Difference in medication trajectories of people with type 2 diabetes according to sex and presence of depression

Kitty Ha (1851942)

First supervisor: Daniala Weir

Second supervisor: Helga Gardarsdottir

July 1, 2022

Master Applied Data Science, Utrecht University

Abstract

Objective

Determine if medication trajectories differ between subgroups of individuals among those with type 2 diabetes and analyse what those differences are. Specifically, look at the differences between patients according to their sex and if they have depression or not.

Methods

The data used in this study is extracted from the CPRD, and medication trajectories for patients were created based on this data. Then, a similarity metric was used to determine if there is a difference in medication trajectories and to create clusters using Ward's method. Furthermore, the most occurring trajectories for each subgroups and the top 3 clusters are used to analyse what the differences between the subgroups are.

Results

For both the sex and depression subgroups, all the similarity scores are between $0.23 - 0.33$. Also, for all the subgroups, the proportion of the most occurring trajectories are all below 0.1. All the representative trajectories of the top 3 clusters take up only 20% of the subgroups. So, there is a lot of variability between the trajectories in the subgroups.

Conclusion

There is a difference in medication trajectories between men and women, and between patients with and without depression. However, even within the subgroups there is a lot of difference. For the sex subgroups, both the male and female groups consists of medication from the Cardiovascular System and Endocrine System. The female subgroup with a diabetes duration of $15 - 20$ years also contains medication from the Infection category. For the depression subgroups, all groups consist of Cardiovascular System and Endocrine System medication. The groups with depression also contain medication from the Central Nervous System, where the position of the medication in the trajectory depends on when depression was diagnosed.

1. Introduction

Diabetes is one of the most common chronic non-communicable diseases, and in North America it affects 7% of individuals and it is a leading cause of death.[1] Furthermore, most patients with diabetes have multimorbidity (i.e., at least one other chronic condition),[2] and as many as 25% have four or more concurrent chronic conditions.[3] Polypharmacy is the simultaneous and prolonged prescription of multiple medications (usually five or more) to a single individual and is the norm in multimorbidity.[4] Common comorbidities of diabetes typically require treatment with one or more medications for adequate control, resulting in a significant total number of possible medications for these complex patients.[5] Furthermore, patients with diabetes using multi-drug regimens are at especially high risk of experiencing adverse drug events.[6] Prescribing for patients with multimorbidity is not a straightforward process given that physicians are often faced with the dilemma to either prescribe a recommended medication that may lead to adverse effects, or not to prescribe a medication that may have potential benefits (e.g., statins).[7]

In addition to the clinical challenges associated with treating complex patients with limited evidence-based guidance, there is a large volume of research suggesting that the treatment of diabetes, hypertension, and cardiovascular differs between patient subgroups.[8,9] More specifically, several studies have revealed disparities in diabetes treatment depending on the sex of the patient.[10-12] One study reported that women tend to be prescribed more medications than men and were also less likely to be prescribed medication that is recommended by current guidelines.[10] Another study reported that men with diabetes were more likely to receive oral combination drugs, ACE inhibitors and calcium channel blockers than women.[12] Furthermore, other research have reported that there are differences in treatment for diabetic patients with and without serious mental illness.[13,14] One study found that patients with serious mental illness were less likely to be prescribed certain medication (e.g. cholesterol-lowering statin medications and angiotensin-converting enzyme inhibitors) than patients without serious mental illness.[13] Another study found that people with serious mental illness were 30% less likely to be prescribed medications known to be effective in the treatment of CVD and diabetes.[14]

To optimize and personalize drug regimens in diabetes, it is important to characterize, and understand current trajectories of treatment patterns among different subgroups of patients. Previous studies have focused on diabetes specific medication trajectories alone.[15-17] However, little attention has been devoted to analysing medication trajectories for diabetic people with multiple chronic conditions and specifically the differences in these trajectories between subgroups of patients according to sex and depression. Therefore, the research question of this paper is: Is there a difference in medication trajectories between subgroups according to sex and the presence of depression among those with type 2 diabetes and what are those differences?

1.1 Review of methods that could be used to answer the research question

There has not been a lot of research published in the area of analysing medication trajectories and the differences of these trajectories between subgroups of patients. However, there are several studies in other research areas that have focussed on analysing trajectories and determining if there is a difference between trajectories. One study evaluated whether intergenerational continuity exists in the demographic trajectories of parents and children. This was done by analysing how similar the trajectories are using optimal matching (OM).[18] Optimal matching is an algorithm that determines the dissimilarity between two sequences by calculating the 'cost' of transforming one sequence into another. This method was specifically developed to analyse sequences in the social sciences[19] and is similar to the edit distance, which is a term used to describe a group of distant metrics used to determine how dissimilar for example two words are. This study reported that intergenerational transmission of demographic trajectories does exists, meaning that children have similar life trajectories (e.g. with regards to marriage and parenthood) as their parents. Similarly, another study analysed the sequences of psychiatric disorders diagnoses of people with eating disorders using optimal matching to compute a sequence dissimilarity matrix and then clustered the trajectories using hierarchical agglomerative clustering with Ward's algorithm. [20] Ward's method clusters in a way that minimizes the total within-cluster variance. This study identified five clusters among individuals with an eating disorder. Furthermore, one study looked at (differences in) trajectories of men and women in the context of employment quality (measured using employment stability, material rewards, working-

time arrangements, collective organization, and interpersonal power relations).[21] This study used Dynamic Hamming distances and Ward's method to obtain a similarity matrix and cluster trajectories, respectively. Dynamic Hamming distances is one of the edit distance metrics that measures the distance between two sequences. This study found that women tended to have worse employment quality than men. Another study looked at patterns in patient disease trajectories using dynamic time warping to cluster trajectories for both men and women.[22]

The methods used in previous studies consists of determining the difference between trajectories by using a dissimilarity measure (either optimal matching or Dynamic Hamming distances) and then clustering these trajectories using Ward's method. The use of a dissimilarity measure and Ward's method is also applicable in this study to determine if there is a difference in medication trajectory and clustering the trajectories. However, a different dissimilarity measure than optimal matching or Dynamic Hamming distances might be more appropriate to determine the differences in medication trajectories. This is because optimal matching does not necessarily focus on whether the values in the sequences are consecutive or not, which might be important in medication trajectories. Furthermore, Dynamic Hamming distances assumes that all trajectories are the same length which is not always the case for medication trajectories.

2. Data

2.1 Study design

This retrospective cohort study uses longitudinal, representative patient data collected from general practitioner practices across the UK, the Clinical Practice Research Datalink (CPRD).

2.2 Data source

The data used in this research is obtained from The Clinical Practice Research Datalink GOLD (CPRD GOLD) and contains primary care data from general practitioners in the United Kingdom. This data contains information of 60 million patients, including 16 million currently registered patients as of 2021.[23] Furthermore, the patients included in the data are representative of the UK general population

in terms of age, sex and ethnicity.[24] The CPRD includes data on demographic information about the patients, clinical information (e.g. diagnosis), symptoms, referrals and prescriptions of the patients. This data is recorded using READ codes, which is a hierarchical clinical classification system containing over 96,000 codes.[24] The prescription data consists of both drug and appliance prescription; further information about these products, e.g. the drug substance of the product, are also included in the data. These prescriptions are recorded with a product name and British National Formulary (BNF) code. The BNF is a reference book that provides guidance on the administrating, dispensing and prescribing of medication.[25] The medications in the book are encoded using hierarchical codes, which provides information about the drug. An example of a BNF code is shown in Table 1.

Chapter	Section	Paragraph	Sub paragraph	Chemical substance	Product	Strength & formulation	Generic equivalent
04	07	02	0	40	BI	AC	AM
Central nervous system	Analgesics	Opioid analgesics	Opioid analgesics	Tramadol hydrochloride	Tradorec	Tradorec XL_tab 300 mg	Refers to strength & formulation

Table 1. Example of BNF code structure

The BNF code in the CPRD data only includes the first seven digits, so the sections Chapter, Section, Paragraph and Sub paragraph.

2.3 Study population

For this study, the population of interest is people who had at least one oral antidiabetic prescribed after 2013. Records prior to 2013 for these patients are also included. Furthermore, patients were excluded if they were prescribed any medication before 2000 to reduce computational intensity.

2.4 Study measure

For this study, all the products that are not medication are excluded (e.g. bandages and catheters) as the focus is on the drug trajectories of each patient. The trajectories are formed based on the BNF chapter of the products (e.g. 1. Gastro-Intestinal System), which is the first digit of a BNF code. These

chapters relate the medications to either a certain body system (e.g. Central Nervous System), or to a certain element of medical care (e.g. Infections).[26] Furthermore, only the medications that belong to the first 10 out of the 23 chapters are included. This was done to further simplify the cohort, and also because the chapters beyond chapter 10 are mostly non-systemic medications. The full list of all BNF chapters can be found in Appendix A.

2.5 Drug exposure

Furthermore, as this study is specifically focused on chronic conditions, for every patient only chronic medication are kept which is defined as a medication that the patient takes for at least 90 days consecutively. This was done by taking the difference between the date of the first occurrence of a prescription of a certain medication and the date of the last prescription of this medication. The assumption is made here that the medication is used daily in between those dates.

2.6 Age and sex measurement

To create subgroups of patients, the data was stratified according to sex and the presence of depression. The subgroup with people who do have depression, include patients who have had a READ code for depression (see Appendix B). Patients who have never had a READ code for depression are grouped in the 'No depression' subgroup. Furthermore, patients who have had a depression READ code, but are missing the date of diagnosis are excluded.

The depression subgroup is further divided into three groups:

1. Existing depression: these are the patients who were diagnosed with depression prior to the date of their first diabetic prescription.
2. Early depression: these are the patients who were diagnosed with depression in the first half of the time period that they were prescribed diabetes medication (e.g. if a patient has been prescribed medication for 10 years, they were diagnosed with depression within the first 5 years).

3. Late depression: these are the patients who were diagnosed with depression in the second half of the time period that they were prescribed diabetes medication.

2.7 Diabetes duration

All the subgroups are then further divided according to their duration of diabetes. Specifically, the patients in these subsets are divided into four intervals: 0 – 5 years, 5 – 10 years, 10 – 15 years and 15 – 20 years. This is so that patients with similar diabetes duration are compared. The subgroups are mutually exclusive categories, meaning that patients in one category do not appear in another.

2.8 Ethical and legal considerations of the data

The data in the CPRD is anonymized and no information that can identify a patient is ever sent to CPRD.[24] Furthermore, 'the CPRD has broad National Research Ethics Service Committee (NRES) ethics approval for purely observational research using the primary care data and established data linkages.'[27]

3. Methods

3.1 Relate research question and data science question

The research question is: Is there a difference in medication trajectories between subgroups according to sex and the presence of depression among those with type 2 diabetes and what are those differences?

Translating the research question to a data science question results in the following two questions:

1. How to establish if there is a difference between sequences?

2. How to create representations of a large amount of sequences and compare them?

3.2 Data preparation

Any patients who had a medication prescription after the date of death or after the exit date out of the database were removed. Furthermore, patients with either an extremely young age (< 5 years) or old

age (>120 years) were excluded. Finally, for every patient, the trajectories are constructed by chronologically combining the medications that are prescribed to the patients; where only the first prescription of a certain medication is kept. Every patient now has one row which consist of their medication trajectory using the BNF chapters (e.g. $(1-5-2-7)$).

3.3 Trajectory description

To get a better understanding of the medication trajectories, the trajectories are described for each of the sex and depression subgroups. First, count how many unique trajectories there are, meaning the number of trajectories excluding the duplicates of these trajectories; so only counting each medication trajectory once. Then, compute the median number of medications in the trajectories. Furthermore, examine which medications are usually prescribed first, this is done by counting how many times a certain medication appears in the first position of a trajectory.

3.4 Trajectory similarity

To determine if there is a difference between medication trajectories of the sex subgroups and depression subgroups, the pairwise similarity of the trajectories in the subsets is measured which will result in a similarity matrix. These measures determine the similarity of each pair of trajectories based on the cost it takes to transform one sequence into another. To determine these costs, the edit distance is used which determines how similar two strings are based on the minimum number of operations it takes to go from one string (e.g. a word) to another. These operations are deletion, insertion and substitution of a character and each of these operations has a certain cost. The edit distance can not only be used to determine how dissimilar two strings are, but can also compare two sequences. In this study, a version of the edit distance is used where only the operations insertion and deletion are allowed with each a cost of one in combination with the Longest Common Substring distance. This method essentially determines the length of the longest consecutive subsequence and then uses that to determine how many insertions and deletions are needed to transform one sequence into another. An example of this method is shown below.

1. A – B – C – D

2. P – B – C – F – G

The sequences have two letters in common which also appear consecutively (B and C). To transform the first sequence into the second, edits need to be made at position 1, 2 and 5. For example, at the first position, the letter 'A' needs to be deleted and then letter 'P' is inserted. The cost of insertion and deletion is 1, so the cost of this edit is 2. The total cost of transforming the first sequence into the second is then 5.

This method is applied to the pairwise trajectories of the two subgroups to be compared. So, each trajectory of one subset is compared to each trajectory of the other subset and a similarity matrix is created which shows the cost of all the pairwise comparisons. Then, similarity score S(a,b) for each trajectory pair can be obtained by

$$S(a,b) = 1 - \frac{d(a,b)}{\text{length a } + \text{ length b}}$$

where d(a,b) is the distance between trajectory a and b, and is divided by the sum of the lengths of both sequences as this is the maximum number of edits it can take to transform one trajectory into the other. The overall average similarity between two subgroups is then the average of these similarity scores, where a higher similarity score means that the trajectories are more similar. Furthermore, a similarity score of 1 means that the trajectories are exactly the same, so 0% of the trajectory needs to be changed. A similarity of 0 means that two trajectories are completely different, and 100% of the trajectory needs to be changed to transform one trajectory into the other. This is also done for each subgroup separately, so all the trajectories within one subgroup is compared to the trajectories of that same subgroup. This will show how similar the trajectories are within a subgroup. The similarity matrix that is computed for each subgroup separately, will also be used to cluster the trajectories within a subgroup.

3.5 Trajectory difference

To determine how these trajectories differ, assess which medication trajectories occur the most in each of the sex and depression subgroups and compare them. Furthermore, also cluster the trajectories of each subgroup based on the similarity matrix. Specifically, hierarchical agglomerative clustering with Ward's algorithm is used. With hierarchical agglomerative clustering, each observation starts in its own cluster and then pairs of clusters are combined based on a linkage criteria until all observations are in one cluster. The linkage criteria determines which clusters are combined based on the distance between these clusters. In this case, the Ward's method is used as linkage, which combines clusters together that minimizes the variances within the clusters. The optimal number of clusters is decided by looking at the dendrogram, which is a diagram that shows the distances at which clusters are merged (see Appendix C). Figure 1 shows an example of a dendrogram, which shows that a good distance to cut the tree would be around 30, as the distance between the last two clusters to be merged is suddenly way higher than when merging the previous clusters. Cutting the tree at this distance then results in a total of two clusters.
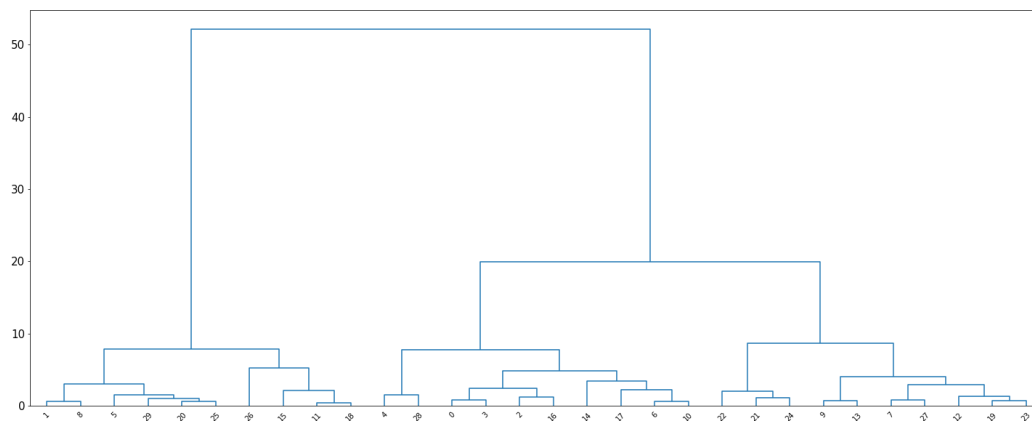
Figure 1. Example of a dendrogram with on the x axis the observations and on the y axis the distances between clusters

After obtaining the clusters, the top 3 clusters are selected based on two criteria's: 1. The average distance of the cluster and 2. The number of trajectories in the cluster. The average distance of the cluster is the sum of the pairwise distances between all the trajectories in the cluster divided by the total number of trajectories in that cluster. Here, a cluster with a lower average distance is preferred

over a cluster with a higher average distance, because this suggests that the trajectories within that cluster are more similar. The second criteria computes the total number of trajectories in a cluster (also accounting for trajectories that occur multiple times). A cluster with a higher number of trajectories is preferred, as too small of a cluster does not capture the data as well as larger clusters. Then, for each of these 3 clusters, a representative trajectory is selected similarly to how the best clusters are selected. So, a trajectory is selected based on how many times this trajectory appears in the data, and the average distance of this trajectory to all the other trajectories.

4. Results

Variable

Total patients	181,322
Age mean (Q1 – Q3)	65 (55-77)
Sex n (%)	
Male	101842 (56.2%)
Female	79480 (43.8%)
Depression n (%)	
Existing depression	28015 (15.5%)
Early depression	13979 (7.7%)
Late depression	11071 (6.1%)
No depression	128181 (70.7%)
Follow-up years mean (Q1 – Q3)	9.15 (4-14)
Diabetes duration n (%)	
0-5 years	60930 (33.6%)
5-10 years	41258 (22.8%)
10-15 years	45974 (25.4%)
15-20 years	33160 (18.3%)

Table 2. Cohort description

As can be seen in Table 2, there are a total of 181,322 patients in this cohort with an average age of 65. A large percentage (70%) of the patients do not have depression, and about half of the patients who do have depression were diagnosed with depression before their first diabetic medication prescription (Existing depression).

4.1 Trajectory description

4.1.1 Sex groups

		Total number of trajectories n (%)		Number of unique trajectories		Number of medications per trajectory median (Q1-Q3)	
		Male	Female	Male	Female	Male	Female
Diabetes duration	0-5	32664 (32.1%)	28266 (35.6%)	8448	9469	3 (2-5)	4 (3-5)
	5-10	23108 (22.7%)	18150 (22.8%)	9233	10298	4 (3-6)	5 (4-7)
	10-15	26881 (26.4%)	19093 (24.0%)	12204	12430	5 (3-6)	6 (4-7)
	15-20	19189 (18.8%)	13971 (17.6%)	10730	10519	5 (4-7)	6 (5-8)

Table 3. Medication trajectory description according to sex

Table 3 shows the medication trajectory description according to sex for different diabetes duration. The female trajectories have a higher median number of medications per trajectory across the different diabetes durations than male trajectories. So, women are prescribed slightly more medication than men. Furthermore, male groups have a higher number of total of trajectories compared to the female groups respective to the diabetes duration. In contrast, all female groups except the group with a duration of 15-20, have a higher number of unique trajectories than the male groups. This means that in the female groups fewer people have a trajectory that is exactly the same as another trajectory in that group.
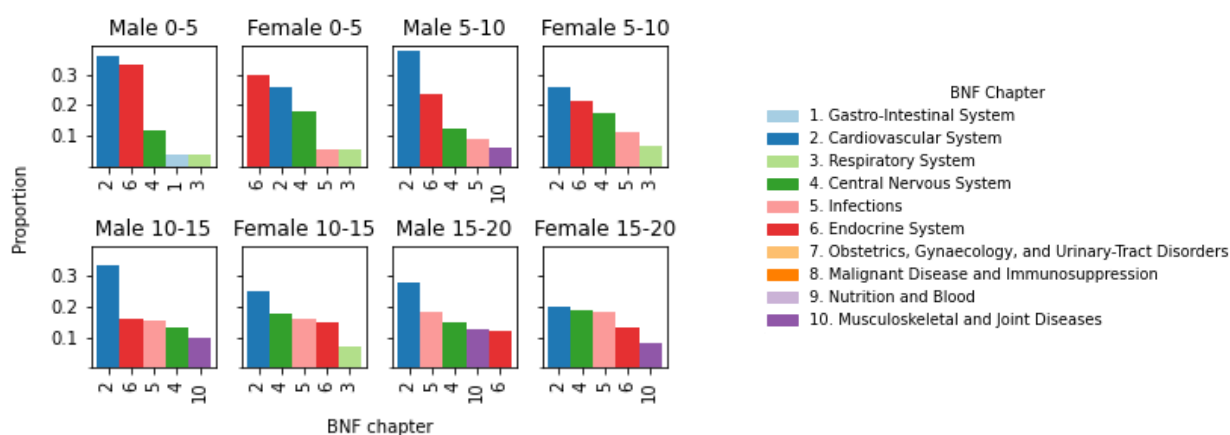
Figure 2. The top 5 BNF chapters that are prescribed first according to sex

As shown in Figure 2, for all the male subgroups, the medication that is prescribed the most as first medication belongs to the Cardiovascular System category. This is the same for all the female subgroups as well, except for the subgroup with a diabetes duration of 0-5 years where the medication that is prescribed the most as first medication belongs to Endocrine System category.

4.1.2 Depression groups

		Total number of trajectories n (%)				Number of unique trajectories				Number of medications per trajectory median (Q1-Q3)			
		ND	ExD	EaD	LD	ND	ExD	EaD	LD	ND	ExD	EaD	LD
Diabetes duration	0-5	43780 (34.2%)	13382 (47.8%)	2104 (15.1%)	1626 (14.7%)	10544	5985	1439	1097	3 (2-5)	4 (3-6)	5 (3-6)	4 (3-6)
	5-10	29060 (22.7%)	6653 (23.7%)	3064 (21.9%)	2468 (22.3%)	11621	4629	2416	1925	4 (3-6)	6 (4-7)	6 (4-7)	5 (4-7)
	10-15	32483 (25.3%)	5061 (18.1%)	4686 (33.5%)	3729 (33.7%)	14950	4016	3817	3077	5 (3-6)	6 (5-7)	6 (5-7)	6 (5-7)
	15-20	22858 (17.8%)	2919 (10.4%)	4125 (29.5%)	3248 (29.3%)	12998	2528	3606	2882	5 (4-7)	6 (5-8)	7 (5-8)	6 (5-8)

Table 4. Medication trajectory description according to the depression groups

Note: ND = No depr., ExD = Exist. depr., LD = Late depr,, EaD = Early depr.

The medication trajectory description of the depression groups across the different diabetes durations are shown in Table 4. The trajectories of patients without depression have a lower median number of medications compared to the patients with depression (existing, early and late depression). So, people with no depression are prescribed less medication than people with depression. Also, the group with patients who do not have depression has the highest number of total trajectories and also the highest number of unique trajectories (for all diabetes durations).
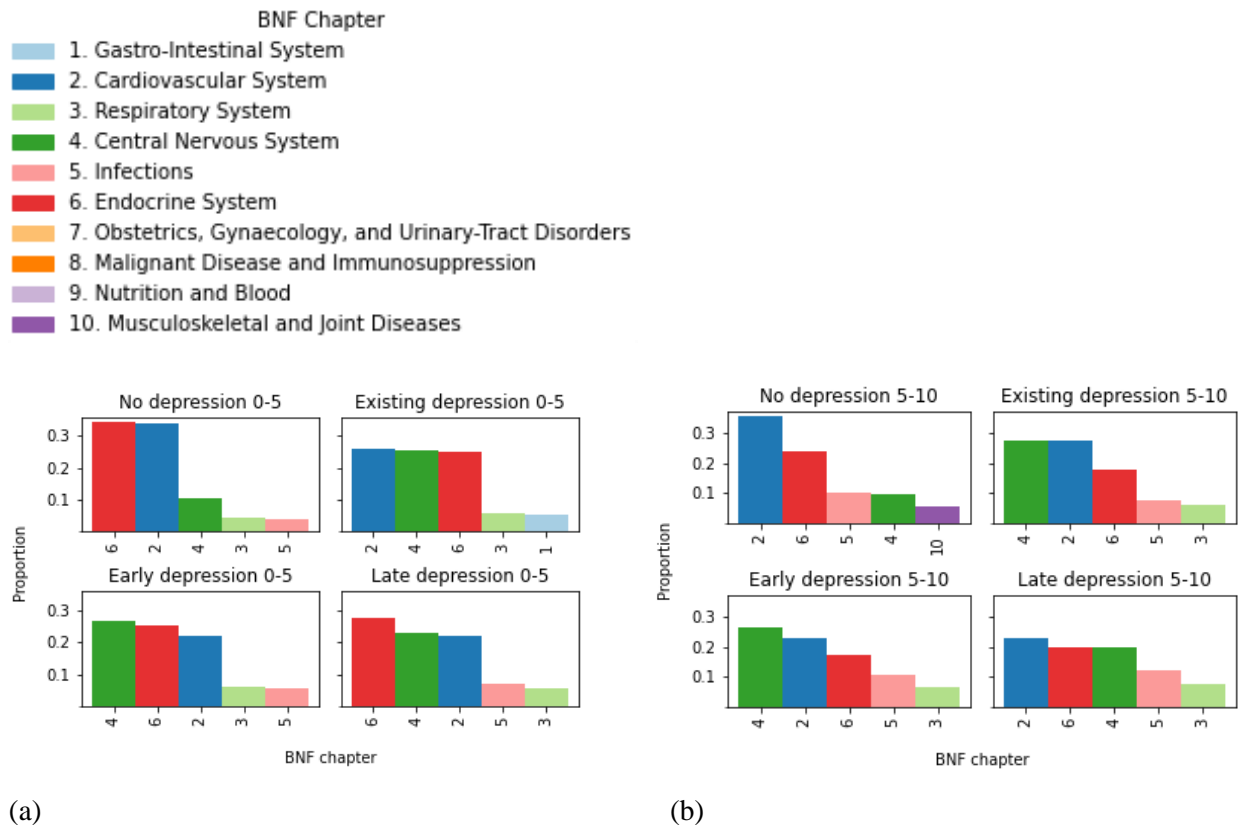
Figure 3. (a). The top 5 BNF chapters that are prescribed first according to the depression groups with duration 0-5 years, (b). The top 5 BNF chapters that are prescribed first according to the depression groups with duration 5-10 years.

Figure 3a shows that the no depression subgroup and the late depression subgroup both have medication belonging to the Endocrine System category as the most prescribed first medication. The existing depression and early depression subgroups have medication from the Cardiovascular System and Central Nervous System as the most prescribed first medication, respectively. Figure 3b shows that for both the no depression subgroup and the late depression subgroup, the medication that is prescribed the most as first medication belongs to the Cardiovascular System category. The existing depression and early depression subgroups both have medication from Central Nervous System category as the most prescribed first medication.

Figure 4a shows that for the no depression subgroup the medication that is prescribed the most as first medication belongs to the Cardiovascular System category. The subgroups with patients who do have depression all have medication of the Central Nervous System as the most prescribed first medication. Figure 4b shows that this is the same for the subgroups with a diabetes duration of 15-20 years.
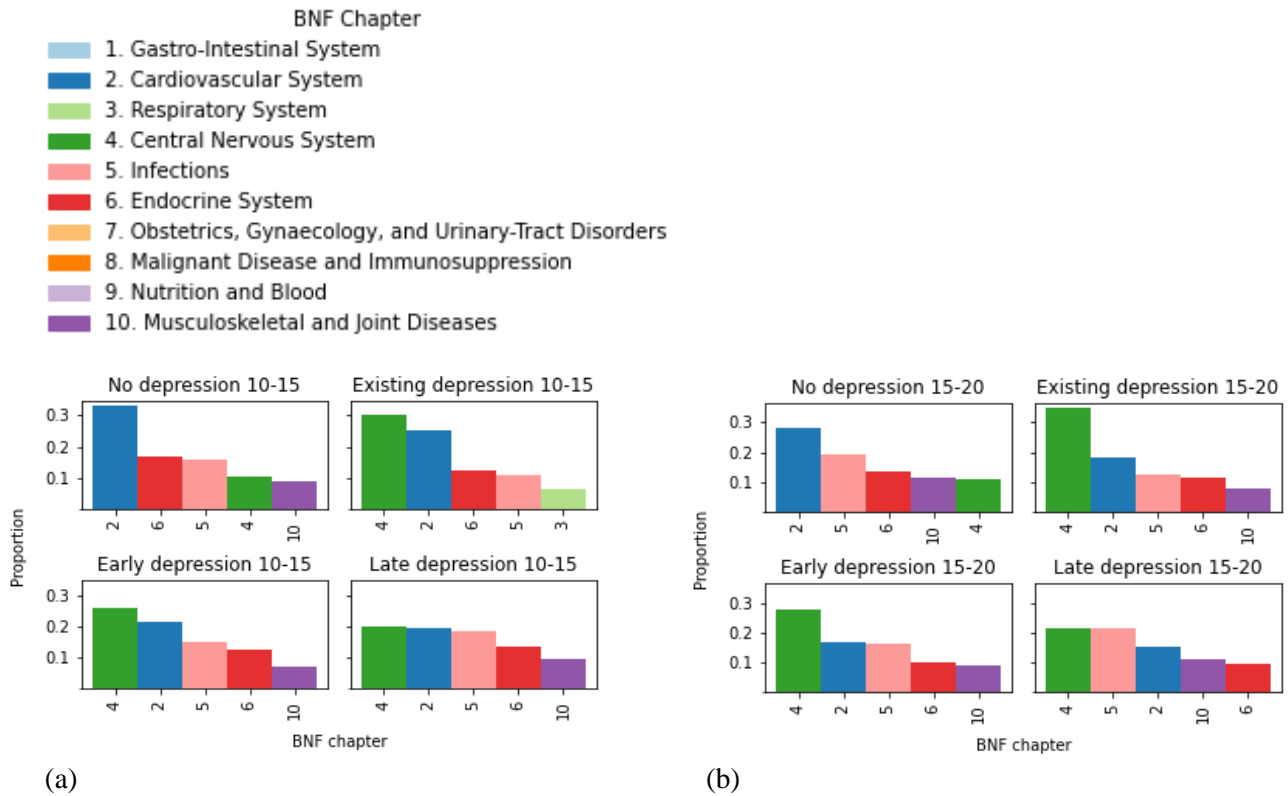
Figure 4. (a). The top 5 BNF chapters that are prescribed first according to the depression groups with duration 10-15 years, (b). The top 5 BNF chapters that are prescribed first according to the depression groups with duration 15-20 years.

Overall, an Cardiovascular System medication is the most prescribed first medication for the no depression subgroups. For the early depression and existing depression subgroups, the most prescribed first medication belongs to the Central Nervous System. For the late depression subgroups, the medication that is prescribed first differs depending on the diabetes duration.

4.2 Similarity scores

Groups	Diabetes duration			
	0-5	5-10	10-15	15-20
Male – Female	0.33	0.28	0.26	0.25
Male – Male	0.36	0.31	0.29	0.27
Female – Female	0.31	0.27	0.25	0.23

Table 5. Similarity scores of the medication trajectories according to sex

Table 5 shows the similarity scores for the subgroups male and female (Male – Female), and also the similarity scores of the trajectories within the two sex group separately (Male – Male and Female – Female). The Male – Male similarity scores show the similarity between the trajectories in the male subgroup and the Female – Female similarity scores show the similarity of the trajectories for the female subgroup. Here, a lower similarity score means that the trajectories are less similar and more changes need to be made to transform one trajectory into another. The similarity scores for the male-female groups ranges from 0.25-0.33, which means that on average about 70% of a male trajectory needs to be changed to transform it to a female trajectory. So, the medication trajectories between men and women do differ. If we look at the Male – Male similarity scores, it shows that the trajectories within the male groups are more similar to each other than the Male – Female trajectories. However, still about 60-70% of a trajectory needs to be changed, which shows that even within the male group the male trajectories are quite different from each other. Finally, the female similarity scores are lower than both the Male – Male and Male – Female similarity scores. This means that there is a lot of variability among the female trajectories. Furthermore, the similarity scores of the lower diabetes duration groups are higher than the groups with a higher diabetes duration, which means that the trajectories of the subgroups with a lower diabetes duration are more similar than the trajectories of the subgroups with a higher diabetes duration. So, there is a difference between medication trajectories according to sex, however there is also a difference within the sex groups.

Groups	Diabetes duration			
	0-5	5-10	10-15	15-20
No depr. – Exist. depr.	0.32	0.27	0.26	0.25
No depr. – Early depr.	0.30	0.26	0.25	0.24
No depr. – Late depr.	0.31	0.27	0.25	0.24
Exist. depr.– Early depr.	0.29	0.26	0.24	0.23
Exist. depr – Late depr.	0.29	0.26	0.24	0.23
Early depr. – Late depr.	0.29	0.26	0.24	0.23
No depr. – No depr.	0.35	0.30	0.28	0.26
Exist. depr. – Exist. depr.	0.30	0.26	0.24	0.24
Early depr. – Early depr.	0.29	0.25	0.24	0.23
Late depr. – Late depr.	0.29	0.26	0.24	0.23

Table 6. Similarity scores of the medication trajectories according to the depression groups

In Table 6 the similarity scores are shown of the medication trajectories according to the depression groups. Here also, the similarity scores of the trajectories within each of the depression groups separately are shown. The similarity scores of the subgroups where the no depression group is being compared to the groups with depression ranges from 0.24-0.32, which means that on average more than 70% of a trajectory needs to be changed to transform a no depression trajectory into a trajectory with depression. Furthermore, the similarity scores of the comparisons of the subgroups with depression (e.g. comparing Late depr. – Exist. depr) are all the same (0.23 – 0.29). So, about 75% of a trajectory needs to be changed to transform it to another trajectory. For example, to turn a late depr. trajectory into an exist. trajectory, 75% of the late depr. trajectory needs to be changed. So, there is a difference between the trajectories for these subgroups. Furthermore, these similarity scores are close to the similarity scores of the within subgroups (e.g. Late depr. – Late depr.). This suggests that the existing depression, early depression and late depression group differ just as much from each other as they differ in their own respective groups. So, there is a difference between the trajectories according to the presence of depression. However, there is also a relatively large difference within the subgroups.

4.3 Trajectory difference

4.3.1 Most occurring trajectories

For every subgroup, the most occurring trajectories were selected. Furthermore, the proportion of those most occurring trajectories are also shown, which is calculated by counting how many times that trajectory appears in the subgroup and dividing it by the total number of trajectories in that subgroup.

Figure 5 shows the most occurring trajectories for the sex subgroups. For all the sex subgroups, the most common trajectories consists of medication belonging to the Cardiovascular System and the Endocrine System. This is not unexpected, as these medications are likely used to treat diabetes. The subgroup female with a diabetes duration of 15-20 years also includes a medication belonging to the Infections category. Overall, looking at the most occurring trajectories, the trajectories do not seem to differ much between the male and female subgroups. However, Figure 5 also shows the proportion of the most occurring trajectory for each sex subgroup. It shows that all have a proportion lower than 0.1, and that the top 2 most occurring trajectories do not take up a large percentage of the subgroup. There is also more variation in subgroups with longer diabetes duration, as the proportion decreases with longer duration. So, according to these most occurring trajectories, there is not a big difference between male and female trajectories. However, this is based on a relatively small part of the subgroups as the most occurring trajectories only take up a small part of the subgroups.
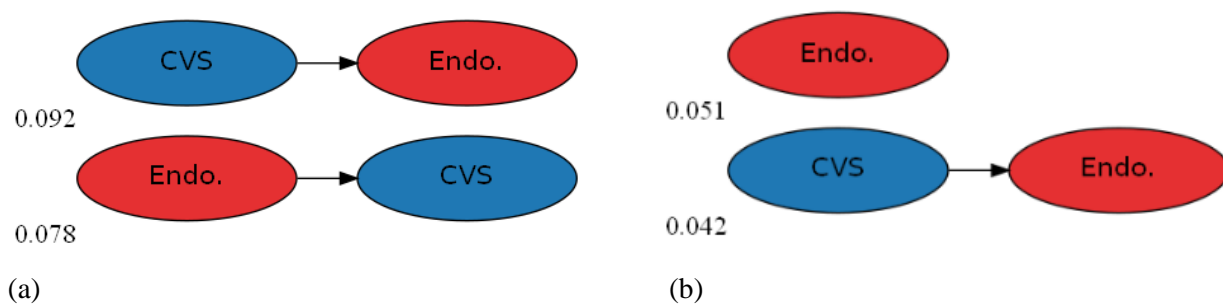
Figure 5. Most common trajectories for (a). male 0-5, (b). female 0-5
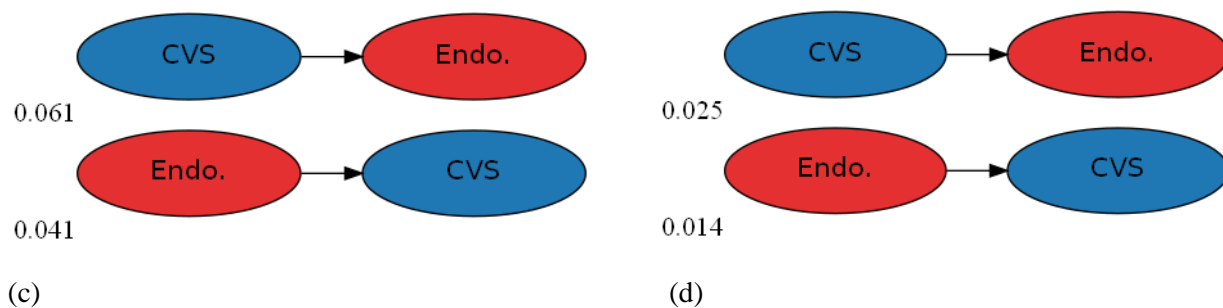
Figure 5. Most common trajectories for (c). male 5-10, (d). female 5-10
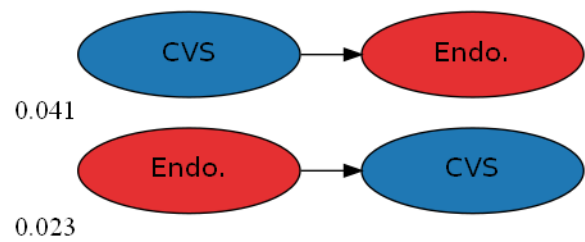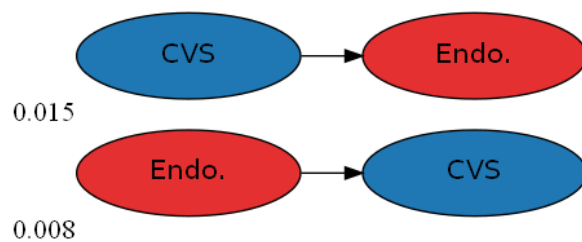
(e)   (f)

Figure 5. Most common trajectories for (e). male 10-15, (f). female 10-15
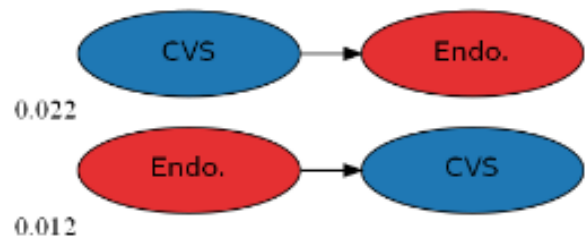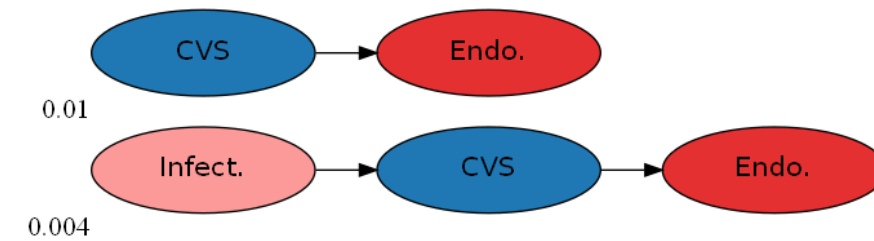
(g)   (h)

Figure 5. Most common trajectories for (g). male 15-20, (h). female 15-20

In Figures 6 – 9, the most occurring trajectories of the depression subgroups are shown for the different diabetes durations. Overall, all trajectories include medication from the Cardiovascular System and the Endocrine System. The trajectories of the with depression groups (Exist. depr., Early depr. and Late depr.) also have medication from the Central Nervous System. This is to be expected, as these are most likely the medication to treat depression. Overall, the Central Nervous System medication appears at the beginning of the trajectories in the Exist. depr. and Early depr. groups; and appears more later in the Late depr. groups. The proportion of the trajectories are also shown, which are all below 0.1 as well. This means that the most occurring trajectories do not take up a large part of the subgroups. Interestingly, the later depression is diagnosed, the lower the proportion is and thus the more variability there is among the trajectories,. So, all subgroups contain medication used to treat diabetes. The difference between the no depression group and the groups with depression, is that the latter groups contain medication used to treat depression.

Depression diabetes duration 0 – 5 years
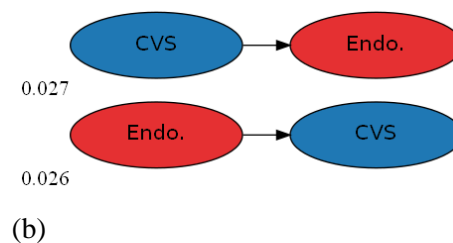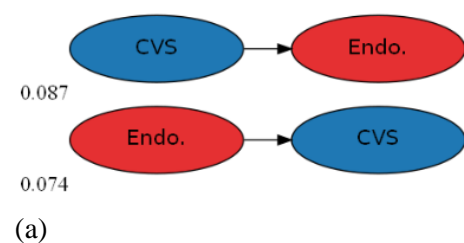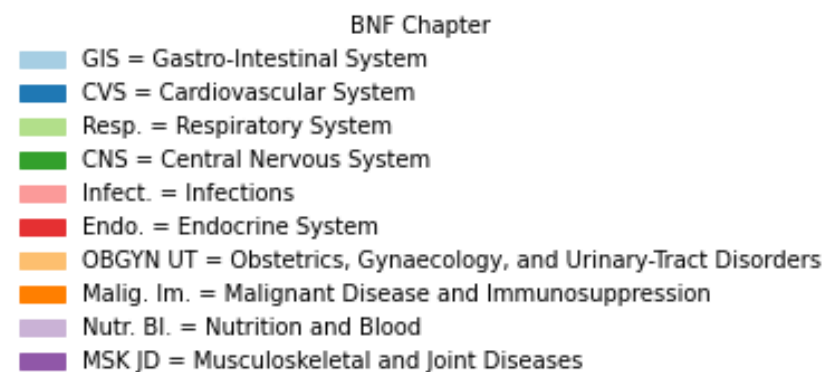
BNF Chapter

▬ (light blue)	GIS = Gastro-Intestinal System
▬ (blue)	CVS = Cardiovascular System
▬ (light green)	Resp. = Respiratory System
▬ (green)	CNS = Central Nervous System
▬ (pink)	Infect. = Infections
▬ (red)	Endo. = Endocrine System
▬ (light orange)	OBGYN UT = Obstetrics, Gynaecology, and Urinary-Tract Disorders
▬ (orange)	Malig. Im. = Malignant Disease and Immunosuppression
▬ (light purple)	Nutr. Bl. = Nutrition and Blood
▬ (purple)	MSK JD = Musculoskeletal and Joint Diseases

(a)

(b)

Figure 6. Most common trajectories for subgroups (a). No depr., (b). Exist. depr.

(c)

(d)
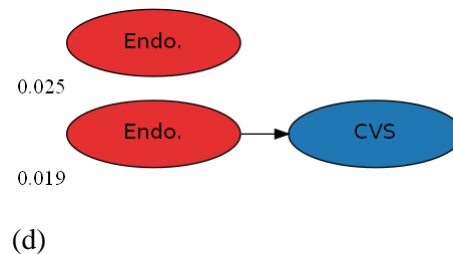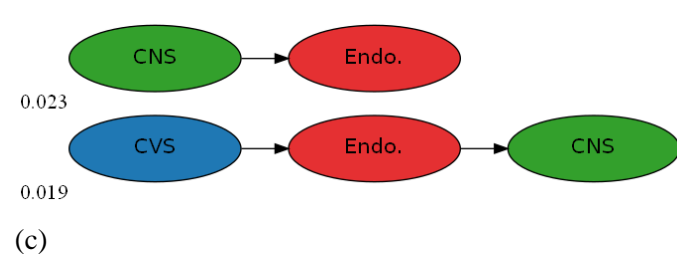
Figure 6. Most common trajectories for subgroups (c). Early depr., (d). Late depr.

Depression diabetes duration 5 – 10 years

BNF Chapter

GIS = Gastro-Intestinal System
CVS = Cardiovascular System
Resp. = Respiratory System
CNS = Central Nervous System
Infect. = Infections
Endo. = Endocrine System
OBGYN UT = Obstetrics, Gynaecology, and Urinary-Tract Disorders
Malig. Im. = Malignant Disease and Immunosuppression
Nutr. Bl. = Nutrition and Blood
MSK JD = Musculoskeletal and Joint Diseases

CVS → Endo.
0.059

Endo. → CVS
0.038

(a)

CVS → Endo.
0.015

Endo. → CVS
0.012

(b)

Figure 7. Most common trajectories for subgroups (a). No depr., (b). Exist. depr.

CNS → Endo. → CVS
0.009

CNS → Endo.
0.009

(c)

Endo. → CVS → CNS
0.011

CVS → Endo. → CNS
0.01

(d)

Figure 7. Most common trajectories for subgroups (c). Early depr., (d). Late depr.

Depression diabetes 10 – 15 years

BNF Chapter
- GIS = Gastro-Intestinal System
- CVS = Cardiovascular System
- Resp. = Respiratory System
- CNS = Central Nervous System
- Infect. = Infections
- Endo. = Endocrine System
- OBGYN UT = Obstetrics, Gynaecology, and Urinary-Tract Disorders
- Malig. Im. = Malignant Disease and Immunosuppression
- Nutr. Bl. = Nutrition and Blood
- MSK JD = Musculoskeletal and Joint Diseases
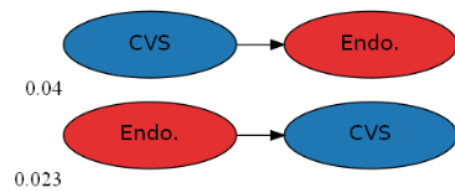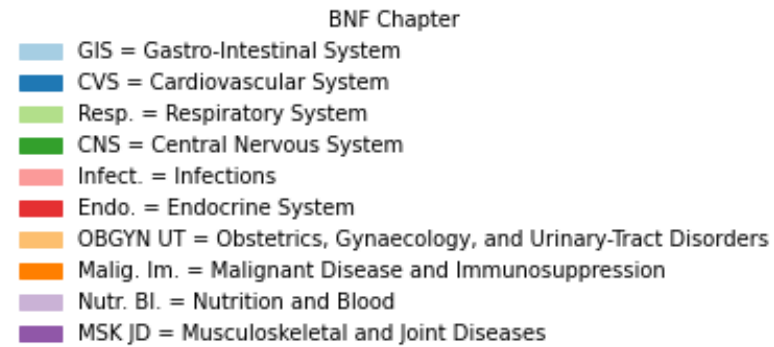
(a)
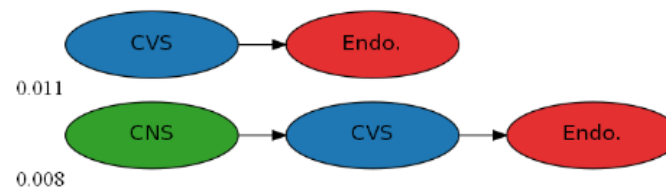
(b)

Figure 8. Most common trajectories for subgroups (a). No depr., (b). Exist. depr.
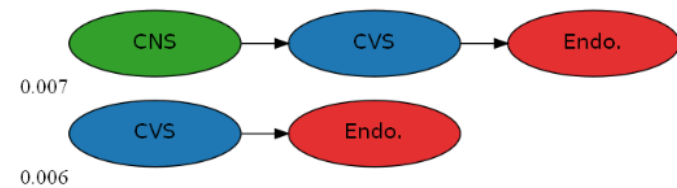
(c)
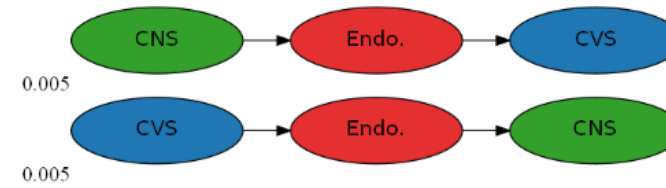
(d)

Figure 8. Most common trajectories for subgroups (c). Early depr., (d). Late depr.

Depression diabetes duration 15 – 20 years
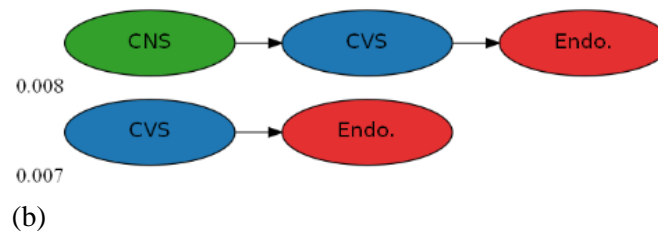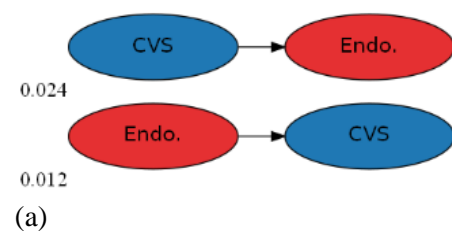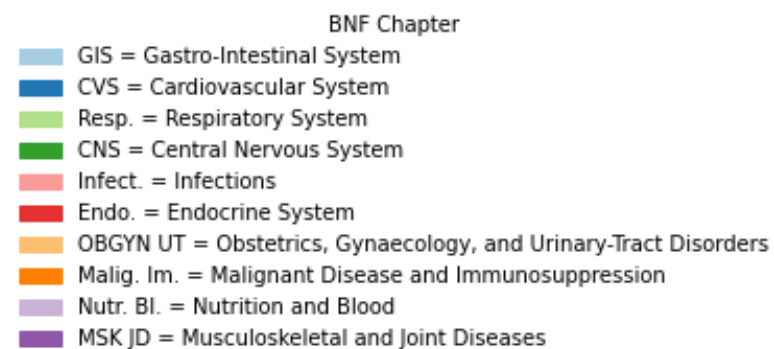
BNF Chapter
- GIS = Gastro-Intestinal System
- CVS = Cardiovascular System
- Resp. = Respiratory System
- CNS = Central Nervous System
- Infect. = Infections
- Endo. = Endocrine System
- OBGYN UT = Obstetrics, Gynaecology, and Urinary-Tract Disorders
- Malig. Im. = Malignant Disease and Immunosuppression
- Nutr. Bl. = Nutrition and Blood
- MSK JD = Musculoskeletal and Joint Diseases

(a)

(b)

Figure 9. Most common trajectories for subgroups (a). No depr., (b). Exist. depr.

(c)

(d)

Figure 9. Most common trajectories for subgroups (c). Early depr., (d). Late depr.

25
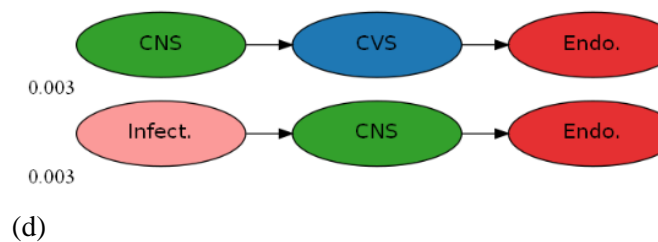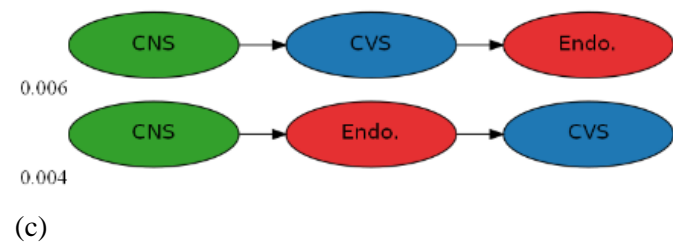
4.3.2 Clusters

The top 3 clusters with their representative trajectory for each sex and depression subgroup can be found in Appendix D. The three clusters for each subgroup are the best clusters selected based on the number of trajectories in the cluster and the average distance of the cluster. Furthermore, the representative trajectory of each cluster is also selected based on how many times this particular trajectory appears in the subgroup, and the average distance to all the other trajectories in the cluster. These representative trajectories are different from the most occurring trajectories shown in Section 4.1.1 and 4.1.2, because the representative trajectories are selected based on all the other trajectories in the subgroup. Whereas for the most occurring trajectories, simply the top 2 trajectories that appear the most in the subgroup were selected. Furthermore, for every representative trajectory, the proportion of that trajectory is also shown. These are calculated by counting how many times this trajectory appears in that cluster and divide it by the total number of trajectories in that cluster.

In Appendix E, the total proportion of the top 3 representative trajectories are shown and it can be seen that these representative trajectories take up in total less than 20% of the subgroups. This means that even after clustering and selecting the best clusters, there is still a lot of variability between the trajectories. Moreover, there is more variability among the trajectories of the groups with longer diabetes duration. Table 7 and Table 8 shows the average distance of the top 3 clusters in the form of a similarity score. As can be seen, even after clustering and selecting the best clusters, there is still a lot of difference between the trajectories and the similarity score is even lower than before clustering the subgroups.

Groups	Diabetes duration			
	0-5	**5-10**	**10-15**	**15-20**
Male	0.24	0.25	0.24	0.24
Female	0.25	0.24	0.27	0.25

Table 7. Average similarity score of the top 3 clusters according to sex

Groups	Diabetes duration			
	0-5	5-10	10-15	15-20
No depr.	0.25	0.24	0.24	0.23
Exist. depr.	0.27	0.26	0.23	0.24
Early depr.	0.30	0.26	0.23	0.23
Late depr.	0.29	0.27	0.24	0.24

Table 8. Average similarity score of the top 3 clusters according to the presence of depression.

5. Discussion

5.1 Answering the data science question

1. How to establish if there is a difference between sequences?

2. How to create representations of a large amount of sequences and compare them?

To answer the first data science question, a similarity score was calculated to determine if there is a difference between the medication trajectories. This score is based on the dissimilarity matrix that is computed by using the Longest Common Substring as edit distance. Then, for the second data science question, first the most occurring trajectories for each subgroup is selected and compared. Also, the trajectories are clustered using Ward's method and selecting the optimal number of clusters by visually inspecting the dendrograms. Then, for each subgroup the top 3 clusters are selected based on the number of trajectories in a cluster and the variance within the cluster. Finally, for each of these three clusters a representative medication trajectory is selected and their proportion is calculated.

5.2 Answering the research question

The research question of this paper is: Is there a difference in medication trajectories between subgroups according to sex and the presence of depression among those with type 2 diabetes and what are those differences?

The similarity scores show that there is a difference between medication trajectories according to sex and the presence of depression. This result is expected and in line with previous findings which stated that there were differences in diabetes treatment depending on sex and the presence of serious mental illness.[10-14] What is surprising, however, is that a comparable amount of difference is also found within the subgroups. This is unexpected as this suggests that even within the individual subgroups, the trajectories differ a lot. A possible explanation for this might be that the subgroups are too wide. So, the patients within the subgroups (e.g. males with diabetes duration 0 – 5 years) are too different from each other and thus their trajectories are also very different.

The most occurring trajectories are used to describe the trajectories and to analyse what the difference are between subgroups. For both the sex subgroups and the depression subgroups, the trajectories all include medication from the Cardiovascular System and the Endocrine System. This is expected as these medications are most likely used to treat diabetes. Furthermore, the female subgroup with a diabetes duration of 15 – 20 years also has a medication from the Infections category, which the other sex subgroups do not have. As for the depression subgroups, the subgroups with depression also include medication belonging to the Central Nervous System. This is expected as these medication are likely used to treat depression. For the Exist. depr. and the Early depr. groups, the Central Nervous System medication appears at the beginning, and for the Late depr. groups it appears more later in the trajectories. The timing of the medication prescription coincides with the timing of the depression diagnosis.

So, the most occurring trajectories show that there are some differences between certain subgroups. However, most of the male and female subgroups have very similar most occurring trajectories, which seems contradicting with the similarity scores. This may be explained by the fact that even the most occurring trajectories only take up a small part of the subgroups. Other differences between the subgroups may be found in the remaining trajectories that occur less often. Furthermore, when clustering all the trajectories and selecting the top 3 representative trajectories, it shows that they only take up a small percentage of the subgroup as well. Also, the trajectories within a cluster also differ quite from each other, further proving that there are a lot of differences within the subgroups.

This study has several limitations. The dissimilarity matrix uses the Common Longest Substring as metric to compute the distances between the trajectories. This metric defines that two trajectories have a low distance if they have a lot of consecutive medication in common. However, it does not take into account if there is another medication that the two trajectories have in common but is not consecutive. Thus, the metric assigns the same distance regardless if there are other common non-consecutive medication in the trajectory. A different, more sophisticated metric may be used to capture the distances between the trajectories. Furthermore, within each subgroup there are a lot of unique trajectories, which means there is still a lot of variability between the subgroups in terms of trajectories. So, purely looking at the most occurring trajectories and also clustering these trajectories based on these subgroups therefore might not lead to the best representation of all the trajectories. Finally, in this study the trajectories are based on the BNF chapter of the medications, which does not give insight into what specific medication are different between subgroups of patients.

Further research might look at smaller subgroups based on for example sex, diabetes duration and age to further stratify patients. This will likely result in more similar trajectories within a subgroup, which might help further understanding the difference between different subgroups. Furthermore, the methodology used in this study can also be applied to analyse the medication trajectories based on for example the drug substance of the medication. Finally, additional research is needed to analyse why there is a difference between subgroups and how to create personalized medication trajectories that fits individual people with type 2 diabetes.

In conclusion, there is a difference in medication trajectories between men and women, and between patients with and without depression. However, even within the subgroups there is a lot of difference. For the sex subgroups and the depression subgroups, all the most occurring trajectories consists of medication from the Cardiovascular System and Endocrine System. The female subgroup with a diabetes duration of 15 – 20 years also contains medication from the Infection category. For the depression subgroups, the groups with depression also contain medication from the Central Nervous System, where the position of this medication in the trajectory depends on the time when depression was diagnosed.

References

1.   Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr*. 2010;8:29.

2.   Griffith LE, Gruneir A, Fisher K, Panjwani D, Gafni A, Patterson C, Markle-Reid M, Ploeg J. Insights on multimorbidity and associated health service use and costs from three population-based studies of older adults in Ontario with diabetes, dementia and stroke. *BMC Health Serv Res*. 2019;19:313.

3.   Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, Meinow B, Fratiglioni L. Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev*. 2011;10:430–9.

4.   Kadam UT, Roberts I, White S, Bednall R, Khunti K, Nilsson PM, Lawson CA. Conceptualising multiple drug use in patients with comorbidity and multimorbidity: proposal for standard definitions beyond the term polypharmacy. *J Clin Epidemiol* [Internet]. 2018 [cited 2018 Oct 31];Available from: https://linkinghub.elsevier.com/retrieve/pii/S0895435617314427

5.   Austin RP. Polypharmacy as a Risk Factor in the Treatment of Type 2 Diabetes. *Diabetes Spectr*. 2006;19:13–16.

6.   AL-Musawe L, Martins AP, Raposo JF, Torre C. The association between polypharmacy and adverse health consequences in elderly type 2 diabetes mellitus patients; a systematic review and meta-analysis. *Diabetes Res Clin Pract*. 2019;155:107804.

7.   Boyd C. Will Running the Numbers First Violate the Principles of Patient-Centered Care? *Ann Intern Med*. 2008;838–839.

8.   Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, Michos ED, Miedema MD, Muñoz D, Smith SC, Virani SS, Williams KA, Yeboah J, Ziaeian B. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140:e596–e646.

9. Burt VL, Whelton P, Roccella EJ, Brown C, Cutler JA, Higgins M, Horan MJ, Labarthe D. Prevalence of Hypertension in the US Adult Population. *Hypertension*. 1995;25:305–313.

10. Manteuffel M, Williams S, Chen W, Verbrugge RR, Pittman DG, Steinkellner A. Influence of patient sex and gender on medication use, adherence, and prescribing alignment with guidelines. *J Womens Health (Larchmt)*. 2014;23(2):112-119.

11. Gouni-Berthold I, Berthold HK, Mantzoros CS, Böhm M, Krone W. Sex disparities in the treatment and control of cardiovascular risk factors in type 2 diabetes. *Diabetes Care*. 2008;31(7):1389-1391.

12. Krämer HU, Raum E, Rüter G, et al. Gender disparities in diabetes and coronary heart disease medication among patients with type 2 diabetes: results from the DIANA study. *Cardiovasc Diabetol*. 2012;11:88.

13. Kreyenbuhl J, Dickerson FB, Medoff DR, et al. Extent and management of cardiovascular risk factors in patients with type 2 diabetes and serious mental illness. *J Nerv Ment Dis*. 2006;194(6):404-410.

14. Scott D, Platania-Phung C, Happell B. Quality of care for cardiovascular disease and diabetes amongst individuals with serious mental illness and those using antipsychotic medications. *J Healthc Qual*. 2012;34(5):15-21.

15. van den Heuvel JM, Farzan N, van Hoek M, Maitland-van der Zee A-H, Ahmadizar F. Mining treatment patterns of glucose-lowering medications for type 2 diabetes in the Netherlands. *BMJ Open Diabetes Res Care*. 2020;8:e000767.

16. Guo J, Parkin L, Zeng J, Barson D, Horsburgh S. Treatment pathways in people with type 2 diabetes mellitus: a nationwide cohort study of new users of metformin monotherapy in New Zealand. *BMJ Open*. 2021;11:e051884.

17. Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, Suchard MA, Schuemie MJ, DeFalco FJ, Perotte A, Banda JM, Reich CG, Schilling LM, Matheny ME, Meeker D, Pratt N, Madigan D. Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci*. 2016;113:7329–7336.

18.   Liefbroer AC, Elzinga CH. Intergenerational transmission of behavioral patterns: how similar are parents' and children's demographic trajectories? *Advances in Life Course Research*. 2012;17(1):1-10.

19.   Abbott, A., & Forrest, J. Optimal Matching Methods for Historical Sequences. *The Journal of Interdisciplinary History*. 1986; *16*(3), 471–494.

20. Van Alsten SC, Duncan AE. Lifetime patterns of comorbidity in eating disorders: An approach using sequence analysis. *Eur Eat Disord Rev*. 2020;28(6):709-723.

21.   Eisenberg-Guyot J, Peckham T, Andrea SB, Oddo V, Seixas N, Hajat A. Life-course trajectories of employment quality and health in the U.S.: A multichannel sequence analysis. *Social Science & Medicine*. 2020;264:113327.

22.   Giannoula A, Gutierrez-Sacristán A, Bravo Á, Sanz F, Furlong LI. Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study. *Sci Rep*. 2018;8(1):4216.

23. Data [Internet]. CPRD. [cited 2022Jun26]. Available from: https://cprd.com/data

24. Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*, *44*(3), 827–836.

25. British National Formulary (BNF) [Internet]. MedicinesComplete. 2021 [cited 2022Jun29]. Available from: https://about.medicinescomplete.com/publication/british-national-formulary/

26. How to use your british national formulary for children (BNFC) [Internet]. How to use the British National Formulary for Children (BNFC). [cited 2022Jun27]. Available from: https://www.rpharms.com/resources/quick-reference-guides/how-to-use-the-british-national-formulary-for-children-bnfc

27. Safeguarding patient data [Internet]. CPRD. [cited 2022Jun26]. Available from: https://cprd.com/safeguarding-patient-data

Appendices

Appendix A

Chapter	Name
1	Gastro-Intestinal System
2	Cardiovascular System
3	Respiratory System
4	Central Nervous System
5	Infections
6	Endocrine System
7	Obstetrics, Gynaecology and Urinary-Tract Disorders
8	Malignant Disease and Immunosuppression
9	Nutrition and Blood
10	Musculoskeletal and Joint Diseases
11	Eye
12	Ear, Nose and Oropharynx
13	Skin
14	Immunological Products and Vaccines
15	Anaesthesia
18	Preparations used in Diagnosis
19	Other Drugs and Preparations
20	Dressings
21	Appliances
22	Incontinence Appliances
23	Stoma Appliances

A.1 All BNF chapters

Appendix B

Medical codes
324, 636, 655, 543, 4639, 462, 1131, 2639, 4069, 4534, 2970, 9211, 4167, 6932, 5987, 6950, 595, 4323, 6482, 6874, 4659, 1758, 10610, 1055, 3292, 5879, 11717, 11913, 6939, 9667, 3076, 10344, 15099, 3291, 3208, 3489, 2300, 25638, 6546, 10667, 5726, 2741, 15155, 14709, 9386, 2030, 5385, 4634, 7953, 7749, 1723, 1907, 9055, 16506, 23838, 1531, 8567, 29520, 7604, 8205, 15220, 962, 12838, 14784, 2560, 14656, 1510, 12099, 9944, 2366, 50191, 7737, 14729, 1533, 10825, 8826, 10455, 8902, 14728, 29342, 16638, 9183, 6854, 15219, 44300, 8851, 29784, 22806, 12173, 25697, 7011, 16632, 98414, 17770, 15566, 21540, 98252, 27584, 25563, 4678, 33469, 31316, 8478, 34390, 6710, 5678, 8584, 2972, 37070, 35825, 18909, 6071, 16808, 11329, 47009, 6221, 20110, 13024, 29907, 12831, 21065, 11596, 9521, 34064, 101054, 24112, 33751, 98346, 19696, 70000, 26161, 22116, 4677, 24171, 10390, 3702, 35671, 17385, 10290, 100211, 37090, 28248, 18510, 23731, 31535, 31522, 103915, 55384, 16562, 98417, 16861, 32159, 18032, 28938, 10720, 20773, 27986, 11252, 24230, 16199, 32941, 60178, 101153, 33425, 43324, 26227, 19967, 44513, 24117, 46415, 47731, 17687, 30688, 23854, 44321, 19054, 56273, 27491, 31633, 63583, 38543, 26839, 24066, 31957, 35594, 28167, 42857, 28863, 41992, 51032, 28277, 37178, 28106, 41989, 57409, 24351, 36246, 28008, 44693, 73924, 55064, 28677, 53840, 28756, 23713, 27685, 29451, 56609, 54607, 4732, 31672, 18603, 57465, 49763, 35738, 37296, 36611, 15923, 57605, 37764, 50998, 27890, 63784, 103677, 26299, 35734, 33426, 59386, 52678, 36126, 46434, 55829, 63150, 39767, 54195, 29921, 31757, 32088, 104051, 50218, 104065, 24689, 36616, 32295, 29579, 37102, 35607, 27739, 65811, 66153, 24640, 16347, 63284, 70399, 46425, 48632, 63701, 73991, 68647, 68326, 54848, 63698, 43093, 59011, 72026, 50243, 63651, 70721, 47365, 73423, 70925, 109485, 58863

Table B.1. Medication codes for depression:

Appendix C
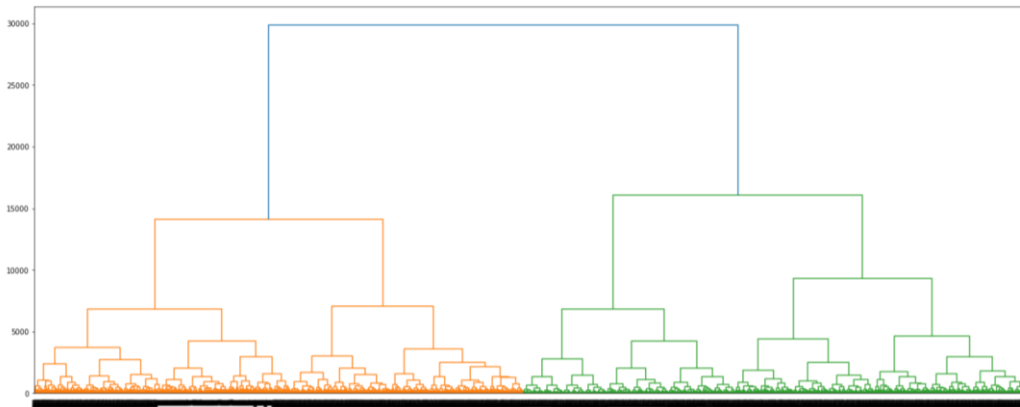
Figure C1. Dendrogram male 0-5
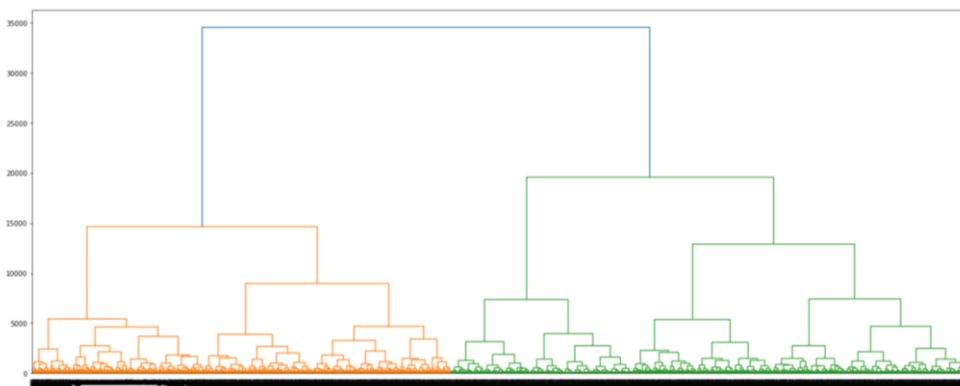
Figure C2. Dendrogram male 5-10
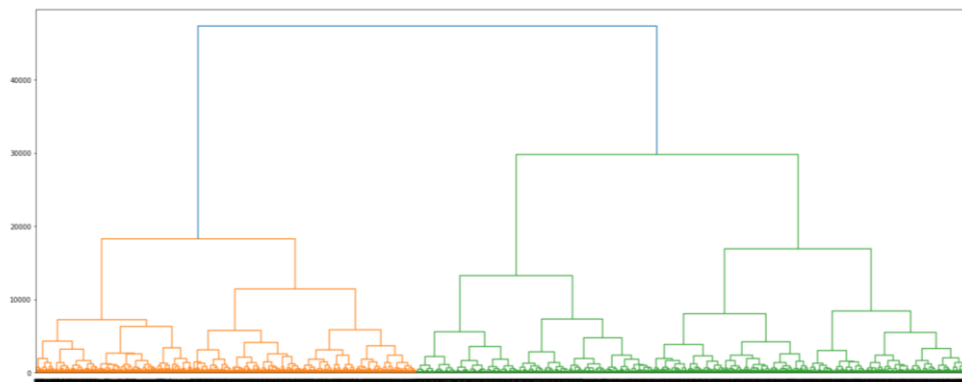
Figure C3. Dendrogram male 10-15
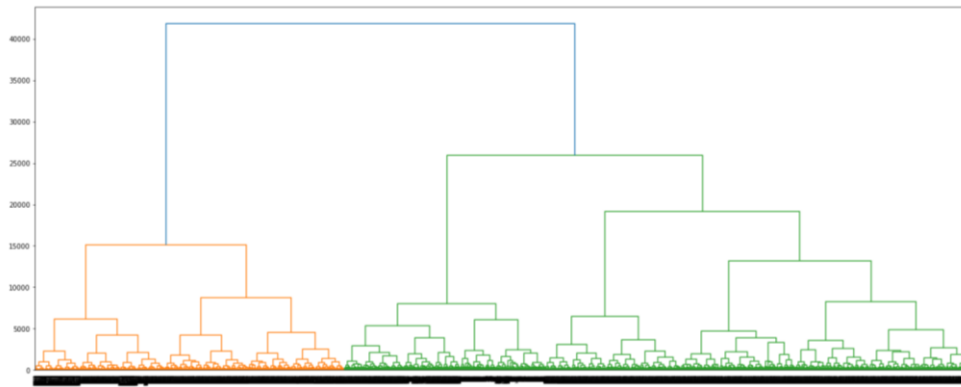
Figure C4. Dendrogram male 15-20
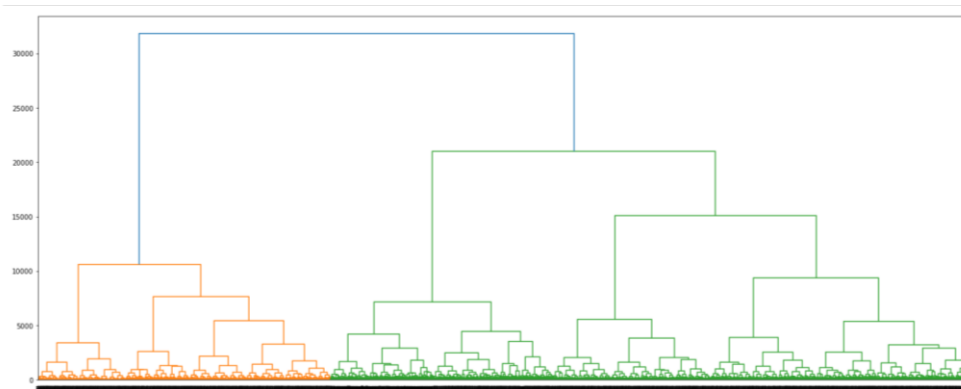
Figure C5. Dendrogram female 0-5
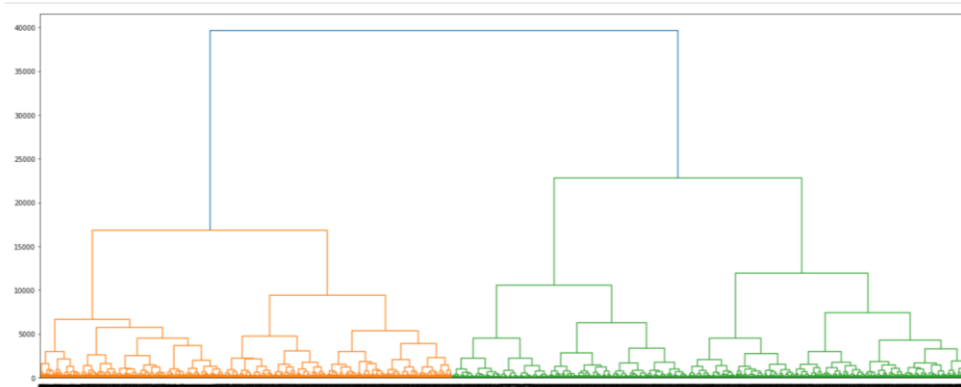
Figure C6. Dendrogram female 5-10
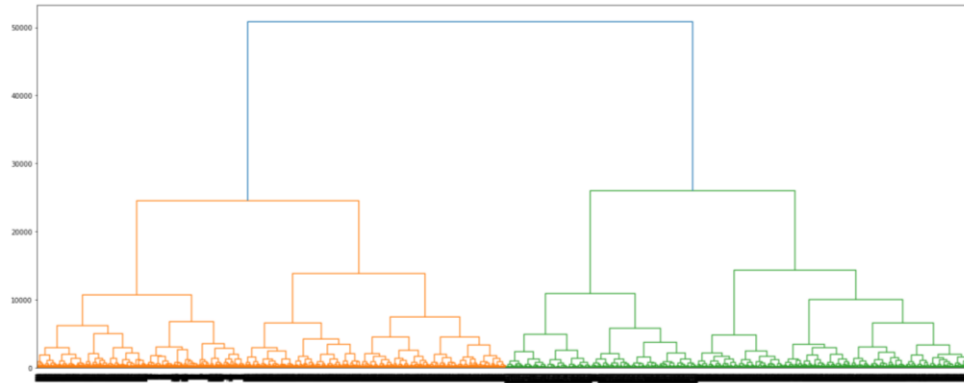
Figure C7. Dendrogram female 10-15
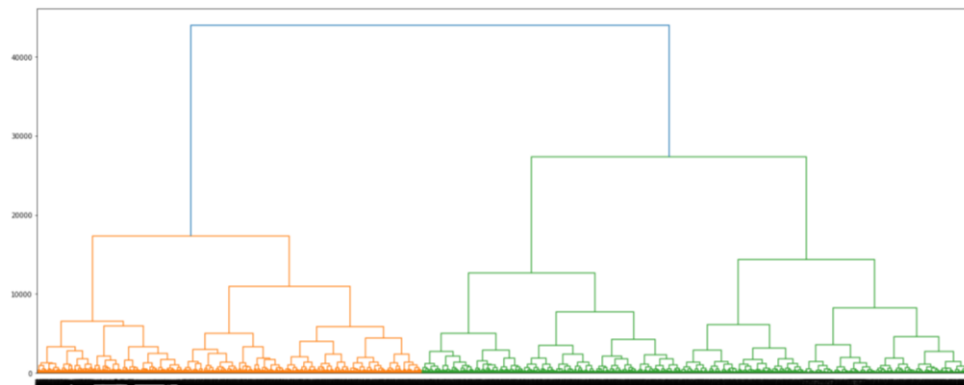
Figure C8. Dendrogram female 15-20
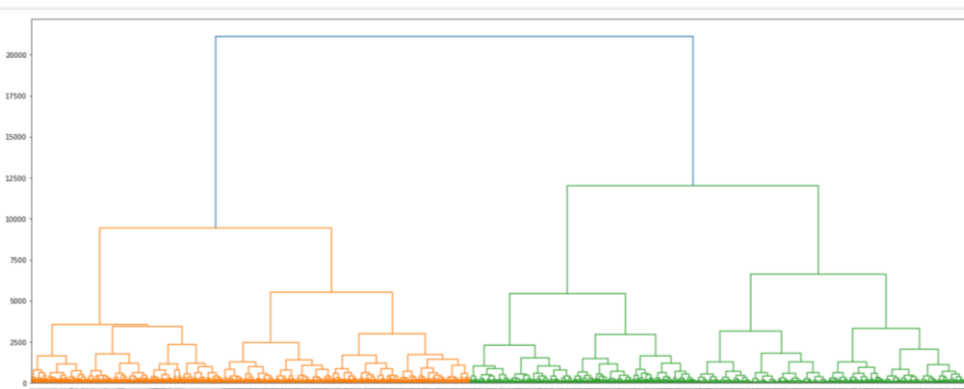
Figure C9. Dendrogram no depr. 0-5
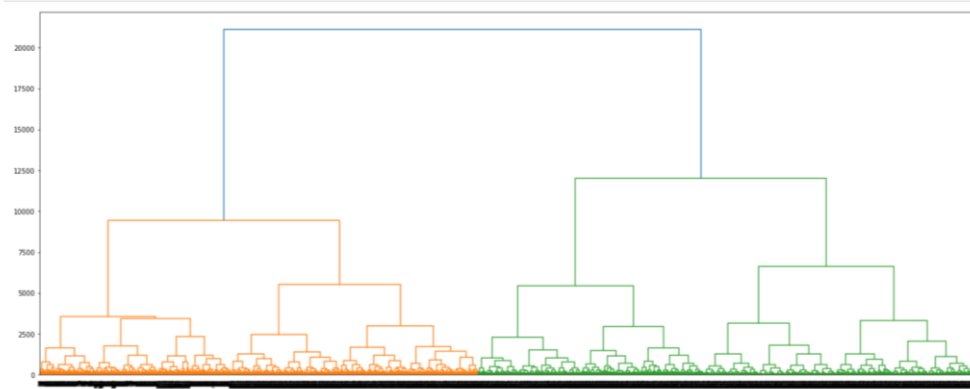
Figure C10. Dendrogram no depr. 5-10
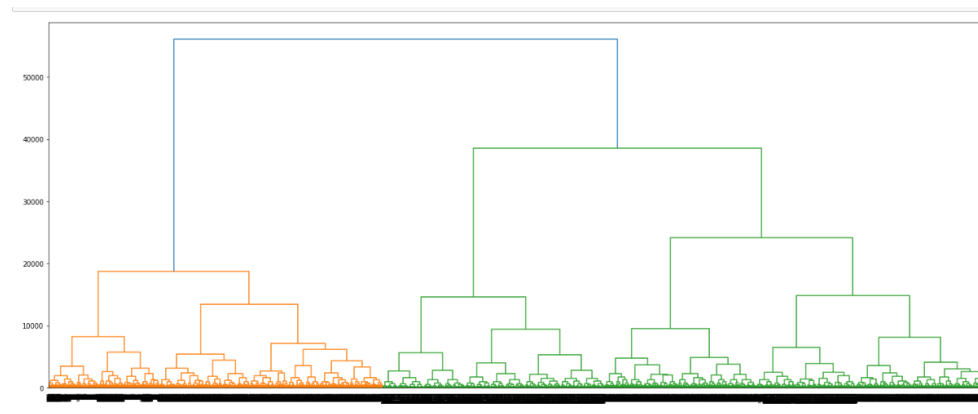
Figure C11. Dendrogram no depr. 10-15
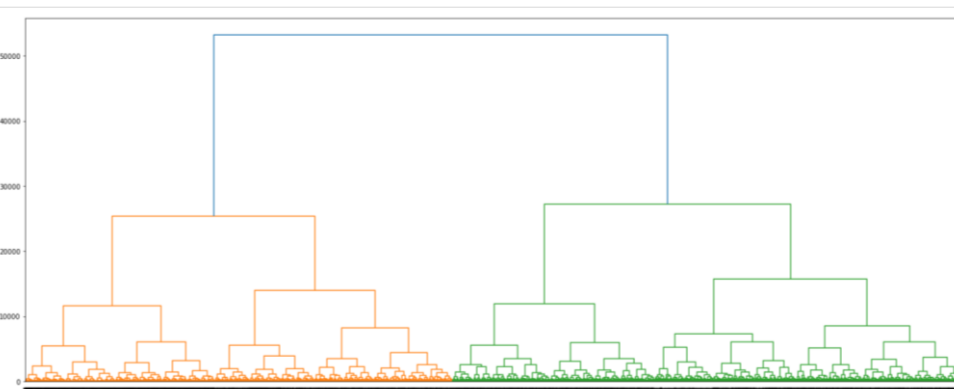
Figure C12. Dendrogram no depr. 15-20
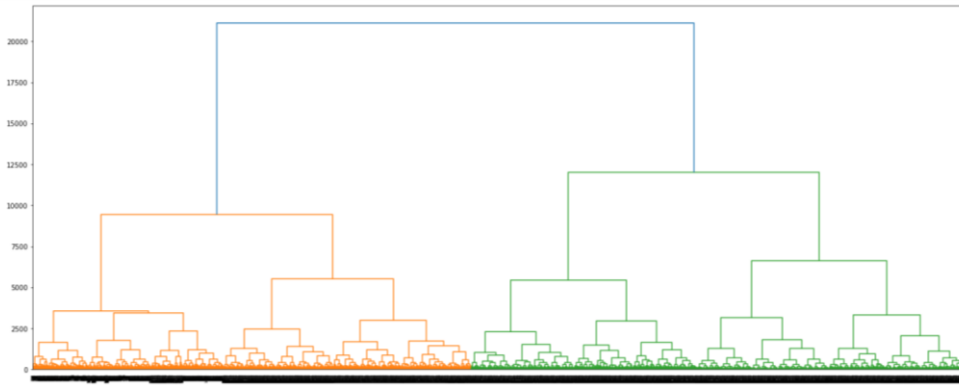
Figure C13. Dendrogram exist depr. 0-5
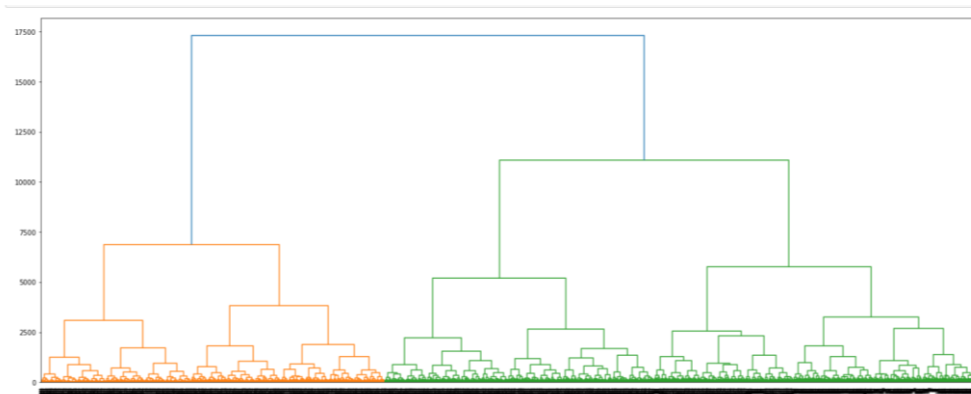
Figure C14. Dendrogram exist depr. 5-10
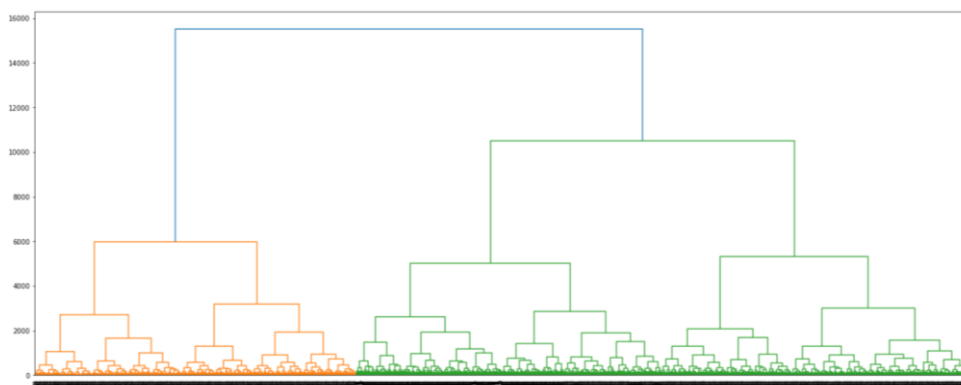
Figure C15. Dendrogram exist depr. 10-15
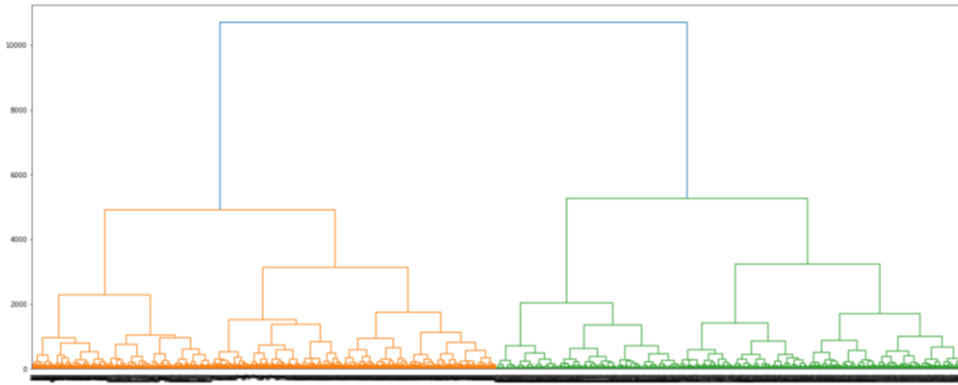
Figure C16. Dendrogram exist depr. 15-20
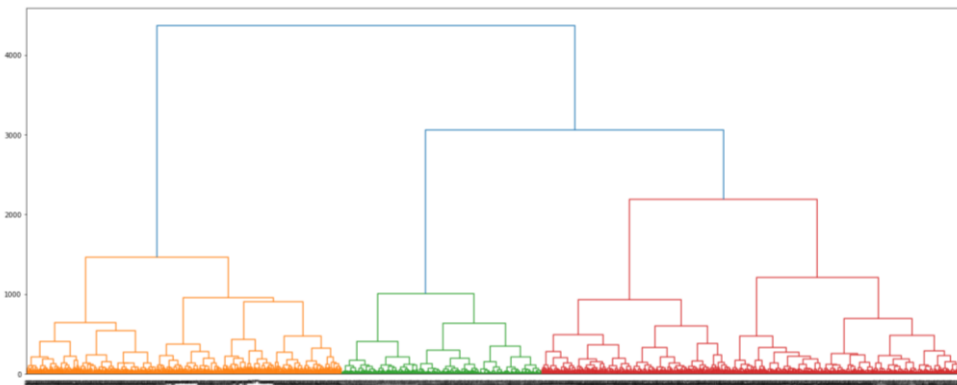
Figure C17. Dendrogram early depr. 0-5
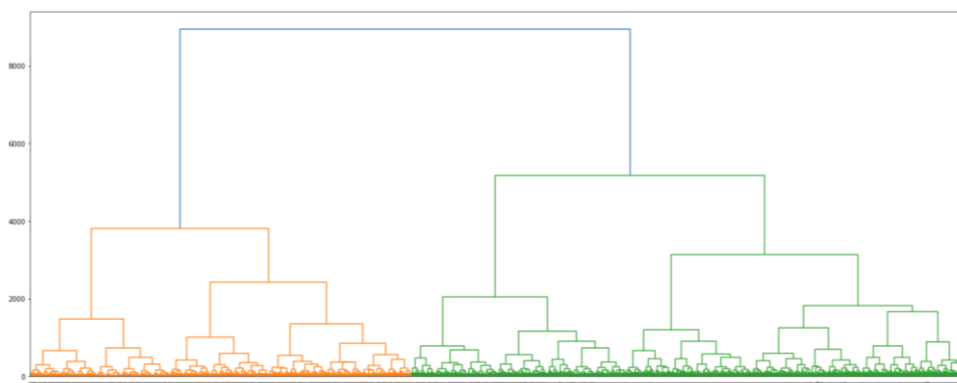
Figure C18. Dendrogram early depr. 5-10
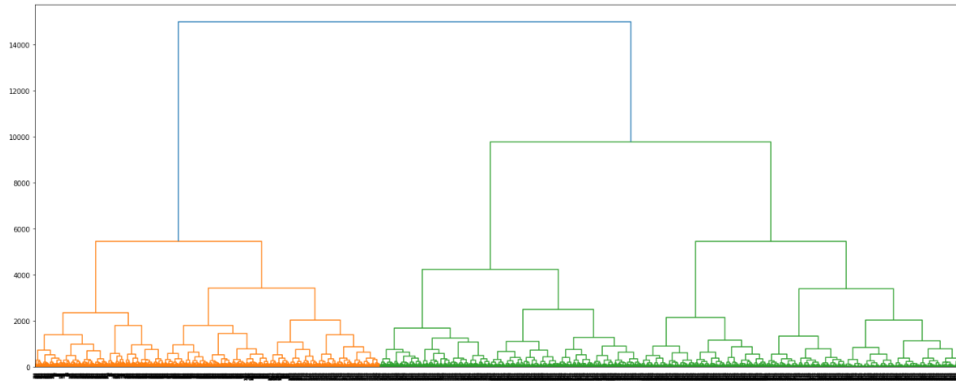
Figure C19. Dendrogram early depr. 10-15
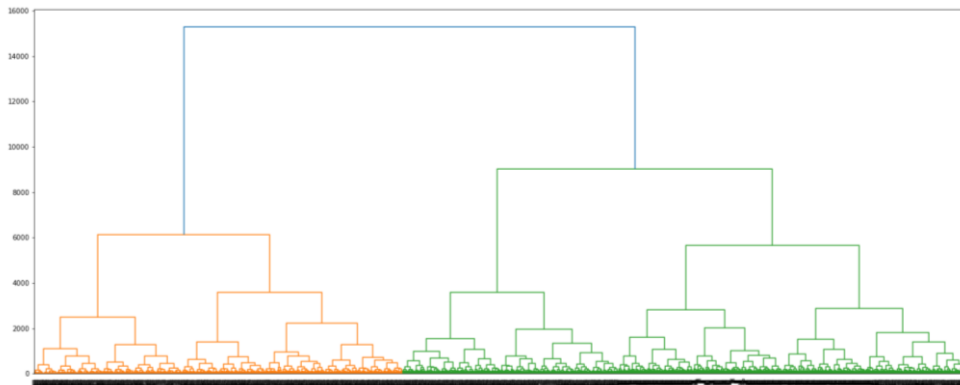
Figure C20. Dendrogram early depr. 15-20
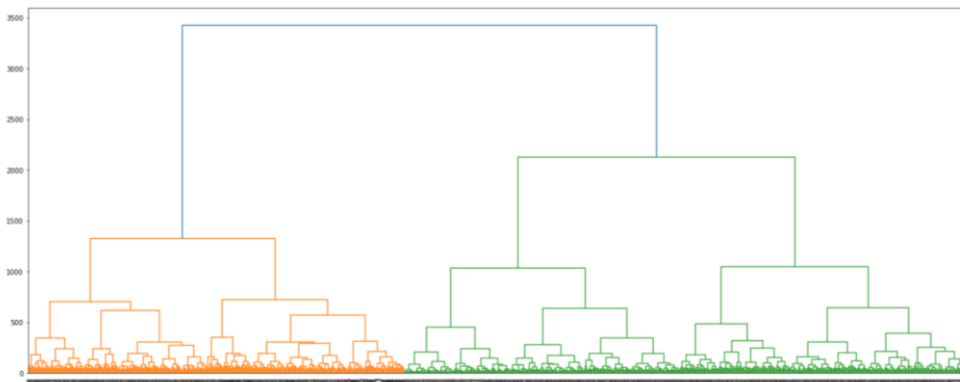
Figure C21. Dendrogram late depr. 0-5
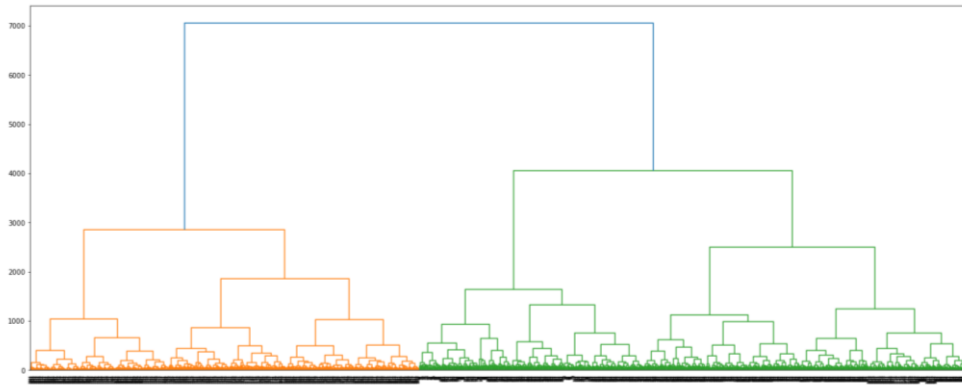
Figure C22. Dendrogram late depr. 5-10
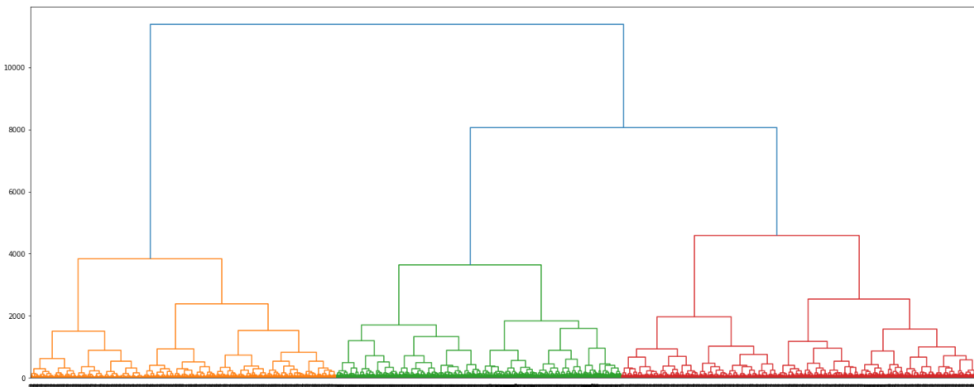
Figure C23. Dendrogram late depr. 10-15
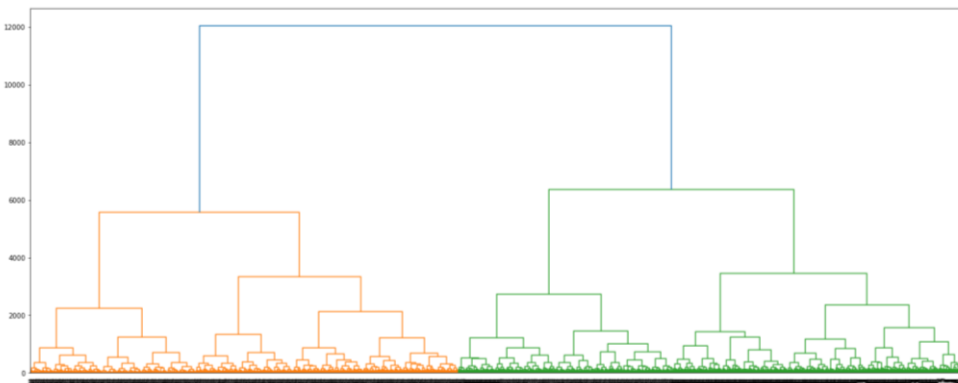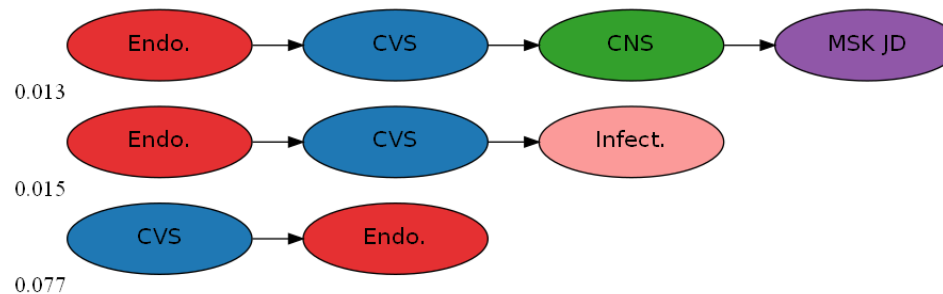

Figure C24. Dendrogram late depr. 15-20

Appendix D

(a)

(b)
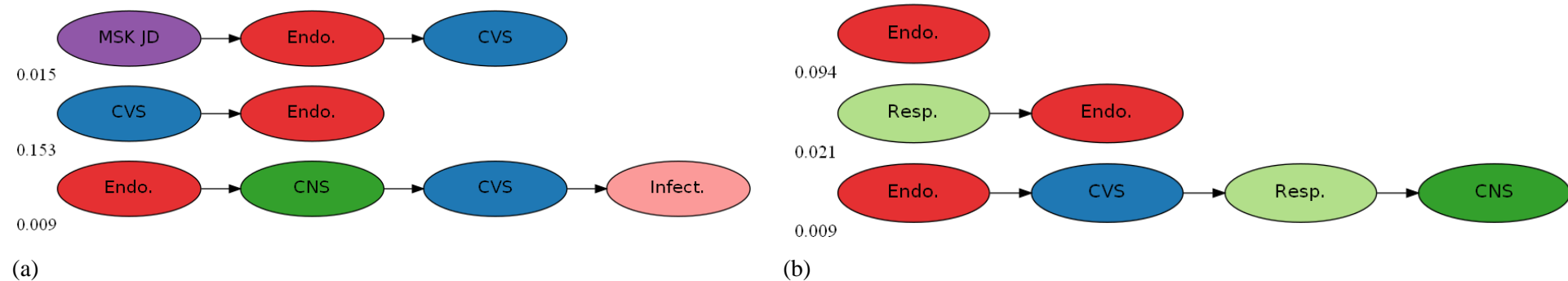
Figure D1. Top 3 clusters, their representative trajectory and the proportion of that trajectory in the cluster for (a). male 0-5 and (b). female 0-5
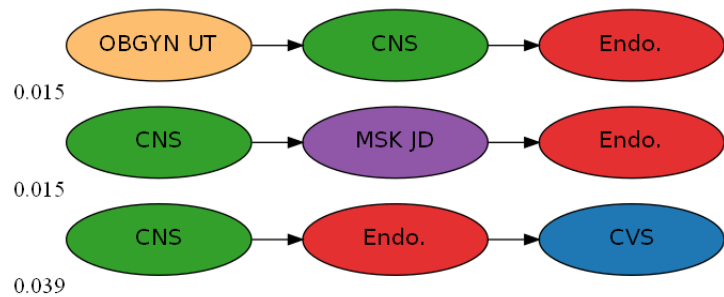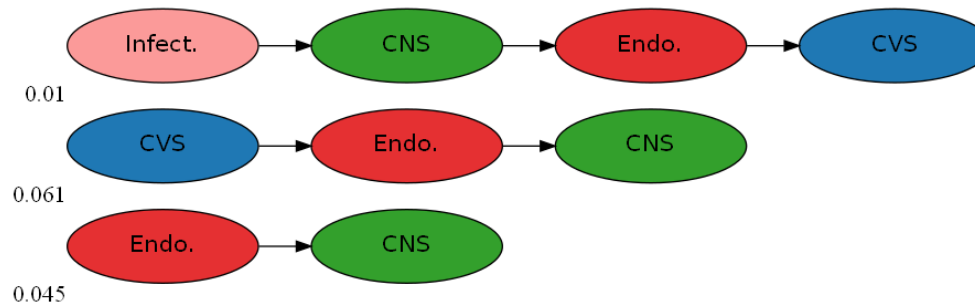

(a)

(b)
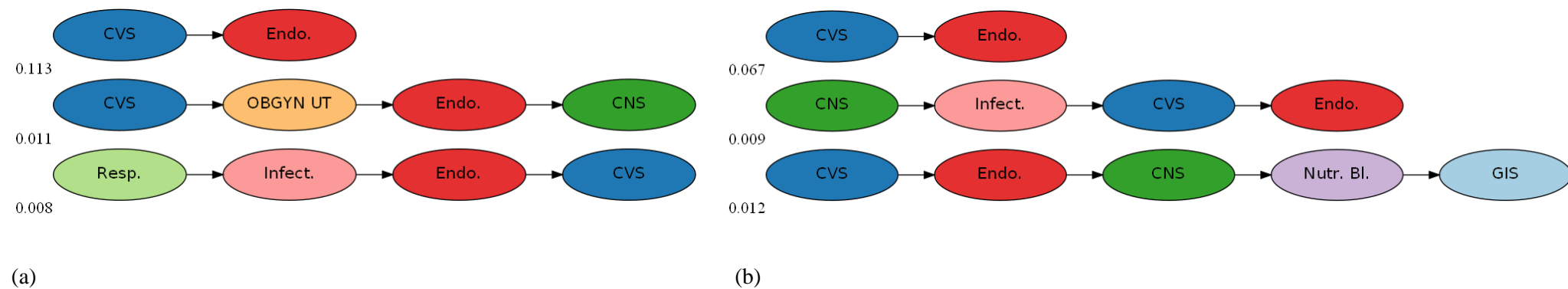
Figure D2. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). male 5-10 and (b). female 5-10
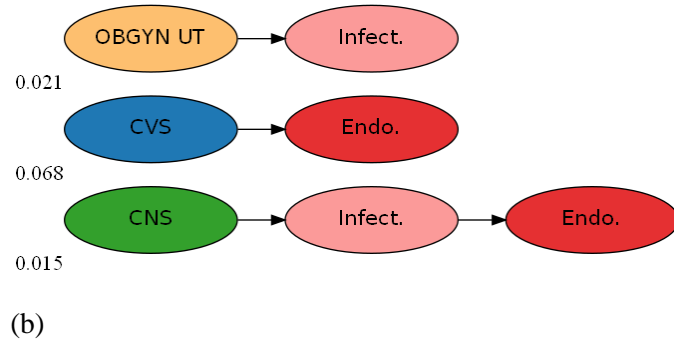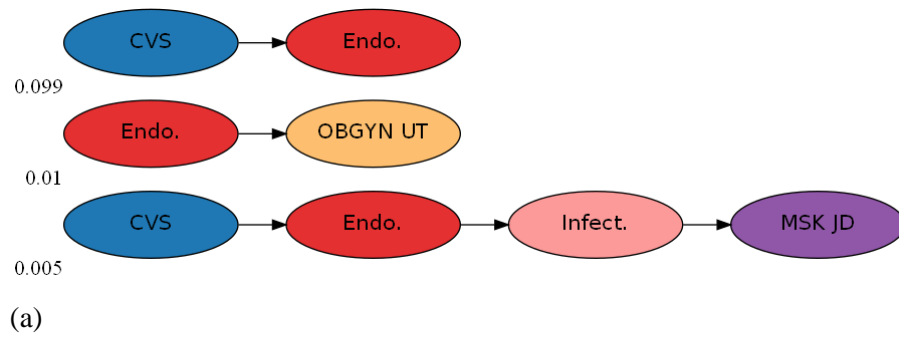
Figure D3. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). male 10-15 and (b). female 10-15
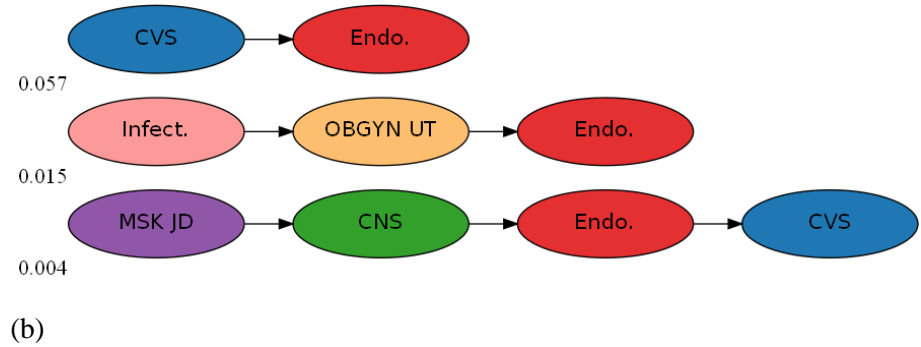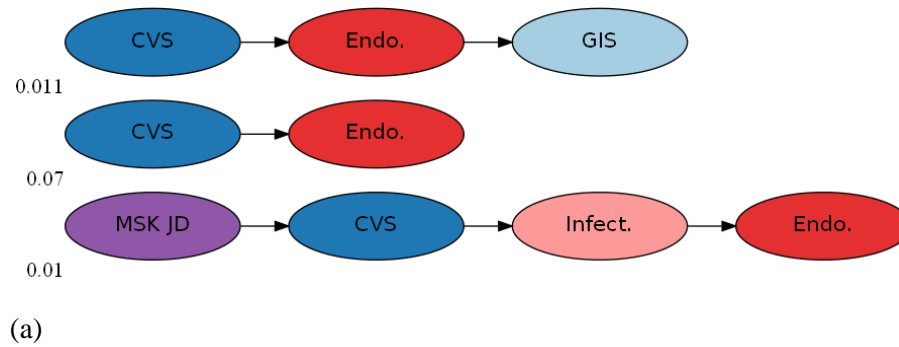
Figure D4. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). male 15-20 and (b). female 15-20
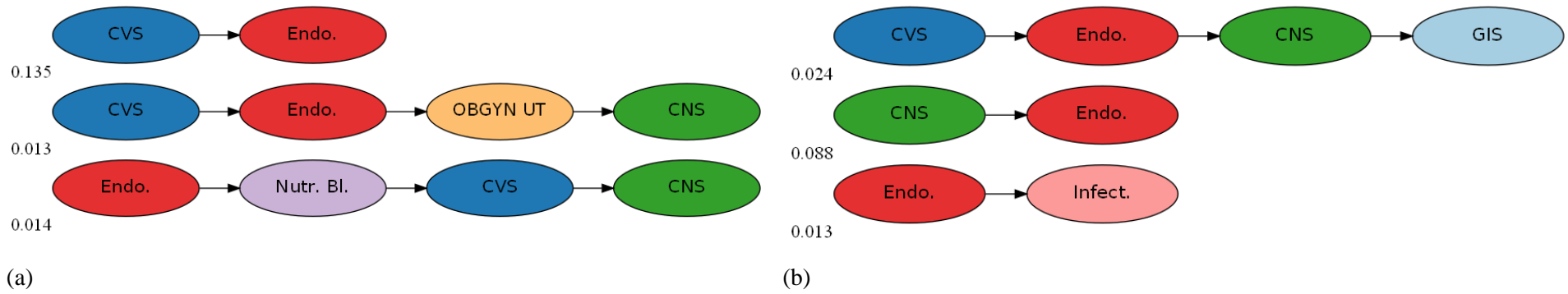
Figure D5. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). No depr. 0-5, (b). Exist. depr. 0-5
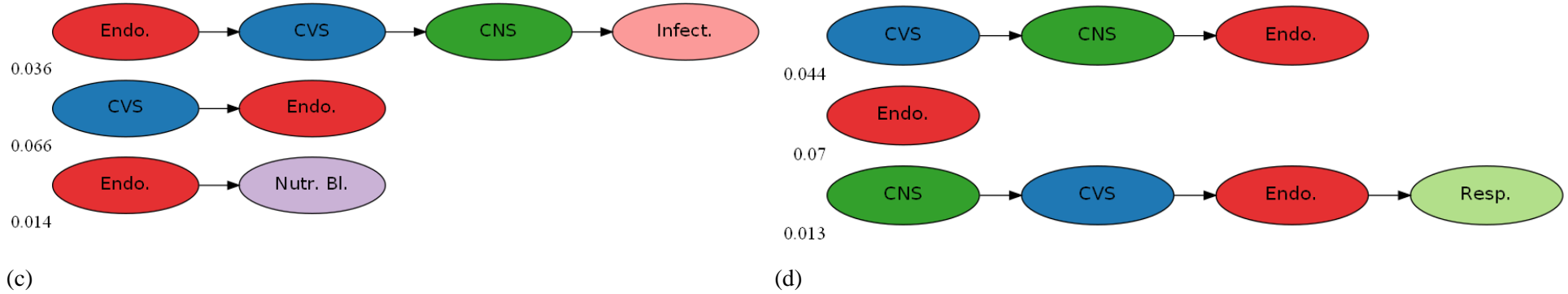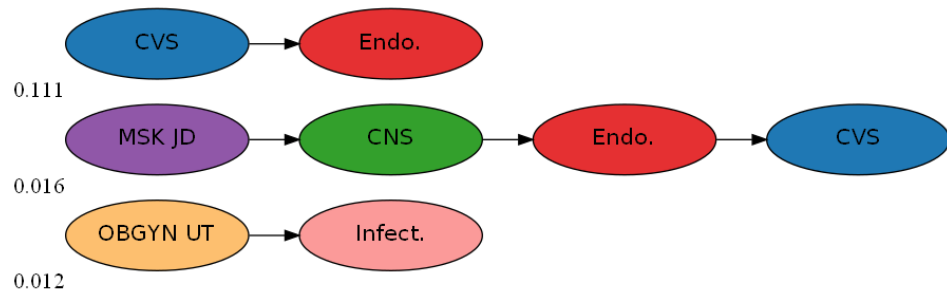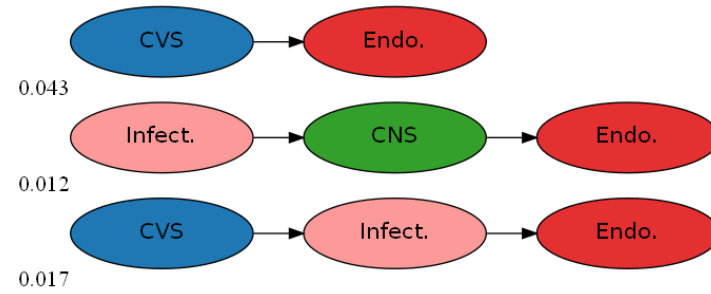

Figure D5. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (c). Early depr. 0-5, (d) Late depr. 0-5
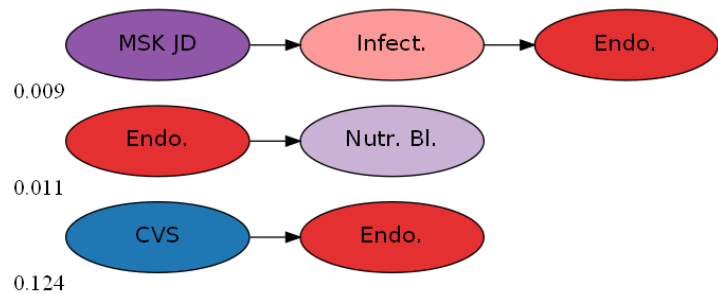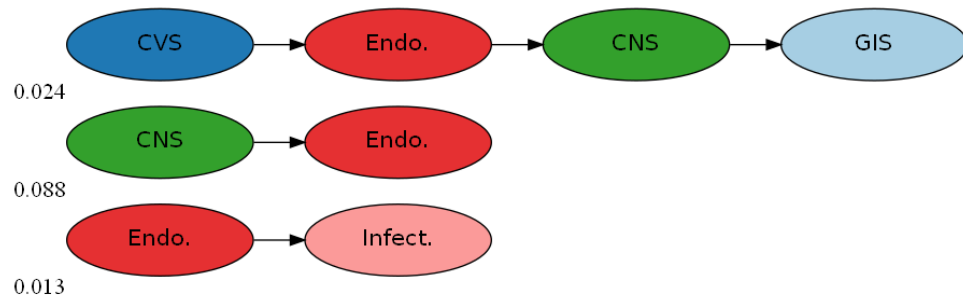
(a)

(b)

Figure D6. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). No depr. 5-10, (b). Exist. depr. 5-10,

(c)

(d)

Figure D6. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (c). Early depr. 5-10, (d) Late depr. 5-10
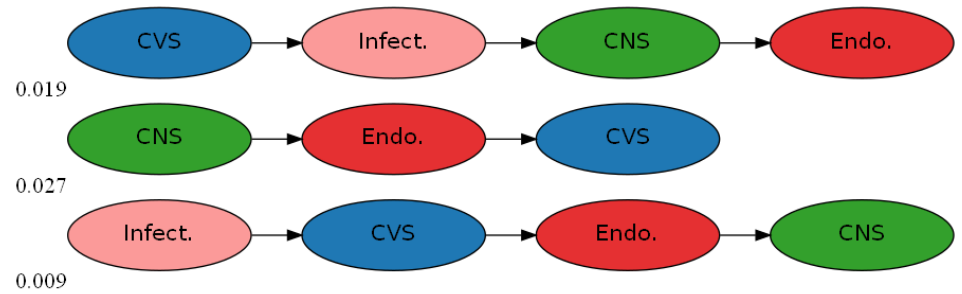
(a)

(b)

Figure D7. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). No depr. 10-15, (b). Exist. depr. 10-15,
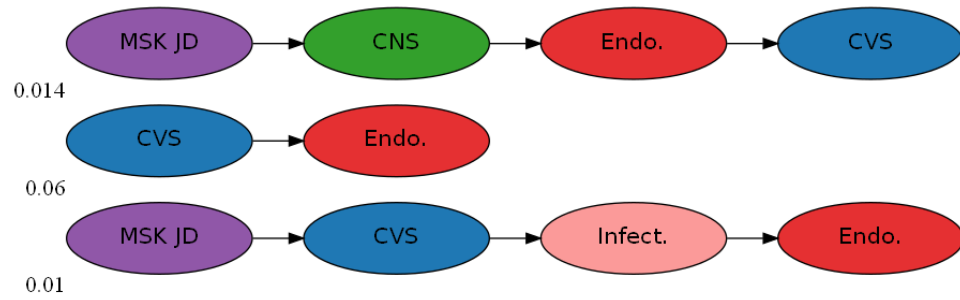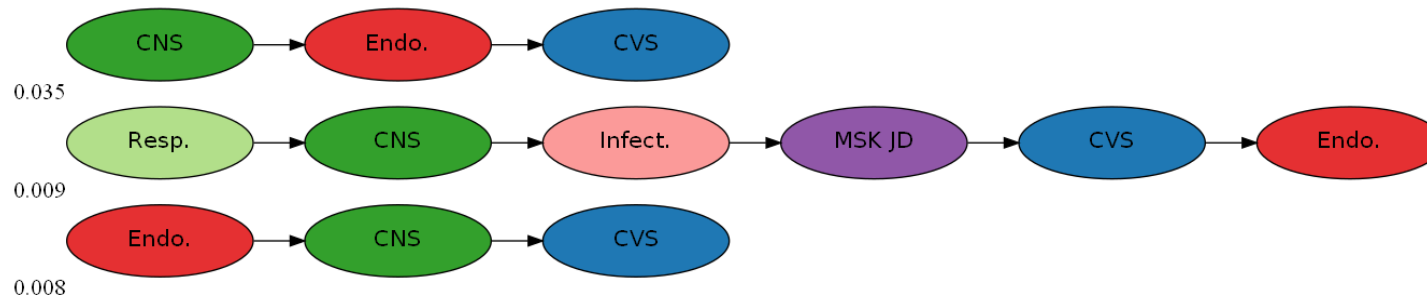

(c)

(d)

Figure D7. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (c). Early depr. 10-15, (d) Late depr. 10-15
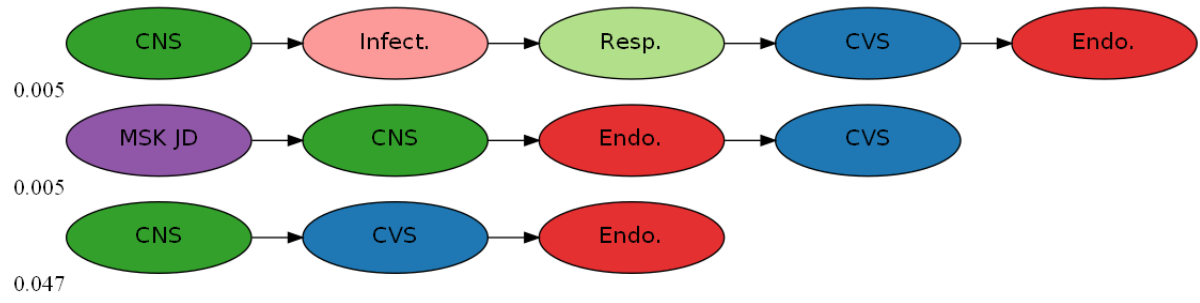
(a)

Figure D8. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (a). No depr. 15-20,
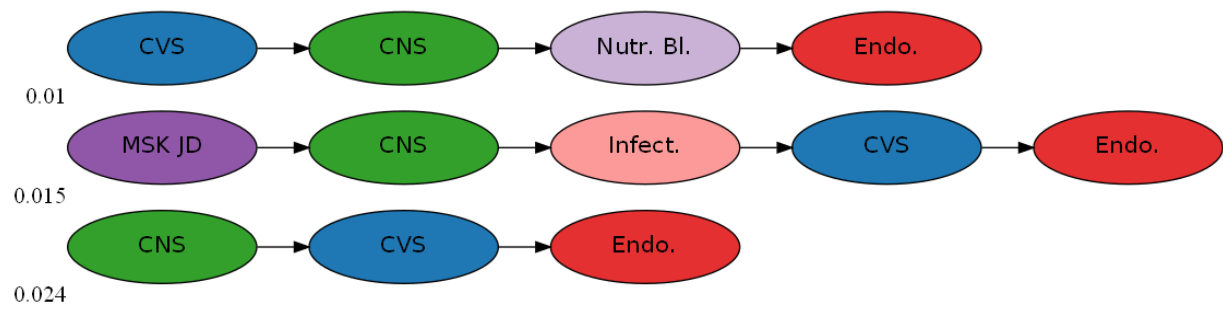

(b)

Figure D8. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (b). Exist. depr. 15-20

0.005

0.005

0.047

(c)

Figure D8. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (c). Early depr. 15-20

0.01

0.015

0.024

(d)

Figure D8. Top 3 clusters including the representative trajectory and the proportion of that trajectory in the cluster for (d) Late depr. 15-20

Appendix E

Groups	Diabetes duration			
	0-5	**5-10**	**10-15**	**15-20**
Male	17.7%	13.2%	11.4%	9.1%
Female	12.4%	8.8%	10.4%	7.6%

Table E1. The total percentage of the top 3 representative trajectories according to sex

Groups	Diabetes duration			
	0-5	**5-10**	**10-15**	**15-20**
No depr.	16.2%	14.4%	13.9%	8.4%
Exist. depr.	12.5%	10.5%	7.2%	5.2%
Early depr.	11.6%	6.9%	12.5%	5.7%
Late depr.	12.7%	11.6%	5.5%	4.9%

Table E2. The total percentage of the top 3 representative trajectories according to depression