School of Engineering and Applied Science (SEAS), Ahmedabad University

B.Tech(ICT) Semester IV: Probability and Random Processes (MAT 202)

- **Project Title:** ***Alzheimer Disease Prediction Model***
- Group No : H_B33
- Name (Roll No) :

1. Parth N Patel (AU1841028)
2. Shivam Lakhtariya (AU1841084)

1 Introduction

1.1 Background

Alzheimer's disease is also known as a disease of Old Age. There is no specific drug to cure Alzheimer's Disease(AD) or to slowdown it. Generally AD is caused between age of 60-75 years. A Recent discovery had shown that the presence of $\epsilon 4$ [1] of the ApoE Gene indicates the presence of the onset of Alzheimer Disease. The Presence of AD Diseases can be known by the following facts:

1. Large Amount of Deposits of β-amyloid protien
2. Neurofibrillary tangles (connections between neurones)
3. Amyloid Protien in the arteries of brain.
4. Loss of Neurones

AD is the most commonest form of cognitive impairement, it is hard to diagnose AD with certainty except by post mortem examination. But as the ApoE gene is present in it, we can use to diagnose with it. The ApoE gene has three common alleles - ϵl, $\epsilon 3$ and $\epsilon 4$. possible genotypes ($\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 3$, e3/e4 and e4/e4). The ApoE$\epsilon 4/\epsilon 4$ has the highest risk factor of AD then any other combinations of the gene. So we are using the proportionn of $\epsilon 4/\epsilon 4$ and $\epsilon 3/\epsilon 4$ [2] to calculate the incidence rate of AD in man and woman. We use a [3] continuous-time multiple state model(continuous time Markov Nikov model).We have used a Markov Models because as this disorder can be described as multi stage progression process.

We have used a Continuous-time multiple state model to predict the Alzheimer Disease when the given set of information about the patient is given in advance.

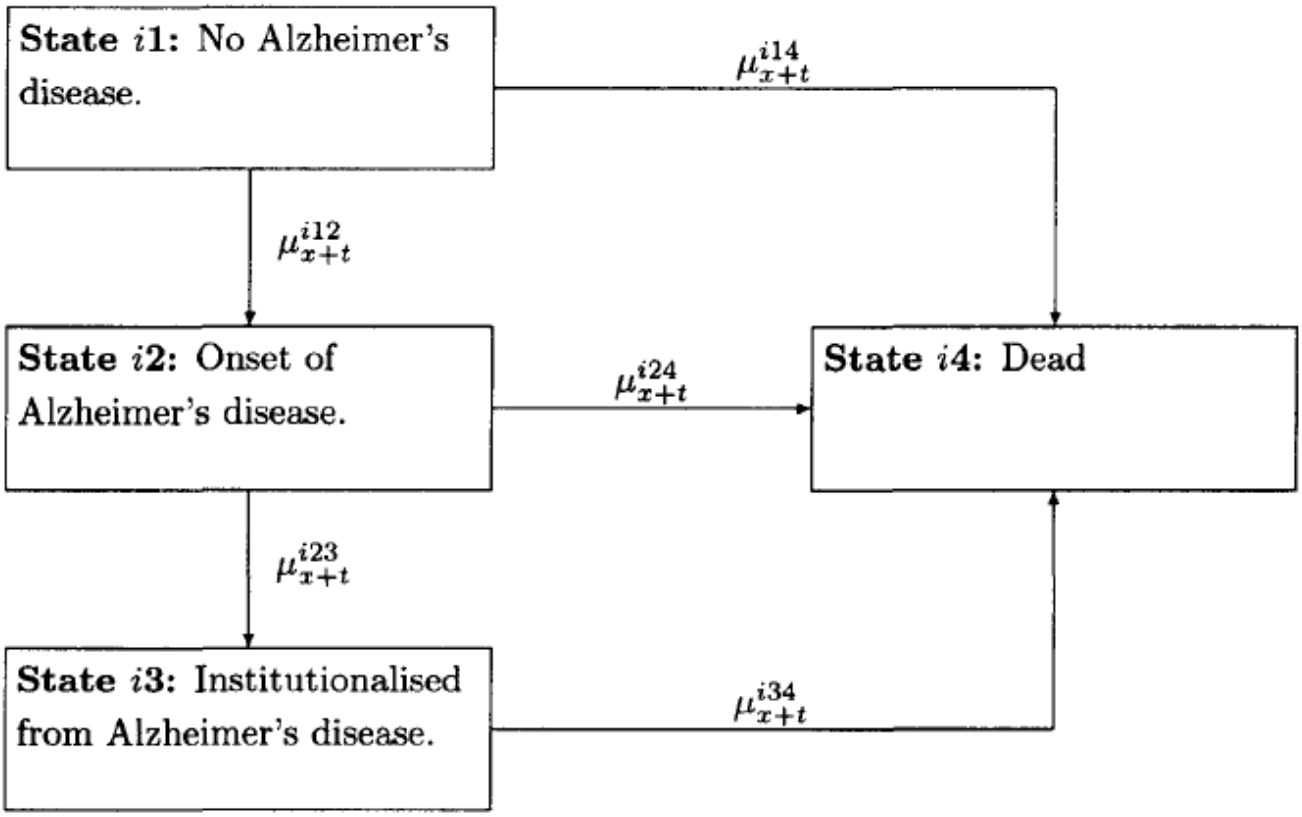

Figure 1: A simple model of Alzheimer's disease in the rth of M subgroups, each representing a different ApoE genotype, x is the age at outset, and t the elapsed duration.

1.2 Motivation

We have been motivated to developed and do some research on the Probabilistic Model to Predict the Alzheimer Disease as it connects to aspects of life i.e It connects mathematics to the life sciences part of our life. It display that mathematics is an important part of life and almost nothing can be learned without having basic knowledge or Mathematics. In our model it determines how the subject probability is connected to our daily life and how it can be used to Predict a severe disease like Alzheimer Disease. Developing this Probabilistic model we have applied all the basics concept of Probability which we have learned in this course throughout the whole semester.

1.3 Problem Statement/ Case Study

Alzheimer Disease is a disease of old age and presently it has no cure for it. So here we are developing a Continuous Time Markov model to predict whether Alzheimer Disease is present or not so that we can take

necessary measures regarding it. We are using Evidence [4] APoE Gene E4 evidences to predict whether Allzheimer Disease is present or not. We have specified a simple continuous-time Markov model of AD allowing for the variability of the ApoE gene. We allow all the intensities in the model to be estimated simultaneously. We know that Alzheimer's is a more genetic-based problem and if we can estimate it by our gene types then we can solve it better way. So that we made a model based on gene type in our project.

2 Data Acquisition

- No our Special Assignment is not Data Dependent.

3 Probabilistic Model Used

- Block diagram is as given in Figure 1
- We use a continuous-time four state Markov model. Figure 1 shows a simple model of AD. Each genotype is represented by such a model. The transition intensities in each model will differ, representing the different genetic risks, x denotes the age at outset and t the elapsed duration. Gompertz curve will be the best fit, despite trying a number of more complex models.
  Using Gompertz equation and Markov chain,transition intensity of incident of AD was found out from Mortality table

$$\mu_{x+t}^{AD} = 1.31275 \times \mu 10^{-7} \times \exp^{0.1459(x+t)}$$
$$\mu_{x+t}^{jk} = A + DB \exp^{C(x+t)}$$

  Here, μ_{x+t}^{jk} is denotes transition intensity from j state to k state in x+t time.

- Assumptions about the general level of mortality: We use parametric approximation to the AM80 and AF80 Ultimate mortality tables as bases for mortality assumptions,y assumptions; for use in the model they are adjusted in a variety of ways. Gompertz curve was ploted to μ_{x+t} at age range of 65-120, using log-linear least square method.

$$AM80\mu_{x+t} = 0.000094116 \exp^{0.084554(x+t)}$$

$$AF80\mu_{x+t} = 0.000025934 \exp^{ea093605(x+t)}$$

  As see that Gompertz approximations had a negligible effect in long-term care applications; we use them because they are sometimes useful in numerical work.

- we estimate the aggregate incidence of AD, denoted μ_{x+t}^{AD},

$$\mu_{x+t}^{AD} = 1.31275 \times 10^{-7} \times \exp^{0.1459(x+t)}$$

- from Onset of Alzheimer's Disease to Institutionalisation or Death. We don't have sufficient data to analyse μ_{x+t}^{i23},μ_{x+t}^{i24},μ_{x+t}^{i34}, by genotype.

  So we wiil write μ_{x+t}^{23},μ_{x+t}^{24} in place of μ_{x+t}^{i23} and μ_{x+t}^{i24} respectively because we are not differentiating with respect to genotype. Here we define two function (I_j) and simple path function (N_{jk}) for finding the years that patient will live the life.P_{xy}^{ij} is probability that life will in stage i at age x and life will in stage j at age y.

$$I_j(t) = \begin{cases} 1 & \text{if life is in state } j \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$
$$dN_{jk}(t) = \begin{cases} 1 & \text{if life transfers from state } j \text{ to state } k \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$
$$N_{jk}(T) = \int_0^T dN_{jk}(t) = \text{No. of transfers from state } j \text{ to state } k$$

- The mean age at onset of AD, given that the life was eventually institutionalised with AD.

$$E[x + \int_x^w I_1(t)dt | N_{23}(w-x) = 1 \; and \; I_1(x) = 1] =$$
$$= x + \frac{\int_x^w (t-x)\mu_t^{12} P_{xt}^{11} [\int_x^w \mu_s^{23} P_{ts}^{22} ds]dt}{\int_x^w \mu_t^{12} P_{xt}^{11} [\int_x^w \mu_s^{23} P_{ts}^{22} ds]dt}$$

- The mean time from onset of AD to institutionalisation:

$$E[\int_x^w I_2(t)dt | N_{23}(w-x) = 1 \; and \; I_1(x) = 1] =$$
$$= \frac{\int_x^w \mu_t^{12} P_{xt}^{11} [\int_x^w (s-t)\mu_s^{23} P_{ts}^{22} ds]dt}{\int_x^w \mu_t^{12} P_{xt}^{11} [\int_x^w \mu_s^{23} P_{ts}^{22} ds]dt}$$

  Here w denotes upper bound age which we took at 120 years age.

  Using the Gompertz approximation . Although it is appropriate to allow for future improvements in mortality in applications, it is not appropriate to do so in estimation based on past data. The values of constant do not depend strongly on the baseline mortality.

$$\mu_{x+t}^{12} = A + \mu_{x+t}^{AD}$$
$$\mu_{x+t}^{23} = D$$
$$\mu_{x+t}^{24} = P\mu_{x+t}^{14}$$
$$\mu_{x+t}^{14} = AM\mu_{x+t}$$

  μ_{x+t}^{14}baseline mortality, was taken as AM 80 mortality, using the Gompertz approximation.Here A,D and P is constant which we derive by extending Gompertz approximation. SO,

$$\mu_{x+t}^{jk} = A + DB\exp^{C(x+t)}$$

- model the incidence of AD for the ith genotype as: For use in our model, these odds ratios have to be converted into relative risks.transition intensities that are consistent with the odds ratios and together are consistent with the aggregate incidence of AD.

$$\mu_{x+t}^{i12} = r_1 f_{x+t}^{i} \mu_{x+t}^{AD}$$

where f_{x+t}^{i} is a parametric function representing the relative risk. $f_{x+t}^{i} = 1$ in the case of the $\epsilon 3/\epsilon 3$ genotype.
μ_{x+t}^{AD} is a aggregate incident rate of AD.
r_1 is a constant.

- To determine the parameter r_1, we calculated the aggregate incidence of AD in the whole model, and fitted this to μ_{x+t}^{AD} by least square, If ${}_iP_{xt}^{11}$ is the probability that a life with genotype i, healthy at age x, is unaffected by AD at age x + t,and if P_i^x is the population frequency of the ith genotype at age x then the aggregate incidence of AD
aggregation intensity at x+t = $r_1 \left\{ \sum_{n=1}^{5} P_x^i P_{xt}^{11 f_{x+t}^{i}} \right\} \mu_{x+t}^{AD}$

- To allow for this possibility we will also consider models assuming that the true relative risks are a proportion $m < 1$ of those estimated above. We do this by adjusting equation so that for genotype i:

$$\mu_{x+t}^{i12} = r_m[(f_{x+t}^{i} - 1)m + 1]\mu_{x+t}^{AD}$$

Here, r_m is chosen as above so that the aggregate incidence of AD in the model is consistent with μ_{x+t}^{AD}

4 Pseudo Code/ Algorithm

- Pseudo Code for Figure 2
begin
declare x1 from 65 to 120
calculate y1 equal from the formula (10^{-5})*9.4116 * exp(0.084554 * x1)
declare x2 from 65 to 120
calculate y2 from the formula (10^{-5})*2.593 * exp(0.084554 * x2)
plot(x1,y1,x2,y2)
end

- Pseudo Code for Figure 3
begin
declare x from 40 to 90
assign E1 equal to 2.98

```
assign F1 equal to 0.00312
assign G1 equal to 0
assign H1 equal to 1
assign k11 equal to 62
assign k12 equal to 0
assign P1 equal to exp(-F1 .* ((x − k11).^2) - G1.*(x-k12))
assign fi1 equal to E1 .* P1 + H1
assign E2 equal to 13.5
assign F2 equal to 0.00529
assign G2 equal to 0
assign H2 equal to 1
assign k21 equal to 60
assign k22 to 0
assign P2 equal to exp(-F2 .* ((x − k21).^2) - G2.*(x-k22))
assign fi2 equal to E2 .* P2 + H2

assign E3 equal to 2.87
assign F3 equal to 0.00938
assign G3 equal to 0
assing H3 equal to 1
assign k31 equal to 68
assign k32 equal to 0

assign P3 equal to exp(-F3 .* ((x − k31).^2) - G3.*(x-k32))
assign fi3 E3 .* P3 + H3
assign E4 0.754
assign F4 equal to 0
assign G4 equal to 0.00859
assign H4 equal to 0
assign k41 equal to 0
assign k42 equal to 60
assign P4 equal to exp(-F4 .* ((x − k41).^2) - G4.*(x-k42));
assign fi4 equal to E4 .* P4 + H4
plot(x,fi1,x,fi2,x,fi3,x,fi4)
end
```

5 Coding and Simulation

5.1 Simulation Framework

f^i_{x+t} is a parametric function representing the risk relative to the aggregate incidence rate.

$$f^i_{x+t} = E \exp^{-F((x+t)-k2)^2-((x+t)-k2)} + H$$

By considering the form of the OR, we set either F = 0 for giving an exponential function or G = 0 for giving a bell-curve function.

Value of k1 or k2 is the nearest integer by inspection.

Whole result dependent is odd rations, they are controlling parameters of the all result.

5.2 Reproduced Figures

- Used Tools (MATLAB)
- Reproduced Figures

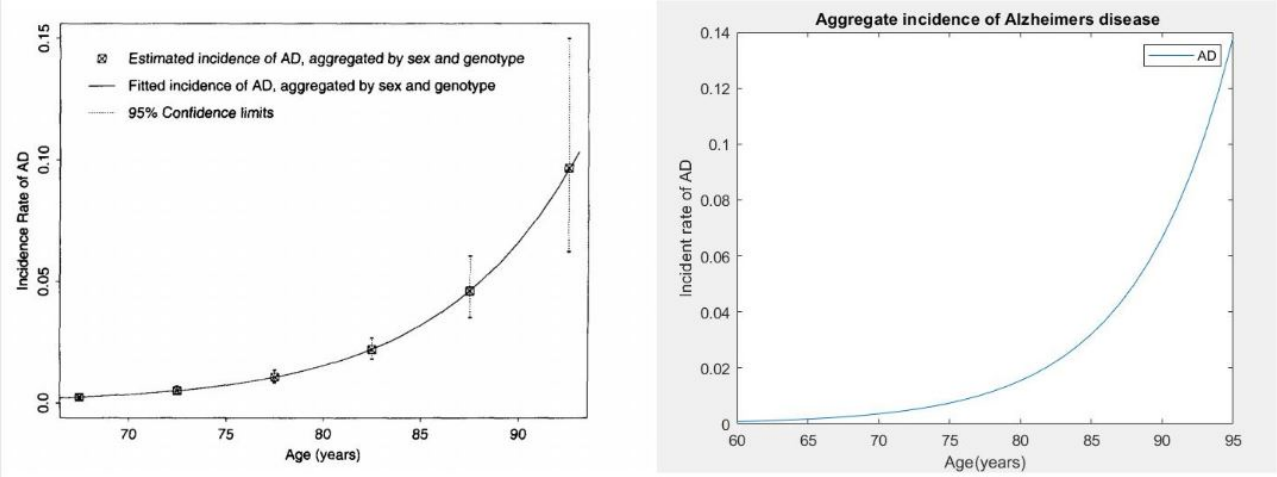

Figure 2: Aggregate incidence of Alzheimer's disease

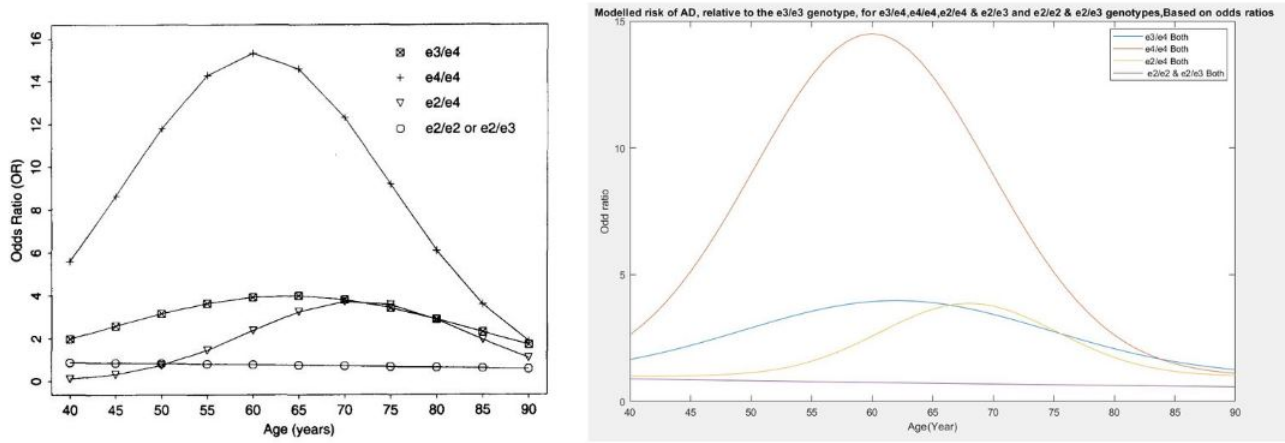

Figure 3: Odds ratios (ORs) of AD relative to $\epsilon 1/\epsilon 3$ genotype for males and females combined.

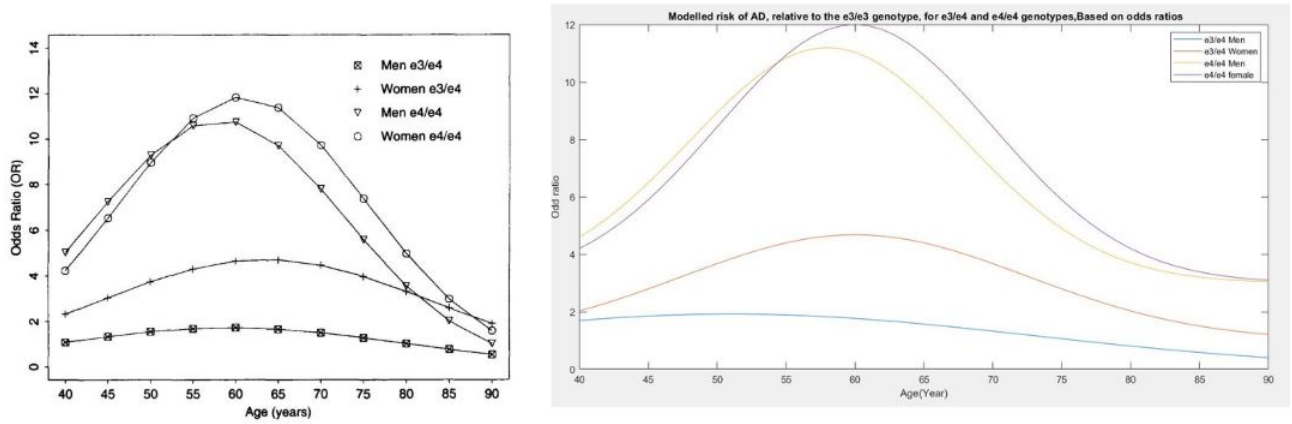

Figure 4: Odds ratios (ORs) of AD relative to $\epsilon 3/\epsilon 3$ genotype for $\epsilon 3/\epsilon 4$ and $\epsilon 4/\epsilon 4$ genotypes

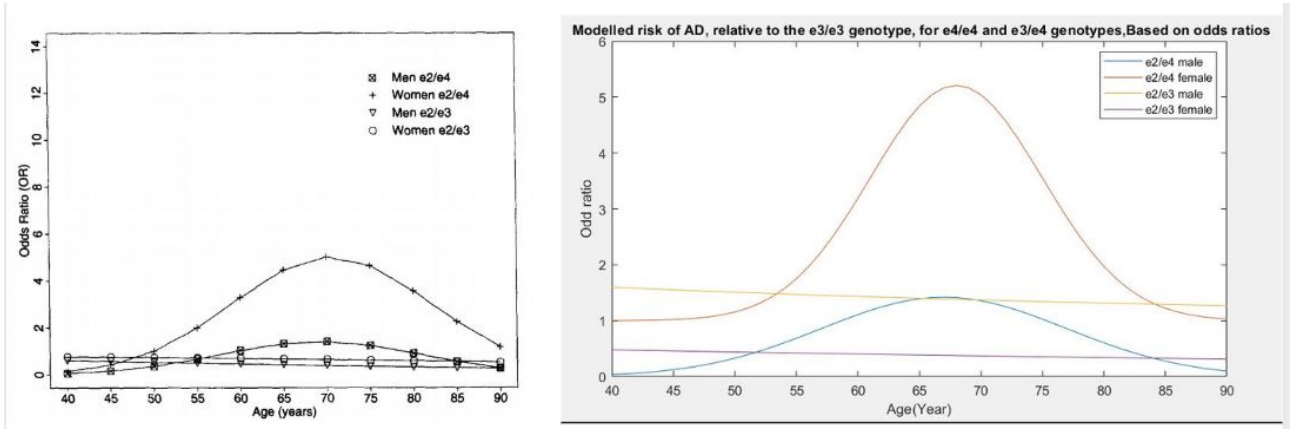

Figure 5: Odds ratios (ORs) of AD relative to $\epsilon 3/\epsilon 3$ genotype for $\epsilon 2/\epsilon 2$ or $\epsilon 2/\epsilon 3$ and $\epsilon 2/\epsilon 4$ genotypes.

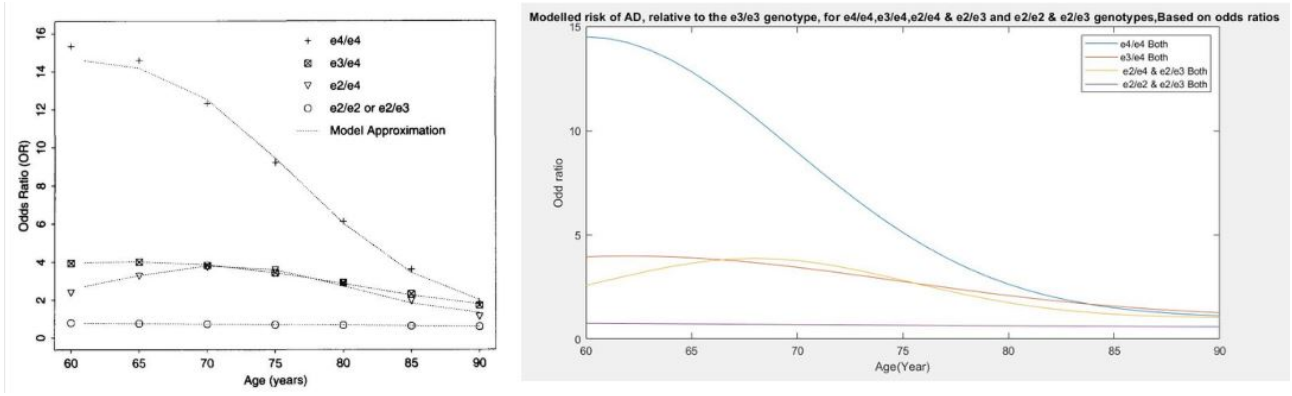

Figure 6: Odds ratios (ORs) of AD relative to $\epsilon 3/\epsilon 3$ genotype, compared with ORs computed using modelled relative risk functions.

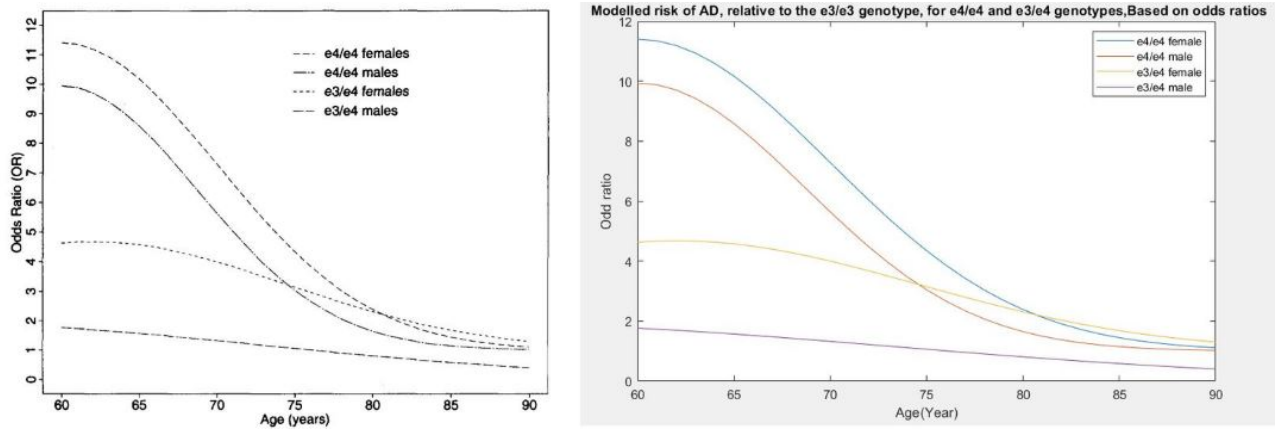

Figure 7: Modelled risk of AD, relative to the $\epsilon 3/\epsilon 3$ genotype, for $\epsilon 4/\epsilon 4$ and $\epsilon 3/\epsilon 4$ genotypes. Based on odds ratios

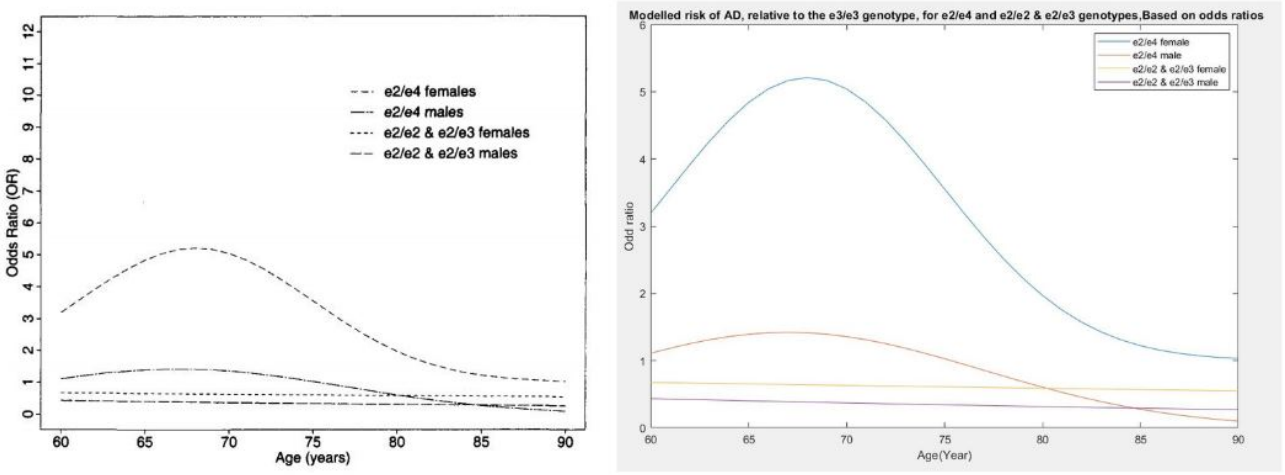

Figure 8: Modelled risk of AD, relative to the $\epsilon 3/\epsilon 3$ genotype, for $\epsilon 2/\epsilon 4$ and $\epsilon 2/\epsilon 2$ $\epsilon 2/\epsilon 3$ genotypes. Based on odds ratios

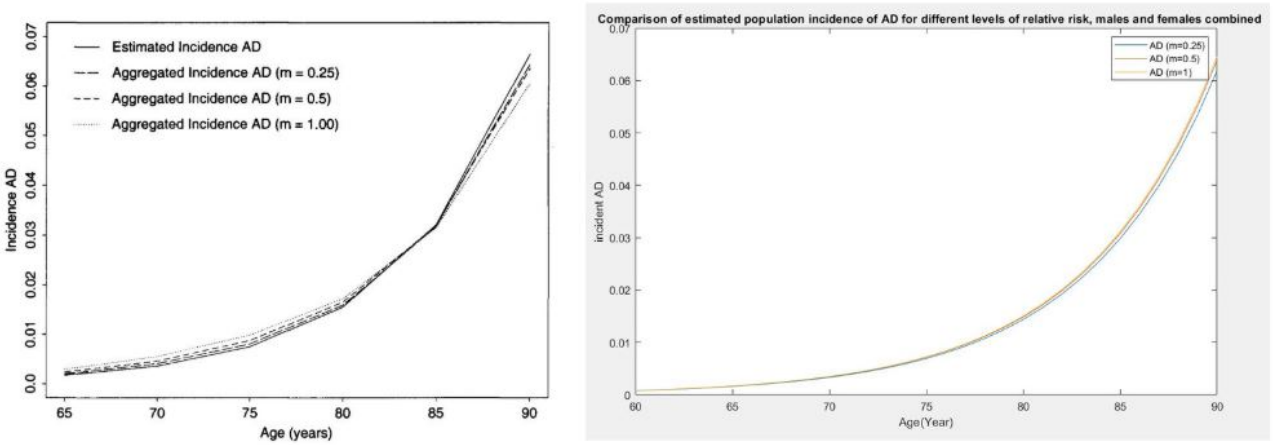

Figure 9: Comparison of estimated population incidence of AD μ_{x+t}^{AD}, with the aggregated incidence of AD for different levels of relative risk, males and females combined.

5.3 New Work Done

5.3.1 New Analysis

- A simple model of Alzheimer's disease,which preditct the probablity of changing the states
- Assumption:-

  (1)All states are independent of each other.

  (2)All makehom constants are independent to each other.

  (3)D and P are not dependent on μ_{x+t}^{AM80} and A is a nuisance parameter.

  (4)Here j is a lower limit which is 60 and w is a upper limit which is 95 in our case.

$$AM80\mu_{x+t} = 0.000094116 \exp^{0.084554(x+t)}$$

$$\mu_{x+t}^{AD} = 1.31275 \times 10^{-7} \times \exp^{0.1459(x+t)}$$

$$\mu_{x+t}^{12} = A + \mu_{x+t}^{AD}$$

$$\mu_{x+t}^{23} = D$$

$$\mu_{x+t}^{24} = P\mu_{x+t}^{14}$$

$$\mu_{x+t}^{14} = AM80\mu_{x+t}$$

$$TransitionIntensity = \int_j^w \mu_{x+t}^{12}\mu_{x+t}^{23}\mu_{x+t}^{24}\mu_{x+t}^{14}dt$$

$$= \int_j^w (A + \mu_{x+t}^{AD})D(P\mu_{x+t}^{14})(AM80\mu_{x+t})dt$$

$$= \int_j^w DP(AM80\mu_{x+t}^{24})dt + \int_j^w DP(AM80\mu_{x+t})\mu_{x+t}^{14}dt$$

- Probability of changing state from No Alzheimer to on set Alzheimer for different genotype

  (1)f^i_{x+t} is parametric function representing the risk relative to the aggregate incidence rate.

  (2) r_i is a constant chosen so that the aggregate incidence of AD based on the modelled intensities is consistent with the aggregate incidence

  (3)Here j is a lower limit which is 60 and w is a upper limit which is 95 in our case.

$$\mu^{AD}_{x+t} = 1.31275 \times 10^{-7} \times \exp^{0.1459(x+t)}$$

$$f_i = E \exp^{-F((x+t)-k2)^2-((x+t)-k2)} + H$$

$$\mu^{i12}_{x+t} = r_1 f^i_{x+t} \mu^{AD}_{x+t}$$

$$= \int_j^w \mu^{i12}_{x+t} dt$$

$$= \int_j^w r_1 f^i_{x+t} \mu^{AD}_{x+t} dt$$

$$= \int_j^w r_1 (E \exp^{-F((x+t)-k2)^2-((x+t)-k2)} + H) 1.31275 \times 10^{-7} \times \exp^{0.1459(x+t)}$$

5.3.2 New Results

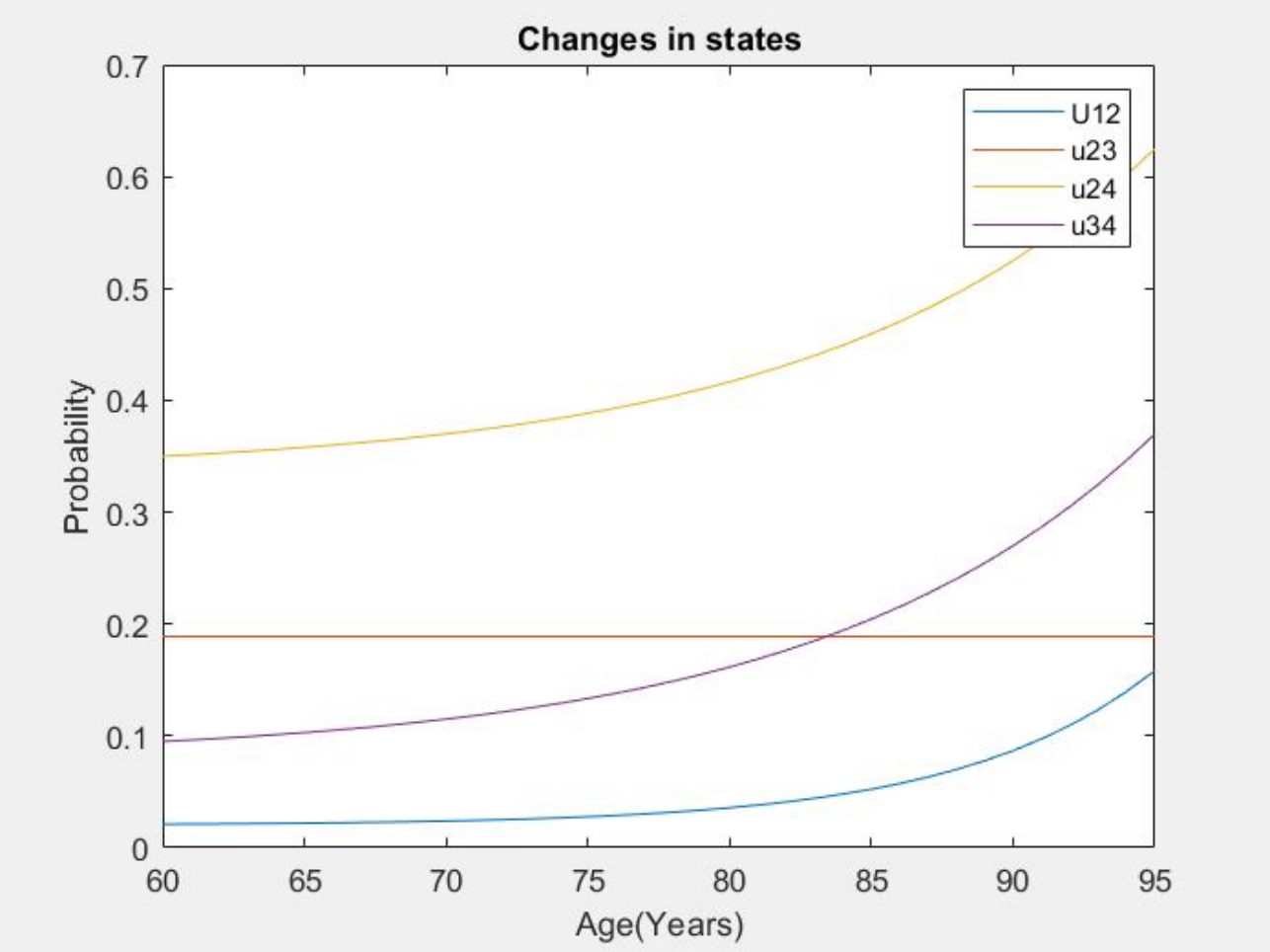

Figure 10: A simple model of Alzheimer's disease,which preditct the probablity of changing the states

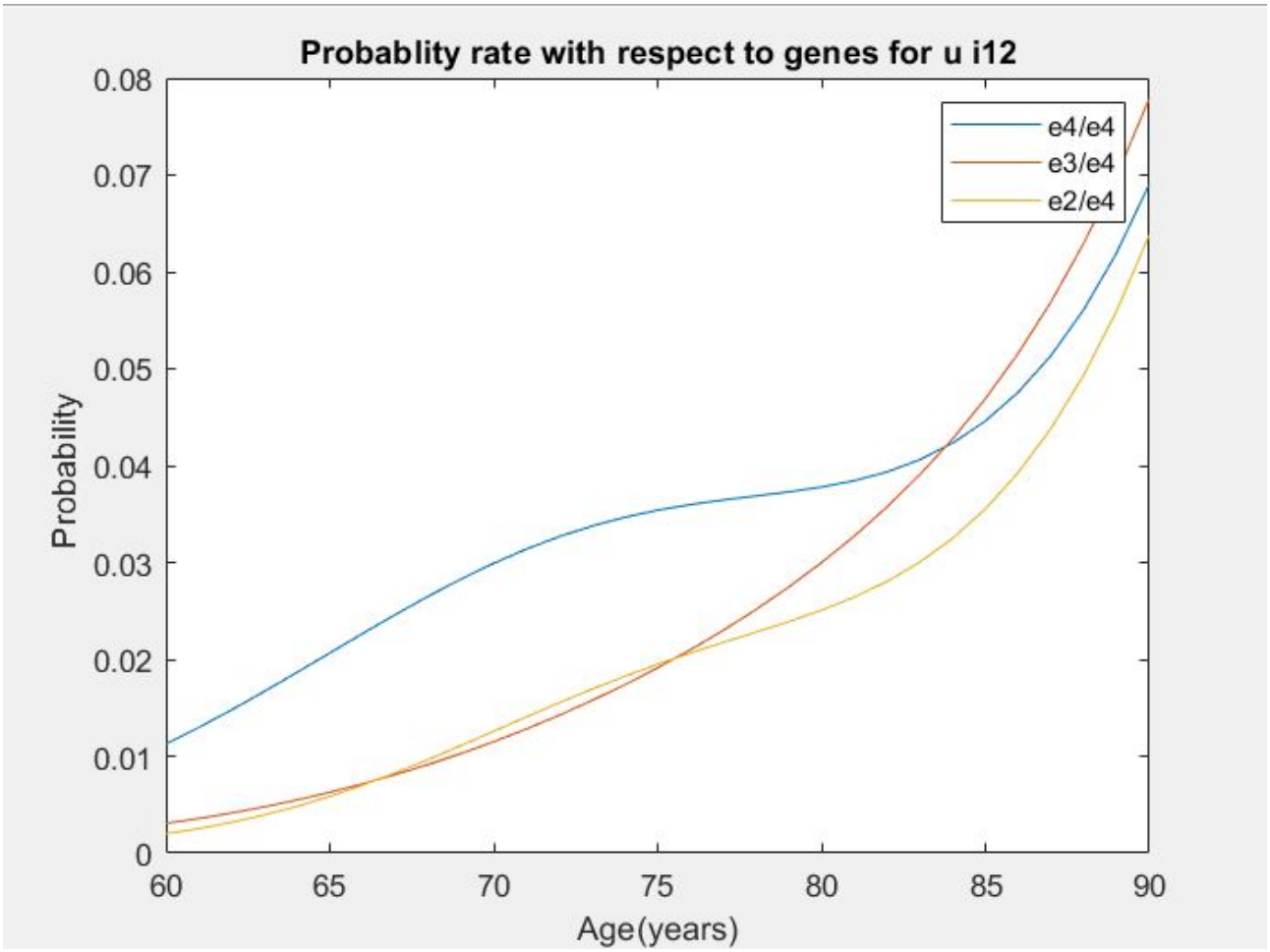

Figure 11: Probability of changing state from No Alzheimer to on set Alzheimer for different genotype

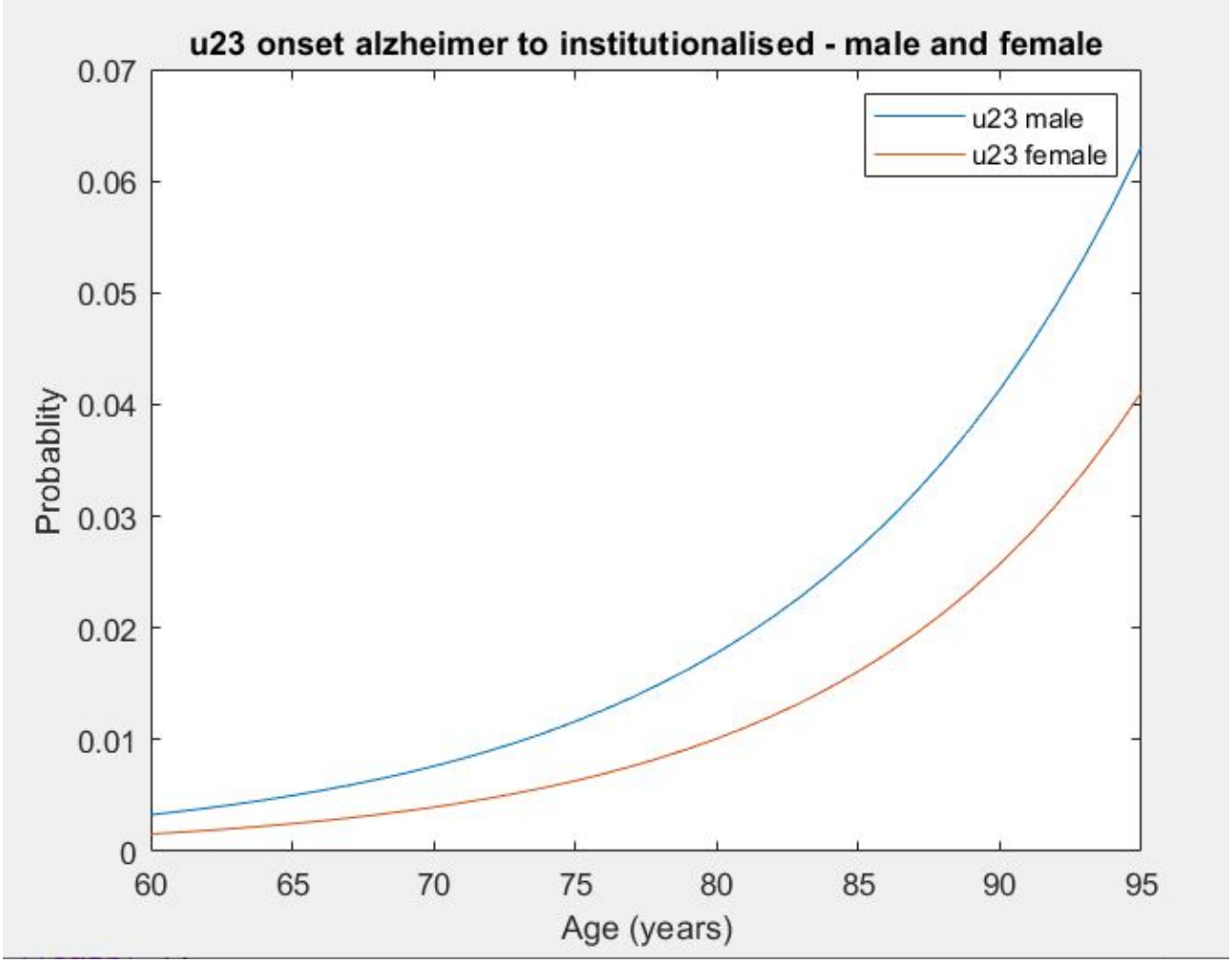

Figure 12: A simple model of Alzheimer's disease,which preditct the probablity of changing the states

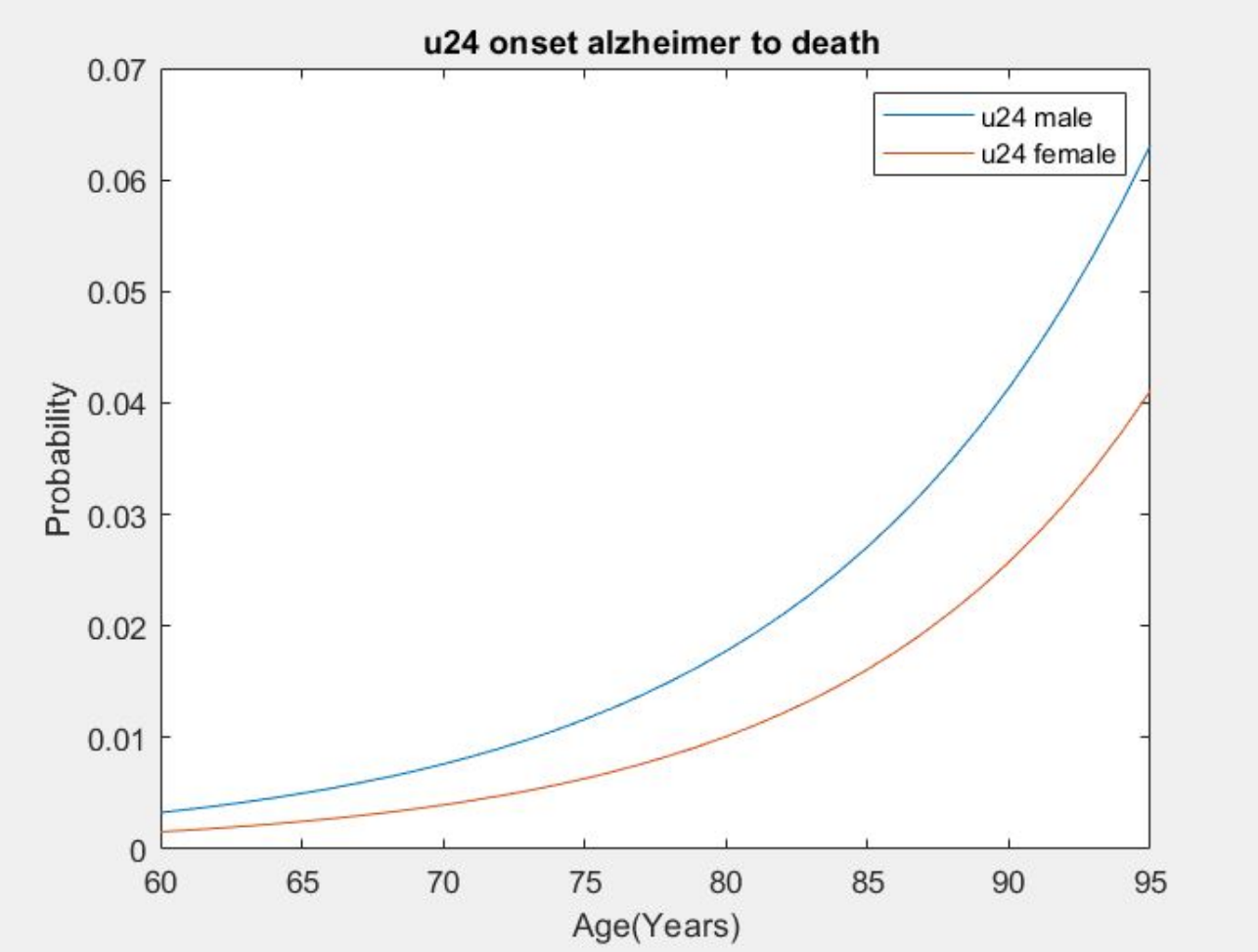

Figure 13: Probability of changing state from Institutionalised Alzheimer to Death for different genotype

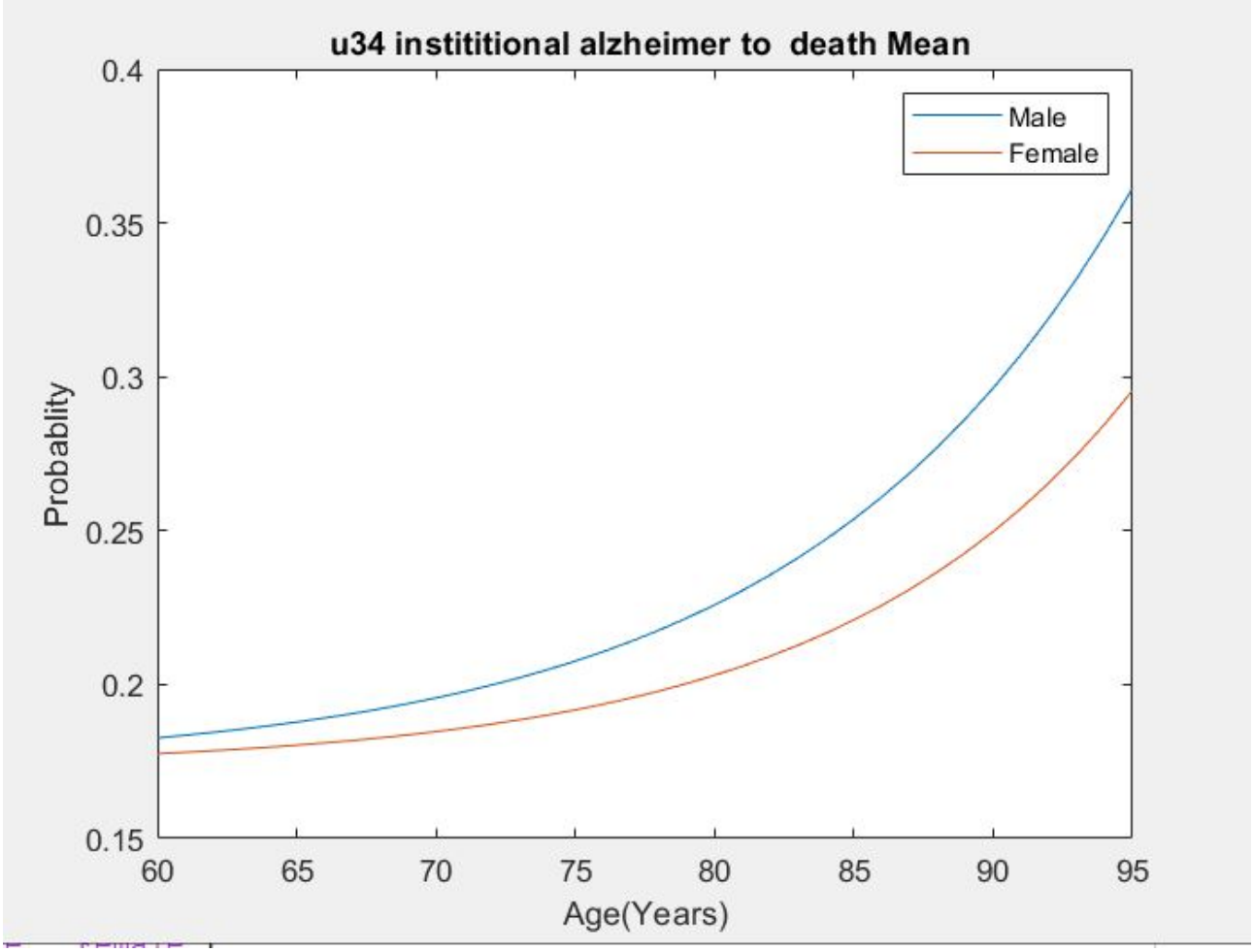

Figure 14: A simple model of Alzheimer's disease,which preditct the probablity of changing the states

5.3.3 New Inferences

- A simple model of Alzheimer's disease,which preditct the probablity of changing the states
  By the Fig.10, We conclude that probability of any person who has Alzheimer.He/She is in onset Alzheimer state then their chances of moving from onset Alzheimer to to die is higher than on set Alzheimer to institutionalised Alzheimer for every age.That shows that more peoples die after on set Alzheimer than going to institutionalised Alzheimer state.
- We can see that transition from onset Alzheimer to institutionalised is constant throughout all ages.
- Transition from no Alzheimer to Institutionalised Alzheimer is low through initial age(till age of 80) and after that it will increases exponentially.So the chances for getting Alzheimer is high after the age of 80 years.

- Probability of changing state from No Alzheimer to on set Alzheimer for different genotype
- If we further describe No Alzheimer to On set Alzheimer in terms of genes,Because The peoples who has any combination with E4 genes has higher chances of getting affected by Alzheimer.
- So,We analysed that peoples who have combination of e2 and e4 genes(e2/e4) has lowest chances of getting affected by Alzheimer. Which shows that e2 allele gene work against e4 allele gene.And which shows that e2 allele genes helps peoples for recovering the Alzheimer.
- After the initial age (Age of 85 years) e3/e4 combination has high probability to getting affected with Alzheimer then e4/e4 genes.Which shows us that e3 allele gene with combination of e4 allele gene is more influenced factor then e4/e4 gene.

- Probability of changing state to Institutionalised from onset Alzheimer for different genotype
- If we analyse the fig.12 we clearly see the transition from on set Alzheimer to die(μ^{i24}_{x+t}), in which male are affected more than female. After initial age it will increases exponential.

- Probability of changing state from Institutionalized Alzheimer to Death for different genotype.
- If we analyse the fig.13 we clearly see the transition from institutionalised Alzheimer to die(μ^{i34}_{x+t}), in which male are affected more than female. After initial age it will increases exponential.

6 Contribution of team members

6.1 Technical contribution of all team members.

Tasks	Parth Patel	Shivam Lakhtariya
Modeling	Done	Done
Coding	Done	Done
Inference	Done	Done

6.2 Non-Technical contribution of all team members

Enlist the non-technical contribution of members in the table. Redefine the tasks (e.g Task-1 as report writing etc.)

Tasks	Parth Patel	Shivam Lakhtariya
Mathematical Work	Done	Done
Report Writing	Done	Done
C Map	Done	Done
Inferences	Done	Done

References

[1] L. Cupples, C. V. Duijn, A. Kurz, R. Green, H. Chui, R. Duara, S. Auerbach, L. Volicer, J. Wells, C. V. Broeckhoven, and et al., "749 evidence for major gene inheritance of alzheimer disease in families of patients with and without apoe 4. vs rao," *Neurobiology of Aging*, vol. 17, no. 4, 1996.

[2] C. J. Brainerd, V. F. Reyna, R. C. Petersen, G. E. Smith, A. E. Kenney, C. J. Gross, E. S. Taub, B. L. Plassman, and G. G. Fisher, "The apolipoprotein e genotype predicts longitudinal transitions to mild cognitive impairment but not to alzheimers dementia: Findings from a nationally representative study." *Neuropsychology*, vol. 27, no. 1, p. 86–94, 2013.

[3] A. Macdonald and D. Pritchard, "A mathematical model of alzheimers disease and the apoe gene," *ASTIN Bulletin*, vol. 30, no. 1, p. 69–110, 2000.

[4] L. A. Farrer, "Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease," *Jama*, vol. 278, no. 16, p. 1349, 1997.