A Study of Malaria and Sickle Cell Anemia:
A Hands-on Mathematical Investigation
Student Materials: Reading Assignment

Malaria is a parasitic disease which is spread by the female Anopheles mosquitoes. There are about 2 million deaths from malaria each year, making it one of the world's deadliest diseases. Forty percent of the world's population is at risk of contracting malaria. Most of the fatal cases of malaria are in Sub-Saharan Africa, and most are children under the age of five or pregnant women. There are some areas where up to 40% of the children die of malaria when conditions are at their worst. The most effective prevention of malaria in children, as shown in a 1996 World Health Organization study, is protecting them from mosquito bites by having them sleep under bednets dipped in permethrin. In the WHO's pilot study in The Gambia, the death rate among children between birth and 5 years was reduced by 63% by this method.

Inside the human host, the malaria parasite first invades the liver cells and then the red blood cells. Disease is produced when the parasite is inside the red blood cells. When the parasite has matured inside the red blood cells, the cells burst, producing chills and a very high fever. The infected red blood cells and the burst blood cells can cause failure of the liver or the kidneys, hypoglycemia, or cerebral malaria which can include blocking the blood vessels carrying blood to the brain; these events may lead to death.

Malaria is an ancient disease; descriptions of its pathology are found in Hippocrates' writings. Through the millennium, the distribution of malaria in the world has changed. It is still a significant health problem in South and Central America as well as in Asia and Africa south of the Sahara. In earlier times, it was present in the Mediterranean countries, the Arabian Peninsula, and in India. It was present in the United States as far north as Baltimore at the time of the Civil War. Many soldiers from both the Union and Confederate armies died of malaria.

Sickle cell anemia is caused by a "defective" allele (mutant form) of the gene coding for a subunit of the hemoglobin protein. Hemoglobin binds oxygen within red blood cells, which then transport the oxygen to body tissues where it is released from the hemoglobin molecule. The sickle hemoglobin (in a person with a mutant allele) tends to precipitate, or "clump together", within the red blood cell after releasing its oxygen. If the clumping is extensive, the red blood cell assumes an abnormal "sickle" shape. These sickled red blood cells plug the blood vessels, thus preventing normal red blood cell passage and, consequently, depriving the tissue of needed oxygen.

Each person has two copies of the gene that determines whether that person has sickle cell disease. If both copies are "normal" alleles, then only normal hemoglobin is produced. If one of the two alleles is "defective", then that person has a mixture of normal and sickle hemoglobin--a condition known as "sickle cell trait." Sickle cell trait usually results in no ill health effects. If both copies are the "defective" alleles, essentially only sickle hemoglobin is made and the person has sickle cell anemia.

This project was supported, in part, by the National Science Foundation. Opinions expressed are those of the authors and not necessarily those of the Foundation

Sickle cell anemia is associated with a multitude of medical complications ranging from acute painful crises caused by the plugging of blood vessels to chronic damage to the spleen, kidneys, lungs, heart, muscles and brain. Repeated hospitalization for intravenous pain medication, antibiotic therapy and blood transfusions is undertaken to treat medical problems as they arise. These patients often die early of overwhelming infection or as a consequence of acute or chronic damage to the body organs. Some progress is being made toward the use of drugs that induce the production of "normal" hemoglobin in sickle cell patients in an effort to decrease the frequency of sickle cell crises. However, bone marrow transplantation, an expensive, high-risk medical procedure, remains the only known cure for this disease.

While the allele causing sickle cell anemia is found most often in people of African ancestry, it also occurs in persons of Mediterranean, Arab, east Indian, and South and Central American ancestry, areas where malaria was once prevalent; this is more than coincidence, as we will see. There is actually a group of sickle cell disease variants caused by a number of genetic mutations (different alleles) affecting the hemoglobin protein. To simplify matters, in this article we are going to assume there is only one such allele.

The Sickle Cell - Malaria Relationship.

The allele that causes sickle cell anemia also imparts partial resistance to malaria. In individuals with two "normal" alleles, the malaria parasite can infect the red blood cells. The bursting of these infected cells can cause kidney and liver failure, anemia, hypoglycemia, or block blood vessels to vital organs, such as the brain (causing cerebral malaria); children under the age of 5 have a high risk of death if this occurs. But the red blood cells of individuals with one sickle cell allele are relatively resistant to malaria; furthermore, these individuals do not get sickle cell anemia. In the United States, for example, the sickle cell allele offers no known health advantage, and healthy parents with one sickle cell allele each have the potential to transmit this defective allele to their offspring, possibly resulting in sickle cell anemia. But if a person lives in an area inhabited by mosquitoes that carry the malaria parasite, then the sickle cell allele can be considered positive in the following sense. <u>One</u> sickle cell allele creates a condition in the blood cells that gives some protection from the malaria parasite, a leading cause of early death in those areas.

In this project, we will model both physically and mathematically the effect of the allele that causes sickle cell anemia on the survival of a population. Our goal will be to understand how the genetic process is making the best of a bad situation. The key mathematical idea in the project is the concept of "optimization"; optimization means getting the best possible outcome from the complex circumstances that make up a situation.

For the physical model, beads will represent alleles. Let us designate the color beads representing the normal alleles as N beads and the color representing the sickle cell alleles as S beads. Each person has two alleles, thus each person has either NN, NS, SN, or SS, where the first letter represents the allele received from the mother and the second letter represents the allele received from the father. NS and SN are indistinguishable as far as the sickle cell trait is concerned, and so they will be treated as the same and will be

referred to as "NS". If a person has SS, we will assume that person will develop a lethal case of sickle cell anemia. If a person has NN, that person is susceptible to malaria. We will simulate birth by drawing beads from a cup. To produce one person, two beads must be drawn to represent the two alleles that person receives from his or her parents.

We will take into account the statistical effects of deaths in the population due to sickle cell anemia and malaria. For reasons related to mathematics learning, we are going to assume for part of this project that only one third of the children in a certain population with two normal alleles will survive malaria; however, that is a lower survival rate than would actually be expected. More probable would be that one third of the children with two normal alleles might die due to malaria; thus, two thirds of the children with normal alleles would survive.

It may bother you that deaths from other causes are not taken into account. However, death from other causes would be expected to be distributed more or less equally among the NN, "NS" and SS members of the population since there is no relation between the sickle cell trait and other fatal diseases or conditions. Thus, death from other causes will not affect our model.

Acknowledgments: We would like to thank Dr. Barry Anderson, MD, Ph.D., Assistant Professor of Hematology/Oncology of the Georgetown University Health Sciences Center and Dr. Marcela Parra, Ph.D., Assistant Research Professor of Biology at Georgetown University, for their help in writing this unit.

Modeling a Population where Malaria is a Risk: A Physical Model
Classroom materials

We are going to simulate the birth of children where there is a risk of malaria and sickle cell anemia. Assume that this population is born into an area in which **one-third of the NN children survive malaria**. Also assume that none of the SS children survive sickle cell anemia. In the simulations, we will experiment with different genetic make-ups in a population (that is, different proportions of N and S alleles among the adults) to see how the genetic makeup affects the number of children that survive both malaria and sickle cell anemia.

Complete the following simulation of the genetic process. This will help prepare you to develop a mathematical model of the population. Before you begin, designate one person to hold the cup of beads, one person to draw from the cup, and one person to record the data.

Simulation 1. Put four S beads and six N beads into a cup. This cup represents the initial genetic makeup of an adult population in which the proportion of normal, N, alleles is $6/10 = 0.6$ and the proportion of sickle cell, S, alleles is 0.4. This models the adults of this population. You can now simulate the birth of the children of this population. The designated drawer in your group should draw one bead at random from the cup, its type should be recorded, and the bead should then be returned to the cup. Draw another bead from the cup, record its type, and return this second bead to the cup. At this point, you have recorded one of NN, "NS" (for either NS or SN), or SS. These are the alleles of the first child. Repeat this process for a total of 30 "births."

When you have completed the 60 draws (30 "births"), you have some number of NN's, some number of "NS"'s and some number of SS's. These represent the 30 children that were born. Now each SS dies of sickle cell anemia. The NN's are in danger of dying of malaria. In this simulation, assume two-thirds of the NN children die from malaria. The number of children that survive to adulthood equals the number of "NS" children plus one-third of the NN children. When we did this simulation, we got a total of $13\frac{1}{3}$ survivors, which we recorded in Table 1. Record your total in Table 1, in the space under the 0.6. (Because your population represents an **average** number of survivors, your population may have a fractional number of "people", as ours did.)

Table 1: Results of simulation with 1/3 of NN's surviving malaria		
fraction of N alleles in adult population	0.6	0.3
total number of 30 births that survive to adulthood in our group		
total number of 30 births that survive to adulthood, class average		

Simulation 2: Repeat the process of Simulation 1, but this time use a cup with 7 S and 3 N beads. This cup represents the genetic makeup of an adult population in which the proportion of normal, N, alleles is 0.3. Again, assume that $\frac{2}{3}$ of the NN children die from malaria and all of the SS children die from sickle cell anemia. Record your result in the space under the 0.3 in Table 1.

1. You can see that the size of the population of surviving children is a function of the fraction of alleles among the adults that are N. Call this function $f(n)$. In your simulation, what was the value of $f(0.6)$? What value did you get for $f(0.3)$? In the context of this model, what is the domain and range of this function?

Making a Mathematical Model of the Population

You have simulated the birth of a population by looking at 30 random "births" and modeling death events with given probabilities. Study your table and those of other groups. Compare each others' results. The purpose of the simulation was to help you understand the genetic process and the way the incidence of the two diseases affects the survival rates of the children.

In the following, you will investigate how the size of the surviving population P depends on the fraction of alleles in the parent population that are N. For example, suppose 60% of the alleles in the parent population are N and 40% are S. (This could be represented by a cup containing 6 N beads and 4 S beads.) Imagine that you are going to draw beads from the cup at random, replacing beads after each draw, to get 30 births, a total of 60 beads, two beads (thus two alleles) for each new birth.

Instead of simulating births and deaths, you will predict them with numerical expressions based on probabilities using the fractions you know describe the situation. You want to know the number of NN individuals that should be expected in a total of 30 births if the fraction of N alleles in the adult population is $n = 0.6$. To compute this number, multiply the probability that one parent will contribute an N allele times the probability that the other parent will contribute an N allele times the number of births: $0.6 \times 0.6 \times 30 = 10.8$ NN children.

2. Use the tree or the area model in Figure 1a) or b) to help compute the expected number of children born with "NS" and the expected number of children born with SS.

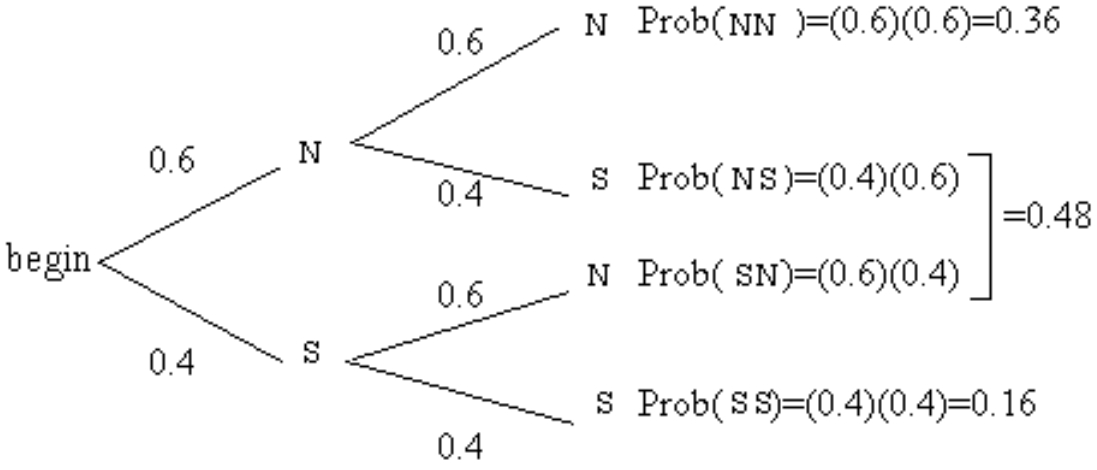


Figure 1a) tree

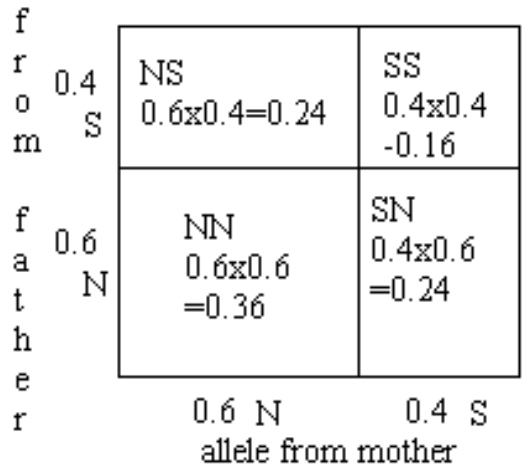


Figure 1b) area

3. Let n represent the fraction of the alleles of the parents that are N. Let $f(n)$ represent the total number of 30 children that are expected to survive both malaria and sickle cell anemia, assuming that two-thirds of the NN children die of malaria and none of the SS children survive sickle cell anemia. Find $f(0.6)$. Completing the column of Table 2 under "0.6" will help you organize your information.

4. Find $f(0.4)$ and $f(0.3)$. (For each, complete a tree diagram similar to Figure 1 and then complete the columns of Table 2 under "0.4" and "0.3".)

Table 2: Results of predictions when 1/3 of NN's survive malaria

fraction of N alleles in adult population	0.6	0.4	0.3	n
fraction of S alleles in adult population				
number of the 30 births that are NN				
number of NN births who survive malaria				
number of the 30 births that are "NS"				
total number of 30 births that survive to adulthood				

Compare the last row of Table 1 to the last row of Table 2. Table 1 is a record of a physical simulation; Table 2 is a mathematical prediction based on probabilities. Assuming you pooled the results of all of the groups in your class on Table 1, you have a reasonable picture of what might happen in some population. Compare that to Table 2. Neither the physical simulation nor the mathematical model should be thought of as a completely accurate picture of what will happen in the real world in a given instance; probabilities tell us what to "expect", but it is unusual for the real world to exactly mirror the expected results.

5. Instead of simulating births and deaths, you will predict them with algebraic expressions. Let n represent the fraction of N beads in the cup and let s represent the fraction of S beads in the cup. Compute **symbolically**, using n and s, the number of the 30 births that will be NN and the number that will be "NS." Use these results to complete the last column of Table 2. Completing the tree diagram in Figure 2 or the area model of Figure 3 may help.

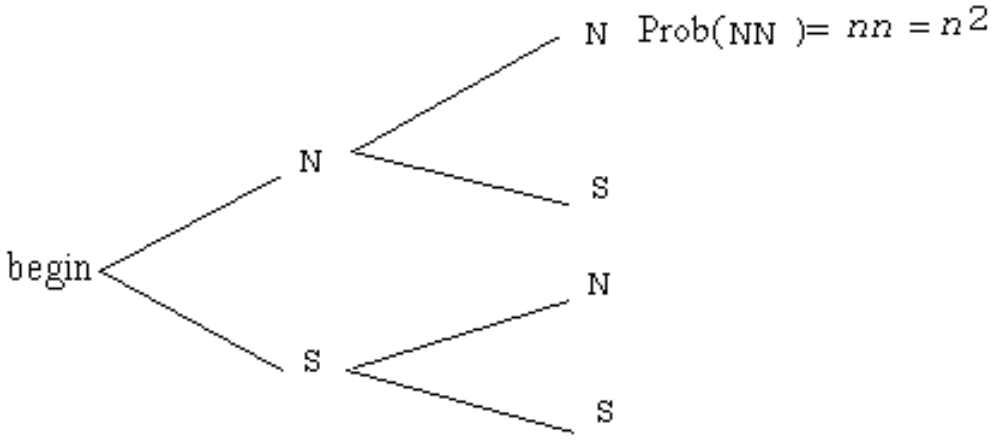


Figure 2

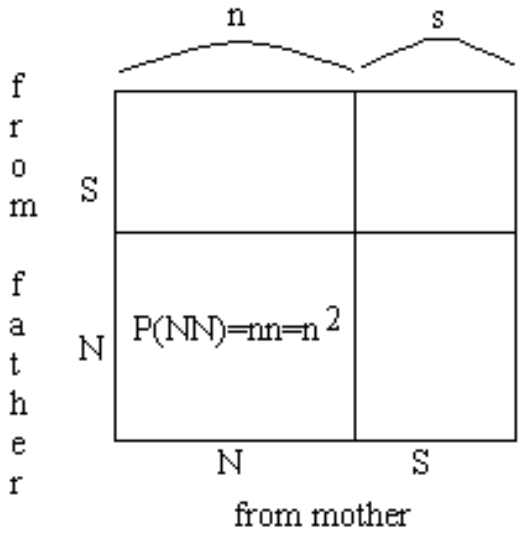


Figure 3

6. Now write the function $f(n)$ for the number of the 30 children who achieve adulthood, where n is the fraction of N alleles in the gene pool. You have probably written your expressions for the last column of Table 2 in terms of both s and n. Rewrite your expressions in terms of n only. This will help you to develop your expression of $f(n)$. Recall that all the alleles are either S or N, so if n is the fraction of one type, you can easily express s in terms of n.

7. Factor the function $f(n)$.

8. Graph the function $f(n)$. Label units on horizontal and vertical axes. What do the numbers on the horizontal axis represent? What do the numbers on the vertical axis represent? How do the n-intercepts relate to the factors?

9. What is the domain of your function $f(n)$, in this context? That is, what values of n are possible in the real world?

10. Notice how the size of the surviving population varies with n. What does this mean? Notice how the slope of the graph of $f(n)$ changes. Explain what it means, in terms of this situation, when $f(n)$ has positive slope. Explain what it means when the slope is negative.

11. What value of n maximizes the number of children who survive to adulthood in our example where two-thirds of the NN's die of malaria? How does this n-value relate to the values where the graph intersects the n-axis? What is the **fraction** of the births that survive to adulthood in this case?

Note that although all SS children are assumed to die before they reach adulthood, having $s = 0$ and $n = 1$ does not maximize the number of children who survive and grow into adults. Malaria is a significant factor. You have shown that the number of children who grow up is greatest when the gene pool has ____% normal alleles and ____% sickle cell trait alleles, assuming two-thirds of the NN's die of malaria.

12. Assume that there are 1000 births.

a. Write a different function, say $g(n)$, for the number of children that survive malaria and sickle cell anemia if the fraction of NN's who die of malaria is $\frac{1}{10}$.

b. What are the n-intercepts for this function?

c. What value of n maximizes the number of children who survive to adulthood where $\frac{1}{10}$ of the NN's die of malaria? What is the fraction of the births that survive to adulthood in this case?

d. Sketch a graph of this function.

13. The two functions you have written, $f(n)$ and $g(n)$, are the same kind of function. (What kind?) In both cases, how does the value of n that maximizes the function relate to the n-intercepts? Why must this be the case in this kind of function?

Summary

One reason that the sickle cell allele occurs with relatively high frequency in some human populations is that, in areas where the malaria parasite thrives, the presence of the sickle cell allele results in the survival of a larger fraction of the population. This is termed a "survival advantage." It is believed that this type of relationship exists for other diseases and genetic traits. For example, there is some evidence that people with just one of the alleles that causes cystic fibrosis have an increased chance of surviving cholera.

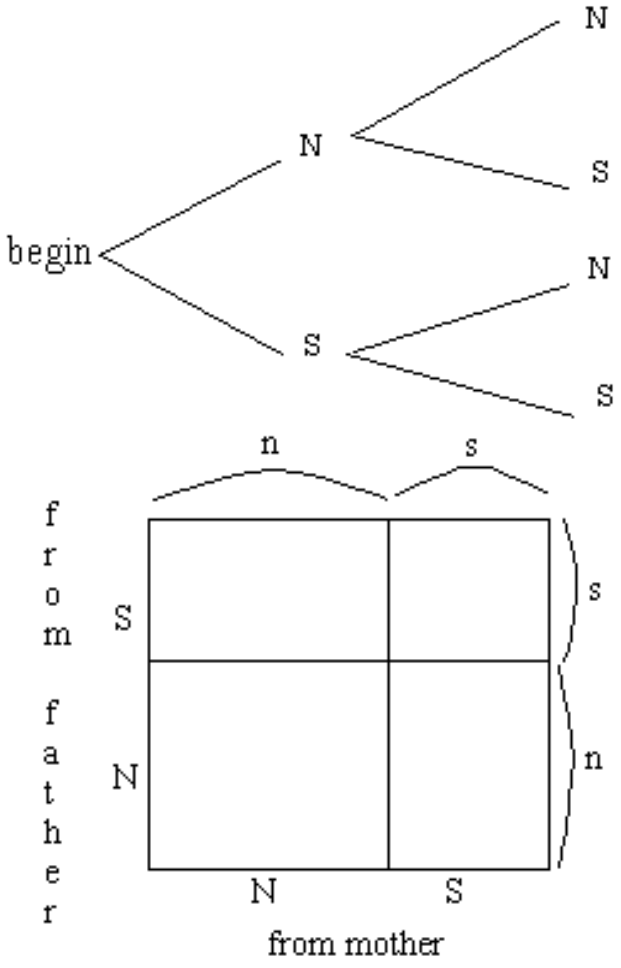
Another important aspect of the relationship between a disease and a new allele for a trait is worth noting. Genetic mutations occur randomly over time, and a large

population can generate a large, diverse number of mutated genes. Most mutations have no beneficial or negative effects on the population, but some mutations may impart protective benefit against a new disease or environmental hazard. When these alleles become useful, they will tend to increase in prevalence among the population allowing a maximum number of individuals to survive the disease or environmental condition; nature tends to optimize. If the population of a species is small, such as in the case of an endangered species, there are fewer potential opportunities for beneficial genetic mutations to occur that could help the species to survive new dangers. Genetic diversity helps a species survive.

A Model with Varying Sickle Cell Survival Rate
Classroom materials

14. Suppose that 90% of NN children survive malaria and that 40% of SS children survive sickle cell anemia. Remember that all of the "NS" children survive both diseases. Assume 1000 children are born.

a. Develop a function $f(n)$ for the total number of children who survive both diseases if n is the fraction of N alleles among the parents and $s = 1 - n$ is the fraction of S alleles among the parents. The following tree diagram and/or area model may help.



b. Find the zeros of $f(n)$. (Note that the roots do not have real world significance.)

c. Use the roots to find the n-value that maximizes the function $f(n)$. This n-value gives the genetic makeup that maximizes the number of children that survive.

d. Graph $f(n)$. Include the roots of $f(n)$ in your graph, even though these values have no physical significance.

A Family of Functions to Model Varying Malaria Rate
Classroom materials

15. The survival rate from malaria for the NN children is dependent on three main factors: the prevalence of malaria-carrying mosquitoes in the region, the amount of protection people have from mosquito bites (such as screened windows and bed nets), and the availability of medical care in the event of a case of malaria. To illustrate the range, in one location in Kenya, people average 10 infective mosquito bites per year; in another location in Ghana, people average from 100 to 1000 infective mosquito bites per year.

a. Let K be the fraction of NN's expected to survive malaria and reach adulthood. Write a function $h(n)$ for the number of children surviving malaria and sickle cell anemia, using the parameter K (that is, you will use K in your function as a constant even though you know it varies from region to region). As before, assume all SS's die of sickle cell anemia and all "NS"'s survive to adulthood. Assume there are 1000 births. What values of K are possible? What values of n are possible?

b. Use a graphing calculator or a computer to graph several of the parabolas in your family of functions, $h(n)$, using $K = 1,\ K = 0.8, K = 0.6,\ K = 0.4,\ K = 0.2,$ $K = 0.$ Put them all on the same coordinate plane. Your graph should use only those values of n that make sense in this situation. The goal of this graphing exercise is to allow you to see all at once a family of functions (in the parameter K) whose equations you wrote. Study this set of graphs. Discuss the patterns. Write down any generalizations you can make.

c. Return to your written function $h(n)$. Find the value of n that maximizes $h(n)$, in terms of K. Use your result to calculate the value of n that maximizes $h(n)$ for each value of $K \epsilon \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Record your answers in a table that shows the value of n that maximizes $f(n)$ for each value of K.

d. Find the maximum survival rate, $\frac{h(n)}{1000}$ where $(n, h(n))$ is the vertex of the parabola. Your population survival rate will be in terms of K, the rate of surviving malaria for the NN adults. Suppose the survival rate for a population is 0.8 meaning that 800 of the 1000 births reached adulthood. Use this to find K, and explain what it means.

Why the Normal and Sickle Cell Alleles Stabilize at the Optimal Proportion

Classroom material

In your study of the interaction of the gene for the sickle cell trait and the presence of malaria in a population, you have seen that the proportion of sickle cell anemia alleles will increase in a population to the point at which the maximum number of individuals survive to adulthood. In this section, we would like to help you understand why because, at first glance, it is not clear that this should actually happen.

Suppose the fraction of normal N alleles in the population is n and the proportion of sickle cell S alleles is $s = 1 - n$. Then you know that the fraction of children born with NN should be n^2, the fraction born with "NS" should be $2ns = 2n(1-n)$, and the fraction born with SS should be $s^2 = (1-n)^2$. If there were 1000 births, there would be $1000n^2$ NN-people, $2000n(1-n)$ "NS"-people, and $1000(1-n)^2$ SS-people.

If all of the SS-people die of sickle cell disease and two-thirds of the NN-people die of malaria, there will be

$$\frac{1000n^2}{3} \text{ NN-survivors and } 2000n(1-n) \text{ "NS"-survivors.}$$

1. Everybody has two alleles. Thus the $\frac{1000n^2}{3}$ NN adults have a total of _____alleles, and the $2000n(1-n)$ NS adults have a total of _____ alleles. So, the total number of alleles in the gene pool for this population is _____

2. Now we want to count the number of N alleles among the survivors. Since each NN-survivor has 2 N alleles, all of their alleles are counted here. NN adults contribute _____ N alleles to the gene pool. However, each "NS"-survivor has only 1 N allele; thus, "NS" adults contribute_____ N alleles to the gene pool. So the total number of N alleles among survivors is _____

3. Use the results of questions 1 and 2 to get the **fraction of N alleles in the next generation**. We call this fraction n_{next}.

$$n_{\text{next}} = \frac{\text{number of N alleles in the gene pool}}{\text{total number of alleles in the gene pool}} = \frac{\qquad\qquad\qquad\qquad\qquad\qquad}{\qquad\qquad\qquad\qquad\qquad\qquad}.$$

If you simplify your fraction as much as possible, you should get $n_{\text{next}} = \frac{3-2n}{6-5n}$.

4. If you begin with a population in which $n = 0.4$, then this fraction implies that for the next generation,

$$n = \frac{3 - 2(0.4)}{6 - 5(0.4)} = 0.55.$$

Use a calculator or a computer to compute the n-value for the next several generations. One way to do this using any graphing calculator is to key in the first value of n, 0.4 in our example, followed by the ENTER key. Then key in $(3 - 2\text{Ans})/(6 - 5\text{Ans})$ followed by repeatedly pressing the ENTER key to get the fraction of N alleles in the population for as many generations as you please. (Note: "Ans" is a **key** on the calculator.) What the calculator is doing is taking its answer at each step and using it as the next value for n. Enter your results in the following table in the column beginning with 0.4. We have entered the generation 2 and generation 3 values for you.

Table 3 : $K = \frac{1}{3}$

Generation	Fraction of N alleles in adult population		
1	0.4	0.6	0.8
2	0.55		
3	0.585		
4			
5			
6			
7			
8			
9			
10			

Repeat the procedure beginning with $n = 0.6$; fill in the next column of Table 3. Then do it once more beginning with $n = 0.8$ to fill in the last column of the table. Do the values approach the value for n that maximizes the number of individuals that survive to adulthood?

5. The value that emerged from your calculator investigation in problem 4 is called an equilibrium value. From your calculator procedure, can you suggest why "equilibrium" is a good name for it? The exact equilibrium value is one of the solutions to the equation

$$n = \frac{3 - 2n}{6 - 5n}$$

Try to explain where this equation comes from.

6. You could now repeat this entire process for a different value of K. If you use K values of $0, 0.2, 0.4, 0.6, 0.8,$ or 1, you can compare your results with the results from question 15 in your previous work.