Unit 5 PROBABILITY, RISK AND ODDS

AIMS

The aims of this session are to introduce and introduce the ideas of probability, risk and odds, and to explain risk and odds ratios.

OBJECTIVES

At the end of Week 5 you should be able to:

- Explain what is meant by the probability of an outcome and calculate and interpret simple probabilities.

- Explain what is meant by the risk of an outcome and calculate and interpret simple (absolute) risks.

- Explain what is meant by odds for an outcome and calculate and interpret simple odds.

- Explain what is meant by a risk ratio and an odds ratio and calculate and interpret both.

Reading: Bland: pp. 129; 235-8.
Bowers: pp. 156-7; 162-3.

Introduction

In medicine we are often interested in the chances of some particular outcome happening, e.g. the chances of dying, of being obese, of having lung cancer, of being born premature, and so on. Medical statistics has a three principal ways of expressing (and calculating) what the chance of any such outcome is: by calculating the probability of the outcome; the risk of the outcome; or the odds in favour of the outcome.

Probability

Suppose you play a dice rolling game. If you roll an even number (2, 4 or 6) you win, if an odd number (1, 3 or 5) you loose. What are your chances of winning? The probability of any particular outcome (or event) happening can be defined as:

> **the probability of a particular event is the number of outcomes which favour the event in question, divided by the total number of possible outcomes.**

In the game there are six possible outcomes altogether (1, 2, 3, 4, 5 or 6), and three favourable outcomes (2, 4 or 6), so the probability of winning (a favourable outcome), is 3/6 = 0.5. In general, the probability of any particular outcome happening can vary between 0 (will never happen) and 1 (is certain to happen). So a probability of 0.5 means that the outcome is as likely to happen as not to happen.

Another way of thinking about probability is as the long-term relative frequency or proportion of times that some particular outcome or event has happened historically. For example, over the long term, the proportion of male babies born has been a half or 0.5. So the probability that the next baby born will be male is equal to this long-term proportion, or 0.5.

Q. 5.1 In the nit lotion study (Figure 1.2, Unit 1), what is the probability that a patient in the Malathion group chosen at random will: (a) be male; (b) be female; (c) have straight hair?

Q. 5.2 Table 5.1 shows the number of years of education for a sample of 59 Chinese women who were subjects in a case-control study of passive smoking as a risk factor in coronary heart disease. What is the probability that a subject chosen at random will have: (a) between 4 and 6 years of education; (b) 10 or more years of education?

Education (years)	Frequency (n=59)
0 - 3	1
4 - 6	15
7 - 9	16
10 - 12	8
≥ 13	19

Table 5.1 Years of education of subjects in passive smoking study. BMJ, 308, 1994.

Risk and the risk ratio

RR = EER / CER.

↳ = Relative Risk

Risk (sometimes referred to as *absolute risk*) means the same thing as probability (and so varies between 0 and 1), but because of its use in epidemiology and the study of risk factors it is labelled and thought of somewhat differently. However, we calculate risk for some event or outcome in exactly the same way as probability, i.e.

> **the risk of a particular event is the number of outcomes which favour the event in question, divided by the total number of possible outcomes.**

Risks and risk ratios are most appropriately calculated for *cohort studies* (in which a group of subjects is followed-up over some period of time and the outcomes and the differential exposure of the subjects to some risk factor form the basis of the anlaysis).

For example, suppose, in a study of the effects of smoking on coronary heart disease (CHD) we follow a cohort of 12000 males for 5 years. At the end of this time we find that 3000 of the subjects had smoked at some time during this period. There were 600 subjects with CHD, 480 among the smokers and 120 among the non-smokers. We want to know what the risk of a smoker developing CHD is (in this situation, smoking would be viewed as a risk factor).

It will help if we convert the above information into what is known as a *contingency table*, (Table 5.2):

		Outcome: CHD?		
		Yes	No	totals
Exposed to risk, i.e a smoker?	Yes	480	2520	3000
	No	120	8880	9000
	totals	600	11400	12000

Table 5.2 Contingency table: smokers/non-smokers v. CHD/not CHD (ficticious data)

There are 3000 smokers and 480 with CHD. If we divide the number of favourable outcomes (CHD) 480, by the total possible number of outcomes (smokers with CHD plus smokers without CHD), 3000, we have:

risk of CHD among smokers = 480/3000 = 0.160

So 16 in every 100 smokers are likely to develop CHD.

In might help if we formulate Table 5.2 in more general terms, as in Table 5.3. The letters in the cells refer to the values in each cell.

		Outcome: with disease?		
		Yes	No	Totals
Exposed to risk?	Yes	(a)	(b)	(a + b)
	No	(c)	(d)	(c + d)
	totals	(a + c)	(b + d)	(a + b + c + d)

Table 5.3 General formulation of risk of disease and exposure to a risk factor in a cohort study

So the risk of disease in those exposed to the risk factor = a/(a + b), and the risk for those not exposed = c/(c + d).

The risk ratio (also known as relative risk)

In practice knowledge of the absolute risk is of little interest on its own. However, we will often want to *compare* the absolute risks between two groups. We can do this by dividing one risk by the other. The result is known as the **risk ratio** (or the relative risk).

For example, the risk ratio of CHD among smokers compared to non-smokers is the risk of CHD for smokers divided by the risk of CHD for non-smokers. The risk of CHD for non-smokers is the number of CHD cases among non-smokers (120) divided by the total number of non-smokers (9000), i.e.

Risk of CHD for non-smokers = 120/9000 = 0.013

And we've already calculated that the risk of CHD in smokers is 0.160.

So the risk ratio = = 0.160/0.013 = 12.31.

So a smoker has about 12 times the risk of developing CHD than a non-smoker.

In terms of the cells in the above contingency table, the risk ratio = [a/(a + b)]/[c/(c + d)], becomes:

$$\textbf{risk ratio} = \frac{a(c+d)}{c(a+b)}$$

Q. 5.3 In a study of the factors affecting mortality in acute renal failure there were a total of 110 patients all with acute renal failure (Source: Renal Failure, 1992, 14, 161-8). Of these, 70 were male, of whom 62 survived, and 40 were female, of whom 33 survived. The rest of the patients died. Transform the above information into a contingency table (in this situation we can think of sex as being the risk factor) and answer the following questions. What is: (a) the risk of dying from acute renal failure for a man; (b) the risk of dying if a woman; (c) the risk ratio? Interpret this last result.

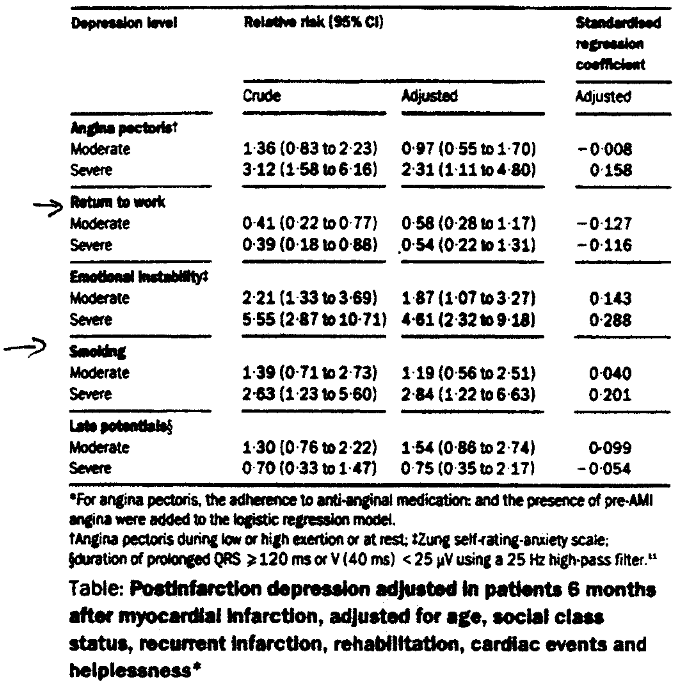
As a further example, Figure 5.1 shows the crude and adjusted* risk ratios (referred to by the authors as relative risks) for, (i) moderate, and (ii) severe,

* Crude risk ratios are for each separate risk factor taking no account of the possible influences of the other factors. The adjusted risk ratio takes the possible mutual influence of the other risk factors (listed in the table footnote) into account.

post-infarction depression, compared to no or low depression, for a number of "risk" factors.

For example, for a patient with the risk factor angina pectoris, the crude risk ratio of moderate depression *compared to a patient without angina* is 1.36 (ignore the numbers in brackets -- we'll return to those in a later unit). For severe depression the risk rises to 3.12.

Q. 5.4 From Figure 5.1 interpret the crude risk ratios for moderate and for severe depression when a patient: (a) has returned to work (compared to one who hasn't); (b) For a patient who smokes (compared to one who doesn't).



Depression level	Relative risk (95% CI)		Standardised regression coefficient
	Crude	Adjusted	Adjusted
Angina pectoris†			
Moderate	1·36 (0·83 to 2·23)	0·97 (0·55 to 1·70)	−0·008
Severe	3·12 (1·58 to 6·16)	2·31 (1·11 to 4·80)	0·158
Return to work			
Moderate	0·41 (0·22 to 0·77)	0·58 (0·28 to 1·17)	−0·127
Severe	0·39 (0·18 to 0·88)	0·54 (0·22 to 1·31)	−0·116
Emotional instability‡			
Moderate	2·21 (1·33 to 3·69)	1·87 (1·07 to 3·27)	0·143
Severe	5·55 (2·87 to 10·71)	4·61 (2·32 to 9·18)	0·288
Smoking			
Moderate	1·39 (0·71 to 2·73)	1·19 (0·56 to 2·51)	0·040
Severe	2·63 (1·23 to 5·60)	2·84 (1·22 to 6·63)	0·201
Late potentials§			
Moderate	1·30 (0·76 to 2·22)	1·54 (0·86 to 2·74)	0·099
Severe	0·70 (0·33 to 1·47)	0·75 (0·35 to 2·17)	−0·054

*For angina pectoris, the adherence to anti-anginal medication: and the presence of pre-AMI angina were added to the logistic regression model.
†Angina pectoris during low or high exertion or at rest; ‡Zung self-rating-anxiety scale; §duration of prolonged QRS ≥120 ms or V (40 ms) <25 μV using a 25 Hz high-pass filter.[11]

Table: **Postinfarction depression adjusted in patients 6 months after myocardial infarction, adjusted for age, social class status, recurrent infarction, rehabilitation, cardiac events and helplessness***

Figure 5.1 Risk ratios (relative risks) for moderate and for severe post-infarction depression for a number of risk factors. Lancet, 343, 1994.

Odds and the odds ratio

We have seen that probability and risk are two ways of expressing or quantifying the chance of some particular outcome happening. The *odds* in favour of some particular outcome happening is an alternative way of expressing this idea. The concept of odds is important in medical statistics because of its applications in epidemiology and its use is widespread in the literature.

We calculate the odds in favour of a particular outcome or event as:

> **the *odds* in favour of a particular event is the number of outcomes favourable to the event in question divided by the number of outcomes not favourable to the event.**

To illustrate the idea, suppose you have 8 patients sitting in your waiting room, of whom 3 are men and 5 women (although, since you can't see them you don't know the composition of those waiting). You call the next patient in . . . .

The *risk* (i.e. the probability) that the next patient is male is the number of males divided by the total number of patients, i.e. 3/8 = 0.375.

However the *odds* that the next patient is male is the number of males divided by the number of females, i.e. 3/5 = 0.600.

Odds and odds ratio calculations are appropriate in case-control studies, where the cases are subjects with the condition in question and the controls are similar to the subjects but do not have the condition. The differential exposure to a risk factor by the two groups of subjects forms the basis of the analysis. We can format such problems as a contingency table, as that in Table 5.4.

		Cases (sick)	Controls (healthy)	Totals
Exposed to risk factor?	Yes	(a)	(b)	
	No	(c)	(d)	
	totals			

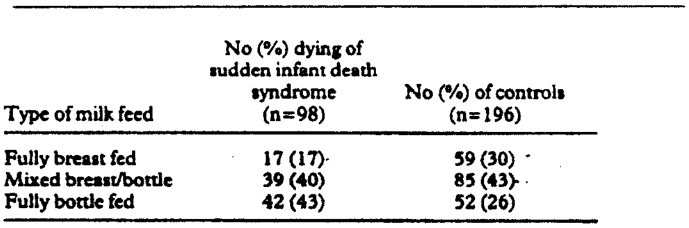
Table 5.4 General formulation of risk of disease and exposure to a risk factor in a case-control study

The odds that a person exposed to the risk is a case (is sick) is the number of cases exposed to the risk factor divided by the number of controls exposed to the risk factor, i.e. = a/b.

The odds that a person *not* exposed to the risk is a case is the number of cases not exposed to the risk factor divided by the number of controls not exposed to the risk factor, i.e. = c/d.

Note that although we might want to calculate risk, we cannot do so in the case-control situation because the row totals (a+b) and (c+d) are meaningless - they depend entirely on how many controls are chosen. The number of controls is decided by the researchers but is usually at least as large as the number of cases and often considerably larger.

As a further example, Figure 5.2 is extracted from a case-control study into whether bottle feeding is a cause of sudden infant death syndrome (SIDS). The method of feeding in a sample of 98 children who suffered SIDS (the cases) was compared with the feeding of 196 healthy matched controls. The method of feeding is the risk factor.



Type of milk feed	No (%) dying of sudden infant death syndrome (n=98)	No (%) of controls (n=196)
Fully breast fed	17 (17)	59 (30)
Mixed breast/bottle	39 (40)	85 (43)
Fully bottle fed	42 (43)	52 (26)

Figure 5.2 Case-control study into feeding and sudden infant death syndrome. BMJ, 1995, 310.

Among fully breast-fed babies (first row of table), the number of outcomes of interest (death from SIDS) is 17. The number of outcomes not SIDS is 59. Therefore the odds of a fully breast fed child dying from SIDS is 17/59 = 0.29. Note that we do not use the % figures supplied by the authors (shown in brackets) since these are column %s and of no practical interest. Even if they were row %s we still couldn't use them since they depend on the number of controls (as explained above).

Q. 5.5 Calculate and interpret: (a) the odds of a mixed breast/bottle fed baby dying from SIDS; (b) the odds for babies fully bottle fed dying from SIDS.

Odds ratio

We are usually more interested in the ratio of two odds, known, not surprisingly, as the odds ratio. For example the odds of mixed breast-bottle fed babies dying of SIDS compared to the odds of fully breast fed babies dying of SIDS. To obtain this odds ratio we divide the odds for the former outcome by the odds for the latter outcome. Thus the odds ratio for mixed breast-bottle fed babies dying of SIDS compared to the same odds for fully breast fed babies is:

$$(39/85)/(17/59) = 0.459/0.288 = 1.593$$

This implies that mixed breast/bottle fed babies have over one and a half times the odds of SIDS as fully breast-fed babies.

Its important to remember that the above odds ratio is a sample-based estimate of the true population odds ratio and might have occurred by chance alone (in the population there may be no such relationship). In a later unit we'll examine this problem further.

In terms of Table 5.4 the expression for the odds ratio for being a case when exposed to the risk factor compared to being a case when not exposed to the risk factor can be written:

$$\textbf{odds ratio} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

Q. 5.6 Calculate the odds ratio for SIDS in babies fully bottle fed compared to SIDS in babies fully breast fed. Interpret your result.

/ EER/CER

Risk ratio or odds ratio?

With cohort studies, i.e. when we can express the problem in the form of Table 5.2, we calculate the risk ratio. In case-control studies (in the form of Table 5.3) we can calculate the odds ratio. For outcomes that are *rare* the odds ratio is approximately the same as the risk ratio[*]. In other words, we can estimate odds ratios (which may be of more interest) in cohort studies by assuming that the risk ratios produced in such studies are roughly the same as the odds ratios.

[*] What constitutes "rare" is questionable. Opinions vary between 5% and 20% of subjects in the population having the condition.

Relationship between probability and odds

We can calculate the odds for a particular outcome, if we know the probability of the outcome happening, and vice versa, using the relationships:

odds = probability/(1 - probability)

probability = odds/(1 + odds)

For example, we found that the odds for SIDS in babies fed with mixed breast/bottle was 0.459. So the probability of mixed breast/bottle fed babies dying from SIDS is:

probability = 0.459/(1 + 0.459) = 0.459/1.459 = 0.314

This result implies that nearly a third of mixed breast/bottle-fed babies will suffer SIDS. We know this result must be nonsense. It is caused by the fact (again) that the number of control subjects in a case-control study (such as this one) is determined by the researchers, and thus risk (i.e. probability) calculations are not appropriate in these circumstances.

Q. 5.7 The post-operative infection rate following a particular surgical procedure is known to be about 15%. Assuming this figure remains the same in future: (a) what is the probability that the next patient to undergo the procedure will suffer post-operative infection? (b) What are the odds for infection compared to non-infection?

Q. 5.8 Table 5.5 is from a cross-section study of deaths following aortic aneurysm in two hospitals (in a cross-section study the population is sampled at some moment in time). Compare the odds for death in the two hospitals. Interpret your answer.

	Died		
	Y	N	totals
Hospital A	7	54	61
Hospital B	10	29	39
totals	17	73	100

Table 5.5 Deaths following aortic aneurysm in two hospitals. Mortality league tables: do they inform or mislead?", Quality in Health Care, 1995.

Referent groups

Figure 5.3 shows the odds ratios for chlamydia taken from a cross-section study comparing two methods of screening for genital chlamydia. For each variable (or risk factor) one of the categories is defined as the referent group - the one with which the other categories will be compared. Any category can in fact be defined as the referent group. (Note: ignore for the moment the numbers in brackets after the odds ratio - we will return to these in the next unit).

Take for example, the risk factor age, for which women aged ≥ 31 is taken as the referent group. Thus for a woman aged ≤ 20 the odds of having chlamydia compared to a woman aged ≥ 31 is 8.64. That is, more than eight times greater.

Q. 5.9 From Figure 5.3 interpret the odds ratios for genital chlamydia for: (a) married women compared to single women; (b) women who have had one or more new sexual partners in the past three months compared to women who have had no new partners in the same time period.

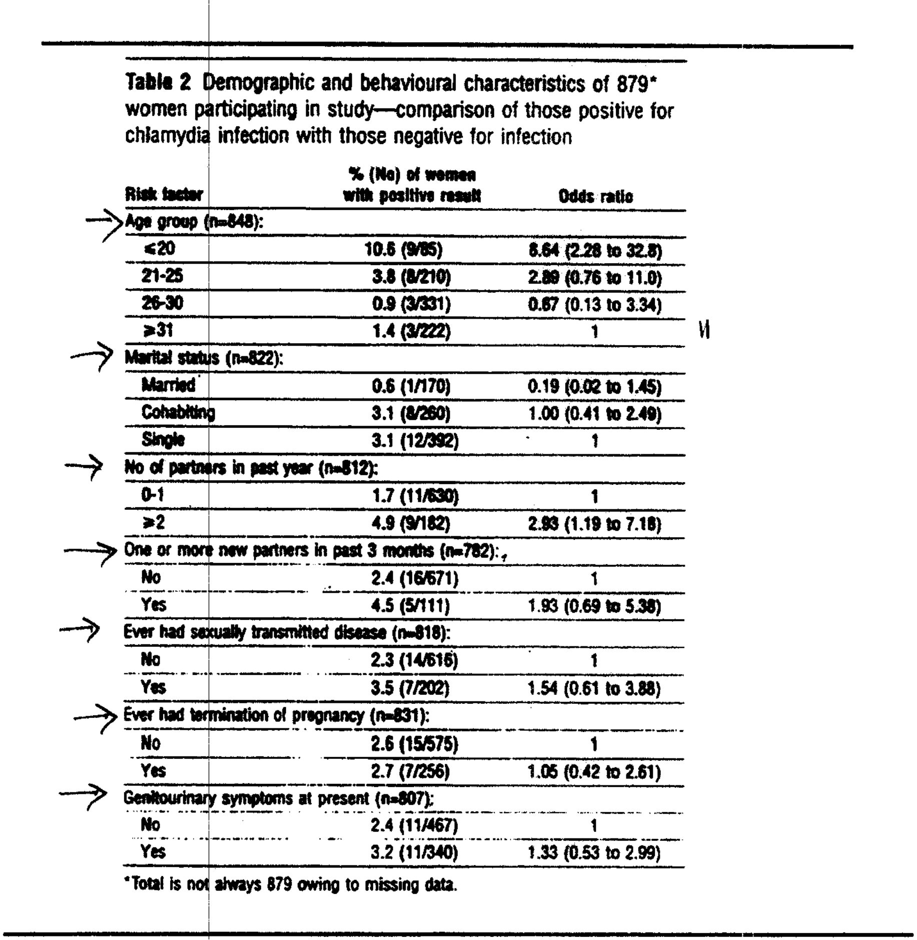


Table 2 Demographic and behavioural characteristics of 879* women participating in study—comparison of those positive for chlamydia infection with those negative for infection

Risk factor	% (No) of women with positive result	Odds ratio
Age group (n=848):		
≤20	10.6 (9/85)	8.64 (2.28 to 32.8)
21-25	3.8 (8/210)	2.89 (0.76 to 11.0)
26-30	0.9 (3/331)	0.67 (0.13 to 3.34)
≥31	1.4 (3/222)	1
Marital status (n=822):		
Married	0.6 (1/170)	0.19 (0.02 to 1.45)
Cohabiting	3.1 (8/260)	1.00 (0.41 to 2.49)
Single	3.1 (12/392)	1
No of partners in past year (n=812):		
0-1	1.7 (11/630)	1
≥2	4.9 (9/182)	2.93 (1.19 to 7.18)
One or more new partners in past 3 months (n=782):		
No	2.4 (16/671)	1
Yes	4.5 (5/111)	1.93 (0.69 to 5.38)
Ever had sexually transmitted disease (n=818):		
No	2.3 (14/616)	1
Yes	3.5 (7/202)	1.54 (0.61 to 3.88)
Ever had termination of pregnancy (n=831):		
No	2.6 (15/575)	1
Yes	2.7 (7/256)	1.05 (0.42 to 2.61)
Genitourinary symptoms at present (n=807):		
No	2.4 (11/467)	1
Yes	3.2 (11/340)	1.33 (0.53 to 2.99)

*Total is not always 879 owing to missing data.

Figure 5.3 Table of risk factors and the odds ratios for genital chlamydia from a study into screening for this disease. BMJ, 1997, 315.

Unit 5 Probability, risk and odds

Solutions to questions

Q. 5.1 (a) probability of male = 31/95 = 0.326; (b) probability of female = 64/95 = 0.674 (or 1 - 0.326); (c) probability of straight hair = 67/95 = 0.705.

Q. 5.2 (a) Probability of education from 4 to 6 years = 15/59 = 0.254; (b) probability of 10 or more years of education = (8 + 19)/59 = 0.458.

Q. 5.3

		Outcome		
		Alive	Dead	totals
Sex	Male	62	8	70
	Female	33	7	40
	totals	15	95	110

(a) risk of death if male = 8/70 = 0.114; (b) risk of death if female 7/40 = 0.175; (c) risk ratio = 0.175/0.114 = 1.535. So the risk of dying if male is about one and a half times that if female.

Q. 5.4 (a) The crude risk ratio for moderate depression for a patient who has returned to work is 0.41. That is, such a patient has about 40% of the risk of moderate depression as a patient who hasn't returned to work. Thus this "risk" factor return to work is *beneficial*. The risk ratio for severe depression is 0.39. Such a patient has about 39% of the risk of severe depression as a patient who hasn't returned to work. Thus again this risk factor is beneficial

(b) The risk ratio for moderate depression for a smoking patient is 1.39 compared to a non-smoking patient, i.e. about one and a third as much risk. For severe depression the risk ratio for a smoking patient is 2.63 compared to a non-smoking patient, i.e. more than two and a half the risk.

Q. 5.5 (a) For babies with mixed breast/bottle feeding, odds for SIDS = 39/85 = 0.46. So these babies have odds of dying of SIDS of just under a half; (b) For babies fully bottle fed, the odds for SIDS = 42/52 = 0.808. So these babies have odds of dying of SIDS of about four-fifths.

Q. 5.6 Odds ratio for SIDS among fully bottle fed babies compared to fully breast fed babies is odds for fully bottle fed divided by odds for fully breast fed

= (42/52)/(17/59) = 0.808/0.290 = 2.79. So a fully bottle fed baby has nearly three times the odds of SIDS as a fully breast fed baby.

Q. 5.7 (a) probability = 0.15; (b) odds = 0.15/(1 - 0.15) = 0.176.

Q. 5.8 odds ratio = $\dfrac{\frac{10}{29}}{\frac{7}{54}}$ = 0.345/0.129 = 2.67. The odds of a patient dying in hospital B are more than two and a half times the odds in hospital A.

Q. 5.9 (a) The odds ratio of 0.19 for married women means that such women appear to have only about a fifth of the odds of genital chlamydia as have single women. Being married appears to be a beneficial "risk" factor (we will have more to say on this particular result in the next unit).

(b) The odds ratio of 1.93 for a women with a new sexual partner in the past three months means that such women have nearly twice the odds of genital chlamydia as women who did not have a new sexual partner in this time period.