

A Three Variable Disease Infection Probability Estimation Model

Oyeka I. C. A., Umeh E. U.*

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Abstract This paper proposes and presents a statistical method for the estimation of joint, marginal and conditional odds that a randomly selected susceptible in a population is infected with some diseases and the numbers of subjects infected assuming that no prior interventionist actions have been taken to influence disease outcomes. The proposed method is intended to help guide health officials make informed decisions on the use of medical resources in the presence of scarcity. Although, it is possible to hypothesis and fit some theoretical probability distribution models to data on disease prevalence rates, and use them in estimating desired odds of infection. The proposed method is however less tedious to apply in practice and the results obtained are relatively easier to interpret and explain. The method is illustrated with some data.

Keywords Variables, Disease Infections, Joint, Marginal and Conditional Odds, Probability Estimation, Models

1. Introduction

At any given point in time there may be many types of diseases existing in a population. A public health worker may in this situation wish to estimate the odds and sizes of the population likely to be infected by either individual diseases or clusters of these diseases assuming that the diseases follow their natural course of occurrence in the population without any prior interventionist remedial actions to guide their management [3-8].

In this paper, we propose a statistical method for estimating the odds that a randomly selected individual or subject from a population of interest is infected by an existing disease and the likely number of subjects in the population so infected, assuming the infections follow the natural course without medical interventions. The method also enables one estimate joint and conditional probabilities that randomly selected subjects or patients are infected and not infected by all or some diseases from three family types of diseases either simultaneously or infected and not infected by some diseases before or after some other diseases operating as predisposing or antecedent diseases in a population. This would enable policy makers and further research either to prove or disprove existing theories and methods. This approach clearly keeps research in a state of continuous flukes and advances knowledge and academics. This is the hope for approach of this present paper.

2. The Proposed Method

Suppose there are several diseases existing in a certain population. Although the proposed method may be used with the diseases treated as individual events, we will here for simplicity but without loss of generality assume that the diseases could be grouped into a finite number of clusters of diseases where the clusters are intra-homogeneous among and inter-heterogeneous between themselves. For example, the diseases may be grouped into cardio-vascular diseases, infectious diseases, congenital diseases, auto immune and nutritional disorder diseases, mental disorder and others. These clusters may be denoted with A, B, C etc. In particular, we will in this paper assume that research interest is in only three existing conditions, diseases, clusters or groups of diseases A, B, and C, for instance greatest epidemiological concern infecting subjects in a given population and also interest is in estimating the joint, marginal and conditional probabilities that a randomly selected subject from the population is infected or not infected with the diseases and the number of subjects so affected to help guide management decisions in the presence of limited medical resources [9-15].

Now a randomly selected individual or subject from this population of current susceptible population size N may suffer from none, one, some combinations of, or all of the diseases in groups A, B and C.

Let A, B and C be the possibility that a randomly selected subject or individual may be infected with disease A, B and C respectively. Let \bar{A} , \bar{B} and \bar{C} be the possibility that a randomly selected subject or individual may not be infected with disease A, B and C respectively. Let ABC be the

* Corresponding author:

editus2002@yahoo.com (Umeh E. U.)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

possibility that a randomly selected subject from the population is infected by all the diseases in the three disease groups. Let $\bar{A}\bar{B}\bar{C}$ be the possibility that a randomly selected subject from the population is not infected with any of the three diseases. Let $AB\bar{C}$ or $\bar{A}B\bar{C}$ or $\bar{A}\bar{B}C$ be the possibility that a randomly selected subject from the population is infected with only one of the three diseases. Let also $AB\bar{C}$ or $\bar{A}B\bar{C}$ or $\bar{A}\bar{B}C$ be the possibility that a randomly selected subject from the population is infected by only two of the three diseases.

Hence possible configurations of the occurrence of these diseases in a subject and in the population may be presented as in Table 1.

Table 1. Possible Configurations of Diseases in a Population

Disease Type A	Disease Type B	Disease Type C		Total
		C	\bar{C}	
A	B	ABC	$AB\bar{C}$	AB
	\bar{B}	$A\bar{B}C$	$A\bar{B}\bar{C}$	$A\bar{B}$
Total		AC	$A\bar{C}$	A
\bar{A}	B	$\bar{A}BC$	$\bar{A}B\bar{C}$	$\bar{A}B$
	\bar{B}	$\bar{A}\bar{B}C$	$\bar{A}\bar{B}\bar{C}$	$\bar{A}\bar{B}$
Total		$\bar{A}C$	$\bar{A}\bar{C}$	\bar{A}
	B	BC	$B\bar{C}$	B
	\bar{B}	$\bar{B}C$	$\bar{B}\bar{C}$	\bar{B}
Total of Grand Total		C	\bar{C}	S(Sample Space)

Of course there are many other possible configurations of the occurrence or non occurrence of these diseases in a population. For example, a randomly selected subject or individual may or may not be infected with disease B say given that he is already infected or not infected with disease A say; the subject may or may not be infected with diseases B and C, say given that he is already infected or not infected with disease A say; the subject may or may not be infected with disease A say, given that he is already infected or not infected with diseases B and C say; the subject may have been infected or not infected with disease B say given that he is free of disease A and not of disease C say; etc.

Research interest may then be in estimating the probabilities of all three and similar events and the number of persons in the population expected to be so infected, if all

things are equal. It is quite possible to hypothesize and fit some theoretical probability distributions such as the Gamma family or other families of distributions to the prevalence values or rates of the diseases or families of diseases of interest and use them to estimate any desired probabilities and the numbers of subjects so affected [1, 2]. This approach would however be rather tedious to use and the estimates so obtained difficult to interpret in practical applications. We will therefore adopt an alternative relatively simpler approach to the problem.

Now suppose that the prevalence rates or the probabilities of the occurrence of events A, B, and C in the population are known, that is suppose that the probabilities that a randomly selected subject from the population is infected with or suffers from any of the diseases in disease groups A, B, and C are known to be 'a', 'b' and 'c' respectively. Furthermore, suppose that it is also known that the relationship between the clusters of diseases and that $P(B/A) = d$; $P(C/A) = e$; $P(C/B) = f$; and $P(C/AB) = g$ for diseases in disease groups A, B and C. With this information, we may also easily calculate other similar conditional and joint probabilities. For example, the conditional probability that a randomly selected subject from the population is infected with disease A given that the subject already has disease B is

$$P(A/B) = \frac{p(B/A).P(A)}{P(B)} = \frac{ad}{b},$$

$$\text{Similarly } P(A/C) = \frac{P(C/A).P(A)}{P(C)} = \frac{ae}{c}$$

$$\text{and } P(B/C) = \frac{p(C/B).P(B)}{P(C)} = \frac{bf}{c},$$

$$\text{also } P(A/BC) = \frac{P(ABC)}{P(BC)} = \frac{P(C/AB).P(B/A).P(A)}{P(C/B).P(B)} = \frac{gad}{bf};$$

$$P(B/AC) = \frac{P(ABC)}{P(AC)} = \frac{P(ABC)}{P(C/A).P(A)} = \frac{gad}{ae} = \frac{gd}{e} \quad (1)$$

However of greater research interest may be in estimating the probabilities of the joint occurrence or non occurrence of all or some combinations of the three disease entities A, B and C in a population.

Now using these known probability values further, we then have that the probability that a randomly selected subject from the population is infected by all the disease in the three disease groups is $P(ABC) = P(C/AB).P(AB) = P(C/AB).P(B/A).P(A) = gad$

The probability that a randomly selected subject from the population is not infected with any of the three diseases, that is in disease condition $\bar{A}\bar{B}\bar{C}$ is

$$P(\bar{A}\bar{B}\bar{C}) = 1 - P(A \cup B \cup C) = 1 - [P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)] \quad (2)$$

which when evaluated gives

$$P(\bar{A}\bar{B}\bar{C}) = 1 - (a + b + c - ad - ae - bf + gad) \quad (3)$$

Other joint probabilities in Table 1 are similarly estimated and the results are presented in Table 2 below

Now the probability that a randomly selected subject is infected with only one of the three diseases; that is disease condition $AB\bar{C}$ or $\bar{A}B\bar{C}$ or $\bar{A}\bar{B}C$ is from Table 2

$$P(AB\bar{C}) + P(\bar{A}B\bar{C}) + P(\bar{A}\bar{B}C) = a + b + c + 3gad - 2(ad + ae + bf) \quad (4)$$

Table 2. Estimated probabilities of occurrence for the joint events in Table 1

Disease Type A	Disease Type B	Disease Type C		Total
		C	\bar{C}	
A	B	ABC gad	$AB\bar{C}$ $(1-g).ad$	AB Ad
	\bar{B}	$A\bar{B}C$ $a(e-gd)$	$A\bar{B}\bar{C}$ $a + gad - ad - ae$	$A\bar{B}$ a-ad
Total		AC ae	$A\bar{C}$ $a - ae$	A A
\bar{A}	B	$\bar{A}BC$ bf-gad	$\bar{A}B\bar{C}$ $b + gad - ad - bf$	$\bar{A}B$ b-ad
	\bar{B}	$\bar{A}\bar{B}C$ $c + gad - ae - bf$	$\bar{A}\bar{B}\bar{C}$ $1 - (a + b + c - ad - ea - bf + gad)$	$\bar{A}\bar{B}$ $1 - a - b + ad$
Total		$\bar{A}C$ $c - ae$	$\bar{A}\bar{C}$ $1 - a - c + ae$	\bar{A} $1 - a$
	B	BC bf	$B\bar{C}$ $b - bf$	B b
	\bar{B}	$\bar{B}C$ $c - bf$	$\bar{B}\bar{C}$ $1 - b - c + bf$	\bar{B} $1 - b$
Total of Totals		C c	\bar{C} $1 - c$	1

Note however that Equation 4 is also easily shown to reduce to

$$P(\bar{A}\bar{B}) + P(\bar{A}\bar{C}) + P(\bar{B}\bar{C}) - 3P(\bar{A}\bar{B}\bar{C}),$$

where

$$\begin{aligned} P(\bar{A}\bar{B}) &= (1 - b) - (1 - d).a; P(\bar{A}\bar{C}) = (1 - c) - (1 - e).a; \\ P(\bar{B}\bar{C}) &= (1 - c) - (1 - f).b \end{aligned} \quad (5)$$

and

$P(\bar{A}\bar{B}\bar{C})$ is given in Equation 3.

The probability that a randomly selected subject from the population is affected by only two of the three diseases; that is the probability of the events $AB\bar{C}$, $A\bar{B}C$, and $\bar{A}BC$ may be similarly obtained as the sum of the values of the probabilities of these events in Table 2 as

$$\begin{aligned} P(AB\bar{C}) + P(A\bar{B}C) + P(\bar{A}BC) &= (1 - g)ad + a(e - gd) + (bf - gad) \\ &= ad + ae + bf - 3(gad) \end{aligned} \quad (6)$$

Furthermore, the conditional probability that a randomly selected subject from the population is infected by disease A say, given that the subject has been treated or has natural or acquired immunity against disease entities B and C say is the probability of the event $A/\bar{B}\bar{C}$, that is

$$\begin{aligned} P(A/\bar{B}\bar{C}) &= (1 - P(\bar{A}/\bar{B}\bar{C})) = 1 - \frac{P(\bar{A}\bar{B}\bar{C})}{P(\bar{B}\bar{C})} = \frac{P(\bar{B}\bar{C}) - P(\bar{A}\bar{B}\bar{C})}{P(\bar{B}\bar{C})} \\ &= \frac{(1 - b - c + bf) - 1 - (a - b + c - ad - ae - bf + gad)}{1 - b - c + bf} \\ &= \frac{a + gad - ad - ae}{1 - b - c + bf} \end{aligned} \quad (7)$$

Hence the conditional probability that a randomly selected subject from the population is infected by any one of the three disease entities or groups given that the subject has been treated or has immunity against the other two disease groups is obtained as the probability of the events $A/\bar{B}\bar{C}$, $B/\bar{A}\bar{C}$, and

$C/\bar{A}\bar{B}$ which is calculated as

$$P\left(\frac{A}{\bar{B}\bar{C}}\right) + P\left(\frac{B}{\bar{A}\bar{C}}\right) + P\left(\frac{C}{\bar{A}\bar{B}}\right) = \frac{a + gad - ad - ae}{1 - b - c + bf} + \frac{b + gad - ad - bf}{1 - a - c + ae} + \frac{c + gad - ae - bf}{1 - a - b + ad} \quad (8)$$

The probability that a randomly selected subject is infected by any two of the family of diseases B and C say given that the subject has been cured or has immunity against the other family of disease A say is the probability of the event BC/\bar{A} which is estimated as

$$\begin{aligned}
 P\left(\frac{BC}{\bar{A}}\right) &= \frac{P(\bar{A}BC)}{P(\bar{A})} = \frac{P\left(\frac{\bar{A}}{BC}\right) \cdot P(BC)}{P(\bar{A})} \\
 &= \frac{\left(1 - P\left(\frac{A}{BC}\right) \cdot P(BC)\right)}{P(\bar{A})} = \frac{\left(1 - \frac{P(ABC)}{P(BC)}\right) \cdot P(BC)}{P(\bar{A})} \\
 &= \frac{P(BC - P(ABC))}{P(\bar{A})} = \frac{bf - gad}{1 - a}
 \end{aligned}$$

Hence the probability that a randomly selected subject is infected by any two of the family of diseases given that the subject has been cured or has immunity against the other family of diseases is the probability of the events BC/\bar{A} , AC/\bar{B} and AB/\bar{C} which is estimated as the probability

$$P\left(\frac{BC}{\bar{A}}\right) + P\left(\frac{AC}{\bar{B}}\right) + P\left(\frac{AB}{\bar{C}}\right) = \frac{bf - gad}{1 - a} + \frac{ae - gad}{1 - b} + \frac{ad - gad}{1 - c} \quad (9)$$

The probability that a randomly selected subject from the population is currently infected with disease C say given that the subject is known to be free from disease A but not free from disease B say is the probability of the event $P(C/\bar{A}\bar{B})$ which is given as

$$P\left(\frac{C}{\bar{A}\bar{B}}\right) = \frac{P(\bar{A}\bar{B}C)}{P(\bar{A}\bar{B})} = \frac{P(\bar{A}\bar{B}C)}{P(B) - P(AB)} = \frac{P(\bar{A}\bar{B}C)}{P(B) - P(B/A)P(A)} = \frac{bf - gad}{b - ad}$$

Hence the probability that a randomly selected subject from the population is currently infected with one of the diseases in three families of diseases given that the subject had earlier been infected with diseases in one of the remaining families is the probability of the events

$C/\bar{A}\bar{B}$, $C/\bar{A}B$, $B/\bar{A}\bar{C}$, $B/\bar{A}C$, $A/\bar{B}\bar{C}$, and $A/\bar{B}C$ which is calculated as

$$\begin{aligned}
 &P(C/\bar{A}\bar{B}) + P(C/\bar{A}B) + P(B/\bar{A}\bar{C}) + P(B/\bar{A}C) + P(A/\bar{B}\bar{C}) + P(A/\bar{B}C) \\
 &= \frac{e - gd}{1 - d} + \frac{bf - gad}{b - ad} + \frac{d(1 - g)}{1 - e} + \frac{bf - gad}{c - ae} + \frac{ad(1 - g)}{b(1 - f)} + \frac{a(e - gd)}{c - bf}
 \end{aligned} \quad (10)$$

The probability that a randomly selected subject from the population is currently not infected with diseases in any one of the three families of diseases given that the subject had earlier been infected with diseases in one of the two remaining disease families is of course the complement of the estimated probability given in Equation 10. Also the probability that a randomly selected subject from the population who is known to have in the recent past not been affected with A and B say of the three families of the diseases is currently not free from infection with the remaining one C say is the probability of the event $\bar{C}/\bar{A}\bar{B}$, $\bar{C}/\bar{A}B$, $\bar{C}/\bar{A}\bar{C}$, $\bar{B}/\bar{A}\bar{C}$, $\bar{B}/\bar{A}C$, $\bar{A}/\bar{B}\bar{C}$, $\bar{A}/\bar{B}C$, $\bar{A}/\bar{B}\bar{C}$

The probability is easily obtained using the estimates in Table 2 and Equation 5 as

$$\begin{aligned}
 &\frac{P(\bar{A}\bar{B}\bar{C})}{P(\bar{A}\bar{B})} + \frac{P(\bar{A}B\bar{C})}{P(\bar{A}B)} + \frac{P(\bar{A}\bar{B}C)}{P(\bar{A}\bar{B})} + \frac{P(A\bar{B}\bar{C})}{P(A\bar{C})} + \frac{P(\bar{A}\bar{B}C)}{P(\bar{A}C)} + \frac{P(\bar{A}\bar{B}\bar{C})}{P(\bar{A}\bar{C})} + \frac{P(\bar{A}B\bar{C})}{P(B\bar{C})} + \frac{P(\bar{A}\bar{B}C)}{P(\bar{B}C)} + \frac{P(\bar{A}\bar{B}\bar{C})}{P(\bar{B}\bar{C})} \\
 &= \frac{a + gad - ad - ae}{a(1 - d)} + \frac{b + gad - ad - bf}{b - ad} + \frac{1 - (a + b + c - ad - ae - bf + gad)}{1 - a - b + ad} + \frac{a + gad - ad - ae}{a(1 - e)} + \frac{c + gad - ae - bf}{c - ae} \\
 &+ \frac{1 - (a + b + c - ad - ae - bf + gad)}{1 - a - c + ae} + \frac{b + gad - ad - bf}{b(1 - f)} + \frac{c + gad - ae - bf}{c - bf} + \frac{1 - (a + b + c - ad - ae - bf + gad)}{1 - b - c + bf}
 \end{aligned} \quad (11)$$

The probability that a randomly selected subject who is already infected by any two of the diseases is currently free from infection with the third disease is the probability of the events AB/\bar{C} , AC/\bar{B} , and BC/\bar{A} which is given as

$P(AB/\bar{C}) * P(AC/\bar{B}) * P(BC/\bar{A})$ or equivalently using the result of Table 2

$$\frac{P(\bar{A}\bar{B}C)}{P(\bar{A})} + \frac{P(\bar{A}B\bar{C})}{P(\bar{B})} + \frac{P(AB\bar{C})}{P(\bar{C})} = \frac{bf - gad}{1 - a} + \frac{ae - gad}{1 - b} + \frac{ad - gad}{1 - c} \quad (12)$$

Similarly the probability that a randomly selected subject who is known to be free from infection with any two of the three diseases is currently infected with only one of them is obtained as

$$P(C/\bar{A}\bar{B}) + P(B/\bar{A}\bar{C}) + P(A/\bar{B}\bar{C}) = \frac{c + gad - ae - bf}{1 - a - b + ad} + \frac{b + gad - ad - bf}{1 - a - c + ae} + \frac{a + gad - ad - ae}{1 - b - c + bf} \quad (13)$$

Also the probability that a randomly selected subject from the population who is known to have in the recent past been infected with any two say A and B of the three families of the diseases is currently free from infection with the remaining one C say is the probability of the event \bar{C}/AB which is calculated using the value in Table 2 as

$$P(\bar{C}/AB) = \frac{P(AB\bar{C})}{P(AB)} = \frac{P(AB\bar{C})}{P(B/A) \cdot P(A)} = \frac{ad - gad}{ad} \quad (14)$$

A researcher may of course be interested in estimating the odds of infection with only two of the three diseases identified in a population. For example, interest may be in estimating the probabilities that a randomly selected subject from the population is infected or not infected simultaneously with diseases A and B say; the probabilities that the subject is infected with one of the two diseases A and not with the other B say or vice versa or the probabilities that a subject who is known not to have been infected with disease B is currently infected with disease A or the probability that the subject who is known not to have been infected with disease B is currently also free from infection with disease A, say; etc.

These are the probabilities of the events $AB, \bar{A}\bar{B}, A\bar{B}, \bar{A}B, A/\bar{B}, \bar{A}/\bar{B}$, etc. The probabilities can easily be obtained from Table 2 as the sum of the joint probabilities of occurrence or non occurrence of events or diseases A and B summed over events C and \bar{C} , that is the marginal probabilities of occurrence or non occurrence of the compound event AB, say. Thus

$$\begin{aligned} P(AB) &= ad = P(ABC) + P(AB\bar{C}); \\ P(\bar{A}\bar{B}) &= 1 - a - b + ad = P(\bar{A}\bar{B}C) + P(\bar{A}\bar{B}\bar{C}) \end{aligned} \quad (15)$$

similarly

$$\begin{aligned} P(A\bar{B}) &= a - ad = P(A\bar{B}C) + P(A\bar{B}\bar{C}); \\ P(\bar{A}B) &= b - ad = P(\bar{A}BC) + P(\bar{A}B\bar{C}) \end{aligned} \quad (16)$$

The conditional probabilities are similarly estimated. For example

$$P(A/\bar{B}) = \frac{a-ad}{1-b} = \frac{P(A\bar{B}C)+P(A\bar{B}\bar{C})}{1-P(B)} \quad (17)$$

and

$$P\left(\frac{\bar{A}}{\bar{B}}\right) = \frac{1-a-b+ad}{1-b} = \frac{P(\bar{A}\bar{B}C)+P(\bar{A}\bar{B}\bar{C})}{1-P(B)} \quad (18)$$

which are the weighted sums of the corresponding marginal probabilities of occurrence of events $A\bar{B}$ and $\bar{A}\bar{B}$, respectively; etc.

Having obtained estimates of the probabilities or odds of infection in the population, the researcher may then also wish to estimate the number of the susceptible population of interest that would be expected to be in the various disease infection classification levels. To do this, assuming that the total number of susceptible in the population is 'N', then the researcher would only need to apply 'N' to each of the estimated probabilities to obtain the desired number of subjects likely to be infected in each disease classification level; assuming all other factors affecting such diseases in the population have been partialled out.

3. Illustrative Example

The proposed method is ideally used especially where the prevalence rates of conditions such as diseases in a population are known and some of the interrelationships between the diseases are also known or can be determined. As already noted above, it is of course also possible to assign and fit some probability distributions to the diseases or families of diseases and then estimate desired marginal, joint, and conditional probabilities of occurrence of these diseases. This approach would however be rather tedious in most applications.

We will here therefore for illustrative purposes only, assume that the diseases or families of diseases of interest A, B, and C can without loss of generality be assigned probability of occurrence in a population and that some of the conditional probabilities of occurrence of the diseases are also known or can be determined in a population with 'N' equal 1,000,000 subjects susceptible to the diseases. Specifically, suppose it is known that

$$\begin{aligned} P(A) &= a = 0.12; P(B) = b = 0.10; P(C) = c = 0.08; P(B/A) = d = 0.06; \\ P(C/A) &= e = 0.05; P(C/B) = f = 0.03 \text{ and } P(C/AB) = g = 0.01 \end{aligned}$$

Now using this information, we would estimate that the probability that a randomly selected subject from the population is infected simultaneously by the three diseases A, B, and C is

$$P_r = P(ABC) = gad = (0.01)(0.12)(0.06) = 0.000072$$

The corresponding number of subjects in the hypothetical population expected to be infected simultaneously by the three diseases is therefore

$$N_1 = N_{p1} = (1,000,000)(0.000072) = 72 \text{ subjects}$$

The probability that a randomly selected subject from the population is not infected simultaneously by the three diseases is

$$\begin{aligned}
 p_8 &= P(\bar{A}\bar{B}\bar{C}) = 1 - (a + b + c - ad - ae - bf + gad) \\
 &= 1 - (0.12 + 0.10 + 0.08 - (0.12)(0.06) - (0.12)(0.05) - (0.10)(0.03) + (0.01)(0.12)(0.06)) \\
 &= 1 - 0.286872 = 0.713128
 \end{aligned}$$

Hence the number of subjects in the population expected not to be infected by the three diseases simultaneously is

$$N_8 = N.p_8 = (1,000,000)(0.713128) = 713128 \text{ subjects}$$

Other estimated joint probabilities of infection shown in Table 2 and the corresponding number of subjects in the population expected to be so infected are similarly calculated and the results are shown in Table 3, where the number of infected subjects are shown in parenthesis.

Table 3. Estimated Joint Probabilities of Infection and Numbers of Infected Subjects for Table 2 under the assumed Disease Prevalence Rates in a Population

Disease Type A	Disease Type B	Disease Type C		Total
		C	\bar{C}	
A	B	0.000072	0.007128	0.0072 (1)
		(72)	(7128)	(7200)
	\bar{B}	0.005928	0.106872	0.1128 (2)
		(5928)	(106872)	(112800)
Total		0.006	0.114	0.12 (3)
		(6000)	(114000)	(120000)
\bar{A}	B	0.002928	0.089872	0.0928 (4)
		(2928)	(89872)	(92800)
	\bar{B}	0.071072	0.713128	0.7872 (5)
		(71072)	(713128)	(787200)
Total		0.074	0.806	0.88(6)
		(74000)	(806000)	(880000)
	B	0.003	0.097	0.10(7)
		(3000)	(97000)	100,000
	\bar{B}	0.077	0.823	0.900 (8)
		(77000)	(823000)	(900,000)
Total of Totals		0.08	0.92	1.00
		(80000)	(920000)	(1000000)

Now the probability that a randomly selected subject is infected with only one of the three diseases, from Tables 2, and 3

$$\begin{aligned}
 a + b + c + 3gad - 2(ad + ae + bf) &= 0.30 + 3(0.000072) - 2(0.0072 + 0.006 + 0.003) \\
 &= 0.300216 - 2(0.0162) = 0.267816
 \end{aligned}$$

With the estimated number of infected subjects given as

$$1000,000(0.267816) = 267816$$

The probability that a randomly selected subjects from the population is affected by only two of the three diseases is similarly obtained from Tables 2 and 3 as

$$\begin{aligned}
 ad - gad + ae - gad + bf - gad &= ad + ae + bf - 3gad \\
 &= 0.0072 + 0.006 + 0.003 - 3(0.000072) = 0.0162 - 0.000216 = 0.015984
 \end{aligned}$$

With the estimated number of subjects so infected in the population as

$$1000,000(0.015984) = 15984$$

Also the probability that a randomly selected subject from the population who is known to have in the recent past not been infected with two A and B say of the three families of the diseases is currently not free from infection with any of the remaining disease C say is obtained using the entries in Table 3 in Equation 11 as

$$P(C/\bar{A}\bar{B}) = \frac{0.08 + 0.000072 - 0.006 - 0.003}{1 - 0.12 - 0.10 + 0.0072} = \frac{0.071072}{0.7872} = 0.090285$$

And the expected number of subjects so affected is 90285

We have also estimated using Equation 12 and the values in Table 3 that a randomly selected subject who is already infected by any two of the diseases is currently free from infection with the third disease is

$$\begin{aligned} & \frac{0.003 - 0.000072}{1 - 0.12} + \frac{0.006 - 0.000072}{1 - 0.10} + \frac{0.0072 - 0.000072}{1 - 0.08} \\ &= 0.003327 + 0.006587 + 0.007748 = 0.017662 \end{aligned}$$

The corresponding number of subjects in the population so infected is

$$3327 + 6587 + 7748 = 17662$$

Use of Equation 13 and Table 3 enables the calculation of the probability that a randomly selected subject who is known to be free from infection with any two of the three diseases is currently infected with only one of them as

$$\begin{aligned} & \frac{0.08 + 0.000072 - 0.006 - 0.003}{1 - 0.12 - 0.10 + 0.0072} + \frac{0.10 + 0.000072 - 0.0072 - 0.003}{1 - 0.12 - 0.08 + 0.006} + \frac{0.12 + 0.000072 - 0.0072 - 0.006}{1 - 0.10 - 0.08 + 0.003} \\ &= 0.090285 + 0.111504 + 0.129857 = 0.331646. \end{aligned}$$

The corresponding number of affected subjects is

$$90285 + 111504 + 129857 = 331646$$

Also from Equation 14 and Table 3, we calculate that the probability that a randomly selected subject from the population who is known to have in the recent past been infected with any two A and B say of the three families of the diseases is currently free from infection with the remaining one C say is

$$P(\bar{C}/AB) = \frac{0.0072 - 0.000072}{0.0072} = \frac{0.007128}{0.0072} = 0.990$$

with 990000 subjects so affected.

Other odds of infection of interest and the corresponding number of subjects so affected in the population may be similarly estimated.

prevalence rates, the proposed method is illustrated with assumed prevalence rates of three diseases that may exist in a population.

4. Summary and Conclusions

In this paper, we have developed and presented a statistical method for the estimation of the joint, marginal and conditional probabilities that a randomly selected subject from a given population would be infected or not infected with two or three diseases or families of diseases either simultaneously or in progression as well as the expected numbers of subjects so infected or not infected in the population assuming no prior remedial actions are taken to influence disease outcome.

Although an alternative approach to this problem could be fitting of some hypothesized probability distributions to available data on disease prevalence in a population and then estimating desired probabilities, the proposed method is rather less tedious to apply in practice and the results relatively easier to interpret and explain.

In the absence of readily available data on disease

REFERENCES

- [1] Andersons, R. M ed (1982): Population dynamic of Infectious Diseases: Theory and Applications. Chapman and Hall, London. New York.
- [2] Andersons, R. M and May, R.M (1991): Infectious Diseases of Humans: Oxford. Oxford Uni. Press.
- [3] Bernoulli, D and Blower, S. (2004): An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. Review in Medical Virology, 14, 225-288.
- [4] Blower, S.M, Mclean, A.R., Panco, T.C; Small, P.M; Hopewell, P.C; Sanchez, M.A; et al (1995): The intrinsic transmission dynamics of tuberculosis epidemics. "Nature Medicine", 815-821.
- [5] Brauer, F and Castillo-Chavez, C. (2001): Mathematical models in Population Biology and Epidemiology, N.Y.

Springer.

- [6] Commun. Statist. TheoryMeth.A9(17), 1749-1874.
- [7] Conover, W.J (1980): Practical Non-Parametric Statistics 2nd ed. John Wiley and Sons Inc, New York, N.Y.
- [8] Daley, D.T and Gani, J. (2005): Epidemic modeling; An Introduction. N. Y. Cambridge Uni. Press.
- [9] Hastie, J.J, Tibshirani, R.J. (1987): Generalised Additive Models. Some Applications. J. Am. Stat. Assoc. , 82, 371-86.
- [10] Hastie, JJ, Tibshirani, RJ. (1990): Generalised Additive models. London, UK Chapman and Hall ltd.
- [11] Heshcote, H.W (2000): The mathematics of infectious diseases. Society for industrial and applied mathematics, 42, 599-653.
- [12] Imen, R.L. and Conover, W.J (1982): Small sample sensitivity analysis techniques for computer models with an application to risk assessment.
- [13] Kendall, M and Stewart, A. (1979): The advance theory of statistics, Vol 11. Hafner publishing company, New York.
- [14] Trotier, H.C and Philippe, P. (2001): Deterministic modeling of infectious diseases theory and methods. The Internet Journal of Infectious Disease.
- [15] Zhao LP, Kristal AR, White E. (1996) Estimating relative risk functions in case-control studies using a non parametric logistic regression. Am J. Epidemiology 144, 598-609.