males, 141 females, 0 unknown gender; age M=45.74, SD=11.52)

# Patients whose symptoms of joint stiffness have decreased/disappeared
Stiff%>%filter(`Comfortable`==1)%>%count() #2593
Stiff%>%filter(`Less aware of pain`==1)%>%count()#1027
Stiff%>%filter(`Pain free`==1)%>%count()#1039

NoStiff <- Stiff%>%filter(`Comfortable`==1)
NoStiff <- Stiff%>%filter(`Less aware of pain`==1)
NoStiff <- Stiff%>%filter(`Pain free`==1)
as.data.frame(table(NoStiff$`Strain Type`))

# Ingestion method
Stif_Flower <- NoStiff%>%filter(`Strain Type`=='Flower')
as.data.frame(table(Stif_Flower$`Ingestion Method`))

Stif_Oil <- NoStiff%>%filter(`Strain Type`=='Oil')
as.data.frame(table(Stif_Oil$`Ingestion Method`))

#####
# (6) Symptom: Insomnia #####
# After medication: Sleepy=1
Insomnia <- Cannabis%>%filter(Symptom=='Insomnia')
Insomnia%>%count() #4291
Insomnia%>%select(`Patient ID`)%>%distinct()%>%count() # 249
Insomnia%>%filter(`Patient Gender`=='Male')%>% select(`Patient

```

```

ID`)%>%distinct()%>%count() #95
Insomnia%>%filter(`Patient Gender`== 'Female')%>%select(`Patient ID`)%>%distinct()%>%
count() #151
Insomnia%>%filter(`Patient Gender`== 'Unknown')%>%select(`Patient ID`)%>%distinct()%>%
count() #3
mean(Insomnia$`Patient Age`) #46.57
sd(Insomnia$`Patient Age`) #12.07
# 249 patients (95 males, 151 females, 3 unknown gender; age M=46.57, SD=12.07)

# Patients whose insomnia symptoms disappear:
Insomnia%>%filter(Sleepy==1)%>%count() #1185

1185/4291*100 #27.6% of patients with insomnia reported improved sleep

# Frequency of use of Strain Type for patients whose insomnia symptoms disappeared
as.data.frame(table(NoInsomnia$`Strain Type`))
# Most frequently used: Flower 780, Oil 380

# Ingestion method
Inso_Flower <- NoInsomnia%>%filter(`Strain Type`=='Flower')
as.data.frame(table(Inso_Flower$`Ingestion Method`))

Inso_Oil <- NoInsomnia%>%filter(`Strain Type`=='Oil')
as.data.frame(table(Inso_Oil$`Ingestion Method`))

#####
# Dosage Measure plot #####
# According to the classification of dosage units, plot dosage measure vs weighted efficacy

# (1) Drops
# Only include ingestion method = Tincture
drop <- Cannabis%>%filter(`Dosage Units`=='Drops')

ggplot(drop,aes(`Dosage Measure`, `Weighted Efficacy`))+
  geom_point(size=4,color="steelblue2",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dosage Measure",y="Weighted Efficacy",title="Dosage Units = Drops")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# (2) Finger Tips
# Only include ingestion method = Topical
tip <- Cannabis%>%filter(`Dosage Units`=='Finger Tip')

ggplot(tip,aes(`Dosage Measure`, `Weighted Efficacy`))+

```

```

geom_point(size=4,color="steelblue2",alpha=0)+
geom_smooth(color="orange",method="gam",se=FALSE)+
labs(x="Dosage Measure",y="Weighted Efficacy",title="Dosage Units = Finger Tips")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# (3) Inhalations
# Include ingestion method = concentrate/Dab bubbler/Dab Portable/Smoke/Vape
joint <- Cannabis%>%filter(`Symptom` == 'Joint Stiffness')

pain <- Cannabis%>%filter(Symptom=='Arm or Leg Pain' |
                           Symptom== 'Back Pain - Lower' |
                           Symptom== 'Back Pain - Upper' |
                           Symptom== 'Joint Pain' |
                           Symptom== 'Muscle Pain' |
                           Symptom== 'Nerve Pain')

inf <- Cannabis%>%filter(Symptom=='Inflammation')
ins <- Cannabis%>%filter(Symptom=='Insomnia')
ment <- Cannabis%>%filter(Symptom=='Anxiety' | Symptom=='Depression' )

joint_smoke <- joint%>%filter('Ingestion Method'=='Smoke')
pain_smoke <- pain%>%filter('Ingestion Method'=='Smoke')
inf_smoke <- inf%>%filter(`Ingestion Method`=='Smoke' & `Pre-medicating Score` >3)
ins_smoke <- ins%>%filter(`Ingestion Method`=='Smoke' & `Pre-medicating Score` >= 4 & `Pre-medicating Score` <= 6)
ment_smoke <- ment%>%filter(`Ingestion Method`=='Smoke')

# Joint Stiffness patients use smoke intake medicine
ggplot(joint_smoke,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="steelblue2",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of inhalations)",y="Weighted Efficacy",title="Joint stiffness patients use smoke for ingestion")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# pain patients use smoke intake medicine
ggplot(pain_smoke,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="steelblue2",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of inhalations)",y="Weighted Efficacy",title="Pain patients use smoke for ingestion")

```

```

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# moderate&severe inflammation patients use smoke intake medicine
ggplot(inf_smoke,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="steelblue2",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of inhalations)",y="Weighted Efficacy",title="Moderate & Severe
inflammation patients use smoke for ingestion")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# moderate insomniacs use smoke intake medicine
ggplot(ins_smoke,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="steelblue2",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of inhalations)",y="Weighted Efficacy",title="Moderate insomnia patients use
smoke for ingestion")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# mental disorders use smoke intake medicine
ggplot(ment_smoke,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="steelblue2",alpha=0.1)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of inhalations)",y="Weighted Efficacy",title="Mental disorders patients use
smoke for ingestion")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# (4) mg
# Include ingestion method = edible/oral/Pill/Suppository/Transdermal

# inflammation patients use oil(oral) intake medicine
ggplot(inf_oil,aes('Dosage Measure','Weighted Efficacy'))+
  geom_point(size=4,color="orange",alpha=0)+
  geom_smooth(color="orange",method="gam",se=FALSE)+
  labs(x="Dose(# of mg)",y="Weighted Efficacy",title="Inflammation patients use      for
ingestion")+

theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# (5) ml
# Only include ingestion method = Oil
inf<- Cannabis%>%filter(`Symptom` == 'Inflammation')

```

```

inf_oil <- inf%>%filter(`Ingestion Method` == 'Oral' & `Dosage Units` == 'ml' & `Pre-medicating Score`<=3)
ins_oil <- ins%>%filter(`Ingestion Method` == 'Oral' & `Dosage Units` == 'ml' & `Pre-medicating Score`>=7)

ggplot(inf_oil,aes(`Dosage Measure`,`Weighted Efficacy`))+  

  geom_point(size=4,color="orange",alpha=0)+  

  geom_smooth(color="orange",method="gam",se=FALSE)+  

  labs(x="Dose(# of ml)",y="Weighted Efficacy",title="Mild inflammation patients use oral to ingest")+
  theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

ggplot(ins_oil,aes(`Dosage Measure`,`Weighted Efficacy`))+  

  geom_point(size=4,color="orange",alpha=0)+  

  geom_smooth(color="orange",method="gam",se=FALSE)+  

  labs(x="Dose(# of ml)",y="Weighted Efficacy",title="Severe insomnia patients use oral to ingest")+
  theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

# (6) Sprays  

# Only include ingestion method = Spray
spray <- Cannabis%>%filter(`Dosage Units`=='Sprays')

ggplot(spray,aes(`Dosage Measure`,`Weighted Efficacy`))+  

  geom_point(size=3,color="orange",alpha=0)+  

  geom_smooth(color="orange",method="gam",se=FALSE)+  

  labs(x="Dosage Measure",y="Weighted Efficacy",title="Dosage Units = Sprays")+
  theme_bw() + theme(plot.title=element_text(hjust=0.5,size=16),axis.title=element_text(size=14))

```

Appendix 3.3 Sustainability code

Python code:

```

import pandas as pd
import numpy as np

cannabis = pd.read_excel("fd.xlsx")

count = {}

for _, row in cannabis.iterrows():
    pid = row["Patient_ID"]
    if pid in count:
        count[pid] += 1
    else:
        count[pid] = 1

split_3 = {}
for pid in count.keys():
    if count[pid] < 3:
        continue
    if count[pid] % 3 == 0:
        i = count[pid] / 3
        split_3[pid] = [i] * 3
    elif count[pid] % 3 == 1:
        i = (count[pid] - 1) / 3
        split_3[pid] = [i+1, i, i]
    else:
        i = (count[pid] - 2) / 3
        split_3[pid] = [i+1, i+1, i]

pid = list(split_3.keys())[0]
a = cannabis.loc[cannabis['Patient_ID'] == pid]
f = int(split_3[pid][0])
s = int(split_3[pid][1]) + f
t = int(split_3[pid][2]) + s
first = a[0:f]
second = a[f:s]
third = a[s:t]
for pid in list(split_3.keys())[1:]:
    a = cannabis.loc[cannabis['Patient_ID'] == pid]
    f = int(split_3[pid][0])
    s = int(split_3[pid][1]) + f
    t = int(split_3[pid][2]) + s
    frames = [first, a[0:f]]
    first = pd.concat(frames)

```

```

frames = [second, a[f: s]]
second = pd.concat(frames)
frames = [third, a[s: t]]
third = pd.concat(frames)
first.to_csv(r'/Users/macbook/Desktop/DATA3001/first.csv', index=False)
second.to_csv(r'/Users/macbook/Desktop/DATA3001/second.csv', index=False)
third.to_csv(r'/Users/macbook/Desktop/DATA3001/third.csv', index=False)

R code:
library("lubridate")
library("tsibbledata")
library("ggplot2")
library("GGally")
library("forecast")
library("feasts")      # for autoplot()
library("tidyverse")
library("magrittr")    # for %>%
library("tsibble")     # for yearquarter()
library("readxl")
library("fpp3")
library("pander")
library("ggpubr")
library("ggfortify")
library("broom")
first<-read_csv("/Users/macbook/Desktop/DATA3001/first.csv")
second<-read_csv("/Users/macbook/Desktop/DATA3001/second.csv")
third<-read_csv("/Users/macbook/Desktop/DATA3001/third.csv")
first %>% group_by('Patient_ID') %>% mean(first$Weighted_Efficacy)
# First time zone
Age_1_1 <- first %>% filter(`Patient_Age_Group` == "18-24")
Age_2_1 <- first %>% filter(`Patient_Age_Group` == "25-34")
Age_3_1 <- first %>% filter(`Patient_Age_Group` == "35-44")
Age_4_1 <- first %>% filter(`Patient_Age_Group` == "45-54")
Age_5_1 <- first %>% filter(`Patient_Age_Group` == "55+")

# second time zone
Age_1_2 <- second %>% filter(`Patient_Age_Group` == "18-24")
Age_2_2 <- second %>% filter(`Patient_Age_Group` == "25-34")
Age_3_2 <- second %>% filter(`Patient_Age_Group` == "35-44")
Age_4_2 <- second %>% filter(`Patient_Age_Group` == "45-54")
Age_5_2 <- second %>% filter(`Patient_Age_Group` == "55+")

# Third time zone
Age_1_3 <- third %>% filter(`Patient_Age_Group` == "18-24")
Age_2_3 <- third %>% filter(`Patient_Age_Group` == "25-34")

```

```

Age_3_3 <- third %>% filter(`Patient_Age_Group` == "35-44")
Age_4_3 <- third %>% filter(`Patient_Age_Group` == "45-54")
Age_5_3 <- third %>% filter(`Patient_Age_Group` == "55+")
WE_1_1<-aggregate(Age_1_1$Weighted_Efficacy, list(Age_1_1$Patient_ID), FUN=mean)
WE_2_1<-aggregate(Age_1_2$Weighted_Efficacy, list(Age_1_2$Patient_ID), FUN=mean)
WE_3_1<-aggregate(Age_1_3$Weighted_Efficacy, list(Age_1_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_1, WE_2_1, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_1, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")

# WE-1824
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)
print(WE_1)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)
print(WE_1)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Weighted Efficacy", title = "Average Weighted Efficacy in 18-24 age group in 3 periods")

WE_1_1<-aggregate(Age_1_1$Safety_Score, list(Age_1_1$Patient_ID), FUN=mean)
WE_2_1<-aggregate(Age_1_2$Safety_Score, list(Age_1_2$Patient_ID), FUN=mean)
WE_3_1<-aggregate(Age_1_3$Safety_Score, list(Age_1_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_1, WE_2_1, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_1, by = "Group.1")

```

```

colnames(WE_1) <- c("Patient ID", "Period 1", "Period 2", "Period 3")

# SS-1824
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)
print(WE_1)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)
print(WE_1)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Safety Score",
  title = "Average Safety Score in 18-24 age group in 3 periods")
WE_1_2<-aggregate(Age_2_1$Weighted_Efficacy, list(Age_2_1$Patient_ID), FUN=mean)
WE_2_2<-aggregate(Age_2_2$Weighted_Efficacy, list(Age_2_2$Patient_ID), FUN=mean)
WE_3_2<-aggregate(Age_2_3$Weighted_Efficacy, list(Age_2_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_2, WE_2_2, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_2, by = "Group.1")
colnames(WE_1) <- c("Patient ID", "Period 1", "Period 2", "Period 3")

# Running a t-test
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

```

```

t.test(preTreat, postTreat, paired = TRUE)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Weighted Efficacy", title = "Average Weighted Efficacy in 25-34 age group in 3 periods") + theme(legend.position = "none")

WE_1_2<-aggregate(Age_2_1$Safety_Score, list(Age_2_1$Patient_ID), FUN=mean)
WE_2_2<-aggregate(Age_2_2$Safety_Score, list(Age_2_2$Patient_ID), FUN=mean)
WE_3_2<-aggregate(Age_2_3$Safety_Score, list(Age_2_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_2, WE_2_2, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_2, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")

# T- test
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Safety Score",

```

```

title = "Average Safety Score in 25-34 age group in 3 periods") + theme(legend.position =
"none")
WE_1_3<-aggregate(Age_3_1$Weighted_Efficacy, list(Age_3_1$Patient_ID), FUN=mean)
WE_2_3<-aggregate(Age_3_2$Weighted_Efficacy, list(Age_3_2$Patient_ID), FUN=mean)
WE_3_3<-aggregate(Age_3_3$Weighted_Efficacy, list(Age_3_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_3, WE_2_3, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_3, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")

preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Weighted
Efficacy", title = "Average Weighted Efficacy in 35-44 age group in 3
periods") + theme(legend.position = "none")

WE_1_3<-aggregate(Age_3_1$Safety_Score, list(Age_3_1$Patient_ID), FUN=mean)
WE_2_3<-aggregate(Age_3_2$Safety_Score, list(Age_3_2$Patient_ID), FUN=mean)
WE_3_3<-aggregate(Age_3_3$Safety_Score, list(Age_3_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_3, WE_2_3, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_3, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")

```

```

# SS-1824
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Safety Score",
  title = "Average Safety Score in 35-44 age group in 3 periods") + theme(legend.position =
  "none")
WE_1_4<-aggregate(Age_4_1$Weighted_Efficacy, list(Age_4_1$Patient_ID), FUN=mean)
WE_2_4<-aggregate(Age_4_2$Weighted_Efficacy, list(Age_4_2$Patient_ID), FUN=mean)
WE_3_4<-aggregate(Age_4_3$Weighted_Efficacy, list(Age_4_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_4, WE_2_4, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_4, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")

preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

```

```

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Weighted Efficacy", title = "Average Weighted Efficacy in 45-54 age group in 3 periods") + theme(legend.position = "none")

WE_1_4<-aggregate(Age_4_1$Safety_Score, list(Age_4_1$Patient_ID), FUN=mean)
WE_2_4<-aggregate(Age_4_2$Safety_Score, list(Age_4_2$Patient_ID), FUN=mean)
WE_3_4<-aggregate(Age_4_3$Safety_Score, list(Age_4_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group for three time zone
# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_4, WE_2_4, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_4, by = "Group.1")
colnames(WE_1) <- c("Patient ID","Period 1","Period 2","Period 3")
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)
preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

```

```

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Safety Score", title = "Average Safety Score in 45-54 age group in 3 periods") + theme(legend.position = "none")

WE_1_5<-aggregate(Age_5_1$Weighted_Efficacy, list(Age_5_1$Patient_ID), FUN=mean)
WE_2_5<-aggregate(Age_5_2$Weighted_Efficacy, list(Age_5_2$Patient_ID), FUN=mean)

```

```

WE_3_5<-aggregate(Age_5_3$Weighted_Efficacy, list(Age_5_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone

# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_5, WE_2_5, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_5, by = "Group.1")
colnames(WE_1)<- c("Patient ID","Period 1","Period 2","Period 3")
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`
t.test(preTreat, postTreat, paired = TRUE)
preTreat <- WE_1$`Period 2`
postTreat <- WE_1$`Period 3`

t.test(preTreat, postTreat, paired = TRUE)

ID<- WE_1$`Patient ID`
df<- WE_1[,-1]
library(reshape2)
df<- melt(df) #the function melt reshapes it from wide to long
df$Patiene_ID <- ID

ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Weighted Efficacy", title = "Average Weighted Efficacy in 55+ age group in 3 periods") + theme(legend.position = "none")

WE_1_5<-aggregate(Age_5_1$Safety_Score, list(Age_5_1$Patient_ID), FUN=mean)
WE_2_5<-aggregate(Age_5_2$Safety_Score, list(Age_5_2$Patient_ID), FUN=mean)
WE_3_5<-aggregate(Age_5_3$Safety_Score, list(Age_5_3$Patient_ID), FUN=mean)
# What we have know is the average weighted efficacy for each patient in the first age group
for three time zone

# plot each patient's average weighted efficacy and compare them across different period
WE_1_0 <- merge(WE_1_5, WE_2_5, by = "Group.1")
WE_1 <- merge(WE_1_0,WE_3_5, by = "Group.1")
colnames(WE_1)<- c("Patient ID","Period 1","Period 2","Period 3")

# SS-1824
preTreat <- WE_1$`Period 1`
postTreat <- WE_1$`Period 2`

t.test(preTreat, postTreat, paired = TRUE)

preTreat <- WE_1$`Period 2`
```

```
postTreat <- WE_1$`Period 3`  
  
t.test(preTreat, postTreat, paired = TRUE)  
  
ID<- WE_1$`Patient ID`  
df<- WE_1[,-1]  
library(reshape2)  
df<- melt(df) #the function melt reshapes it from wide to long  
df$Patiene_ID <- ID  
  
ggplot(df, aes(variable, value, group=factor(Patiene_ID))) +  
  geom_line(aes(color=factor(Patiene_ID))) + labs(x = "Period", y = "Average Safety Score",  
  title = "Average Safety Score in 55+ age group in 3 periods") + theme(legend.position =  
  "none")
```

Appendix 3.4 Modelling code

Python code

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
import numpy as np
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import StandardScaler

df=pd.read_excel('model.xlsx')
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57152 entries, 0 to 57151
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Patient Gender    57152 non-null   object  
 1   Strain Type       57152 non-null   object  
 2   Symptom           57152 non-null   object  
 3   Ingestion Method  57152 non-null   object  
 4   Dosage Measure    57152 non-null   float64 
 5   Dosage Units      57152 non-null   object  
 6   Patient Age       57152 non-null   int64  
 7   Pre-medicating Score 57152 non-null   int64  
 8   THC                57152 non-null   float64 
 9   CBD                57152 non-null   float64 
 10  Weighted Efficacy 57152 non-null   float64 
dtypes: float64(4), int64(2), object(5)
memory usage: 4.8+ MB

df.isna().sum()
Patient Gender      0
Strain Type         0
Symptom             0
Ingestion Method    0
Dosage Measure      0
Dosage Units        0
Patient Age         0
Pre-medicating Score 0
THC                 0
CBD                 0
Weighted Efficacy   0
dtype: int64

corr=df[['Dosage Measure','Patient Age','Pre-medicating Score','THC','CBD','Weighted Efficacy']].corr()
plt.figure(figsize=(4,3),dpi=150)
sns.set(font_scale=0.5)
ax = sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns,
                  linewidths=0.1, cmap="YlOrRd_r", annot=True, annot_kws={'size': 5},
                  center=0)
plt.title("Correlation between variables")
plt.show()
df.columns
df.head(1)
cate=['Patient Gender', 'Strain Type', 'Symptom', 'Ingestion Method']
num=list((df.columns).difference(cate))

```

```

# One-hot encoded
df1=df.copy()
for col in cate:
    if df1[col].nunique() == 2:
        df1[col] = pd.factorize(df1[col])[0]
    else:
        df1 = pd.get_dummies(df1,columns=[col])
for i in num:
    df1[i]=df[[i]]

# train test split
X = df1.drop("Weighted Efficacy", axis=1)
y = df1["Weighted Efficacy"]
X_train, X_test, y_train, y_test =train_test_split(X,y,random_state=1)

# MSE, MAE, R2, RMSE
import numpy as np
from sklearn.metrics import mean_squared_error #MSE
from sklearn.metrics import mean_absolute_error #MAE
from sklearn.metrics import r2_score#R 2

# train
from sklearn import tree
tree = tree.DecisionTreeRegressor(max_depth=5)
# To prevent overfitting, set the highest depth
tree.fit(X_train, y_train)
# predict
y_pred = tree.predict(X_test)
# calculate mse
tree_mse = mean_squared_error(y_test, y_pred)
# calculate rmse
tree_rmse = np.sqrt(tree_mse)
# R2-score
tree_r2=r2_score(y_test,y_pred)

print('RMSE',tree_rmse,'MSE',tree_mse,'\nR2',tree_r2,'MAE',tree_mae)

#random forest
from sklearn.ensemble import RandomForestRegressor

forest = RandomForestRegressor(n_estimators=100)
forest.fit(X_train, y_train)
y_pred = forest.predict(X_test)
mse = mean_squared_error(y_test, y_pred)

```

```

rmse = np.sqrt(mse)
# R2-score
r2=r2_score(y_test,y_pred)
print('RMSE',rmse,'MSE',mse,'nR2',r2,'MAE',mae)

x = [i for i in X.columns]
y1=map(lambda x:round(x,4), forest.feature_importances_)
# The numerical value of feature importance
y=[i for i in y1]
plt.figure(figsize=(90,10),dpi=350)
plt.title("The importance of different features")
plt.bar(x,y)
for i in range(len(y)):
    plt.annotate(y[i],xy=(i,y[i]),xytext=(i-0.15,y[i]+0.001))

```



R code

```

#Code new variables within the dataset: 1 if a customer is ; 0 if not
dta.customer$Gender_Female <- ifelse(dta.customer$Patient_Gender== "Female", 1, 0)
dta.customer$Gender_male <- ifelse(dta.customer$Patient_Gender== "Male", 1, 0)
dta.customer$Time_of_Day_Morning <- ifelse(dta.customer$Time_of_Day== "Morning", 1, 0)
dta.customer$Time_of_Day_Evening<- ifelse(dta.customer$Time_of_Day== "Evening", 1, 0)
dta.customer$Time_of_Day_OverNight<- ifelse(dta.customer$Time_of_Day== "Over Night", 1, 0)
dta.customer$medicating_Score = dta.customer$Pre_medicating_Score-
dta.customer$Post_medicating_Score
dta.cus_clustering <- select(dta.customer, c(Patient_Age,
Gender_Female,Gender_male,Time_of_Day_Morning,Time_of_Day_Evening,Time_of_Day_OverNight,THC,CBD,Pre_medicating_Score))
sample <- sample(c(TRUE, FALSE),nrow(dta.customer), replace=TRUE, prob=c(0.7,0.3))
train_data<- dta.customer[sample,];
test_data <- dta.customer[!sample,]

```

#MLS of effecacy

```

fit1<- lm(Weighted_Efficacy~Patient_Age+ Gender_Female + Gender_male +
Time_of_Day_Morning+Time_of_Day_Evening + Time_of_Day_OverNight + THC +
CBD+Dosage_Measure+Ingestion_Method ,data = train_data)
summary (fit1)

```

```

```
fit2<- lm(Weighted_Efficacy~Patient_Age+
Gender_Female+Gender_male+Time_of_Day_Morning+Time_of_Day_Evening
+Time_of_Day_OverNight+THC+CBD+medicating_Score+Post_medicating_Score+
Dosage_Measure +medicating_Score+Ingestion_Method,data = train_data)
summary (fit2)
#
train_data$Patient_Age_square=train_data$Patient_Age^2
train_data$THC_square=train_data$THC^2
train_data$CBD_square=train_data$CBD^2
train_data$Pre_medicating_Score_square=train_data$Pre_medicating_Score^2

fit4<- lm(Weighted_Efficacy~Patient_Age_square+
Gender_Female+Gender_male+Time_of_Day_Evening
+Time_of_Day_OverNight+THC_square+CBD_square+medicating_Score+Post_medicating
_Score+Ingestion_Method+Pre_medicating_Score_square + Pre_medicating_Score
*Post_medicating_Score+Pre_medicating_Score,
 data = train_data)

summary (fit4)
##check multicollinearity
if (!"car" %in% installed.packages()) install.packages("car")
library(car)
vif(fit1)
vif(fit2)
#vif(fit4)
train_data$pos_eff<-ifelse(train_data$Weighted_Efficacy > 0,1,0)
calculate fitted probabilities
Prob.Fit <- fitted(fit1, type = "response")

predict choice based on the probabilities
Choice.Pred <- ifelse(Prob.Fit<0.5, 0, 1)

calculate success rate of prediction
Success <- 1 - mean((Choice.Pred - train_data$pos_eff)^2)
Success

fit5<- lm(Safety_Score~Patient_Age+ Gender_Female + Gender_male +
Time_of_Day_Morning +Time_of_Day_Evening + Time_of_Day_OverNight + THC +
CBD+Dosage_Measure+Ingestion_Method ,data = train_data)
summary (fit5)

fit6<- lm(Safety_Score~Patient_Age+ Gender_Female + Gender_male +
Time_of_Day_Morning +Time_of_Day_Evening + Time_of_Day_OverNight + THC +
CBD+Dosage_Measure+Ingestion_Method +medicating_Score

```

```

,data = train_data)
summary (fit6)

fit7<- lm(Safety_Score~Patient_Age+ Gender_Female + Gender_male +
Time_of_Day_Morning +Time_of_Day_Evening + Time_of_Day_OverNight + THC +
CBD+Dosage_Measure+Ingestion_Method
+medicating_Score+Pre_medicating_Score+Post_medicating_Score,data = train_data)
summary (fit7)

vif(fit5)
vif(fit6)
#vif(fit7)

train_data$pos_safe<-ifelse(train_data$Safety_Score > 0,1,0)
calculate fitted probabilities
Prob.Fit <- fitted(fit6, type = "response")

predict choice based on the probabilities
Choice.Pred <- ifelse(Prob.Fit<0.5, 0, 1)

calculate success rate of prediction
Success <- 1 - mean((Choice.Pred - train_data$pos_safe)^2)
Success

library(rpart);
model_rpart <-rpart(Weighted_Efficacy~Patient_Age+
Gender_Female+Gender_male+Time_of_Day_Morning+Time_of_Day_Evening
+Time_of_Day_OverNight+THC+CBD+medicating_Score
,data = train_data)
summary(model_rpart)
model_rpart_safety<-rpart(fit6)
summary(model_rpart_safety)

plotcp(model_rpart)
plotcp(model_rpart_safety)

plot(model_rpart,margin = 0.1)
text(model_rpart,all = T,use.n = T)
plot(model_rpart_safety,margin = 0.1)
text(model_rpart_safety,all = T,use.n = T)

min(model_rpart$cptable[, "xerror"])
which.min(model_rpart$cptable[, "xerror"])
prune.tree<-prune(model_rpart,cp=model_rpart.cp)

```

```
plot(prune.tree,margin = 0.1)
text(prune.tree,all = T,use.n = T)
printcp(model_rpart)
printcp(model_rpart_safety)
```
if (!"rcompanion" %in% installed.packages()) install.packages("rcompanion")
library(rcompanion)
accuracy(list(fit1,fit2,model_rpart))
accuracy(list(fit5,fit6,model_rpart_safety))
```