A Study of Sickle Cell Anemia:
A Hands-on Mathematical Investigation
Teacher Materials

Mathematics Contained in the Lesson

This investigation of the genetics of the Sickle Cell trait via a mathematical model uses probability and teaches properties of quadratic functions and the concept of optimization of a function.

The properties of quadratic functions brought out by this investigation are
- the relationship between the zeros of the function and its factors,
- the relationship between the zeros and the location of its vertex,
- the symmetry of its graph, and the location of its extreme point,
- factors of quadratics of the form $ax^2 + bx$.

More complicated quadratics are studied in the section **A Family of Functions to Model Varying Malaria Risk**, including the *possible use of the quadratic formula* and the *effect that changing the coefficients in a function has on the location of the maximum value.*

Students gain further experience in this lesson in
- developing a mathematical model from given information,
- using function notation.

The last (optional) section, **Why the Normal and Sickle Cell Alleles Stabilize at the Optimal Proportion**, explores the dynamical system that represents the genetic process. This is an extension of the lesson from which students will be able to see that the equilibrium value of the system is the very value of n that maximizes the number of people who survive to adulthood. This will help to clarify the biological process behind the algebraic results. Because this last section is not directly related to quadratic functions, you may prefer to omit it or hand it out to students who have an interest in following-up on this activity.

The lesson can be used at the Intermediate Algebra level; it is also appropriate at the Pre-calculus level.

Set-up

To do the simulation in the section of the Student Materials called **Modeling a Population where Malaria is a Risk: A Physical Model**, each group of three or four students needs 1 cup, 7 beads of one color and 6 beads of another color. The beads must be identical except for color. They should be somewhat flat rather than perfectly round so that they do not roll off students' work tables. If such beads are not available, you could use, for example, pennies with one set colored with a marker.

We assumed that two-thirds of the NN's die of malaria in our simulation. A more realistic choice would be that one-third of the NN's die of malaria. You should inform your students of this fact. We chose two-thirds because the results of the two simulations

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

are more likely to be different using this death rate than using the more appropriate one of one-third.

Organization

The first three pages of the Student Materials is a reading assignment intended as homework before you begin the lesson in class. It provides the background information about the Sickle Cell trait and Malaria.

Page 4 of the Student Materials is a simulation of the genetics process described in the Introduction. The purpose of the simulation is to help students understand the problem. It is not intended to teach the mathematics and it might seem expendable, but we encourage you to do it because the hands-on experience is very helpful to students' visualization of the process for which they will afterward need to develop a mathematical model. (It is possible to assign the simulation as homework for the same night as the reading assignment. Each student could conduct a simulation using pennies instead of beads. If you make this assignment, you should demonstrate the process before assigning it so each student is sure how it is done.) When they have completed the simulation, ascertain that they know:

- Each person is born with two genes, one from each parent, that determine whether they have the sickle cell trait or have normal blood cells.
- The normal allele is labeled N; the sickle cell allele is labeled S.
- The probability of an allele being normal, N, is the fraction of N alleles in the parent generation gene pool; we call it n. The probability of an allele being "defective", S, is the fraction of S alleles in the parent generation gene pool; we call it s.
- All of the alleles are either N or S, so $n + s = 1$.
- Every person is either NN, "NS", or SS. (We do not distinguish between NS and SN.)
- In our example, $\frac{2}{3}$ of the NN children die from malaria before they are old enough to reproduce; all of the SS children die from Sickle Cell Anemia before they are old enough to reproduce; all of the NS children live to be old enough to reproduce. A single sickle cell allele does not result in Sickle Cell Anemia, and the single allele creates a condition that affords protection from malaria.

The third section, **Making a Mathematical Model of the Population**, guides students in developing the algebraic function for the expected number of adult survivors from a certain birth population. The function they develop is quadratic, with one root at the origin. Students find both roots and use symmetry to find the vertex.

In the fourth section, **A Model with Varying Sickle Cell Survival Rate**, students solve a problem using what they have learned. They develop a quadratic with nonzero roots and use symmetry to find the vertex. This quadratic can easily be factored, but you can give them variations in which they need the quadratic formula to find the roots.

In the fifth section, **A Family of Functions to Model Varying Malaria Rate**, students solve a problem in which the quadratic function involves a parameter. Students then explore how the parameter effects the shape and position of the parabola.

The fourth and fifth sections do not need to be covered immediately after the third section. They can be omitted or covered at an appropriate later time during the course, say when you are studying the quadratic formula.

The last section of this lesson, **Why the Normal and Sickle Cell Alleles Stabilize at the Optimal Proportion**, is optional. It provides an opportunity to view the situation from a quite different mathematical perspective. Much of the mathematics is done for the students in this section; however, at the end of the section of the Teacher Materials titled **Answers and Teaching Suggestions**, we offer two additional questions that you might assign to students if you want to give more emphasis to this optional section.

Answers to Problems and Teaching Suggestions for:

Modeling a Population Where Malaria is a Risk: A Physical Model

The simulation is valuable but it should not consume too much time. To speed it up, you can introduce it rather than have students read how to do it on page 4.

Each group's Table 1 will differ according to the number of N-beads and S-beads they happened to draw, but each one should fit this pattern:

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	$\frac{1}{3}$(#NNs) + #NSs	$\frac{1}{3}$(#NNs) + #NSs

1. Domain = $\{n : 0 \leq n \leq 1\}$and Range $= \{y : 0 \leq y \leq 30\}$

Reflection point: Have students pool their simulation results by averaging all groups' entries in Table 1. Seek student comment on the range of outcomes.

Making a Mathematical Model of the Population

2. Expected number of "NS" is $0.48 \times 30 = 14.4$; expected number of SS is $0.16 \times 30 = 4.8$

Table 2: Results of predictions when $K = 1/3$				
fraction of N alleles in adult population	0.6	0.4	0.3	n
fraction of S alleles in adult population	0.4	0.6	0.7	s or $1-n$
number of the 30 births that are NN	10.8	4.8	2.7	$30n^2$
number of NN adults who survive malaria	3.6	1.6	0.9	$10n^2$
number of the 30 births that are "NS"	14.4	14.4	12.6	$60ns$ or $60n(1-n)$
total of 30 births surviving to adulthood	18	16	13.5	$10n(n+6s)$ or $10n(6-5n)$

If students have trouble with the tree diagram and the area model, and thus this table, refer them to the tree diagram and area model in figure 1 for assistance. We included both a tree diagram and the area model because different students will respond to different approaches.

3. $f(0.6) = 18$

4. $f(0.4) = 16$; $f(0.3) = 13.5$

Reflection point: *Simulations and theoretical expected results.* If possible, take time to compare the pooled results from Table 1 to the last line of Table 2. Discuss why neither the physical simulation nor the theoretical 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 produce exactly the expected results.

5. The number of NN $= 30n^2$. The number of "NS" $= 60ns$ or $60n(1-n)$ (see last column of Table 2). *Be sure every group has the correct results and every student understands them. The rest of the lesson depends on this understanding.*

6. $f(n) = 10n^2 + 60ns = 10n^2 + 60n(1-n) = 60n - 50n^2$
If students wrote the last column of Table 2 in terms of n only, this question may confuse them by being "too easy." Help them to see that they already have the expression they need and all that remains is to rewrite it as an equation.

7. $f(n) = 10n(6 - 5n)$

8. The graph of $f(n) = 10n(6-5n)$ is a parabola with intercepts at (0,0) and ($6/5$, 0) and with a maximum point at ($3/5$, 18). Each number on the horizontal axis represents a population with that n-value being the fraction of the alleles in that population that are N. The numbers on the vertical axis represent the number of people per thirty births who are expected to live to adulthood given that genetic make-up of the population. Each n-intercept makes one of the factors equal to 0.

9. The domain of $f(n)$ in this context is $0 \leq n \leq 1$.

10. For $f(n) = 10n(6-5n)$, the size of the adult population $f(n)$ increases and then decreases. This means that at first, as n increases, more and more people live long enough to grow up; but after a certain point, as n increases further the number of people who survive to adulthood is smaller and smaller. When $f(n)$ has positive slope ("goes uphill to the right"), it means that when n is in that interval, if n increases more people will live. When the slope is negative, it means that when n is in that interval, if n increases fewer people will live to adulthood.

11. When $n = \frac{3}{5}$, the greatest number of children survive to adulthood in our example in which two-thirds of the NN's die from Malaria before reaching adulthood. Algebra students should find the value of n that maximizes $f(n)$ by using their knowledge of the symmetry of the parabola. If $f(n) = 10n(6-5n)$, then they know that the line of symmetry, and thus the vertex of the parabola, is halfway between any two n values that give the same value for $f(n)$; in particular, halfway between the two horizontal intercepts. They can solve

$$0 = 10n(6 - 5n), \text{getting } n = 0 \text{ or } n = \frac{6}{5}.$$

Thus, at the vertex of the parabola, $n = \frac{3}{5}$. The number of the 30 births that survive to adulthood is 18 so that the fraction of the births that survives is $\frac{18}{30} = \frac{3}{5}$.

The number of children who grow up is greatest when the gene pool has 60% normal alleles and 40% sickle cell trait alleles.

12. $g(n) = \frac{9}{10}(1000n^2) + 2000ns = 900n^2 + 2000n(1-n) = 100n(20-11n)$. To find the value of n that maximizes the number of children who survive to adulthood where $\frac{1}{10}$ of the NN's die from Malaria, we need to know the value of n that maximizes $g(n)$; that is. the value of n at the vertex of the parabola. Again using the symmetry of a parabola, solving $0 = 100n(20-11n)$ gives $n = 0$ and $n = \frac{20}{11}$ as the horizontal intercepts. Thus the vertex of the parabola is at $n = \frac{10}{11}$. Students may wish to compare the fraction of the births that live to adulthood in each case. Note that $f(\frac{3}{5})/30 = \frac{3}{5}$ and $g(\frac{10}{11})/1000 = \frac{10}{11}$.

13. $f(n)$ and $g(n)$ are quadratic functions. The value of n that maximizes each of these functions is halfway between the n-intercepts on the horizontal axis since the function is symmetrical about the vertical line through its vertex.

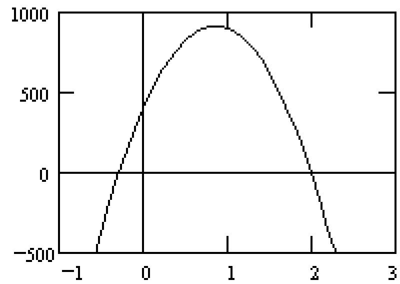
Reflection point: *Quadratic functions and optimum values.* When all students have completed at least through problem 11, initiate a whole class discussion for students to summarize basic points: (1) What is a quadratic function? (2) Why is the function describing the number of people in the adult population in the sickle cell/malaria problem quadratic? (3) How are the factors of a function related to its x-intercepts and why? (4) How can you find the coordinates of the vertex of a quadratic function? (6) What is meant by the word "maximize"? Other useful discussion questions: Does every function have a maximum? Does every function have a minimum?

A Model with Varying Sickle Cell Survival Rate

Reflection Point: This section is to introduce factoring quadratic functions. The factors can then be used to find the vertex. This can be covered slightly later during the course, depending on the organization of your course.

14a. $f(n) = 0.9(1000n^2) + 1000(2ns) + 0.4(1000s^2)$
$= 900n^2 + 2000n(1-n) + 400(1-n)^2$
$= 400 + 1200n - 700n^2$

b. $n = -2/7,\ n = 2$
c. $\frac{1}{2}(-\frac{2}{7} + 2) = \frac{6}{7}$
d.



Note that although this function can be factored over the integers, doing so may be difficult for some students. If so, this is an excellent opportunity to use the quadratic formula.

Reflection Point: If you give problem 14 again, but use different values for the survival rates for malaria and sickle cell anemia, the resulting quadratic function may require the quadratic formula to factor. In fact, if $m =$ the fraction of NN that survive malaria and $s =$ the fraction of SS that survive sickle cell anemia, then the quadratic function is

$$g(n) = (m + s - 2)n^2 + (2 - 2s)n + s$$

and the two roots are

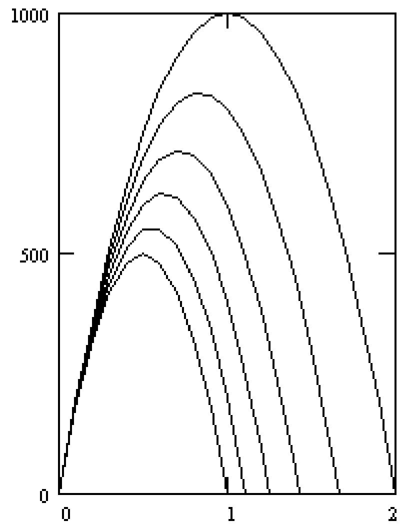
$$n = \frac{s - 1 \pm \sqrt{1 - ms}}{m + s - 2}$$

A Family of Functions to Model Varying Malaria Rate

Reflection Point: This problem helps students understand how parameters can effect the shape of a parabola. In this problem, the parameter is K, the fraction of NN individuals that survive of malaria. Two important results will be 1) as K increases, the size of the population that survives increases, as seen by the parabolas moving upward, and 2) as K increases, the optimal proportion of n-alleles increases, as seen by the vertex of the parabolas moving to the right.

15. a. $h(n) = 1000Kn^2 + 2000n(1 - n)$, $0 \le K \le 1$, $0 \le n \le 1$.
"Simplifying" the expression serves no useful purpose.

b.



Using a List function to get all 6 graphs from inputting the function once as $1000\mathrm{L}_1 x^2 + 2000x(1-x)$ is a satisfying experience and gives students a useful tool for exploring the effects of parameters on functions in your future lessons. Students should note that as K, the fraction of NN individuals who survive malaria, increases, the value of n that maximizes the size of the adult population (minimizes the number of childhood deaths) also increases. As K increases, the position of the vertex of the parabola moves to the right. Note that as K increases, the coefficient of the quadratic term increases. Students will also note that as K increases, the position of the vertex moves upward and to the right; this reflects that more people achieve adulthood at higher values of K, and the optimal n-value increases.

c. Where $h(n) = 1000Kn^2 + 2000n(1-n)$, we can find the turning point by using the symmetry of the parabola in the same way we did it for $K = 1/3$ and $K = 9/10$ in the last section. Locate the horizontal intercepts, the two n values where $h(n) = 0$. The vertex of the parabola is on the line of symmetry, so it must be halfway between these two points, which would be mirror images of each other in the line of reflection. Thus, students can solve

$$1000Kn^2 + 2000n(1-n) = 0$$
$$1000n([K-2]n + 2) = 0$$
$$n = 0 \text{ or } n = \tfrac{2}{2-K}$$

Thus, the vertex of the parabola is at $n = \frac{1}{2}(0 + \frac{2}{2-K}) = \frac{1}{2-K}$

K	0	.2	.4	.6	.8	1
n	.5	.555	.625	.714	.833	1

d. By part c, the function $h(n) = 1000Kn^2 + 2000n(1-n)$ has its vertex at

$$h(1/(2-K)) = 1000\left(\frac{1}{2-K}\right)([K-2]\left(\frac{1}{2-K}\right)+2) = 1000\left(\frac{1}{2-K}\right)(-1+2)$$

$$= \frac{1000}{2-K}$$

Therefore, $h(\frac{1}{2-K})/1000 = \frac{1}{2-K}$. You are given that the survival rate is 0.8, so

$$\frac{1}{2-K} = 0.8 \quad \text{or} \quad K = 0.75$$

Reflection Point: Part d is a difficult problem, both conceptionally and computationally. This part can be omitted. The reason we include it is three fold. First, it gives students practice manipulating variables. Second, it is interesting that the optimal level of the N allele, $n = \frac{1}{2-K}$, equals the survival rate of the population, $h(\frac{1}{2-K})/1000 = \frac{1}{2-K}$. Third, the students should observe that they are able to find the survival rate from malaria for the NN individuals indirectly, while it would be difficult to find their survival rate directly since they are indistinguishable from the NS individuals.

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

Recall that this supplementary section is optional.

1. NN's contribute $\frac{2000n^2}{3}$ alleles to the gene pool and NS's contribute $4000n(1-n)$. Thus thee total number of alleles $= \frac{2000n^2}{3} + 4000n(1-n)$.

2. NN's contribute $\frac{2000n^2}{3}$ N-alleles to the gene pool and NS's contribute $2000n(1-n)$ N alleles. Thus, the total number of N alleles among survivors $= \frac{2000n^2}{3} + 2000n(1-n)$.

3. $n_{\text{next}} = \frac{\frac{2000n^2}{3}+2000n(1-n)}{\frac{2000n^2}{3}+4000n(1-n)} = \frac{\frac{n}{3}+(1-n)}{\frac{n}{3}+2(1-n)} = \frac{3-2n}{6-5n}$.

4.

Table 3 : K = 1/3			
Generation	Fraction of N alleles in adult population		
1	0.4	0.6	0.8
2	0.55	0.6	0.7
3	0.5846	0.6	0.64
4	0.595	0.6	0.6143
5	0.5983	0.6	0.6049
6	0.5995	0.6	0.6016
7	0.5998	0.6	0.6005
8	0.59993	0.6	0.6002
9	0.59997	0.6	0.6001
10	0.59999	0.6	0.6000

Assure that students note that in each column, the fraction of N alleles in the population approaches 0.6, the value of n that maximizes the number of individuals who survive to adulthood.

5. It is hoped that students will explain that the equation $n = \frac{3-2n}{6-5n}$ comes from assuming that the n values have achieved "equilibrium" so that the **next** n, $\frac{3-2n}{6-5n}$, is now equal to the old n.

In this optional section, you might wish to include the two additional questions given below for some students.

1. Repeat the process you used to make Table 3, but this time let the fraction of NN's expected to survive Malaria be $K = \frac{9}{10}$. You can use your work from problem 12 of **Making a Mathematical Model of the Population** and the process used in this section. Complete Table 4 using your results for $n(5)$, $n(10)$, $n(15)$, $n(20)$ and $n(25)$. To get these, just count the number of times you press the ENTER key; the first time gives you $n(2)$.

Table 4 : $\boldsymbol{K = 9/10}$			
Generation	Fraction of W alleles in adult population		
1	0.4	0.6	0.8
5			
10			
15			
20			
25			

How many years, on the average, would you expect five generations to cover?

2. Study the two tables you made for the outcomes of your investigations, Table 3 and Table 4. Do you find any relationships between any line in the tables and the outcomes of your investigations of the functions $f(n)$ and $g(n)$ from your work in **Making a Mathematical Model of the Population**?

Answers to the additional questions:

1. The fraction of N alleles in the adult gene pool when $\frac{1}{10}$ of the NN people die of malaria is $\frac{100n(20-2n)}{200n(20-11n)} = \frac{10-n}{20-11n}$. If you use a graphing calculator, first enter the initial value of n, followed by the ENTER key. Then enter $(10 - ANS)/(20 - 11ANS)$ followed by repeatedly hitting the ENTER. If students miscount the ENTERs and thus put the wrong numbers in the table, they will still see the trend, which is the point.

Table 4 : $\boldsymbol{K = 9/10}$			
Generation	Fraction of W alleles in adult population		
1	0.4	0.6	0.8
5	.7949	.8156	.8584
10	.8646	.8702	.8847
15	.8872	.8896	.8961
20	.8973	.8985	.9019
25	.9025	.9031	.9045

2. Students may notice that all three columns of Table 3, where $K = 1/3,$ are approaching the value of n they found in problem 11 to maximize the size of the adult population. Continuing the iteration on the calculator several more times after completing the table will provide convincing evidence.

All three columns in Table 4 approach the value of n they found in problem 12 to maximize the size of the adult population when $K = 9/10$. Thus, the equilibrium value in each case is the value of n that will maximize the number of people who live to adulthood. Restated, the proportion of the gene pool that is "normal" is the value of n that minimizes the number of childhood deaths from a combination of Sickle Cell Anemia and Malaria.