TYPE-1 DIABETES SPREAD ON POPULATION OVER GENERATIONS

A Simulation Study

Trinity College Dublin
ST2004 – Applied Probability
January, 2014

Eduardo Vasconcelos

Introduction

According to the American Diabetes Association (ADA), diabetes inheritance seems to be more complex than other genetic diseases'[i]. Depending on elements such as which parent has the disease and the parents' ages, the probability that a child will develop diabetes may fluctuate.

The variation of diabetes covered on this paper is type-1 diabetes. Given the odds that a child will develop this type of diabetes, the intent here is to explore how the disease spreads in a simulated population of initially 100,000 individuals given changes on population attributes such as:

1. The initial proportion of diabetics;
2. Women's fertility;
3. The proportion of mothers under the age of 25 (according to the ADA children born from diabetic mothers under the age 25 are more likely to develop the disease);
4. An exchange factor, which represents population dynamics and
5. The average number of partners each woman may have children with over her life.

Data structures and application meant to serve for this purpose where coded in the Java programming language and the results of the simulation exported to .csv files so the data could be analyzed on Microsoft Excel. An algorithm to determine how likely a child born from parents from this population was conceived using the ADA odds for inheritance. The algorithm is explained in the next section.

Inheritance algorithm

Profiles

The ADA states that the odds that a child will develop diabetes are related, primarily, to the father's situation. There are 3 possible profiles for a father. He may:

A. Have developed the disease under the age 11;
B. Have developed the disease after the age 11 or
C. Not be diabetic.

For means of organizing the simulation, profile trees were drawn. A fathers' profile tree is shown below.

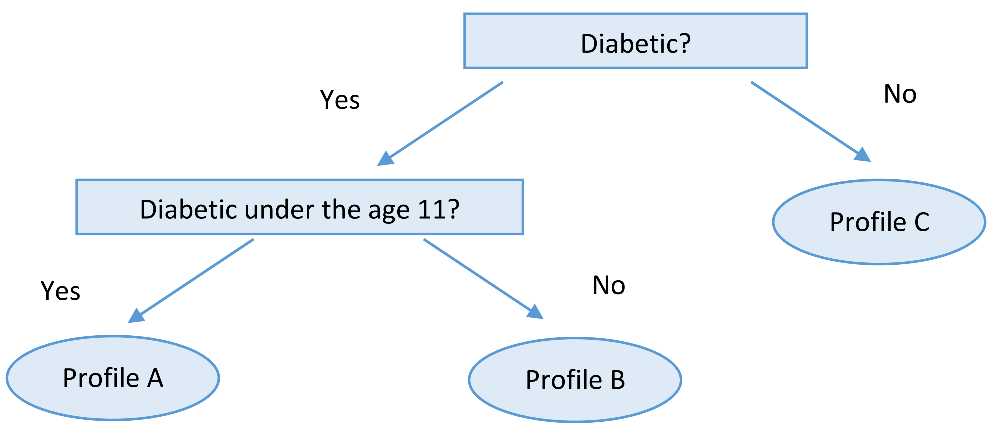

Figure 1 Profile tree for fathers.

Similarly, a profile tree for mothers was drawn, which has 5 profiles. A mother may either:

D. Have developed the disease under the age 11 and be under 25;
E. Have developed the disease under the age 11 and be over 25;
F. Have developed the disease after the age 11 and be under 25;
G. Have developed the disease after the age 11 and be over 25 or
H. Not be diabetic.

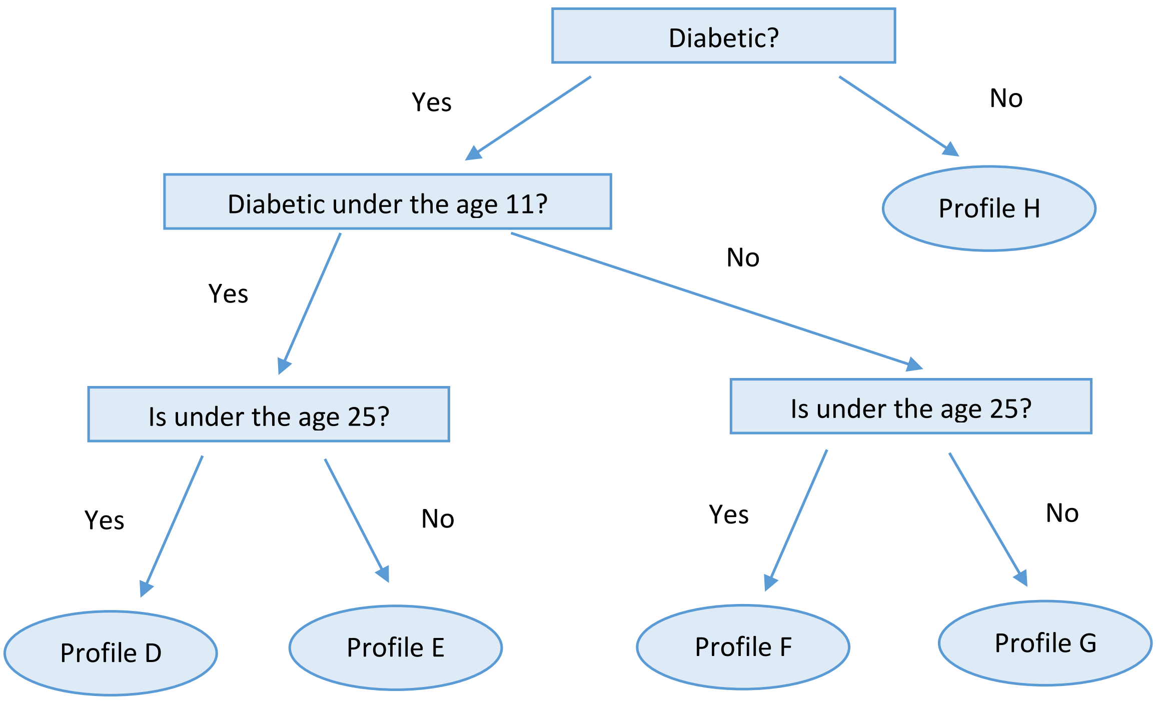

Figure 2 Profile tree for mothers.

Odds

According to the ADA, a child whose father is of Profile A has a chance of $\frac{1}{17}$ of developing diabetes. This chance is doubled if the father had developed the disease before the age 11. Hence, a child whose father is of Profile B has a chance of $\frac{2}{17}$ of developing type-1 diabetes. A child whose father is of Profile C may not inherit diabetes from this parent.

When it comes to the mother, ADA states that a child born from a mother who is of Profile F has a chance of $\frac{1}{25}$ of developing diabetes. A child whose mother is of Profile G has a chance of $\frac{1}{100}$. These odds are doubled if the mother has developed the disease under the age 11. Hence, for a child born from a mother of Profile D the odds are of $\frac{2}{25}$ and for a child whose mother is of Profile E odds are of $\frac{2}{100}$. A child whose mother is of Profile H may not inherit diabetes from this parent.

Furthermore, chances of developing diabetes for a child whose parents both have the disease stand somewhere between $\frac{1}{4} = 0.25$ and $\frac{1}{10} = 0.1$. For this case, a random number in such interval was picked on simulation. It is important to stress that the robustness of the random number generator used for this purpose was verified and it was found to be proven robust. The verification proceeded as follows.

The random number generator in question uses the formula:

$$R = 0.1 + r \times 0.15$$

In order for this to return a number between 0.1 and 0.25, r must be such that $r \in [0, 1]$. The Java programming language offers a method called Math.random() that returns a random real number between 0 and 1 inclusively. Math.random() was then used to generate r. Hence, the robustness verification would simply consist of analyzing the probability distribution of R. The problem is that R is not a discrete random variable. Instead, R is a continuous random variable, which implies that we may not calculate the probability that R will assume any certain predetermined real value: such probability would always be equal to 0[ii].

However, we may speak of probabilities of a certain random variable like R assuming values in predetermined *intervals*. The interval [0.1, 0.25] was then divided in 14 subintervals (below), a set of 1,000,000 R's was generated in Java using the formula in question, each of these R's was classified as belonging to a certain subinterval and the relative frequency of occurrence of each subinterval was calculated. These frequencies, by the Law of Large Numbers, are a good approximation of the odds of occurrence of each subinterval. The results are shown below. These were extracted directly from Excel.

Subinterval	Elements	Relative frequency
[0.1, 0.11)	66473	0.066473
[0.11, 0.12)	67159	0.067159
[0.12, 0.13)	67049	0.067049
[0.13, 0.14)	67092	0.067092
[0.14, 0.15)	66789	0.066789
[0.15, 0.16)	65800	0.0658
[0.16, 0.17)	66710	0.06671
[0.17, 0.18)	66818	0.066818
[0.18, 0.19)	66735	0.066735
[0.19, 0.20)	66498	0.066498
[0.20, 0.21)	66508	0.066508
[0.21, 0.22)	66511	0.066511
[0.22, 0.23)	66694	0.066694
[0.23, 0.24)	66795	0.066795
[0.24, 0.25]	66369	0.066369
Sum	**1000000**	**1**

Table 1 Random number generator relative frequencies.

Since the relative frequencies are so close to each other (they differ only from the third decimal), we may infer that the occurrence of each of these subintervals is equally likely. Hence, the random number generator may be said to be robust.

Pseudocode

Knowing the possible profiles and their respective odds of perpetuating the disease, the inheritance algorithm was written. The method in question receives two objects, one of which represents the father and the other, the mother – it must be clear that both classes representing father and mother have the necessary attributes to perform the classification task – and returns the odds that a child of these two would develop diabetes given the couple. After an individual is assigned a probability of developing diabetes, he is "looped through the years", as will be exposed in detail in the next section, in order to determine if he will, indeed, develop diabetes. The selection of odds of inheritance for a child given the parents happens as follows:

- If the father has diabetes and mother does not then:
    - If the father was diabetic before age 11 then:
        - The odds for this child are of $\frac{2}{17}$.
    - Otherwise, if the father was diabetic after the age 11 then:
        - The odds for this child are of $\frac{1}{17}$.
- Otherwise, if the father does not have diabetes and the mother does then:
    - If the mother was diabetic before age 11:
        - If the mother is under than 25:
            - The odds for this child are of $\frac{2}{25}$.
        - Otherwise, if the mother is over 25:
            - The odds for this child are of $\frac{2}{100}$.
    - Otherwise, if the mother was diabetic after the age 11:
        - If the mother is under than 25:
            - The odds for this child are of $\frac{1}{25}$.
        - Otherwise, if the mother is over 25:
            - The odds for this child are of $\frac{1}{25}$.
- Otherwise, if both the father and the mother have diabetes then:
    - The odds for this child are a random number between $\frac{1}{4}$ and $\frac{1}{10}$ inclusively.
- Otherwise, if neither the father nor the mother have diabetes then:
    - This child may not inherit diabetes.

Table 2 Selection of odds of inheritance for a child given the parents.

Creation of new generations

To create an entire new generation based on the particular odds each individual has to inherit diabetes from their parents; the new generation goes through a process of exchange, maturing and then procreation. The exchange stage is the one in which population dynamics is simulated and is explained in the next section.

The maturing part is a phase in which the individuals are brought to a relevant procreation age, that is, the age in which no more implications upon the chances they have to transmit diabetes to their children exist. This is also the phase in which they may develop diabetes themselves. In maturation, each individual is “looped through the years”, that is, a repetition structure is used to simulate his growth. At each year of age until the age 11, a random number is picked using the method Math.random() from the Java platform. If that number turns out to be lower than the chance that the individual will develop diabetes, that individual is then flagged as being diabetic. After the age 11, a new number is picked once more. Again, if it turns out to be lower than the chance that the individual will develop diabetes, that individual is flagged as being diabetic. This procedure was adopted because it was made the assumption that growing-up individuals are more susceptible to illness. Furthermore, after the “loop through the years”, several women, based on the percentage of mothers under the age 25, are flagged as being under 25. The others are assumed to be over 25.

After maturation, the individuals may then procreate. In this stage, their generation will conceive a new one. The process then repeats until the wanted number of generations is reached. The very first generation is created by instantiating individuals with random attributes but guaranteeing that more or less a certain percentage of them has diabetes (real world percentage is exposed in the next section) and that approximately 50% of them are men and 50% are women, which is a reasonable approach to real world gender ratio[iii].

The very first generation and the one immediately after that were not considered on the results. The reason for that is that the first one is generated with random attributes, hence, it works simply as a seed. The one after that, in practice, showed a huge decrease in the initial percentage of diabetics, which was interpreted as a period of instability in the system. Hence, the first generation is then taken as the second generation after the seed generation. This practice is known as burn-in and consists of discarding outputs of a recently started system until it is noticed that the output is stationary[iv].

Simulation

Real world data

After the Java data structures and application were ready, the simulations could be started right away. However, before running our Control simulation, real world data was gathered so that this first run could be more realistic to be compared with the subsequent ones, in which the population attributes were to be changed.

Some research upon relevant both diabetic and non-diabetic American population attributes was made and these are displayed below.

Estimate proportion of diabetics in the US[v]	8.3%
Average fertility in the US[vi]	2.03
Proportion of mothers under 25 in the US (2009)[vii]	34.47%

Table 3 Real world data upon American population.

As previously stated, the Exchange factor represents population dynamics. In our simulation, this value will be used to calculate the fraction of the population that will be replaced between two generations. Let us say that we have a population of 9,000 individuals and the Exchange factor was set to 3. Then, for the next generation, $\frac{1}{3}$ of the population (3,000 individuals randomly chosen) will be replaced by 3,000 new individuals. These new individuals are instantiated in such a way that they have random attributes. Still, it is guaranteed that:

1. Approximately 8.3% of them are diabetic to match the Estimate proportion of diabetics in the US (this is what would be expected from an immigrant group as well given that it is a mixture of people from all over the country);
2. Approximately 50% of them are men and
3. Approximately 50% of them are women.

With these attributes set, the Control run was finally made and the results are shown in the next section.

Control run

A Control run was made with the following attributes:

Estimate proportion of diabetics	8.3%
Fertility	2
Proportion of mothers under 25	34.47%
Initial population	100,000
Number of generations	1,000
Exchange factor	3
Highest possible number of partners with which a woman may have children with	1

Table 4 Control run attributes.

The graph below displays the proportion of diabetics over 1,000 generations that resulted from this simulation.

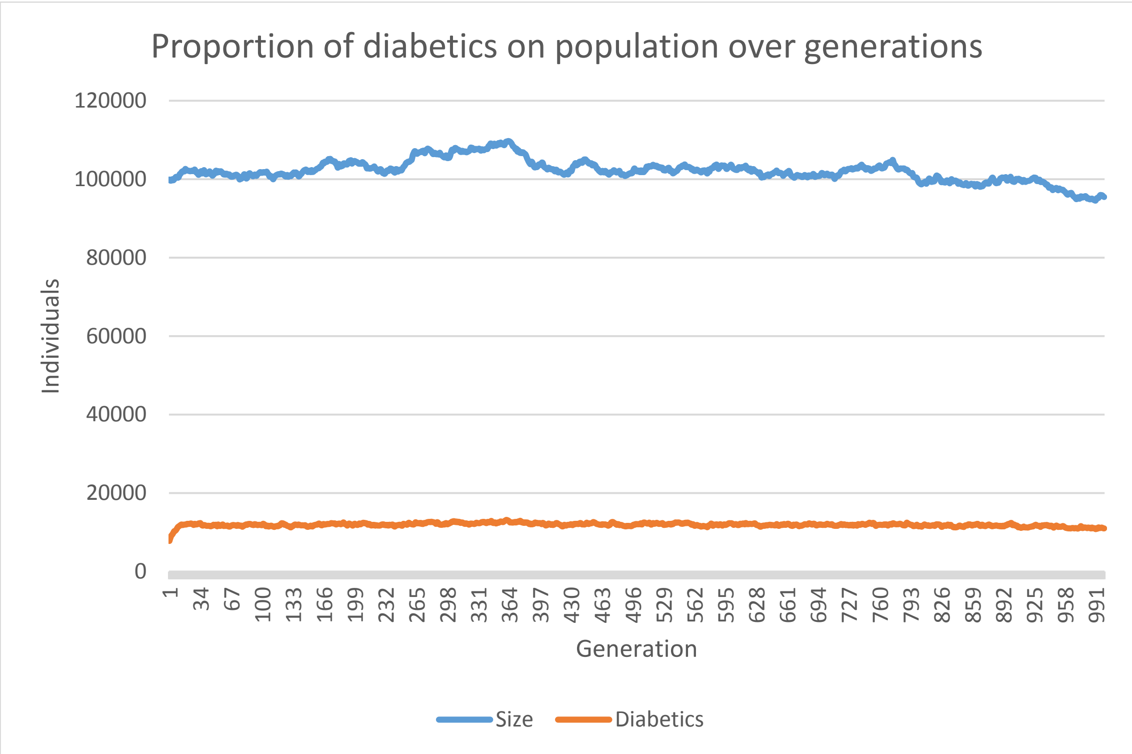

Figure 3 Proportion of diabetics on population over generations.

It may be observed that after approximately 20 generations the number of diabetics seems to stabilize. The stabilization is shown in detail below (from generations 1 to 33).

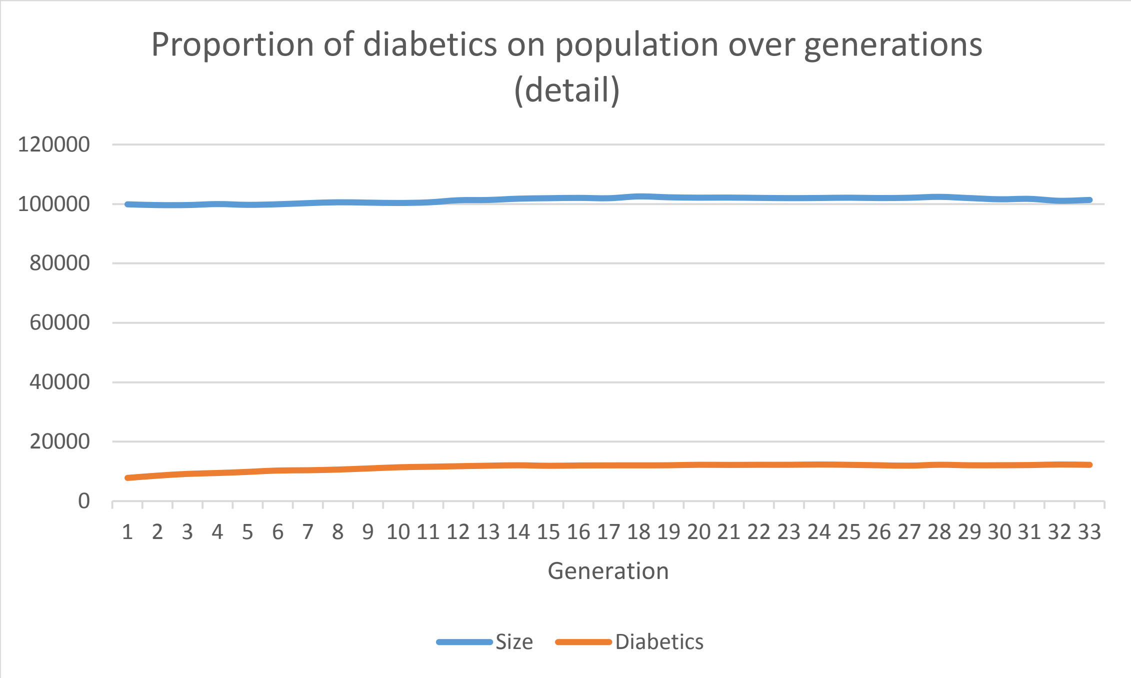

Figure 4 Proportion of diabetics on population over generations (detail).

Furthermore, the graph below displays the percentage of diabetics on population over all 1,000 generations. It may be clearly seen the transition to the steady state of the system right after the burn-in period.

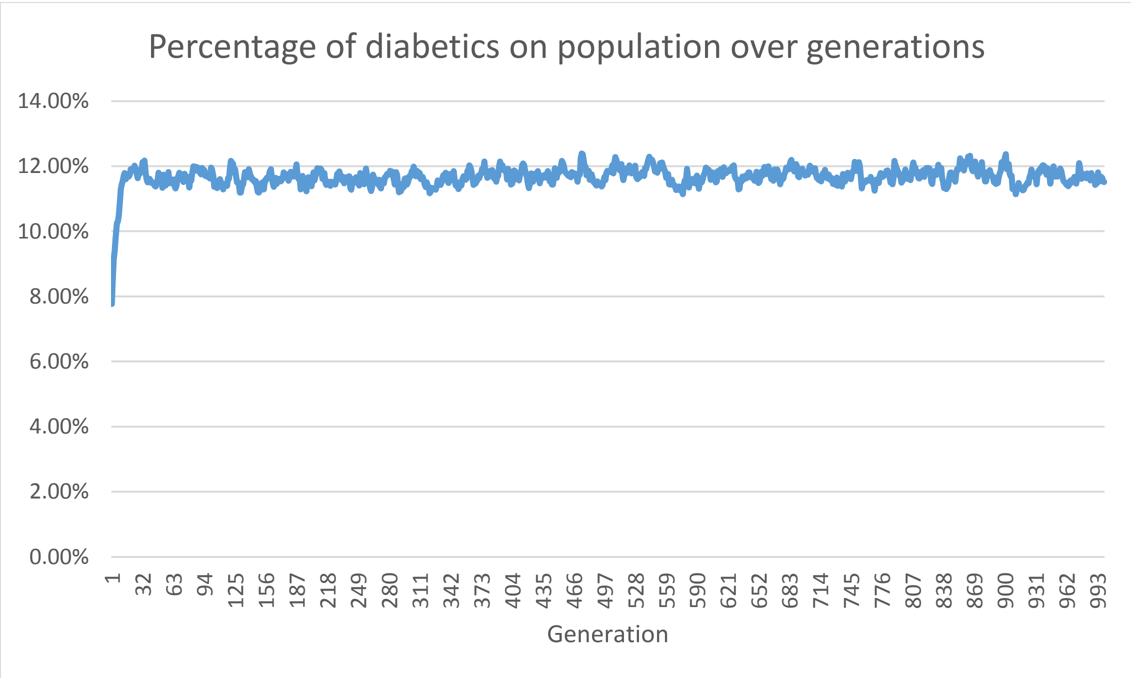

Figure 5 Percentage of diabetics on population over generations.

Thus, the number of diabetics in such a population stabilizes, roughly, around 12%.

Subsequent runs

Variation on initial Estimate proportion of diabetics

The first change in attributes to be analyzed was the variation on the initial proportion of diabetics. Three cases were simulated: changes to 1%, 50% and 90% of the population. All the other attributes remained the same. In all cases, stabilization occurred and the results are shown below.

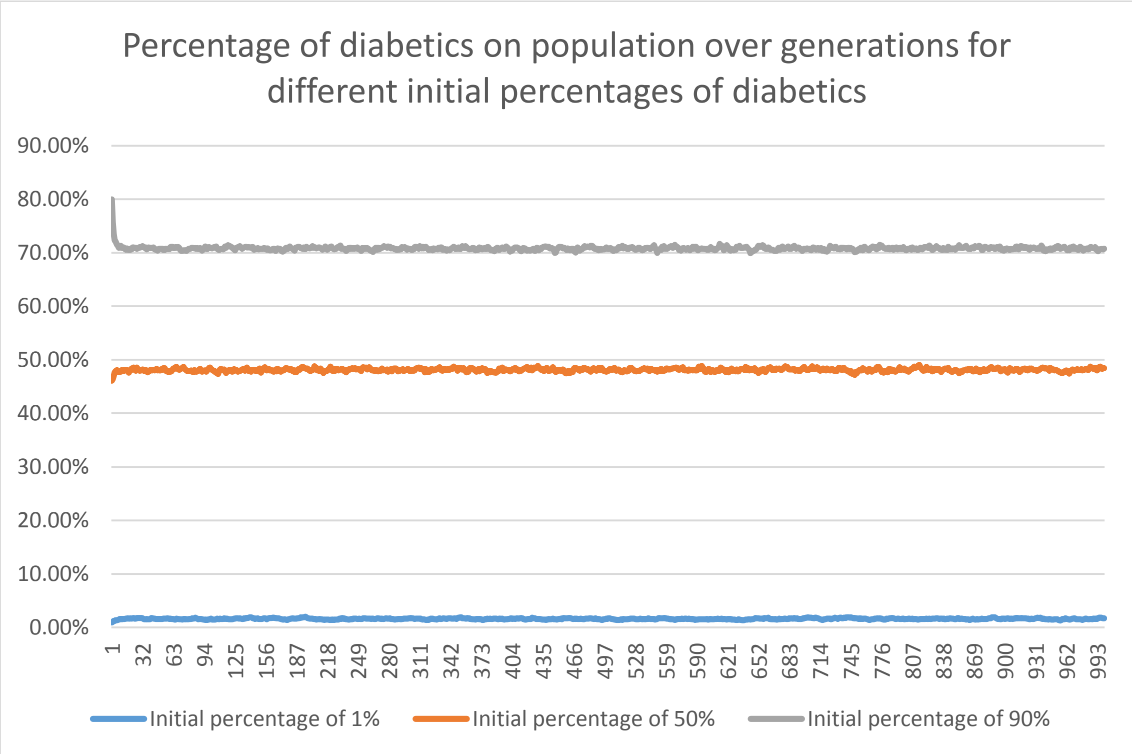

Figure 6 Percentage of diabetics on population over generations with different initial percentages of diabetics.

With an initial percentage of 1% of diabetics in the seed generation, the percentage of diabetics did not suffer considerable changes. The percentage stabilized around 1.6%, more or less.

Similarly, with an initial percentage of 50% of diabetics in the seed generation, the percentage of diabetics stabilized around 48%. Thus, with such initial percentage, no big changes are to be expected in a long run.

On the other hand, with an initial percentage of 90% of diabetics in the seed generation, the percentage of diabetics stabilized around 71%. Thus, with such initial percentage, a big decrease in the percentage of diabetics is to be expected in a long run. Such a decrease may be explained by being due to population dynamics.

Variation on Fertility

Variation on fertility could not be simulated. The change to a fertility of 3 resulted in a very slow simulation, which would not come to an end in feasible time. The reason for that is the enormous growth of population size resulting from such fertility associated with the lack of computational tools available for such a heavy data processing. Variation to a fertility of 1 could not be simulated as well since this resulted in an extinguishment of the population, which was evidenced in practice.

Variation on Proportion of mothers under 25

In order to evaluate the impact of changes in the percentage of first time mothers under the age 25, the real world percentage of 34.47% was substituted by 17% and 68%, respectively, in two distinct simulations. The results are shown and discussed below.

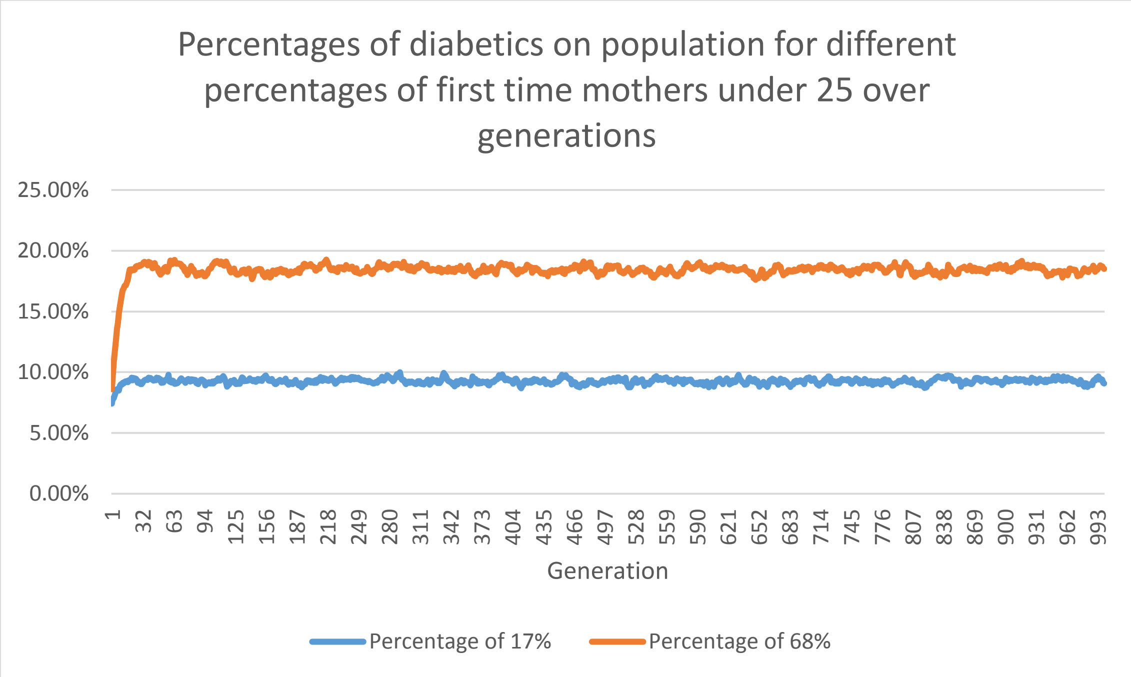

Figure 7 Percentages of diabetics on population for differnet percentages of first time mothers under 25.

From this graph, we conclude that the proportion of mothers under the age 25 plays a significant role on the spread of type-1 diabetes. Confronting the simulations for older and younger mothers, one may observe a difference of roughly 10% between the two scenarios: for a percentage of mothers under age 25 of 68%, the stabilization on the percentage of people with diabetes stabilized around 20% whereas for a percentage of mothers under age 25 of 17% this number stabilized around 10%.

For a percentage of mothers under the age 25, the expected proportion of diabetics was calculated. This was achieved by computing the long-run average of diabetics on population. The respective long-run average graph and final value follow.

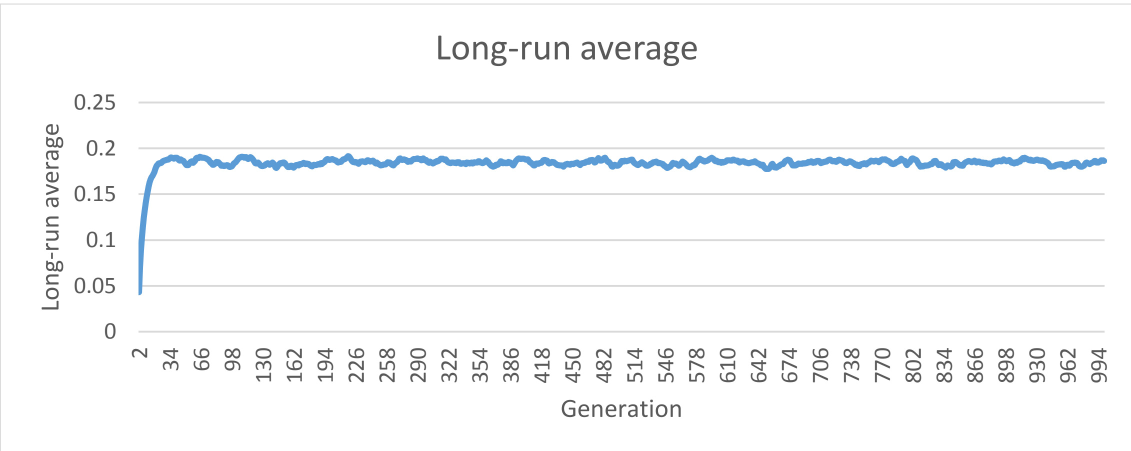

Figure 8 Long-run average of diabetics on population for initial percentage of mothers under the age 25 of 68%.

The final expectation is:

$$E[D] = 0.186331349 \approx 18.6\%$$

Where D is the proportion of diabetics in question.

The variance found for this measure was:

$$Var[D] = 7.29541 \times 10^{-5}$$

Such a small variance indicates that the result found is indeed a good approximation of the expected value.

Variation on Exchange factor

A run was made with the exact same attributes as the Control run except for the Exchange factor. It was taken an infinite Exchange factor, what would correspond to the absence of population dynamics, since:

$$\lim_{f \to \infty} \frac{1}{f} = 0$$

It was observed that the absence of population dynamics increases considerably the percentage of diabetics on population after the number has stabilized. The graphs that follow show the proportion of diabetics on population and the corresponding percentage, respectively, for an infinite Exchange factor.

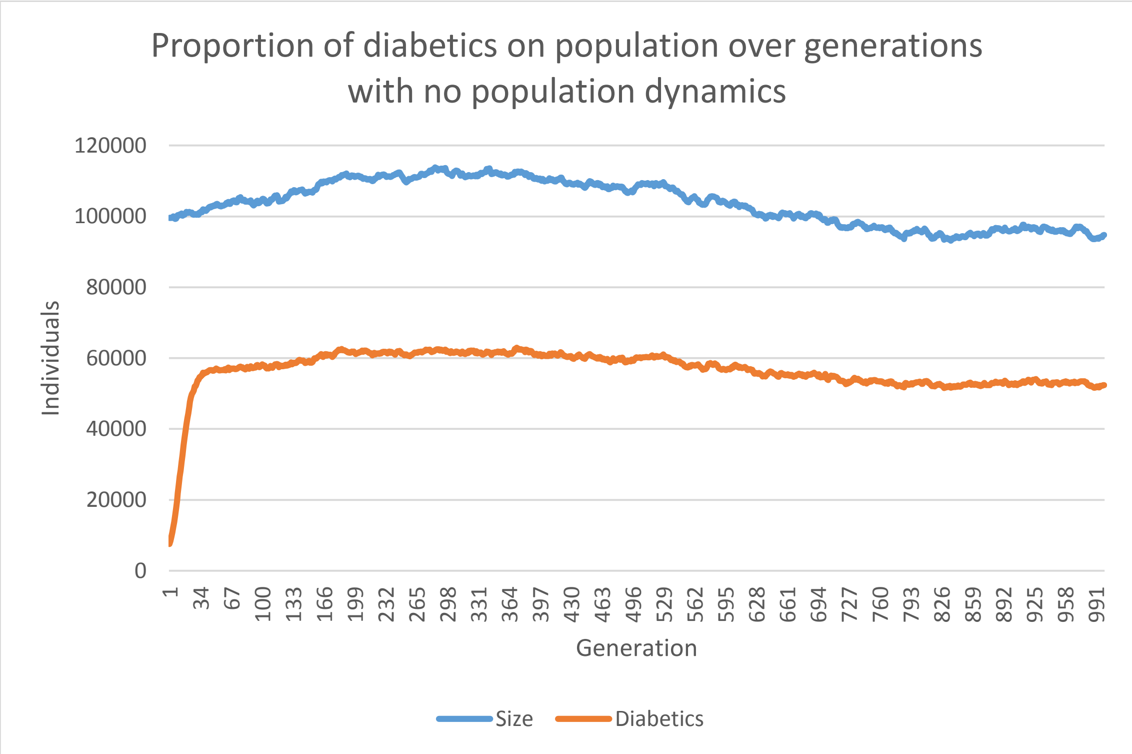

Figure 9 Proportion of diabetics on population over generations without population dynamics.

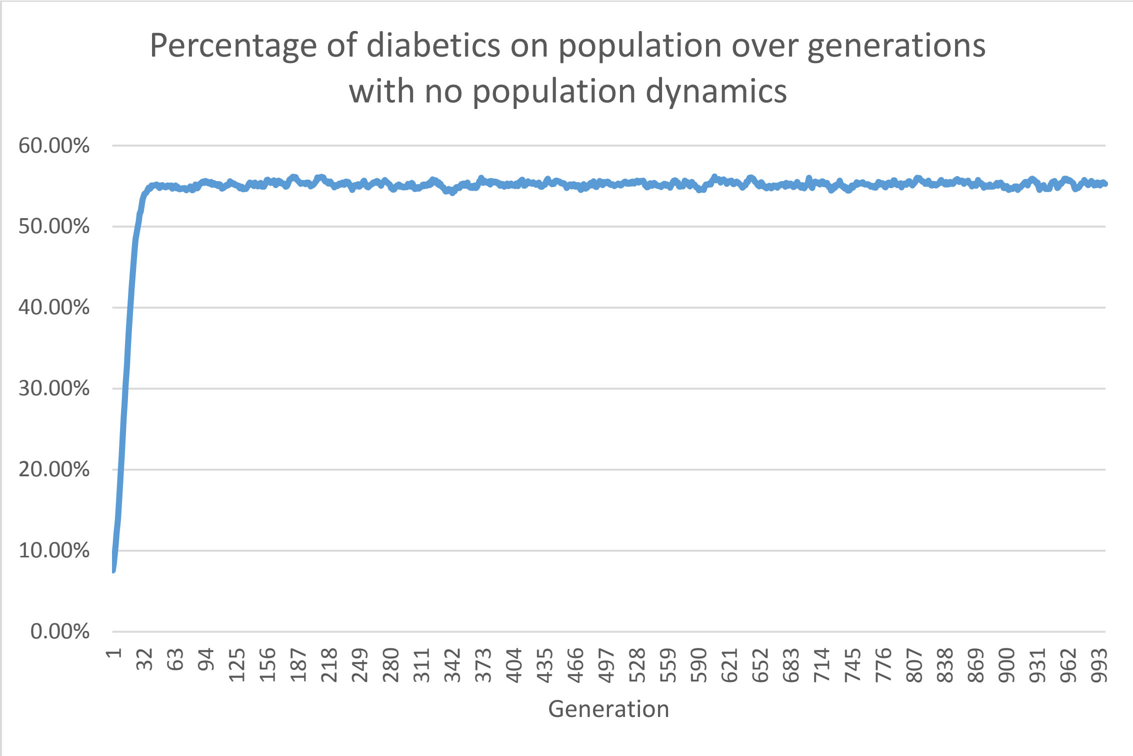

Figure 10 Percentage of diabetics on population over generations without population dynamics.

Historic evidences may reinforce this result. The isolation of a population may result in the increase on the relative amount of diabetic individuals in that population. The Journal of Clinical and Applied Research and Education of the ADA, Diabetes Care, issued an article that discusses upon this[viii].

Variation on Highest possible number of partners with which a woman may have children with

Real world individuals may have children with more than one partner along their lives. The impact of such a factor on the spread of the disease was measured through the designation of randomly chosen fathers for each different child a mother gives birth to, instead of keeping the same partner for all of her children. Results and discussion upon this follow.

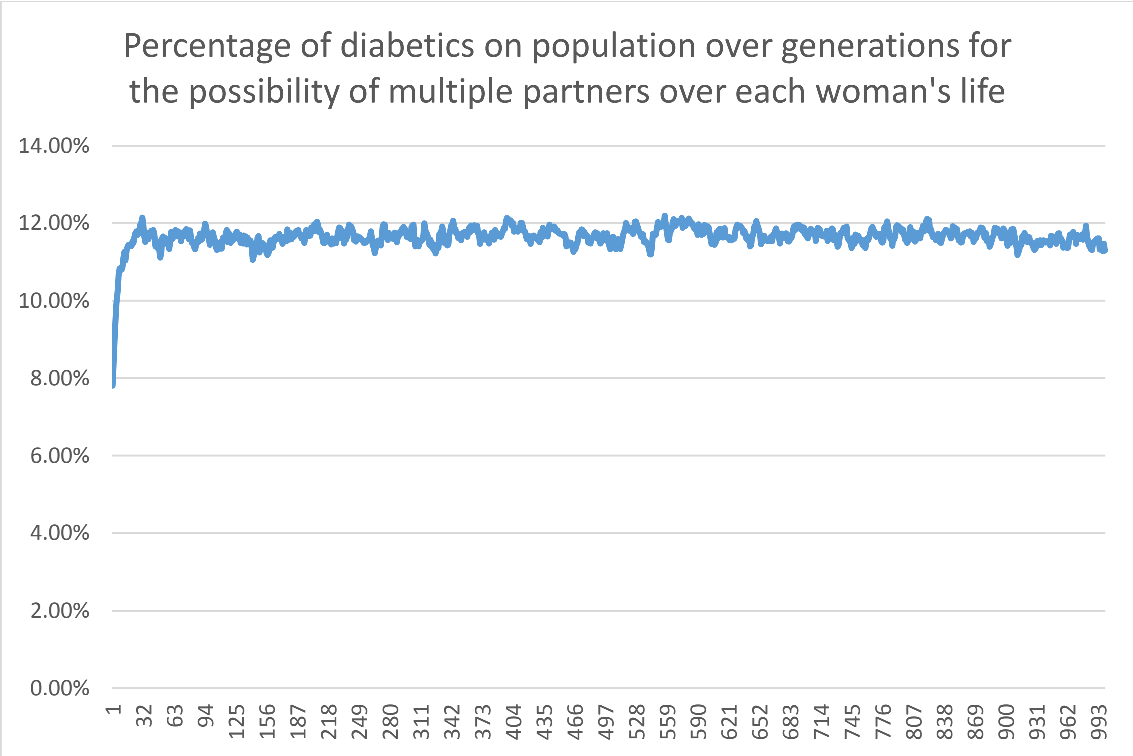

Figure 11 Percentage of diabetics on population over generations for possibility of multiple partners over generations over each woman's life.

The insertion of the possibility of multiple partners over a woman's life on the system did not result in expressive changes on the percentage of diabetics on population over generations. Stabilization on percentage occurred more or less under the same conditions as in the Control run, that is, around the 20th generation and at about 12% of the population.

Hence, this attribute, by itself, may not be considered as determinative for the spread of type-1 diabetes.

Conclusion

Even with a relatively complex set of individual probabilities of inheritance of type-1 diabetes, the behavior of the population in a long run may be taken as being rather predictable.

All feasible changes in population attributes resulted in stabilization on the percentage of diabetics over generations. It may be then assumed that the number of diabetics on a population tends to stabilize over generations, it does not matter how unrealistic the initial population attributes might be, such as an initial percentage of 90% of diabetics. The other attributes will always compensate the changes and conduce the system to a steady state.

Perhaps the most interesting result found on this simulation is the (predictable) fact that population dynamics plays such an important role on the control of genetic diseases' spread.

---

[i] American Diabetes Association Website: http://www.diabetes.org/diabetes-basics/genetics-of-diabetes.html. Accessed: November 21, 2013 at 09:40.
[ii] TIJMS, Henk in Understanding Probability, Cambridge University Press, p. 323.
[iii] GeoHive – Population Statistics Website: http://www.geohive.com/earth/pop_gender.aspx. Accessed: November 24, 2013 at 12:50.
[iv] Wikipedia: http://en.wikipedia.org/wiki/Burn-in. Accessed: January 16, 2014 at 01:00.

---

[v] American Diabetes Association Website – Diabetes Statistics: http://www.diabetes.org/diabetes-basics/diabetes-statistics/. Accessed: November 21, 2013 at 09:38.
[vi] Support US Population Stabilization Website: http://www.susps.org/overview/birthrates.html. Accessed: November 22, 2013 at 10:44.
[vii] InfoPlease Website: http://www.infoplease.com/ipa/A0005074.html. Accessed: Nov 22, 2013 at 10:46.
[viii] The Journal of Clinical and Applied Research and Education of the ADA, Diabetes Care Website: http://care.diabetesjournals.org/content/24/4/650.full. Accessed: November 24, 2013 at 12:10.