EXCELLENT PUBLISHERS

International Journal of Current Research and Academic Review

ISSN: 2347-3215 Volume 3 Number 12 (December-2015) pp. 114-123

www.ijcrar.com

Estimates of Probabilities and odds of Infections in two disease infectious Probabality Model

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

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Corresponding author

KEYWORDS

Conditional Probability, Infection, Odds, Sample

A B S T R A C T

The present study represents a statistical method for the estimation of probabilities, conditional probabilities and odds and odds ratios of infections or non infections of randomly selected subjects by two groups or class types of diseases which may be independent or interacting operating in a population assuming no interventionist control measures are applied. The proposed method is appropriate for use with data obtained from diagnostic screening tests or clinical trials. Test statistics are presented to test some null hypotheses thus; the null hypothesis that the probability that a randomly selected subject is equally likely to be in any of the four mutually exclusive disease conditions; the null hypothesis that the probability of infection by one class type of disease is some specified values including zero which indicates absence of disease in the population; etc. In the event of the rejection of the null hypotheses, estimates of odds and odds-ratios of infection in a two disease infection model are provided. In addition, estimates of probabilities, odds and odds-ratios of infections are also provided in such instances where some diseases maybe the predisposing cause or the anteceded to some other subsequent diseases. Estimates of conditional probabilities, odds and odds-ratios of infection and non-infection of subjects by a disease or diseases of one class type given infection and non-infection by some disease or diseases of another disease class type are also provided. The proposed method is also illustrated with data.

Introduction

At any given point in time, there may be many types of diseases existing in a population. A public health worker may in these situations wish to estimate the probabilities, odds, and odds-ratios that a randomly selected subject from the population is infected and not infected by some disease or combinations of the existing diseases, assuming that the diseases follow their natural course, without some medical intervention (Fulton et al, 2012). We here present a more simplified and easier to use method assuming the existence of only two broad classes of diseases jointly operating in a population. Estimates of conditional probabilities, odds and odds-ratios of infection and non-infection by one type of disease conditional on infection and non-infection by the other class type disease are provided. Test statistics are also provided for testing of any desired hypothesis concerning the probabilities or rates of infection (see James,2001; Gelman et al, 2004 and Lisa, 2010).

The Proposed Method

Suppose a researcher collects a random sample of size n from a certain population. Research interest is to determine whether each subject in the sampled population is infected or not infected with two diseases from n classes of diseases. Specifically, interest is to determine whether or not a randomly selected subject from the population is infected with some disease from within a class of diseases and whether or not the same subject is also infected with some other disease from among a second class of diseases, where these classes of diseases may be mutually dependent and interacting.

Let A and $\bar{A}$ be respectively the events that a randomly selected subject from the sampled population is infected and not infected with some disease from one class of diseases. Similarly, let B and $\bar{B}$ be respectively the events that the randomly selected subject is infected and not infected with some other disease from a second class of diseases. Then, a randomly selected subject from the population would be in any one of the four mutually exclusive disease conditions, A.B, A. $\bar{B}$, $\bar{A}$. B, or $\bar{A}$..$\bar{B}$. (i.e The subject may be infected by both diseases A and B, that is in disease condition AB; the subject may be free of disease A but not of disease B, that is in disease condition $A^{-}B$; the subject may not be free of disease A but of disease B, that is in disease condition AB^{-}; or the subject may be free of both diseases A and B, that is in disease condition $A^{-}B^{-}$).

To estimate the probabilities and odds of occurrence of these disease conditions and others, where diseases A and B may perhaps be the most prevalent or most violent diseases or otherwise in their respective classes of diseases, we may let,

$$uij = \begin{cases} 1, & \text{if the ith subject from the population is in disease condition } j \\ 0 & \text{otherwise} \end{cases} \quad \text{....1}$$

$i = 1,2 \ldots\ldots\ldots, n;$ $j = 1$ (in disease condition AB), j = 2(in disease condition AB^{-}), j =3 (in disease condition A^{-}B) or j = 4(in disease condition $A^{-}B^{-}$)

Let

$$\pi_j = P(uij = 1) \quad \text{..................2}$$

And

$$W_j = \sum_{i=1}^{n} uij \quad \text{..................3}$$

Now the mean or expected value and variance of uij are respectively

$$E(uij) = \pi_j,\ Var(uij) = \pi_j(1-\pi_j) \quad \text{...................4}$$

Similarly, the expected value and variance of W_j are respectively.

$$E(Wj) = \sum_{i=1}^{n} E(uij) = n.\pi_j$$

$$Var(Wj) = \sum_{i=1}^{n} Var(uij) = n.\pi_j(1-\pi_j) \quad \text{...............5}$$

Now π_j is the proportion or the probability that a randomly selected subject from the sample population is in disease condition j , for some

$j = 1$ (*in disease condition AB*),
2 (*in disease condition* AB^{-})
; 3 (*in disease condition* $A^{-}B$); or
4 (*in disease condition* $A^{-}B^{-}$).

Its sample estimate is

$$\hat{\pi}_j = P_j = \frac{f_j}{n} = \frac{W_j}{n} \qquad ......6$$

where $f_j = W_j$ is the total number of subjects in the sample that are in disease condition j. In other words, $f_j = W_j$ *is the* total number of 1"s in the frequency distribution of the n values of 0"s and 1"s in uij,for $i = 1,2,\ldots,n$; and for some $j = 1$ (*in disease condition AB*), 2 (*in disease condition* AB^-) ; 3 (*in disease condition* A^-B); or 4 (*in disease condition* A^-B^-)

The estimated variance of $\hat{\pi}_j$ is from equation (5)

$$Var(\hat{\pi}_j) = Var\frac{(W_j)}{n^2} = \frac{\hat{\pi}_j(1-\hat{\pi}_j)}{n} \qquad ....7$$

The estimated proportions or probabilities and the corresponding frequencies may be presented in a four - fold table as in Table 1

Present (B)	Absent ($B^\wedge -$)	Total ($\hat{\pi}_{j.}$)
Present $P_1(f_1)$ (A)	$P_2(f_2)$	$P_1(f_1 = f_1 + f_2)$
Absent $P_3(f_3)$ $\bar{A}$	$P_4(f_4)$	$P_2(f_2 = f_3 + f_4)$
Total $P_{.1}(f_1 = f_1$ ($\hat{\pi}_{.j} = P_{.j}$)	$P_{.2}(f_2 = f_2$	1.00 ($n = f_2 + f_4 + \_f_3$

With these results, one may proceed to test some null hypothesis about π_j, the probability that a randomly selected subject from the sample population is in some disease condition j, for $j = 1$ (*in disease condition AB*), 2 (*in disease condition* AB^-) ; 3 (*in disease condition* A^-B); or 4 (*in disease condition* A^-B^-).

One may also wish to test the null hypothesis that a randomly selected subject is equally likely to be in any of the four disease conditions. This is equivalent to testing the null hypothesis

$$H_o = \pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi \quad vs \ H_1 \neq \pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi \qquad ....8$$

for some $j = 1, 2, 3$ *or* 4 where π is the common value of π_j ;s under a null hypothesis (H_o)

The null hypothesis H_0 of equation 8 maybe tested using the chi-square using the test statistic for independent in a four-fold table, using the frequencies in Table 1, as,

$$\chi^2 = \frac{n(P_1 P_4 - P_2 P_3)^2}{P_{..}P_{.1}P_{.2}P_{..2}} = \frac{n(f_1 - f_4 - f_2 - f_3)^2}{f_1 f_{.1} f_2 f_{.2}} \qquad ....9$$

which under H_0 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large n.

The null hypothesis H_0 of equation 8 is rejected at the α level of significance if

$$\chi^2 \geq \chi^2_{1-\alpha/2;\,1} \qquad ....10$$

otherwise the null hypothesis H_0 is accepted

One may also wish to test the null hypothesis that $\pi_j$is at most some value π_{j0}, for some $j = 1$ (*in disease condition AB*), 2 (*in disease condition* AB^-) ; 3 (*in disease condition* A^-B); or 4 (*in disease condition* A^-B^-)

That is the null hypothesis

$$H_o = \pi_j \geq \pi_{j0}\, vs\, H_1 = \pi_j < \pi_{j0}\, (0 \leq \pi_{j0} \leq 1) \quad .....11$$

For some $j = 1, 2, 3, or 4$

The null hypothesis H_o of equation 11 is tested using the chi-square test statistic

$$\chi^2 = \frac{\left(W_j - n.\pi_{j0}\right)^2}{Var(W_j)} = \frac{n\left(\hat{\pi}_j - \pi_{j0}\right)^2}{\hat{\pi}_j(1-\hat{\pi}_j)} = \frac{n\left(P_j - \pi_{j0}\right)^2}{P_j(1-P_j)} \quad ....12$$

For $j = 1, 2, 3, 4$

The null hypothesis H_o of equation 11 is rejected at the α level of significance if the chi-square value of equation 12 satisfies equation 10, otherwise H_o is accepted.

If the null hypotheses H_o are rejected or otherwise, one may then also proceed to estimate the odds and odds-ratios of infection in a two disease infection model.
In particular, sometimes some diseases may be the predisposing cause or the antecedent to some other subsequent diseases. For example some diseases of one class type, disease group type A say, may be the predisposing cause and expose subjects to some disease in another class of diseases, disease group type B say, or vice –versa.

This pattern of relationship would make it often desirable to estimate conditional probabilities and odds of infection and non-infection of subjects by a disease or diseases of one class type given infection and non-infection by some disease or diseases of another disease class type.

Now assuming that disease in disease group type A also pre-dispose subjects to infections by some disease group type B, say , then the odds that a randomly selected subject in a population is infected by some diseases in disease group type B in the presence of some disease or diseases on disease group type A say is

$$\frac{\pi_1}{\pi_2} = \frac{P(AB)}{P(\bar{B}A)} = \frac{P(B/A)}{P(\bar{B}/A)} \quad ....13$$

Whose sample estimate is

$$\frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{P_1}{P_2} = \frac{f_1}{f_2} \quad ....14$$

Similarly, the odds that a randomly selected subject in a population infected by some diseases in disease group type B in the absence of some diseases in group type A say, is

$$\frac{\pi_2}{\pi_4} = \frac{P(\bar{A}B)}{P(\bar{B}\bar{A})} = \frac{P(B/A)}{P(\bar{B}/\bar{A})} \quad ....15$$

Whose sample estimate is

$$\frac{\hat{\pi}_2}{\hat{\pi}_4} = \frac{P_2}{P_4} = \frac{f_2}{f_4} \quad ....16$$

Hence the odds that a randomly selected subject from the sampled population is infected by some disease in disease group type B in the presence as well as in the absence of the same subject being infected by some disease in disease group type A, or the so called odd-ratio of infection is

$$\frac{\Omega_A}{\Omega_{\bar{A}}} = \frac{\pi_1/\pi_2}{\pi_3/\pi_4} = \frac{P(AB)/P(A\bar{B})}{P(\bar{A}B)/P(AB)} = \frac{P(A/B).P(\bar{B}/A)}{P(B/\bar{A}).P(\bar{B}/\bar{A})} \quad ....17$$

Whose sample estimate is

$$\widehat{W} = 0 = \frac{\hat{\pi}_1\hat{\pi}_4}{\hat{\pi}_2\hat{\pi}_3} = \frac{P_1 P_4}{P_2 P_3} = \frac{f_1 f_4}{f_2 f_3} \quad ....18$$

The conditional probabilities that a randomly selected subject from the sampled population is infected and not infected by some disease in disease group type B given that the same subject is already infected by some disease in disease group type A are easily estimated as respectively

$$\frac{P(B/A)}{P(A)} = P(AB) = \frac{\hat{\pi}_1}{\hat{\pi}_1+\hat{\pi}_2} = \frac{f_1}{f_1+f_2};$$
$$\frac{P(\bar{B}/A)}{P(A)} = P(A\bar{B}) = \frac{\hat{\pi}_2}{\hat{\pi}_1+\hat{\pi}_2} = \frac{f_2}{f_1+f_2} \quad \text{....19}$$

Similarly, the conditional probabilities that a randomly selected subject from the sampled population is infected and not infected by some disease in disease group type B given that the same subject is not infected by some disease group type A are estimated as respectively

$$\frac{P\left(\frac{B}{\bar{A}}\right)}{P(A)} = P(\bar{A}B) = \frac{\hat{\pi}_2}{\hat{\pi}_2+\hat{\pi}_4} = \frac{f_3}{f_3+f_4};$$
$$\frac{P\left(\frac{\bar{B}}{\bar{A}}\right)}{P(\bar{A})} = P(A\bar{B}) = \frac{\hat{\pi}_4}{\hat{\pi}_3+\hat{\pi}_4} = \frac{f_4}{f_3+f_4} \quad \text{......20}$$

The sample estimated of conditional probabilities and odds of infection by disease B conditional on infection by disease A are presented in Table2

Table 2:
Estimated Conditional Probabilities, Odds and Odds ratio of infection by Type 1 class disease (B) conditional on infection by Type 2 class disease (A)

Some Disease Class Type 2

Present (B)	Absent (B^-)	Total Odds (Ω)

Some Disease class type 1

Present $\frac{\hat{\pi}_4}{\hat{\pi}_3+\hat{\pi}_4} = \frac{f_4}{f_3+f_4}$

(A) $\frac{\hat{\pi}_2}{\hat{\pi}_1+\hat{\pi}_2} = \frac{f_2}{f_1+f_2}$

Absent

$$\Omega A = \frac{\hat{\pi}_3}{\hat{\pi}_4} = \frac{f_3}{f_4}$$

$(\bar{A})$ $\frac{\hat{\pi}_3}{\hat{\pi}_3+\hat{\pi}_4} = \frac{f_3}{f_3+f_4}$

Total

$$\hat{w} = \Omega(\bar{\mathrm{A}}) = \frac{\hat{\pi}_1\hat{\pi}_4}{\hat{\pi}_2\hat{\pi}_3} = \frac{f_1 f_4}{f_2 f_3}$$

If on the other hand, infection by disease class type 1 represented by disease A is conditional to infection by disease class type 2 represented by disease B, then the sample estimates of the conditional probabilities, odds and odds ratio of infection and non infection by disease class type 1 (A) given infection and non infection by disease type 2 (B) can similarly be obtained as presented in Table 3

Table 3
Estimated Conditional Probabilities, Odds and Odds ratio of infection by Type 2 class disease (A) conditional on infection by Type 1 class disease (B)

Some Disease Class Type 2

Present (B)	Absent (B^-)	Odds (Ω)

Some Disease class type 1

Present $\frac{\hat{\pi}_1}{\hat{\pi}_1+\hat{\pi}_3} = \frac{f_1}{f_1+f_3}$

(A)

Absent

$$\Omega A = \frac{\hat{\pi}_1}{\hat{\pi}_3} = \frac{f_1}{f_3}$$

(A^-) $\frac{\hat{\pi}_3}{\hat{\pi}_1+\hat{\pi}_3} = \frac{f_3}{f_1+f_3}$ P$A^- =$ $\frac{\hat{\pi}_3}{\hat{\pi}_1+\hat{\pi}_3} = \frac{f_3}{f_1+f_3}$

Odds

$$\Omega A^- = \frac{\hat{\pi}_2}{\hat{\pi}_4} = \frac{f_2}{f_4} \; (\Omega)$$

Finally, note from Tables 2 and 3 that as expected, the estimated odds-ratio is invariant whether the probabilities and odds are based or conditioned on disease class type 1,A or on disease class type 2, B.

A researcher may wish to test the null hypothesis of independence or no association between infection by diseases in disease group types A and B, which is equivalent to testing the null hypothesis that w is unity $(H_0 : w = 1)$. This null hypothesis is tested using the chi- square test statistic

$$\chi^2 = \frac{n(f_1f_4 - f_2f_3)^2}{(f_1+f_2)(f_3+f_4)(f_1+f_3)(f_2+f_4)}$$

……………..19

which under H_0 has approximately the chi-square distribution with 1 degree of freedom. The null hypothesis H_0 is rejected at the α level of significance if equation 10 is satisfied, otherwise H_0 is accepted.

Illustrative Example.

A researcher clinician screened a random sample of 60 adult males in a certain population for the possible presence $(+)$ and absence $(-)$ of two broad class type of diseases A and B. The screening test results are presented in Table 2

Table.2 Diagnostic Screen Test Results of a Sample of adult males for the possible presence (1) and absence (0) of two types of diseases A and B.

S/N	DISEASE (A)	DISEASE (B)
1	0	0
2	0	1
3	0	0
4	0	1
5	1	1
6	1	0
7	1	0
8	1	0
9	1	9
10	1	1
11	1	1
12	0	1
13	1	0
14	1	1
15	0	0
16	1	1
17	0	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	0
23	1	0
24	1	1
25	0	0
26	0	1
27	1	1
28	1	1

29	1	1
30	0	0
31	0	0
32	1	1
33	0	0
34	1	1
35	1	0
36	0	0
37	0	1
38	0	0
39	0	1
40	0	0
41	1	0
42	1	0
43	0	0
44	0	0
45	0	1
46	0	1
47	1	0
48	0	1
49	1	1
50	1	1
51	0	1
52	1	1
53	0	0
54	1	1
55	1	0
56	0	1
57	0	1
58	1	1
59	1	0
60	1	0

Using equation 1 in Table 2, we obtain values of u_{ij} shown in Table 3

Table.3 Values of u_{ij} (equation 1) for the data of Table 2

S/N	u_{ij}	Events
1	0	$\bar{A}\bar{B}$
2	0	$\bar{A}B$
3	0	$\bar{A}\bar{B}$
4	0	$\bar{A}B$
5	1	AB
6	0	$A\bar{B}$
7	0	$\bar{A}\bar{B}$
8	0	$A\bar{B}$
9	0	$A\bar{B}$
10	1	AB
11	1	AB
12	0	$\bar{A}B$
13	0	$A\bar{B}$
14	1	AB
15	0	$\bar{A}\bar{B}$
16	1	AB
17	0	$\bar{A}B$
18	1	AB
19	1	AB
20	1	AB
21	1	AB
22	0	$A\bar{B}$
23	0	$A\bar{B}$
24	1	AB
25	0	$\bar{A}\bar{B}$
26	0	$\bar{A}B$
27	1	AB
28	1	AB
29	1	AB
30	0	$\bar{A}\bar{B}$
31	0	$\bar{A}\bar{B}$
32	1	AB
33	0	$\bar{A}\bar{B}$
34	1	AB
35	0	$A\bar{B}$
36	0	$\bar{A}\bar{B}$
37	0	$\bar{A}B$
38	0	$\bar{A}\bar{B}$
39	0	$\bar{A}B$
40	0	$\bar{A}\bar{B}$
41	0	$\bar{A}\bar{B}$

42	0	$A\bar{B}$
43	0	$A\bar{B}$
44	0	$\bar{A}\bar{B}$
45	0	$\bar{A}B$
46	0	$\bar{A}B$
47	0	$A\bar{B}$
48	0	$\bar{A}B$
49	1	AB
50	1	AB
51	0	$\bar{A}B$
52	1	AB
53	0	$\bar{A}\bar{B}$
54	1	AB
55	0	$A\bar{B}$
56	0	$\bar{A}B$
57	0	$\bar{A}B$
58	1	AB
59	0	$A\bar{B}$
60	0	$A\bar{B}$

The values of u_{ij} and the corresponding event in Table 3 are summarized in Table 4

Table.4 Summary values of u_{ij} and the Events in Table 3 and other statistics

Disease Group B			
Present (1,B)	Absent(0,$\bar{B}$)	Total(fl.)	Odds
Disease Group A			
Present 20;0.33 (1;A)	15;0.250	35;0.583	1.333
Absent 13;0.217 (0;$\bar{A}$	12;0.200	25;0.417	1.083
Total 33,0.550 (f.j)	27;0.450	60;1	1.250
Odds 1.538			Ω=1.231

Table 4 shows the estimated joint probabilities of the presence and absence of disease group types A and B in the sample 1 population given an estimated odds–ratio of Ω=1.231. To test the statistical significance of the odd-ratio, that is to test whether or not disease groups type A and B are associated in the sampled proportion, we have from equation 19 that

$$\chi^2 = \frac{60(20)(12)-(13)(15)^2}{(33)(25)(35)(27)} = \frac{121500}{779625} = 0.156$$

which at 1 degree of freedom is not statistically significant, leading to the acceptance of the null hypothesis that the two group types of diseases A and B may not be associated in the sampled population.

One may nevertheless still wish to estimate some conditional probabilities of infection.

For example, one may be interested in estimating the probability that a randomly selected subject from the population is infected by some disease in disease group type B even though that the subject has not been infected by some disease in disease group type A. The conditional probability is estimated as

$P(B/\bar{A}) = \frac{15}{27} = 0.556$; and so on.

Conclusion

From the result of the analysis, the odds of selecting from the sample population in disease group B is greater than the odds of selecting subject from disease group A with 0.28 (28%). The estimated joint probabilities of the presence and absence of disease group types A and B gives an estimated odds-ratio of 1.231. Using χ^2 statistic to test for the statistical significance of the odds-ratio gives 0.156 at 1 degree of freedom indicates not statistically significance leading to the acceptance of the null hypothesis that the two group types of diseases A and B may not be associated in the sampled population. The conditional probabilities of infection from group B given that the subject has not be infected by some diseases in group type A is 0.556 (55.6%); this implies that there is a higher chance of being infected when event A has occurred.

References

Fulton L.V., Mendez F.A., Bastian N.D., Musal R.M. (2012). "Confusion Between Odds and Probability, a Pandemic?". *Journal of Statistics Education* Volume 20, Number 3 (2012),

James F. (2001). *The Science of Conjecture: Evidence and Probability Before Pascal*. Baltimore: The Johns Hopkins University Press. pp. 280–281.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2003). *Bayesian Data Analysis* (2nd ed.). CRC Press.

Lisa G. (2010). "Odds of Finding Earth-Size Exoplanets Are 1-in-4". *Wired*.