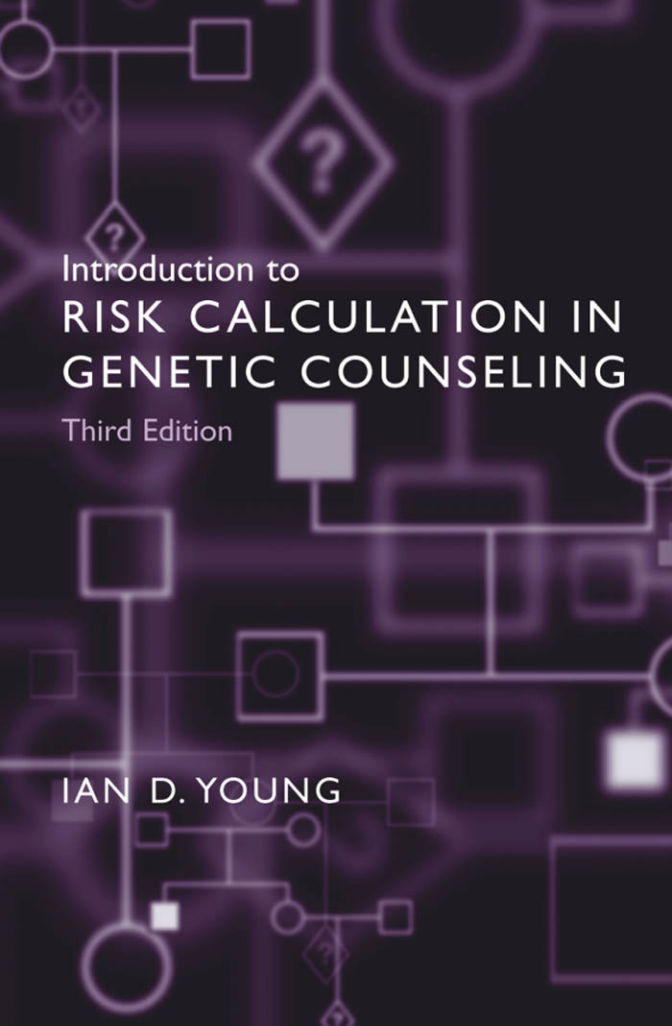

Introduction to RISK CALCULATION IN GENETIC COUNSELING

Third Edition

IAN D. YOUNG

INTRODUCTION TO RISK CALCULATION IN GENETIC COUNSELING

This page intentionally left blank

INTRODUCTION TO RISK CALCULATION IN GENETIC COUNSELING

Third Edition

Ian D. Young

Department of Clinical Genetics
Leicester Royal Infirmary
Leicester, United Kingdom

OXFORD
UNIVERSITY PRESS

2007

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Young, Ian D.
Introduction to risk calculation in genetic counseling / Ian D. Young.—3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-0-19-530527-2
ISBN: 0-19-530527-2
1. Genetic counseling. 2. Health risk assessment. 3. Probabilities. 4. Genetic disorders—
Risk factors—Mathematical models. I. Title.
[DNLM: 1. Genetic Counseling. 2. Risk Assessment. QZ 50
Y72i 2006]
RB155.7.Y68 2006
616'.042–dc22 2006001581

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Preface

Although there has been remarkable progress in medical genetics over the past 10 years, with rapid advances in both molecular and computer technology, many situations still arise in which the providers of genetics services are called upon to undertake risk calculation. Indeed, the profusion of genetic tests has often served to fuel demand and expectations rather than ease this burden of responsibility. Factors such as heterogeneity, the use of linked markers, germline mosaicism, and the current limitations of mutation screening techniques have all served to complicate what was once perceived as a relatively straightforward process. In the present climate of accountability and litigation it is no longer acceptable, either in the clinic or in the laboratory, to conclude that a risk is simply high or low.

Numbers and probability theory are not to everyone's taste. Some genetic risks can be calculated easily. Others, to borrow a term from the popular pastime of Su Doku, can be fiendishly difficult. Those seeking guidance are not always helped by abstruse mathematical papers on the rigorous treatment of already difficult concepts. For most of us mere mortals, the maxim is to keep it short and simple. If it doesn't make sense to us, how can we hope to explain it to our colleagues or our patients? As with the two previous editions, this book has been written to come to the rescue of those who seek to master the basic principles without necessarily achieving grand master status.

Those who have access to the previous edition will note that there have been a number of significant changes. Case scenarios have been included in many chapters to further demonstrate how the theoretical principles can be applied in practice. Chapter 3 has been expanded to show how multiple consanguineous loops can be accommodated, and a new chapter has been included to address the difficult and very important issue of germline mosaicism. The chapters on cancer and multifactorial inheritance have been extensively revised, and a new appendix has been added to demonstrate approaches to the challenging calculations that can arise in prenatal diagnosis. Unfortunately, many publishers (Oxford University Press is a

notable and worthy exception) now levy exorbitant fees to reproduce material from their books and journals, so that in some instances it has not been possible to include helpful illustrations and tables. It is hoped that readers will forgive this occasional omission, brought about by the importance of keeping the purchase price as low as possible.

Once again, I am indebted to the many colleagues who have reviewed relevant chapters and found embarrassing errors. Any that remain are entirely the responsibility of the author. Feedback from readers with comments and criticism, preferably constructive, is always welcome.

Contents

This page intentionally left blank

INTRODUCTION TO RISK CALCULATION IN GENETIC COUNSELING

This page intentionally left blank

1

Genetic Counseling and the Laws of Probability

1.1 Genetic Counseling and the Concept of Risk

Genetic Counseling

Since the introduction of genetics clinics approximately 50 years ago, many attempts have been made to devise a universally acceptable definition of genetic counseling. As the subject has expanded, a consensus has gradually emerged that this should be viewed as a nondirective communication process that addresses an individual's needs and concerns relating to the development and/or transmission of a genetic disorder. This definition implies that the process of genetic counseling involves many important components, including the gathering of family information, the establishment of a diagnosis, the communication of information, and the provision of ongoing support. Equally important is the mathematical assessment of risk, and it is with this particular aspect of genetic counseling that this book is concerned.

It is true that in the past the determination of genetic risks often required little more than a knowledge of the basic principles of Mendelian inheritance. For many disorders, risk assessment simply involved the provision of a straightforward dominant or recessive recurrence risk, and today this still applies in some situations. However, an increasing awareness of the complexity and heterogeneity of genetic disease has focused attention on the importance of taking other factors, such as reduced penetrance and delayed age of onset, into account. In addition, the use of linked DNA markers and the increasing availability of specific mutation analysis often serve to complicate rather than simplify risk calculations that require careful consideration and a relatively high level of numerical competence if the provision of incorrect information is to be avoided.

Expressing a Risk in Mathematical Terms

There are several ways in which a risk can be expressed. Mathematically, probability is usually indicated as a proportion of 1, ranging from 0, indicating that an event cannot occur, up to 1, which implies that an event has to occur. However, many individuals are more comfortable with the concept of *odds*, as illustrated by the recent demonstration that a group of women being counseled for a family history of breast cancer actually expressed a preference for receiving information about risks in the form of "gambling odds" (Hopwood et al., 2003). Thus, a probability of 0.25 can be expressed as a risk of 1 in 4 ("1 chance in 4"), or 25%, that an event will occur, or conversely, as 3 chances out of 4 (75%) that an event will not occur. Note that a risk of 1 in 4 should not be expressed as 1:4, i.e., 1 *to* 4, which actually equals 1 *in* 5. When counseling those who are particularly comfortable with the concepts of gambling, a probability of 0.25 could also be expressed as odds of 3 to 1 against or 1 to 3 in favor of a particular outcome being observed.

Key Point 1

A probability or risk can be expressed

1. as a percentage, e.g., 50%
2. as a proportion of 1, e.g., 0.5
3. as a "chance," e.g., 1 chance in 2
4. as an odds ratio, e.g., 1 to 1 (evens)

Whilst discussing the quantitative nature of a risk, it is both helpful and important to point out that chance does not have a memory. The fact that a couple's first child has an autosomal recessive disorder does not imply that their next three children will be unaffected. The key point is that the risk of 1 in 4 applies to *each* future child. It can be extremely embarrassing to learn that a second affected child has been born to parents who have clearly misinterpreted information given at a previous counseling session.

Qualifying a Genetic Risk

The provision of a genetic risk does not simply involve the conveyance of a risk figure in stark isolation. Genetic risks should be qualified in a number of ways. For example, risks should be placed in context, perhaps by pointing out that 1 in 40 of all babies has a congenital malformation when dealing with a pediatric problem, or by reminding adults at risk of malignancy that there is an extremely high background population risk of approximately 1 in 3 for developing cancer. It is also important that the genetic counselor should not be seen exclusively as a prophet of doom. Thus, it is well worth emphasizing that a risk of 1 in 20 for an adverse outcome such as a neural tube defect means that there are 19 chances out of 20 that a future baby will not be affected.

Having ensured that a risk has been correctly quantified, understood, and placed in context, it is essential that an indication should be given of what the risk is actually for. Does the quoted figure relate to the risk of inheriting the relevant gene or does it give an indication of the probability that a serious complication will occur? The interpretation of severity and risk is subjective and unpredictable. For example, some parents who have undergone extensive surgery in childhood for repair of an abnormality such as cleft lip and palate can be profoundly concerned about a relatively low polygenic/multifactorial risk for offspring. In contrast, other prospective parents with disorders such as achondroplasia or autosomal dominant deafness are often not at all perturbed by much higher risks for offspring. In some genetics text books it is stated that, as an arbitrary guide, risks of 1 in 10 or greater can be viewed as "high," 1 in 20 or less as "low," and intermediate values as "moderate." As a generalization, these values have some merit, but for many individuals the perception and interpretation of risk are based more on emotion and personal experience than on cold objectivity and logic. Studies of the long-term impact of genetic counseling have shown that it is the burden of a disorder rather than its precise risk that is of major concern to counselees.

This brings us to a final point that is often raised in the context of risk calculation. A case can reasonably be made that many counselees are primarily concerned with whether a risk is high or low, and that it is therefore both unnecessary and unhelpful to define a risk to the nearest decimal point. While this may well be true in some situations, it does not detract from the importance of giving correct and precise information, an obligation neatly summarized by Lalouel et al. as long ago as 1977, when they maintained that "the question of whether consultees demand as much specificity should be subordinate to the question of whether counselors are justified in providing less." Almost 30 years later, in an editorial in the *American Journal of Medical Genetics* (Hodge and Flodman, 2004), the point was elegantly made that it can be extremely dangerous to "trust intuition," as illustrated by estimated carrier risks ranging from less than 1% to 50% for a specific female in a cited X-linked pedigree—the correct value was 10.7% (Hodge and Flodman, 2005). There is some evidence that it is more effective to present risks in the form of numbers rather than words (Marteau et al., 2000), and it is obvious that if a numerical risk is being given, then it should be correct.

1.2 The Laws of Probability

Two relatively simple principles are often used when calculating risks in genetic counseling. These are known as the *laws of addition and multiplication*.

Law of Addition

If two (or more) events are mutually exclusive, and the probability of event one occurring is $P1$ and that of event two occurring is $P2$, then the probability of *either* the first event *or* the second event occurring equals $P1 + P2$. An obvious example of the application of this law is the probability that a baby will be male or female. If the

probability of having a boy equals 0.5 and the probability of having a girl also equals 0.5, then the probability that any particular baby will be *either* male *or* female equals 1.

Law of Multiplication

This law is applied when the outcome of two (or more) events is independent. If the probability of one event occurring is $P1$ and that of another event is $P2$, then the probability of *both* the first *and* the second event occurring equals $P1 \times P2$. For example, if parents are considering having two children, then excluding the unlikely event of identical twins, the probability that both children will be boys equals the product of the probabilities that *both* the first child *and* the second child will be male, i.e., $1/2 \times 1/2 = 1/4$.

> **Key Point 2**
>
> Probabilities are added when they relate to mutually exclusive alternative outcomes. Probabilities are multiplied when they relate to independent events or outcomes.

Examples Based on Twin Pregnancies

The demonstration of twins in a pregnancy can generate some difficult counseling problems (Hunter and Cox, 1979). The following two examples show how these simple laws of addition and multiplication can be used to calculate the probability that one or both babies will be affected with a genetic disorder. The underlying principle is to determine the probabilities if the twins are either monozygotic (MZ) or dizygotic (DZ) and then calculate the total risks by adding the probabilities for each weighted on the basis of their relative frequencies. In these examples, it is assumed that the ratio of MZ to DZ twinning is 1 to 2, i.e., that one-third of twin pairs are MZ and two-thirds are DZ.

Example 1

Healthy parents have had a child with a severe autosomal recessive disorder. Aware of the recurrence risk of 1 in 4, they embark upon another pregnancy. During the early stages of this pregnancy, ultrasonography reveals the presence of twins. The parents now wish to know the probability that at least one twin will be affected by the autosomal recessive disorder.

To answer this question, the risks are first calculated for the two mutually exclusive possibilities that the twins are either (1) monozygotic ($P = 1/3$) or (2) dizygotic ($P = 2/3$).

1. Monozygotic. In this situation, the probability that both twins will be affected equals $1/4$, that only one will be affected equals 0, and that both will be unaffected equals $3/4$.

2. Dizygotic. If the twins are dizygotic, then the genotype of one twin does not influence the genotype of the other twin, i.e., these events are independent. Therefore, the probability that both twins will be affected equals 1/16 (i.e., $1/4 \times 1/4$), the probability that only one will be affected equals 3/8 [i.e., $(1/4 \times 3/4) + (3/4 \times 1/4)$], and the probability that both will be unaffected equals 9/16 (i.e., $3/4 \times 3/4$).

Using this combination of mutually exclusive and independent events, the parents' question can now be answered. The overall probability that

1. both twins will be affected equals MZ ($1/3 \times 1/4$) plus DZ ($2/3 \times 1/16$), giving a total probability of 1/8.
2. only one twin will be affected equals MZ ($1/3 \times 0$) plus DZ ($2/3 \times 3/8$), giving a total probability of 1/4.
3. both twins will be unaffected equals MZ ($1/3 \times 3/4$) plus DZ ($2/3 \times 9/16$), giving a total probability of 5/8.

Based on this information, the parents can be reliably informed that there is a probability of $1/8 + 1/4 = 3/8$ that at least one of their unborn babies will be affected.

Example 2

A healthy 40-year-old woman with no relevant family history is found to be carrying twins and wishes to know the probability that one or both of her babies will have Down syndrome. In calculating the answer, it is assumed that the ratio of MZ to DZ twinning remains constant at 1:2 at all ages and that the risk of Down syndrome in a singleton pregnancy conceived by a 40-year-old woman equals 1 in 100.

Once again, the first step is to determine the risks for each of the two mutually exclusive possibilities, i.e., that the twins are either monozygotic or dizygotic.

1. Monozygotic. The probability that both twins will be affected equals 1/100, that only one will be affected equals 0 (this ignores the unlikely possibility of postzygotic nondisjunction), and that both will be unaffected equals 99/100.
2. Dizygotic. The karyotypes of the twins are independent of each other. Therefore, the probability that both will be affected equals 1/10 000 (i.e., $1/100 \times 1/100$), the probability that only one twin will be affected equals 198/10 000 (i.e., $2 \times 1/100 \times 99/100$), and the probability that both will be unaffected equals 9801/10,000 (i.e., $99/100 \times 99/100$).

The concluding step is to weight the probabilities obtained, assuming either MZ or DZ according to their relative frequencies, i.e., 1:2. The overall probability that

1. both twins will be affected equals MZ ($1/3 \times 1/100$) plus DZ ($2/3 \times 1/10{,}000$)), which equals 17/5000 (approximately 1 in 294 or 0.34%).
2. only one twin will be affected equals MZ ($1/3 \times 0$) plus DZ ($2/3 \times 198/10{,}000$), which equals 66/5000 (approximately 1 in 76 or 1.32%).
3. both twins will be unaffected equals MZ ($1/3 \times 99/100$) plus DZ ($2/3 \times 9801/10{,}000$), which equals 4917/5000 or 98.34%.

Based on this information, the prospective mother can be informed that the probability that at least one of the twins will have Down syndrome equals 1.66% or 1 in 60.

1.3 The Binomial Distribution

Use of the binomial distribution enables easy calculation of the probability of obtaining a particular number or distribution of events of one kind (e.g., boys or girls) in a sample, given knowledge of the probability of each event occurring independently. The binomial distribution can be presented in the form of an equation.

$$P = \frac{n!}{(n-r)!r!} p^{n-r} q^r$$

where

P = the probability of observing the particular split of r of one event and $n-r$ of the other
n = the total sample size
r = the number of events of one type observed
p = the probability of this event not occurring
q = the probability of this event occurring
! = factorial, e.g., $4! = 4 \times 3 \times 2 \times 1$, and by convention $0! = 1$

Example 3

Prospective parents who are planning to decorate their spare bedrooms wish to know the probability that their prenatally diagnosed quadruplets will consist of both boys and girls. To simplify the calculation, it is assumed that the babies were conceived with the help of an ovulation-inducing drug such as clomiphene, so that the probability that they are MZ can effectively be ignored.

In this example, the sibship size (n) equals 4, and the probability of a male baby (p) equals the probability of a female baby (q) equals 1/2. Using the binomial distribution, the probability of having zero, one, two, three, or four male babies can be calculated as follows:

1. no male babies

$$P = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

2. one male baby ($r = 1$)

$$P = \frac{4!}{3!\ \ 1} \left(\frac{1}{2}\right)^3 \frac{1}{2} = \frac{1}{4}$$

3. two male babies ($r = 2$)

$$P = \frac{4!}{2!\ \ 2!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

4. three male babies ($r = 3$)

$$P = \frac{4!}{1!\quad 3!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4}$$

5. four male babies

$$P = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

Thus, these parents can be informed that there is a probability of

$$\frac{\frac{1}{4} + \frac{3}{8} + \frac{1}{4}}{\frac{1}{16} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16}} = \frac{7}{8}$$

that their quadruplets will consist of different-sex infants.

Example 4

In this example the same prospective parents are expecting nonidentical quadruplets, having already had a child with an autosomal recessive disorder. They wish to know the probability that at least two of their four babies will be affected.

The calculation proceeds as in Example 3, but now $p = 3/4$ and $q = 1/4$, i.e.:

1. no affected babies

$$P = \left(\frac{3}{4}\right)^4 = \frac{81}{256}$$

2. one affected baby ($r = 1$)

$$P = \frac{4!}{3!\quad 1}\left(\frac{3}{4}\right)^3\frac{1}{4} = \frac{108}{256}$$

3. two affected babies ($r = 2$)

$$P = \frac{4!}{2!\quad 2!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{54}{256}$$

4. three affected babies ($r = 3$)

$$P = \frac{4!}{1\quad 3!}\frac{3}{4}\left(\frac{1}{4}\right)^3 = \frac{12}{256}$$

5. four affected babies

$$P = \left(\frac{1}{4}\right)^4 = \frac{1}{256}$$

Thus, these parents can be informed that the probability that at least two of their four babies will be affected equals

$$\frac{54 + 12 + 1}{81 + 108 + 54 + 12 + 1} = \frac{67}{256}$$

Example 5

The binomial distribution can be used in many other situations, such as that outlined in Chapter 3 (p. 53), in which parents have had children affected with different autosomal recessive disorders. In this example, consider parents who have already had three children, each with a different autosomal recessive disorder. They wish to know the probabilities that their next child will inherit all, some, or none of these conditions. To simplify matters, it is assumed that the loci of the disorders are not linked. The various probabilities can be calculated using the binomial distribution, where

n = the total number of disorders
r = the number of disorders inherited by the fourth child
p = the probability of not inheriting each disease = 3/4
q = the probability of inheriting each disease = 1/4

Thus, the probabilities for the next child will be

1. child inherits none of the conditions

$$P = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$$

2. child inherits one condition ($r = 1$)

$$P = \frac{3!}{2!\ \ 1}\left(\frac{3}{4}\right)^2\frac{1}{4} = \frac{27}{64}$$

3. child inherits two conditions ($r = 2$)

$$P = \frac{3!}{1\ \ 2!}\frac{3}{4}\left(\frac{1}{4}\right)^2 = \frac{9}{64}$$

4. child inherits all three conditions ($r = 3$)

$$P = \left(\frac{1}{4}\right)^3 = \frac{1}{64}$$

Key Point 3

The binomial distribution is obtained by expanding $(p + q)^n$ and can be used to determine the probability that a particular distribution of events (e.g., boys or girls, heads or tails) will occur.

Pascal's Triangle

The binomial distribution is derived by expanding $(p + q)^n$. This can be presented diagrammatically in the form of a triangle, which derives its name from the famous French mathematician and polymath Blaise Pascal.

Sample Size ($=n$)	Expansion of $(p+q)^n$	Number of Possible Combinations
1	$p \quad q$	2
2	$p^2 \quad 2pq \quad q^2$	3
3	$p^3 \quad 3p^2q \quad 3pq^2 \quad q^3$	4
4	$p^4 \quad 4p^3q \quad 6p^2q^2 \quad 4pq^3 \quad q^4$	5
5	$p^5 \quad 5p^4q \quad 10p^3q^2 \quad 10p^2q^3 \quad 5pq^4 \quad q^5$	6
n	$p^n \quad np^{n-1}q \quad \ldots \quad npq^{n-1} \quad q^n$	$n+1$

The number preceding each term in the triangle is known as the *coefficient of the expansion* and can be obtained easily by adding together the two coefficients lying on either side of it in the row above. This coefficient is equal to $\frac{n!}{(n-r)!r!}$ in the equation, which defines the binomial distribution. Thus, those who dislike algebra intensely can use Pascal's triangle rather than the full binomial distribution. For example, if parents who both carry the same autosomal recessive gene have five children, they might wish to know the probability that three will be affected. Using the binomial distribution, the answer will be

$$\frac{5!}{2!\quad 3!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 = \frac{90}{1024}$$

The same answer can be obtained by consulting Pascal's triangle and reading across the line $n=5$. The coefficient for $p^2\ q^3$ (i.e., 2 unaffected and 3 affected) is 10. Thus, the desired probability will be $10\times\left(\frac{3}{4}\right)^2\times\left(\frac{1}{4}\right)^3=\frac{90}{1024}$.

1.4 Bayes' Theorem

This theorem provides an extremely useful means of quantifying genetic risks (Murphy and Mutalik, 1969). It is derived from an essay on the *doctrine of chances* written by an eighteenth-century English clergyman, Thomas Bayes, first published posthumously by one of his friends in 1763 and republished in 1958 (Bayes, 1958). Essentially, it offers a method for considering all possibilities or events and then modifying the probabilities for each of these by incorporating information that sheds light on which is the most likely. The initial probability for each event, such as being a carrier or not being a carrier, is known as its *prior* probability and is based on "anterior" information such as the ancestral family history. The observations that modify the prior probabilities allow *conditional* probabilities to be derived using "posterior" information such as the results of carrier tests.

The resulting probability for each possibility or event is known as its *joint* probability and is calculated by multiplying the prior probability by the conditional probability for each observation (the observations should be independent in that they should not influence each other). The overall final probability for each event is known as its *posterior* or *relative* probability and is obtained by dividing its joint

probability by the sum of all of the joint probabilities. This has the effect of ensuring that the sum of all of the posterior probabilities always equals 1. Alternatively, posterior probabilities can be expressed in the form of odds for or against a particular event occurring or not occurring.

On first reading this can be very confusing, and the reader will not necessarily be helped by the following formal statement of Bayes' theorem:

1. If the prior probability of an event C occurring is denoted as $P(C)$ and
2. the prior probability of event C not occurring is denoted as $P(NC)$ and
3. the conditional probability of observation O occurring if C occurs equals $P(O|C)$ and
4. the conditional probability of observation O occurring if C does not occur equals $P(O|NC)$, then the overall probability of event C given that O is observed equals

$$\frac{P(C) \text{ multiplied by } P(O|C)}{[P(C) \text{ multiplied by } P(O|C)] + [P(NC) \text{ multiplied by } P(O|NC)]}$$

This may be a little clearer if a Bayesian table (Table 1.1) is constructed. The posterior probability of event C occurring equals

$$\frac{P(C) \times P(O|C)}{[P(C) \times P(O|C)] + [P(NC) \times P(O|NC)]}$$

The posterior probability of event C not occurring equals

$$\frac{P(NC) \times P(O|NC)}{[P(C) \times P(O|C)] + [P(NC) \times P(O|NC)]}$$

This is not nearly as complicated or difficult as it seems. If you are not convinced, then consider the following example.

Example 6

A woman, II2 in Figure 1.1, wishes to know the probability that she is a carrier of Duchenne muscular dystrophy. Her concern is based upon her family history, which reveals an affected brother and an affected maternal uncle. This is anterior information that enables the prior probability that she is a carrier $P(C)$ to be de-

Table 1.1.

Probability	Event C Occurs	Event C Does Not Occur
Prior	$P(C)$	$P(NC)$
Conditional O occurs	$P(O\|C)$	$P(O\|NC)$
Joint	$P(C) \times P(O\|C)$	$P(NC) \times P(O\|NC)$

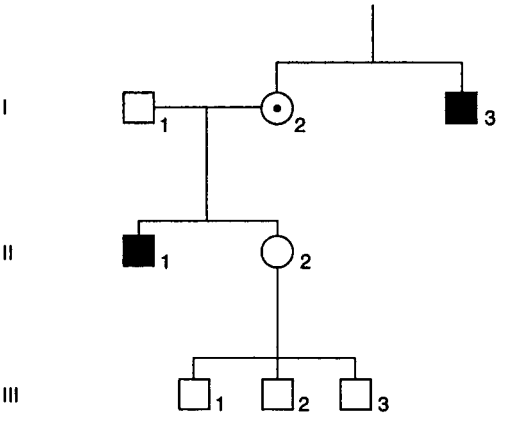

Figure 1.1. When calculating the probability that II2 is a carrier of Duchenne muscular dystrophy Bayes' theorem provides a method for taking into account the fact that she already has three unaffected sons.

termined. As her mother (I2) must be a carrier, there is a prior probability of 1/2 that II2 is a carrier and an equal prior probability of 1/2 that she is not a carrier.

Posterior information is provided by the fact that the consultand already has three unaffected sons. If the consultand is a carrier, then each of her sons will have a risk of 1 in 2 of being affected. Thus, $P(O|C)$ equals $1/2 \times 1/2 \times 1/2$ since the consultand has three unaffected sons, and $P(O|NC)$ equals $1 \times 1 \times 1$, since if the consultand is not a carrier, there is a probability of 1 ($1 - \mu$, to be exact, but μ—the mutation rate—can be ignored since it is less than 1/1000) that a son will be unaffected.

This information is used to construct a Bayesian table (Table 1.2). The posterior probability that the consultand is a carrier equals $1/16/(1/16 + 1/2)$, which equals 1/9. Alternatively, the posterior probability can be stated in the form of odds by indicating that there are 8 chances to 1 that the consultand is not a carrier.

Effectively, in this example Bayes' theorem has been used to quantify the intuitive recognition that the birth of three unaffected sons makes it rather unlikely

Table 1.2.

Probability	Consultand Is a Carrier		Consultand Is Not a Carrier
Prior	$\frac{1}{2}$		$\frac{1}{2}$
Conditional 3 unaffected sons	$\frac{1}{8}$		1
Joint	$\frac{1}{16}$		$\frac{1}{2}$
Odds	1	to	8
Posterior probability that consultand is a carrier $= \frac{\frac{1}{16}}{\frac{1}{16} \times \frac{1}{2}} = \frac{1}{9}$			
Posterior probability that consultand is not a carrier $= \frac{\frac{1}{2}}{\frac{1}{16} \times \frac{1}{2}} = \frac{8}{9}$			

that the consultand is a carrier. The greater the number of unaffected sons, then, the more likely it becomes that the consultand is not a carrier. Obviously, the birth of one affected son would totally negate the conditional probability contributed by unaffected sons by introducing conditional probabilities of 1/2 (carrier) versus μ (new mutation) in the noncarrier column. In other words, the birth of an affected son would make it overwhelmingly likely that the consultand is a carrier.

Key Point 4

Bayes' theorem provides a method for taking into account all relevant information when calculating the probability of an event such as carrier status. Key points to remember are:

1. A table should be drawn up that includes all relevant possibilities.
2. The prior probability for each possibility is derived from ancestral anterior information.
3. The conditional probabilities are obtained from posterior information that sheds light on which initial possibility is more or less likely. Conditional probabilities can be calculated by asking "What is the probability that this observation would be made given that the initial possibility or event occurs or applies?"
4. The joint probability for each possibility is calculated and then compared with the other joint probabilities to give a posterior or relative probability for each possibility or event.
5. All relevant information should be used once and only once.

The concepts introduced in this chapter, and in Bayes' theorem in particular, are not easy to grasp. However, with a little practice, even the most reluctant mathematician can become reasonably proficient at simple probability calculations. More testing examples are provided in the next four chapters. Readers who are still struggling with the underlying principles are invited to consult the review by Ogino and Wilson (2004), which provides a very clear explanation of how to apply Bayesian analysis.

1.5 Case Scenario

A woman, II2 in Figure 1.2, is referred from the antenatal clinic for genetic risk assessment. She is 20 weeks pregnant and is known to have two maternal uncles and a brother with severe learning disability. Investigations undertaken in the past, including Fragile X mutation analysis, have failed to identify a specific cause, prompting a clinical diagnosis of nonspecific X-linked mental retardation. Ultrasonography has revealed that this woman is carrying male twins (III2 and III3) of unknown zygosity. The woman specifically wishes to know the probability that one or both of her unborn sons will be affected.

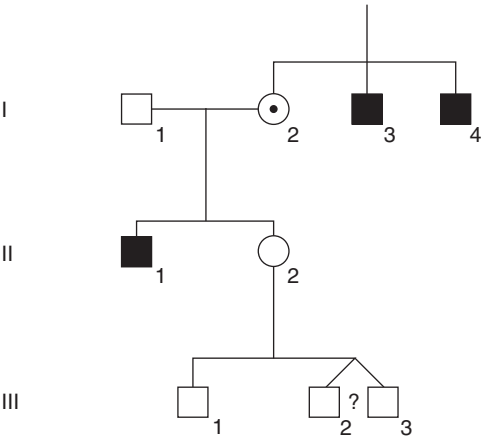

Figure 1.2. II2 is carrying male twins of unknown zygosity. What is the probability that none, one, or both will be affected?

To answer her question, we have to carry out three relatively easy calculations using two Bayesian tables and a simple application of the basic laws of probability.

1. The first Bayesian calculation (Table 1.3) indicates the probability that this woman is a carrier of the X-linked mental retardation that affects her uncles and brother. This yields a figure of 1 in 3, which is less than the ancestral prior pedigree risk of 1 in 2 because the woman already has an unaffected son (III1).
2. The second Bayesian calculation (Table 1.4) determines the probability that the male twins are either MZ or DZ, given that the ratio of MZ to DZ twins is 1:2. As shown in Table 1.4, half of all same-sex twins will be MZ and half will be DZ.

Table 1.3.*

Probability	II2 Is a Carrier		II2 Is Not a Carrier
Prior	$\frac{1}{2}$		$\frac{1}{2}$
Conditional			
1 unaffected son	$\frac{1}{2}$		1
Joint	$\frac{1}{4}$		$\frac{1}{2}$
Odds	1	to	2
Posterior probability that II2 is a carrier $= \frac{1}{4} \Big/ \left(\frac{1}{4} + \frac{1}{2}\right) = \frac{1}{3}$			
Posterior probability that II2 is not a carrier $= \frac{1}{2} \Big/ \left(\frac{1}{4} + \frac{1}{2}\right) = \frac{2}{3}$			

*See Figure 1.2.