JSDSS 日本ソーシャルデータサイエンス論文誌
第 5 巻　第 1 号（2021 年 3 月）

原著論文

Simulation of Severity of Diabetic Nephropathy Using a Markov Chain

Shinya Mizuno[†1], Haruka Ohba[†1], Tatsuo Yanagawa[†2†3], Keiko Koyano[†2†3], Shuhei Iida[†2†3], Tokimune Kou[†4], Hajime Okuno[†4] and Naokazu Yamaki[†4]

Abstract: In Japan, National healthcare expenditure in 2015 was 42,364.4 billion yen, 3.8% more than the previous year, which indicates a significant problem. As diabetes becomes severe, it costs a lot for dialysis and medication. As the population with diabetes is increasing and diabetes is a risk factor causing complications, it is necessary to undertake efforts to ensure that diabetes does not lead to severe illness. In this study, we construct a simulation with a Markov chain on diabetes, which will become an increasingly important issue in the future. First, we create state distribution using eGFR and urine protein. The initial distribution first uses eGFR and urine protein tested values. The final distribution uses the last inspection value existing as data. We calculate the average inspection period from the data and make it the unit period of the Markov chain. We calculate the transition probability matrix from the inspection data and observe the state transition by stationary distribution and simulation. This simulation clarifies the progressive severity of diabetes, making it easier to deal with stages leading to severe illness. Simulations are categorized according to patient attributes and implemented so that they can be applied in many cases.

Keywords: Markov chain, simulation, transition probability, stationary distribution

1. INTRODUCTION

In Japan, national healthcare expenditure in 2015 was 42,364.4 billion yen, an increase of 3.8% from the previous year [1]. With the increase in national medical expenses, the diabetes-affected population and Impaired glucose tolerance in Japan are estimated to be about 10 million people [2], which indicates a significant problem. As diabetes becomes severe, it costs a lot for dialysis and medication [3]. As the population of diabetes is increasing, and diabetes is also a risk factor causing complications, efforts that do not lead to severe illness are necessary [4].

Various studies have been undertaken on diabetes to date. Specifically, many studies have been undertaken on clinical research [5-10], but research using information science has also been increasing recently. Statistical approaches to improvement of diabetic nephropathy patients have been conducted [11-13]. In recent years, research has been undertaken to predict the number of patients in the future using machine learning theory, such as neural network [14,15]. Research using Markov chain to measure transitions per unit time and to predict the number of patients in the future has also been undertaken [16-19]. The application of artificial intelligence to the medical field is increasing, and machine learning is one of the big tools, but at the same time, the calculation process is complicated, and there are drawbacks, for example, it is difficult for humans to grasp the process leading to the result. First, we use statistics to grasp the overall trend to basically analyze clinical data. For that, it is necessary to create an environment with a database structure that can analyze data. We consider it necessary to clarify the temporal trend of each data and to clarify the state transition of the patient. Essentially, a Markov chain is assumed as stationary, but temporal transition is easy to express, and flexible expression can be undertaken by linking with simulation.

In this study, we construct a simulation with a Markov chain on diabetes, which will become an increasingly important issue in the future. In the complications of diabetes, we focus on nephropathy this case. The severity of nephropathy is regulated by eGFR and urine protein. So, we first create state distribution using eGFR and urine protein. From this state distribution, we try to simulate the deterioration of diabetic nephropathy. The initial distribution first uses eGFR and urine protein tested values. The final distribution uses the last inspection value existing as data. We calculate the average inspection period from the data and make it the unit period of the Markov chain. We calculate the transition probability matrix from the inspection data and observe the state transition by stationary distribution and simulation. This simulation clarifies the progressive severity of diabetes, making it easier to deal with stages leading to severe illness. Simulations are categorized according to patient attributes and implemented so that they can be applied in many cases.

†1 Shizuoka Institute of Science and Technology
(Correspondence author: mizuno.shinya@sist.ac.jp)
†2 Nerima General Hospital
†3 Institute for Healthcare Quality Improvement, Tokyo Healthcare Foundation
†4 Research Institute of Systems Planning, Inc.
投稿日：2020 年 12 月 25 日
採録日：2021 年 3 月 13 日

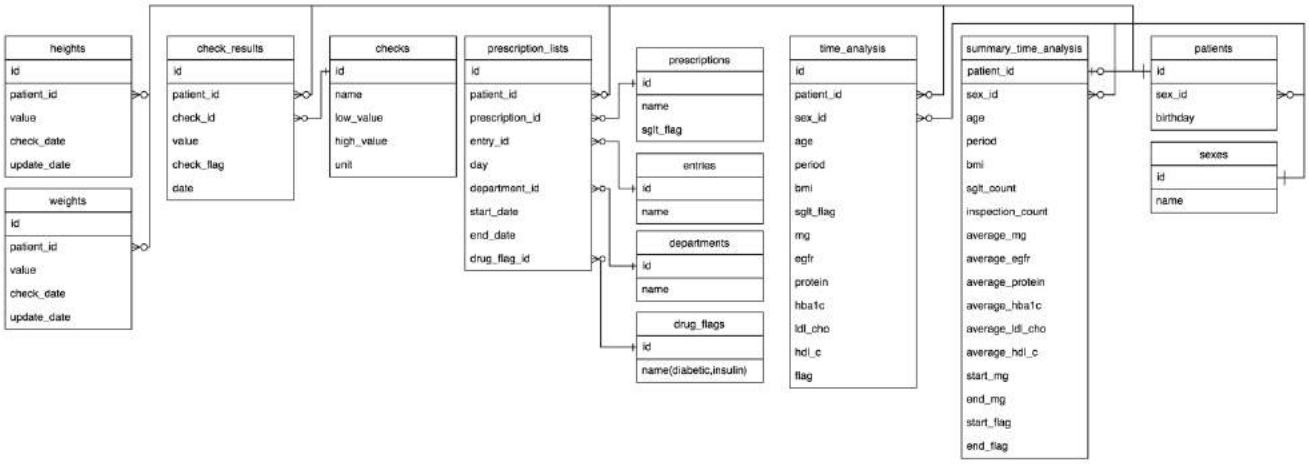

Fig. 1 Database structure for analysis of severe diabetes

Table 1 Explanation of factors used

Factor	Contents
Age	Age at start of examination
Sex	Sex
BMI[kg/m^2]	Body mass index (BMI; weight/height^2; kg/m^2) [21]
Treatment period[day]	Number of days from examination start date to final examination date
eGFR (including creatinine) [$ml/min/1.73m^2$]	Glomerular filtration rate [22]
Urinary protein[g/day]	The function of the kidneys and detects disease [23]
Mg concentration[mg/dl]	Concentration of Mg in serum (mg/dl) [24]
SGLT inhibitors	preventing the kidneys from reabsorbing glucose back into the blood [25]
HbA1c[%]	The term HbA1c refers to glycated haemoglobin. It develops when haemoglobin, a protein within red blood cells that carries oxygen throughout the body, joins with glucose in the blood, becoming "glycated." HbA1c reflects the mean blood glucose level of about the past 1-2 months [26].
LDL cholesterol[mg/dl]	low-density lipoprotein cholesterol. It can be considered less desirable or even lousy cholesterol, because it contributes to fatty buildups in arteries (atherosclerosis) [27, 28].
HDL cholesterol[mg/dl]	high-density lipoprotein cholesterol. Experts believe HDL acts as a scavenger, carrying LDL cholesterol away from the arteries and back to the liver. There, it's broken down and passed from the body [28, 29].

2. BASIC ANALYSIS OF EACH ELEMENT

In order to simulate the severity of diabetes, we first perform a basic analysis of each factor. We also use databases to analyze the data. The database structure constructed in this study is shown in Figure 1. The factor and attribute used for the simulation in this study is shown in Table 1.

In order to distinguish the severity of diabetes in patients, we classify the data as shown in Table 2 using eGFR and urine protein. We used a table with fewer states from the original table [20]. We reduced the number of states to clarify the analysis, but adopt many attributes of patients. State A indicates a normal state. When eGFR decreases, urine protein becomes +, diabetic nephropathy gets worse, and state I shows severe condition.

Table 2 Classification by eGFR and urine protein

eGFR\Urinary protein	(-)	(±)	(+) or more
60 or more	A	B	C
30 or more and less than 60	D	E	F
Less than 30	G	H	I

2.1 Patient state transition by factor

We first compare the first and last data of the patient's examination according to Table 2. Patients are classified as A

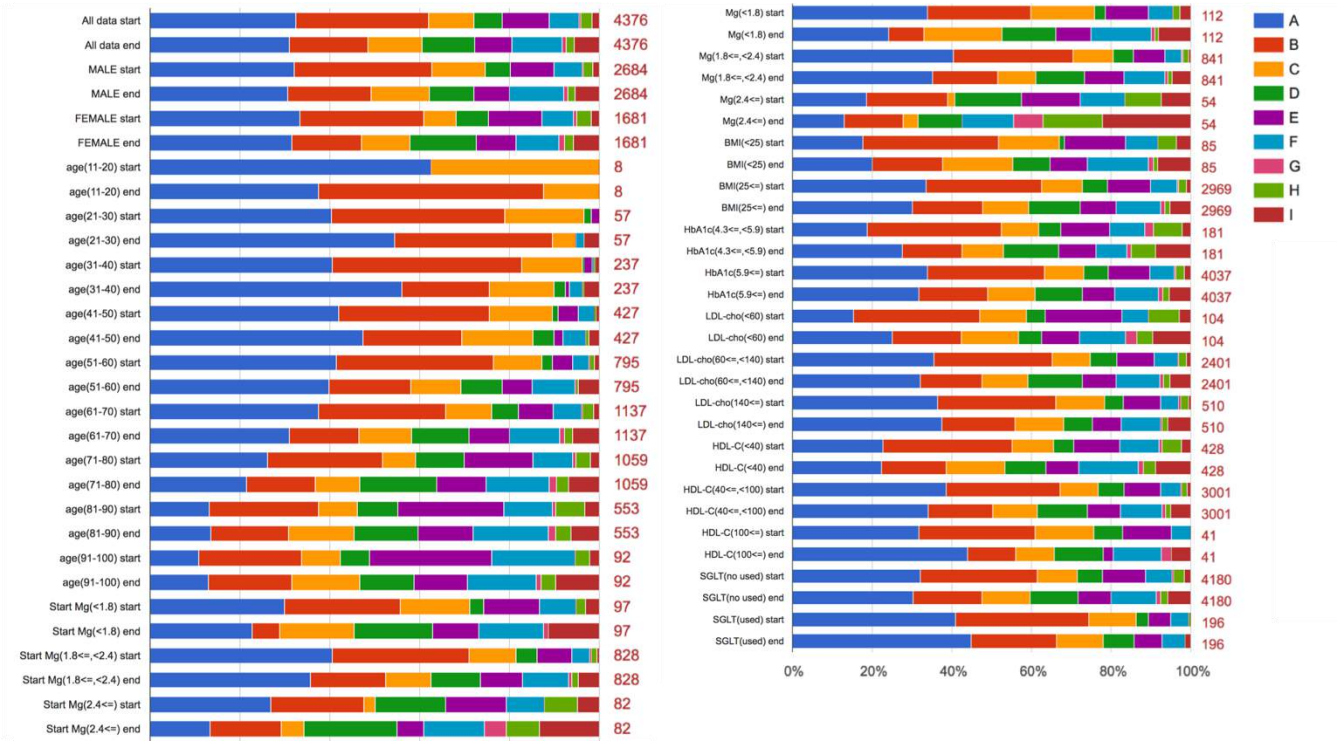

Fig. 2 State transition of patients classified by eGFR and urinary protein

to I by eGFR and urine protein testing. The results are shown in Figure 2 and Appendix A.1. The numerical value on the right-hand side of Figure 2 shows the number of subjects of the element. The top row in Figure 2 represents the state classification in the first examination result of all patients. The second line shows the final status of all patients. We observe that the proportion of state A and B decrease according to the period. On the other hand, in condition I, which is considered to be serious diabetes, the ratio increases from 1.6% to 5.51%, Increase rate of state I is 3.44. The next line is a state transition when divided by sex. There is no big difference by sex. The next line shows state transitions by age and the influence of age is large: as age rises, the severity rate also increases. Next, during classification of the state using the initial Mg concentration, the state does not become severe when it is in the proper range from 1.8 to 2.4, but the proportion of patients who are out of the appropriate range leads to severe cases. The next item classifies the state by the average Mg concentration value. Similarly, patients outside the reference value are likely to become severely ill. When classified by BMI, although the number of people is small, even if the BMI is below the reference value, it leads to serious illness. In addition, we confirm the state transition due to the difference in numerical values in HbA 1 c, LDL, HDL, but we do not obtain a big difference from other factors. Finally, when SGLT inhibitors are used, it has a strong effect on non-severity. The transition to state I is very small, but there is a transition to state A, and improvement is observed.

2.2 Time-series observation of patient's state transition

Next, we observe at the state transition of patients in time series. This graph shows how the patient's condition changes from A to I over time. When the state changes from A to I, the value of eGFR is less than 30 and the value of urine protein is (+) or more. Figure 3 shows how the value of eGFR decreases. Figure 4 shows the change of urinary protein with time series. Figure 5 shows a state transition from state A to state I over time. These figures show that patients' conditions do not suddenly worsen, but tend to worsen over time. For example, the decrease rate of eGFR is -18.4 $[ml/min/1.73m^2/year]$, and although it is affected by aging, the decrease is not sharp. We need to use the patient's state transition as an important signal to prevent patient condition deterioration.

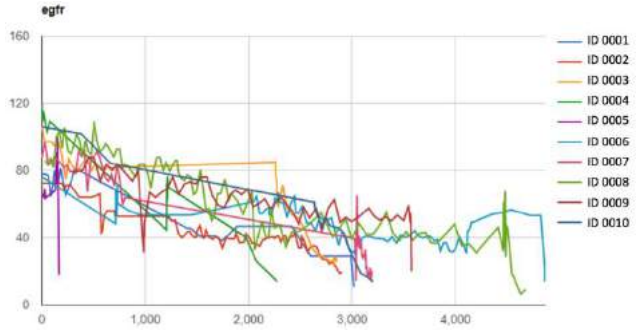

Fig. 3 Transition of eGFR when the state changes from A to I

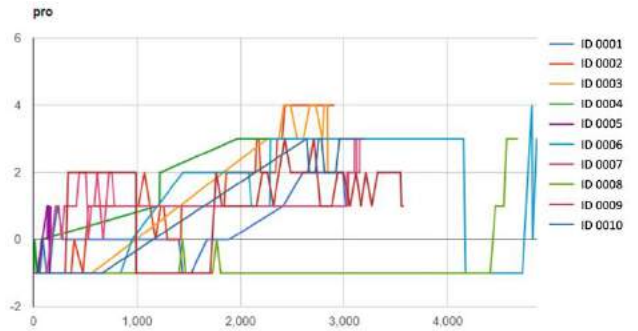

Fig. 4 Transition of protein when the state changes from A to I

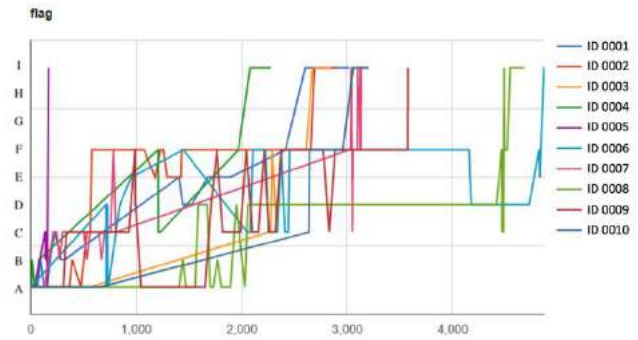

Fig. 5 Transition of state when the state changes from A to I

2.3 Consideration of influence of patient state transition

Next, we consider the data related to patient state transitions. From Figure 2, we observe that the transition to state I is greatly different between patients taking magnesium concentrations within the standard value and patients taking values outside the reference value. Therefore, we check whether there is a difference in ratio of the final state I depending on whether the initial inspection data value of Mg concentration is within the reference value or not. As Table 3 shows, when the Mg concentration is 828 people outside the standard value and 179 people outside the reference value, the number of people in state I at the end is 39 people and 22 people, respectively. In this case, when the mother ratio is tested, the p value becomes 0.00023, and from the initial Mg concentration, the result is obtained that the difference in the ratio of the number of people in state I is significant. Similarly, when SGLT inhibitors are used, the mother ratio is also tested. In this case, the p value is 0.01944, and a significant difference is obtained in the ratio of the number of people in state I depending on whether or not SGLT inhibitors are used.

Table 3 Initial Mg concentration, number of people using SGLT inhibitors, and number of people in state I

		total	state I	p-value
Mg	within the standard	828	39	0.00023
	out of standard	179	22	
SGLT	no use	4180	238	0.01944
	used	196	3	

3. DIABETES SEVERITY SIMULATION

Here, a simulation is carried out to confirm the state of severe diabetes. Let $\{X_n, n = 0, 1, 2, \cdots,\}$ be a stochastic process that takes on a finite or countable number of possible values, such as the set of non-negative integers $\{0,1,2, \cdots\}$, which describes clinical unit time. If $X_n = a$, then the process is said to be in state a at time n. We suppose that X_n takes the state $\{a, b, c, \cdots, i\}$ defined in Table 2. We suppose that whenever the process is in state a, there is a fixed probability p_{ab} that it will next be in state b. $\{X_n\}$ is interpreted as stating that, for a Markov chain, the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state. The value p_{ab} represents the probability that the process will, when in state a, next make a transition to state b [30].

3.1 Simulation of severe diabetes in all patients

First, the transition probability matrix is calculated using all patient data. Every time the patient is examined for eGFR or urinary protein, the state transition is confirmed. If only one of the tests is received, we use the latest one for the data not inspected. Appendix A.2 is a transition probability matrix when all patient data are used. The average treatment period of these data was 1618 days, the average treatment interval was 49 days, and the average treatment number was 33 times. Therefore, we assume an interval of 49 days for one transition. From this transition probability matrix, the transition probability from state I to state I is the highest value of 0.86123. The transition from state H to state I is also 0.10423, which is a very large value compared to the other. This indicates that diabetes is difficult to improve if it becomes severe. We also calculate the transition probabilities for each element.

	A	B	C	D	E	F	G	H	I	Difference
All data	0.386	0.127	0.105	0.140	0.072	0.109	0.007	0.011	0.043	0.010
MALE	0.373	0.135	0.117	0.126	0.072	0.121	0.006	0.008	0.042	0.008
FEMALE	0.407	0.114	0.087	0.161	0.072	0.090	0.009	0.015	0.045	0.011
age(11–20)	0.603	0.239	0.158	0.000	0.000	0.000	0.000	0.000	0.000	0.122
age(21–30)	0.657	0.156	0.114	0.004	0.001	0.028	0.000	0.000	0.039	0.055
age(31–40)	0.591	0.160	0.144	0.023	0.003	0.047	0.000	0.004	0.029	0.002
age(41–50)	0.532	0.181	0.135	0.050	0.034	0.048	0.001	0.006	0.013	0.006
age(51–60)	0.467	0.133	0.108	0.106	0.049	0.100	0.002	0.003	0.031	0.008
age(61–70)	0.391	0.116	0.089	0.150	0.074	0.111	0.007	0.013	0.050	0.010
age(71–80)	0.275	0.108	0.095	0.196	0.101	0.139	0.012	0.016	0.059	0.007
age(81–90)	0.176	0.117	0.119	0.209	0.130	0.151	0.020	0.025	0.053	0.010
age(91–100)	0.175	0.102	0.214	0.152	0.086	0.148	0.018	0.025	0.080	0.015
Start Mg(<1.8)	0.300	0.123	0.150	0.125	0.067	0.130	0.012	0.023	0.072	0.016
Start Mg(1.8,2.4)	0.434	0.115	0.085	0.135	0.073	0.107	0.006	0.008	0.038	0.010
Start Mg(2.4<=)	0.300	0.072	0.050	0.224	0.098	0.136	0.011	0.027	0.082	0.043
Mg(<1.8)	0.332	0.130	0.137	0.103	0.065	0.133	0.013	0.020	0.066	0.016
Mg(1.8,2.4)	0.432	0.113	0.085	0.143	0.074	0.105	0.005	0.008	0.035	0.010
Mg(2.4<=)	0.213	0.049	0.026	0.198	0.089	0.174	0.025	0.050	0.175	0.049
BMI(<25)	0.291	0.104	0.168	0.114	0.049	0.169	0.006	0.015	0.085	0.016
BMI(25<=)	0.389	0.125	0.102	0.151	0.076	0.105	0.006	0.010	0.036	0.012
HbA1c (4.3,5.9)	0.297	0.176	0.110	0.143	0.083	0.104	0.006	0.016	0.065	0.005
HbA1c (5.9<=)	0.393	0.187	0.113	0.101	0.081	0.083	0.005	0.015	0.021	0.008
LDL-cho (<60)	0.268	0.189	0.125	0.081	0.095	0.134	0.013	0.034	0.060	0.003
LDL-cho (60,140)	0.391	0.121	0.100	0.153	0.074	0.103	0.006	0.010	0.041	0.007
LDL-cho (140<=)	0.430	0.186	0.117	0.087	0.073	0.077	0.006	0.013	0.012	0.006
HDL-C (<40)	0.254	0.110	0.145	0.117	0.097	0.185	0.007	0.015	0.070	0.006
HDL-C (40,100)	0.409	0.123	0.097	0.145	0.070	0.100	0.007	0.010	0.039	0.007
HDL-C (100<=)	0.540	0.080	0.064	0.153	0.030	0.096	0.003	0.000	0.033	0.015
SGLT(no used)	0.381	0.125	0.103	0.142	0.074	0.112	0.008	0.012	0.045	0.009
SGLT(used)	0.494	0.162	0.146	0.100	0.039	0.039	0.000	0.003	0.016	0.007

Table 4 Stationery distribution and square error of each element

3.2 Calculation of stationary distribution of each element

Here, we calculate the stationary distribution for each element and compare it with the state ratio actually obtained in the data. Table 4 shows the difference between the stationary distribution and the actual data in that study. The difference from the actual data is calculated by subtracting the value and calculating the sum of squares. We can see that the steady distribution tends to be lower than the actual data. There is also a tendency for errors to tend to be large in Mg-related distributions, and we need attention. However, the correlation coefficient between the stationary distribution of state I and the actual data is 0.97, which can be said to well represent the influence of each element. The probability of state I in the case of using SGLT inhibitors is 0.0157, and from the actual data, it is 0.0156, indicating high accuracy.

3.3 Transition from specific state

Next, we make sure how long it takes for diabetes to shift from state to state I over time. In all patients, 1.3% of the patients transition to state I after 10 unit hours, even in state A in the first diagnosis from Table 5 of “All data” column. Similarly, 3.1% of patients in 20 unit hours and 3.8% of patients in 30 unit hours transition to state I. As age increases, the transition rate to state I increases. The transition from state B is not much different to the transition from state A, but the transition from states C and D to state I greatly increases. Patients who were not in states A and B in the initial diagnosis need to be careful with state transition. Furthermore, in the case of patients whose initial Mg concentration is out of the reference value range, the transition to state I is very high even if the condition is A or B. By contrast, when using SGLT inhibitors, the transition to state I is a very low value. By using SGLT inhibitors, it is considered that the transition to severe diabetes can be prevented.

3.4 State simulation of aggregation

Next, we try to summarize the state to improve the simulation accuracy. In the first transition probability matrix, convergence is also unstable, as shown in left of Figure 6. By aggregating states A and B, the error of the data with the actual data is reduced by 81%, and the accuracy can be increased as right of Figure 6 and Appendix A.3.

Table 5 Time-series probability from a specific state to state I

State	step	All data	Male	Female	age 11–20	age 21–30	age 31–40	age 41–50	age 51–60	age 61–70	age 71–80
A	10	0.013	0.013	0.015	0.000	0.008	0.004	0.005	0.009	0.014	0.020
A	20	0.031	0.030	0.033	0.000	0.021	0.013	0.010	0.022	0.034	0.044
A	30	0.038	0.038	0.041	0.000	0.029	0.020	0.012	0.027	0.044	0.054
B	10	0.020	0.019	0.022	0.000	0.015	0.007	0.006	0.015	0.023	0.029
B	20	0.034	0.034	0.037	0.000	0.026	0.016	0.011	0.025	0.039	0.049
B	30	0.040	0.039	0.042	0.000	0.032	0.021	0.013	0.029	0.046	0.056
C	10	0.035	0.036	0.035	0.000	0.052	0.018	0.013	0.029	0.040	0.047
C	20	0.042	0.043	0.043	0.000	0.050	0.025	0.014	0.033	0.049	0.058
C	30	0.043	0.043	0.045	0.000	0.045	0.027	0.014	0.032	0.050	0.059
D	10	0.028	0.026	0.031	0.000	0.015	0.012	0.013	0.019	0.030	0.034
D	20	0.039	0.038	0.041	0.000	0.026	0.019	0.013	0.027	0.043	0.050
D	30	0.041	0.041	0.044	0.000	0.032	0.023	0.013	0.030	0.048	0.056
State	step	age 81–90	age 91–100	Mg (<1.8)	Mg (1.8,2.4)	Mg (2.4<=)	Mg (<1.8)	Mg (1.8,2.4)	Mg (2.4<=)	BMI (<25)	BMI (25<=)
A	10	0.024	0.036	0.022	0.011	0.032	0.016	0.011	0.052	0.024	0.013
A	20	0.044	0.067	0.049	0.027	0.063	0.039	0.025	0.114	0.053	0.027
A	30	0.051	0.076	0.062	0.034	0.075	0.052	0.031	0.146	0.069	0.033
B	10	0.028	0.064	0.025	0.016	0.044	0.020	0.016	0.069	0.050	0.019
B	20	0.046	0.078	0.050	0.030	0.069	0.041	0.028	0.124	0.072	0.030
B	30	0.051	0.080	0.063	0.035	0.077	0.054	0.032	0.151	0.079	0.034
C	10	0.043	0.076	0.032	0.030	0.088	0.027	0.030	0.164	0.099	0.032
C	20	0.052	0.084	0.055	0.037	0.090	0.046	0.035	0.184	0.105	0.036
C	30	0.053	0.082	0.064	0.038	0.085	0.056	0.035	0.181	0.096	0.036
D	10	0.034	0.050	0.042	0.027	0.044	0.036	0.025	0.080	0.033	0.025
D	20	0.048	0.070	0.062	0.036	0.068	0.054	0.033	0.130	0.058	0.034
D	30	0.052	0.077	0.068	0.038	0.077	0.060	0.034	0.154	0.071	0.035
State	step	HbA1c (4.3,5.9)	HbA1c (5.9<=)	LDL-cho (<60)	LDL-cho (60,140)	LDL-cho (140<=)	HDL-C (<40)	HDL-C (40,100)	HDL-C (100<=)	SGLT (no used)	SGLT (used)
A	10	0.016	0.013	0.017	0.013	0.012	0.023	0.012	0.004	0.014	0.005
A	20	0.043	0.030	0.048	0.030	0.028	0.052	0.028	0.012	0.032	0.011
A	30	0.063	0.037	0.071	0.037	0.034	0.064	0.035	0.019	0.040	0.014
B	10	0.024	0.020	0.030	0.020	0.018	0.034	0.018	0.009	0.021	0.007
B	20	0.054	0.033	0.063	0.033	0.030	0.058	0.031	0.018	0.036	0.013
B	30	0.071	0.039	0.081	0.038	0.035	0.066	0.036	0.023	0.042	0.015
C	10	0.068	0.034	0.059	0.034	0.032	0.053	0.032	0.030	0.037	0.014
C	20	0.091	0.041	0.089	0.041	0.037	0.068	0.038	0.039	0.044	0.016
C	30	0.094	0.041	0.097	0.041	0.037	0.070	0.039	0.039	0.045	0.016
D	10	0.038	0.027	0.042	0.027	0.027	0.048	0.025	0.006	0.029	0.012
D	20	0.061	0.038	0.065	0.037	0.035	0.064	0.035	0.015	0.040	0.015
D	30	0.073	0.040	0.080	0.040	0.037	0.068	0.038	0.022	0.043	0.015

4. CLINICAL FEEDBACK AND APPLICATION

In this section, we consider how to feed back the simulation results of diabetic nephropathy to the clinic.

4.1 Clinical feedback

From the results of this study, we found that various factors affect the severity of diabetic nephropathy. In particular, it was confirmed that the influence of the Mg concentration was large. Patients whose Mg concentration is out of a reference value that introduced the proper range from 1.8 to 2.4 for Mg concentration in Section 2.1 have a significantly higher probability of becoming the state I than those in a reference value. The simulation also shows that the patient has a faster rate of symptoms worsening. Patients whose initial Mg concentration is out of a reference value are also distributed except A and B compared to patients within a reference value, and we need attention to the patient.

From the analysis of this study, there are also interesting results in clinical practice. Normally, HbA1c is considered to have a bad influence on diabetes if it is 5.9 or higher. However, in Appendix A.4, patients with HbA1c smaller than 5.9 tend to get worse. The number of patients with HbA1c less than 5.9 is small, but the proportion of state A is only 18.78%, the proportion of state A is quite small compared to that of HbA1c is more than 5.9 or the whole patient. The condition $(A + B)$ ratio of patients with HbA1c less than 5.9 is 52.49%, less than 62.02% of the whole patient. Also, the ratio of state H is very large, 7.28%. Patients with HbA1c less than 5.9 are considered to have severe diabetes already due to other factors. There are the following reasons : Patients who have deteriorated just before dialysis may sometimes become hypoglycemic without diabetes medicine. As kidney function decreases, doctors often reduce the amount of medication or reduce insulin. Therefore, the numerical value of HbA1c decrease [31]. BMI also tends to be severe for patients under 25. It implies that diabetes worsened and the numerical value of BMI decreased. We think that this analysis result seems

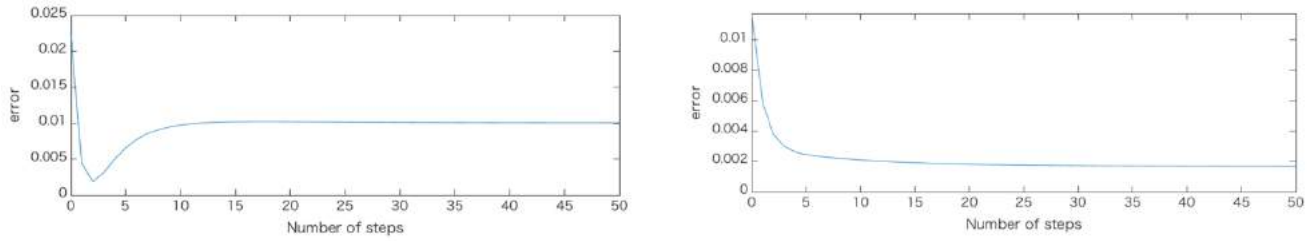

Fig. 6 Convergence of square error, Left : normal, Right : aggregating states A and B

to contain a lot of useful information for clinical doctors.

4.2 Cooperation with medical system

Currently in Japan there is a framework to conduct medical cooperation. As an example, there is Tonet [32]. The core function of diabetes-associated path is disease management of patients. Because the severity prediction function is not the current system, it is possible to support the function using the results of this study.

5. CONCLUSION

In this study, we constructed a simulation with a Markov chain on diabetes to prevent severe diabetes. In order to distinguish the severity of diabetes in patients, we classified the state of patients using eGFR and urine protein. First, using each factor, we showed the distribution of the patient's initial and ending status, and we examined which factors were more influential. We confirmed that age, Mg concentration, and use of SGLT inhibitors have a large influence on the state change of diabetic nephropathy. Temporal changes were also shown to worsen over time.

In this study, we considered how changing the deterioration of diabetic nephropathy with the passage of time is important, and developed a model using a Markov chain. From the actual data, a transition probability matrix based on each factor and a steady distribution were calculated. At this stage, the unit of the transition used the average inspection interval calculated from the actual data. By simulation using the transition probability matrix for each factor, the situation deteriorating to state I became clear. In addition, by aggregating states, we reduced simulation errors and devised measures to improve simulation accuracy. We hope that we can effectively predict the results of the examination and use the model of this study to predict patient condition deterioration.

As a future task, because we used the simple state classification in Table 2, we need to subdivide the state to simulate detailly. In addition, we think a more realistic simulation considering an influence between element items.

REFERENCES

[1] Ministry of Health, Labor and Welfare: “Overview of the National Health Care Fiscal Year in 2005,” (2006). https://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/15/. Accessed Sep. 2018 (in Japanese).

[2] Ministry of Health, Labor and Welfare, Press Release: “As a result of National Health and Nutrition Survey,” (2006). https://www.mhlw.go.jp/stf/houdou/0000177189.html. Accessed Sep. 2018 (in Japanese).

[3] Japan Preventive Association of Life-style related Disease: “Survey and Statistics of Diabetes,” (2016). http://www.seikatsusyukanbyo.com/statistics/2016/009222.php. Accessed Sep. 2018 (in Japanese).

[4] National Center for Global Health and Medicine: “Complications,” http://dmic.ncgm.go.jp/general/about-dm/060/index.html. Accessed Sep. 2018 (in Japanese).

[5] Hata, A., Y. Doi, T. Ninomiya, N. Mukai, Y. Hirakawa, et al.: “Magnesium Intake Decreases Type 2 Diabetes Risk through the Improvement of Insulin Resistance and Inflammation: The Hisayama Study.” *Diabetic Medicine: A Journal of the British Diabetic Association*, Vol. 30, No. 12, pp. 1487–94 (2013) doi:10.1111/dme.12250.

[6] Khan, Abigail May, Steven A. Lubitz, Lisa M. Sullivan, Jenny X. Sun, Daniel Levy, et al.: “Low Serum Magnesium and the Development of Atrial Fibrillation in the Community: The Framingham Heart Study,” *Circulation*, Vol. 127, No. 1, pp. 33–38 (2013). doi:10.1161/CIRCULATIONAHA.111.082511.

[7] Gröber, Uwe, Joachim Schmidt, and Klaus Kisters: “Magnesium in Prevention and Therapy,” *Nutrients*, Vol. 7, No. 9, pp. 8199–226 (2015). doi:10.3390/nu7095388.

[8] Qu, Xinhua, Fangchun Jin, Yongqiang Hao, Huiwu Li, Tingting Tang, et al.: “Magnesium and the Risk of Cardiovascular Events: A Meta-Analysis of Prospective Cohort Studies,” *PloS One*, Vol. 8, No. 3, p. e57720 (2013). doi:10.1371/journal.pone.0057720.

[9] Guasch-Ferré, Marta, Mònica Bulló, Ramon Estruch, Dolores Corella, Miguel A. Martínez-González, et al.: “Dietary Magnesium Intake Is Inversely Associated with Mortality in Adults at High Cardiovascular Disease Risk,” *The Journal of Nutrition*, Vol. 144, No. 1, pp. 55–60 (2014). doi:10.3945/jn.113.183012.

[10]Zinman, Bernard, Christoph Wanner, John M. Lachin, David Fitchett, Erich Bluhmki, et al.: “Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes,” *New England Journal of Medicine*, Vol. 373, No. 22, pp. 2117–2128 (2015). doi:10.1056/NEJMoa1504720.

[11] Maki OKUDAIRA, Yasuko UCHIGATA, Taisuke OKADA, and Yasuhiko IWAMOTO: “Influence of Health Checkup and Previous Intermittent Treatment on Diabetic Complications,” *Journal of the Japan Diabetes Society*, Vol. 46, No.10, pp. 781–785 (2003).

[12] Kazuo KATSUMATA, Yutaka OOISO, Jiro NAKAMURA,

Manabu SHIMIZU, Katsumoto KATO, et al.: “Prevalence and Outcome of Severe Hypoglycemia at a Hospital in Aichi Prefecture, Based on a Questionnaire Survey,” *Journal of the Japan Diabetes Society*, Vol. 49, No. 4, pp. 251–257 (2006). doi:10.11213/tonyobyo.49.251 (in Japanese).

[13] Mitsuru HOSHI, Takeo KotzuMi and Yukio SHIGETA: “Supplementary: Application of Multivariate Analysis on the Estimation of Prognosis and Establishment of Management of Diabetic Nephropathy, *Journal of the Japanese Society of Internal Medicine*, Vol. 68, No. 10, pp. 1239–1241 (1979). doi:10.2169/naika.68.1239 (in Japanese).

[14] Miotto, Riccardo, Li Li, Brian A. Kidd, and Joel T. Dudley.: “Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records,” *Scientific Reports*, Vol. 6, No. 1, p. 26094 (2016). doi:10.1038/srep26094.

[15] Fred Farrell, Orkun S. Soyer, Christopher Quince.: “Machine Learning Based Prediction of Functional Capabilities in Metagenomically Assembled Microbial Genomes,” *BioRxiv* (2018). https://www.biorxiv.org/content/10.1101/307157v1.abstract Accessed Sep. 2018.

[16] Kamado, K.: “Studies on Diabetic Nephropathy Estimation of Renal Glomerular Changes from Clinico-laboratory Data,” *Journal of the Japan Diabetes Society*, Vol. 23, No. 10, pp. 913‑921 (1980). doi:10.11213/tonyobyo1958.23.913 (in Japanese).

[17] Ragnarson Tennvall, G., and Apelqvist, J.: “Prevention of Diabetes-Related Foot Ulcers and Amputations: A Cost-Utility Analysis Based on Markov Model Simulations,” *Diabetologia*, Vol. 44, No. 11, pp. 2077–2087 (2001). doi:10.1007/s001250100013.

[18] Yu, Junhua, Bijal M. Shah, Eric J. Ip, and James Chan.: “A Markov Model of the Cost-Effectiveness of Pharmacist Care for Diabetes in Prevention of Cardiovascular Diseases: Evidence from Kaiser Permanente Northern California,” *Journal of Managed Care Pharmacy*, Vol. 19, No. 2, pp. 102–114 (2013). doi:10.18553/jmcp.2013.19.2.102

[19] Honeycutt, Amanda A., James P. Boyle, Kristine R. Broglio, Theodore J. Thompson, Thomas J. Hoerger, et al.: “A Dynamic Markov Model for Forecasting Diabetes Prevalence in the United States through 2050,” *Health Care Management Science*, Vol. 6, No. 3, pp. 155–164 (2003). doi:10.1023/A:1024467522972.

[20] Japan Society of Nephrology, CKD Clinical Practice Guide: “Clinical Practice Guidebook for Diagnosis and Treatment of Chronic Kidney Disease 2012,” *The Japanese Journal of Nephrology*, Vol. 54, No. 8, pp. 1031–1191 (2013). https://ci.nii.ac.jp/naid/40019514036/ (in Japanese).

[21] Lopes, L., R. Santos, B. Pereira, and V. Lopes.: “Maternal Perceptions of Children’s Weight Status.” *Child: Care, Health and Development*, Vol. 39, No. 5, pp. 728–736 (2013). doi:10.1111/j.1365-2214.2012.01380.x.

[22] National Kidney Foundation: “Estimated Glomerular Filtration Rate (EGFR),” (2015). https://www.kidney.org/atoz/content/gfr.

[23] 24-Hour Urine Protein Test: Purpose, Procedure, and Results. https://www.healthline.com/health/24-hour-urine-protein. Accessed Sep. 2018.

[24] Swaminathan, R.: “Magnesium Metabolism and Its Disorders,” *The Clinical Biochemist Reviews*, Vol. 24, No. 2, pp. 47–66 (2003). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1855626/.

[25] Drugs, Suitability, Benefits & Side Effects. “Sodium-Glucose Co-Transporter-2 (SGLT2) Inhibitors Are Used for Treating Type 2 Diabetes by Helping the Kidneys to Lower Blood Glucose Levels,” *Diabetes*, (2019), https://www.diabetes.co.uk/diabetes-medication/sglt2-inhibitors.html. Accessed Sep. 2018.

[26] Diabetes Digital Media Ltd.: “Glycosylated Haemoglobin & Diabetes. HbA1c Facts, Units, Diagnosis, Testing Frequency, Limitations, Control & Conversion. How Blood Glucose Levels Link to A1c.” *Diabetes* (2019). https://www.diabetes.co.uk/what-is-hba1c.html. Accessed Sep. 2018.

[27] Friedewald, William T, Robert I Levy, and Donald S Fredrickson.: “Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge.” *Clinical Chemistry*, Vol. 18, No. 6, pp. 499–502 (1972), doi:10.1093/clinchem/18.6.499.

[28] “HDL (Good), LDL (Bad) Cholesterol and Triglycerides.” Www.Heart.Org, https://www.heart.org/en/health-topics/cholesterol/hdl-good-ldl-bad-cholesterol-and-triglycerides. Accessed Sep. 2018.

[29] Diabetes Control and Complications Trial Research Group, D. M. Nathan, S. Genuth, J. Lachin, P. Cleary, et al.: “The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus,” *The New England Journal of Medicine*, Vol. 329, No. 14, pp. 977–986 (1993). doi:10.1056/NEJM199309303291401.

[30] Ross, S. M.: *Introduction to Probability Models, 11th Edition*, Academic Press (2014)

[31] Alsahli, M., and Gerich, J. E.: “Hypoglycemia, Chronic Kidney Disease, and Diabetes Mellitus,” *Mayo Clinic Proceedings*, Vol. 89, No. 11, pp. 1564–1571 (2014). doi:10.1016/j.mayocp.2014.07.013.

[32] The Community Medicine Network System in the Tone Medical Care Area in Saitama, Tonet. http://www.saitama-tonet.jp/. Accessed Sep. 2018 (in Japanese).

A. Appendix

A.1 State transition of patients classified by eGFR and urinary protein

	All data	Male	Female	age 11–20	age 21–30	age 31–40	age 41–50	age 51–60	age 61–70	age 71–80
People	4376	2684	1681	8	57	237	427	795	1137	1059
Mean period to I (days)	1362	1354	1384	0	1392	1719	1160	1767	1581	1515
Start rate of I(%)	1.6	1.53	1.73	0	0	0.84	0.7	1.13	1.32	1.98
End rate of I(%)	5.51	5.33	5.77	0	3.51	3.38	2.34	4.65	5.89	6.8
Increase rate of I	3.44	3.48	3.34	-	-	4.02	3.34	4.12	4.46	3.43
	age 81–90	age 91–100	Mg (<1.8)	Mg (1.8,2.4)	Mg (2.4<=)	$\overline{Mg}$ (<1.8)	$\overline{Mg}$ (1.8,2.4)	$\overline{Mg}$ (2.4<=)	BMI (<25)	BMI (25<=)
People	553	92	97	828	82	112	841	54	85	2969
Mean period to I (days)	599	445	2210	2504	1668	1786	2412	2253	974	1704
Start rate of I(%)	3.25	2.17	3.09	0.48	4.88	2.68	0.48	7.41	3.53	1.08
End rate of I(%)	6.33	9.78	11.34	4.71	13.41	8.04	4.76	22.22	8.24	5.15
Increase rate of I	1.95	4.51	3.67	9.81	2.75	3.00	9.92	3.00	2.33	4.77
	HbA1c (4.3,5.9)	HbA1c (5.9<=)	LDL-cho (<60)	LDL-cho (60,140)	LDL-cho (140<=)	HDL-C (<40)	HDL-C (40,100)	HDL-C (100<=)	SGLT (no used)	SGLT (used)
People	181	4037	104	2401	510	428	3001	41	4180	196
Mean period to I (days)	1509	1398	1016	1953	1408	1332	1837	1000	1356	1950
Start rate of I(%)	2.21	1.54	2.88	1.12	0.59	2.34	0.9	0	1.67	0
End rate of I(%)	8.84	5.42	9.62	5.25	5.69	8.88	5.13	4.88	5.69	1.53
Increase rate of I	4.00	3.52	3.34	4.69	9.64	3.79	5.70	-	3.41	-

A.2 Transition probability matrix for all patients

From \To	A	B	C	D	E	F	G	H	I
A	0.84806	0.07209	0.02172	0.04935	0.00581	0.00241	0.00033	5.00E-05	0.00017
B	0.21907	0.62321	0.08369	0.01353	0.04965	0.00976	0.00015	0.00059	0.00035
C	0.07991	0.10068	0.71244	0.0043	0.00807	0.09334	7.00E-05	0.00033	0.00086
D	0.13503	0.01249	0.00333	0.73137	0.08019	0.02396	0.01111	0.00164	0.00087
E	0.03107	0.09284	0.01252	0.14802	0.58258	0.10631	0.00235	0.02034	0.00395
F	0.01031	0.01045	0.08918	0.03212	0.07218	0.74639	0.00042	0.00237	0.03658
G	0.00761	0.0000	0.00109	0.24348	0.02065	0.01413	0.60217	0.07609	0.03478
H	0.00195	0.00456	0.00195	0.01954	0.13941	0.0228	0.06319	0.64235	0.10423
I	0.00254	0.00127	0.00169	0.00148	0.00551	0.09068	0.0053	0.0303	0.86123

A.3 Stationary distribution and square error of each element aggregated state A and B

	A and B	C	D	E	F	G	H	I	Difference
All data	0.466	0.130	0.126	0.084	0.125	0.007	0.012	0.049	0.0007
MALE	0.460	0.141	0.115	0.082	0.139	0.006	0.009	0.048	0.0019
FEMALE	0.469	0.112	0.143	0.088	0.109	0.009	0.018	0.053	0.0002
age(11–20)	0.802	0.198	0.000	0.000	0.000	0.000	0.000	0.000	0.0108
age(21–30)	0.741	0.161	0.003	0.001	0.040	0.000	0.000	0.055	0.0363
age(31–40)	0.684	0.192	0.017	0.003	0.061	0.000	0.004	0.037	0.0085
age(41–50)	0.668	0.169	0.042	0.041	0.058	0.001	0.006	0.015	0.0013
age(51–60)	0.535	0.143	0.095	0.060	0.123	0.002	0.004	0.039	0.0038
age(61–70)	0.450	0.117	0.134	0.086	0.133	0.007	0.015	0.059	0.0008
age(71–80)	0.352	0.110	0.177	0.113	0.154	0.012	0.017	0.064	0.0008
age(81-90)	0.290	0.124	0.200	0.134	0.153	0.020	0.025	0.054	0.0045
age(91–100)	0.268	0.220	0.135	0.095	0.155	0.016	0.027	0.084	0.0079
Start Mg(<1.8)	0.399	0.172	0.114	0.068	0.139	0.011	0.023	0.074	0.0193
Start Mg(1.8,2.4)	0.491	0.115	0.122	0.082	0.129	0.006	0.009	0.045	0.0023
Start Mg(2.4<=)	0.324	0.065	0.198	0.115	0.160	0.012	0.031	0.095	0.0096
Mg(<1.8)	0.429	0.155	0.098	0.070	0.144	0.013	0.021	0.070	0.0134
Mg(1.8,2.4)	0.487	0.118	0.128	0.084	0.128	0.005	0.009	0.041	0.0022
Mg(2.4<=)	0.241	0.029	0.173	0.102	0.185	0.027	0.054	0.188	0.0310
BMI(<25)	0.346	0.189	0.087	0.065	0.192	0.006	0.018	0.097	0.0038
BMI(25<=)	0.465	0.128	0.136	0.089	0.123	0.006	0.011	0.042	0.0006
HbA1c (4.3,5.9)	0.434	0.117	0.135	0.081	0.117	0.006	0.015	0.096	0.0041
HbA1c (5.9<=)	0.467	0.130	0.126	0.085	0.126	0.007	0.012	0.048	0.0011
LDL-cho (<60)	0.402	0.128	0.076	0.082	0.165	0.012	0.027	0.107	0.0043
LDL-cho (60,140)	0.460	0.127	0.138	0.087	0.122	0.006	0.012	0.047	0.0006
LDL-cho (140<=)	0.486	0.120	0.107	0.084	0.140	0.008	0.012	0.043	0.0086
HDL-C (<40)	0.331	0.160	0.107	0.105	0.199	0.007	0.015	0.076	0.0064
HDL-C (40,100)	0.477	0.125	0.131	0.082	0.120	0.007	0.012	0.046	0.0010
HDL-C (100<=)	0.484	0.107	0.151	0.048	0.156	0.003	0.000	0.053	0.0090
SGLT(no used)	0.457	0.127	0.128	0.086	0.131	0.008	0.013	0.052	0.0008
SGLT(used)	0.605	0.191	0.089	0.044	0.048	0.000	0.003	0.019	0.0099

A.4 State distribution when classified by HbA1c and BMI

Factor	Number of Patient	State A (%)	State B (%)	State H (%)	State I (%)
All Patient (start)	4376	32.52	29.50	2.58	1.60
HbA1c $<$ 5.9 (start)	181	18.78	33.70	7.18	2.21
HbA1c $\geq$ 5.9 (start)	4037	33.86	29.43	2.23	1.54
BMI $<$ 25 (start)	85	17.65	34.12	4.71	3.53
BMI $\geq$ 25 (start)	2969	33.48	29.03	1.95	1.08