Relative Risk Analysis of Educational Data

Kaye Stacey
University of Melbourne
<k.stacey@unimelb.edu.au>

Vicki Steinle
University of Melbourne
<v.steinle@unimelb.edu.au>

This paper demonstrates the use of *relative risk*, a statistic which is widely used in other areas but currently under-utilised in education. Relative risk analysis provides a language for comparing educational outcomes as well as statistical tests of significance. We illustrate this statistic with data on students' understanding of decimal notation. In particular, we determine differences in how misconceptions operate at different ages, by analysing the relative risk of primary and secondary students persisting with particular misconceptions, and becoming experts.

This paper illustrates an approach to reporting, comparing and analysing educational outcomes, which is widely used in other fields but which has not been often used in mathematics education. The concept of *relative risk* is widely used in reporting results of medical, environmental and epidemiological research, in both scientific papers and in the popular press. Educators could capitalise on this popular familiarity in reporting their results. As well as being useful for describing results, measures of relative risk are amenable to statistical analysis. There are simple tests of statistical significance, confidence intervals and effect sizes, which are easily calculated; either manually, with a spreadsheet or in standard statistical analysis packages such as SPSS. Analysis of relative risk therefore meets the American Psychological Association standards for reporting research, which require reporting of effect sizes and confidence intervals to supplement significance testing (Capraro, 2004). Because of the widespread use of measures of relative risk in popular and scientific reporting, these statistical ideas are now being introduced in some introductory statistics courses for specialists and non-specialists (see for example Bulmer (2005)).

This paper begins with examples of the concepts related to relative risk taken from articles intended for health professionals and the general public. We have selected an example from public health, not because this field of application is special, but because these issues are likely to be within the common experience of readers. Furthermore, this will remind readers of how often scientific results are presented to the general public in this way. The example is used to introduce the terminology and the concepts of absolute and relative risk, the use of statistical testing, and some important points for appreciating the techniques involved. We will then show how these ideas can be applied to educational data, by re-analysing some data on students' understanding of decimal notation. The analysis gives us a different way of measuring differences, and a more readily understood language for reporting them.

Will a Glass of Red Wine a Day Keep Heart Attack Away?

The website of the American Heart Association (HREF1) provides information for the general public and medical practitioners on the claims in the media that drinking red wine is beneficial in combating heart disease. These claims arose as an explanation of the "French paradox." Researchers noted that compared with other Western countries, in France there was a relatively low incidence of coronary atherosclerosis, the accumulation of fatty plaques in the arteries that supply the heart and which can lead to blood clots, chest pain and heart attacks. This was despite the generally high intake of saturated fat in the French diet. Was it a high consumption of red wine that protected French people from heart disease? The AHA website reports the evidence in the following way; we have used italics to highlight the concepts related to relative risk.

> When the data from 51 epidemiological studies were combined, they showed that the *risk of coronary heart disease decreased by approximately 20%* when 0 to 2 alcoholic drinks were consumed per day. ….. Results from the large Health Professionals Follow-Up Study, a study in which 38,077 male health professionals who were free of cardiovascular disease were observed for 12 years, suggested that drinking 1 to 2 drinks per day, 3 to 4 days per week *decreased the risk of having a heart attack by as much as 32%*. [….]
>
> …. Support for a more pronounced cardioprotective effect for red wine as compared with other alcoholic beverages first emerged from the Copenhagen City Heart Study, in which 13,285 men and women were observed for 12 years. The results from this study suggested that *patients who drank wine had half the risk of dying from coronary heart disease or stroke as those who never drank wine*. […] The additional benefit of red wine is supported further by an analysis of 13 studies involving 209,418 participants. This analysis showed a *32% risk reduction* of atherosclerotic disease with red wine intake, which was greater than the *22% risk reduction* for beer consumption. Other studies and reviews have failed to show a beneficial effect for red wine, however, and hence it could be concluded that other [factors may be operating]. (HREF1)

The AHA website article reports on change in risk in three ways. The comment that "*patients who drank wine had half the risk of dying from coronary heart disease or stroke as those who never drank wine*" shows the concept of relative risk most clearly. A reader infers that the proportion of wine drinkers in the Copenhagen study who died from heart disease was half the proportion of non-wine drinkers who died from heart disease. For example, if 6% of the non-wine drinkers died (i.e. the probability of dying during the study was 6/100), then 3% of the wine drinkers died. If 0.6% of the non-wine drinkers died, then 0.3% of the drinkers died. This is the concept of relative risk: the risks are not given in absolute terms, but relative to each other. We can say that the relative risk (the ratio of one risk to the other) is a half.

The first two references to relative risk in the article describe the *risk being decreased by* 20%, and then by 32%. This quantity is also called the *risk reduction*, as used in the final two references. Here, in the first reference, if the risk of getting coronary heart disease had been 6%, then the new risk is 20% less (i.e. only 80% of 6%, or 4.8%). Using risk reduction is very common in reporting health results. Instead of describing the new risk as 80% of the previous, the risk reduction of 20% is quoted. This is done in order to focus attention on the perceived benefit of the intervention; in this case, the benefit of the moderate drinking described. Another important point to observe in the reporting in the first quoted paragraph is that the base-line risk is implied rather than explicit. The article does not explicitly say what the risk for moderate drinkers (e.g. 0 – 2 drinks per day) is being compared to. The reference group is to be inferred from the context: we presume it is the group of non-drinkers in the study. It is therefore important to note that in all use of the concept of relative risk, there must be a comparison, even if it in not explicitly stated, as in these first two references. Choosing the comparison sensibly is an important decision for using relative risk analysis.

A second important point to observe is the choice of terminology. Having heart disease is bad, so the language of risk and risk reduction is appropriate to the context. This language is also appropriate in many other contexts, such as for environmental assessment of hazard reduction and in some educational contexts. However, in many educational contexts, as we shall see later, although the *concept* of relative risk and the associated statistical analysis can be employed to good advantage, risk terminology is not appropriate and an alternative such as relative *chance* seems better than relative risk.

A third point to observe is in the final two references to relative risk in the website article. The 32% risk reduction from drinking red wine is said to be greater than the 22% risk reduction from drinking beer. This can be said with statistical confidence because the analysis of relative risk is not a yes-no answer about whether there is a difference in proportions, as with a chi-squared test. Instead, confidence intervals can be placed on the relative risks supporting a claim that relative risks for wine and beer drinking (each measured against the same baseline samples of non-drinkers, we presume) are indeed statistically different.

The reporting of results in terms of relative risk is particularly useful when absolute risk is low, even though it may be an important risk. The website above gives no indication of absolute risk – the chance that a person in any of the studies will in fact develop heart disease. An article from the Yale-New Haven Hospital online health information (HREF2) about a study of lifestyle factors, including moderate drinking of alcohol on heart disease, provides a useful example to illustrate this point.

> A new study … suggests people who follow several of the known steps to prevent heart disease benefit more than previously thought. In fact, a healthy lifestyle reduced the risk of heart attack, congestive heart failure and stroke by 82 percent. … The study was conducted by researchers at the Harvard School of Public Health and Brigham and Women's Hospital who surveyed 84,129 women health professionals enrolled in the Nurses' Health Study. [….]
>
> Dr. Frank B. Hu, assistant professor of nutrition at the Harvard School of Public Health, who presented the study, reported 1,129 cases of heart disease among this group. There were 296 fatal heart attacks and 833 nonfatal heart attacks. […..] If none of the other low risk behaviours were considered, non-smokers enjoyed a 74 percent reduction in risk. (HREF2)

The absolute risk of a participant in the Nurses' Health Study suffering heart disease was 1129/84129 = 0.0134 = 1.34% over the 14 years of the study, and the absolute risk of a fatal heart attack was 296/84129 = 0.00352 = 0.35%. A risk reduction of 82% leads to absolute risks of 0.24% and 0.06% respectively. If the results were reported instead in terms of absolute risk reduction of 1.10% (= 1.34% - 0.24%) and 0.29% (= 0.35% - 0.06%) the effect would be to hide the significance of the healthy life style factors. All the absolute risks involved sound trivial and inconsequential, although they are not. On the surface, there seems little point in adopting the recommended series of lifestyle measures, including moderate wine drinking, exercise, diet, giving up smoking etc in order to reduce the risk of a fatal heart attack by 0.29%.

Figure 1 sets out the notation and then the formulae for calculating relative risk. In this case condition 1 (e.g. moderate drinking) is being compared with condition 2 (e.g. non-drinking). There are also two outcomes (e.g. developing heart disease or not). The relative risk of outcome 1 is the ratio of the probability of it occurring under conditions 1 and 2. From the formulae, it can be deduced that the relative risk of comparing condition 2 with condition 1 leads to the reciprocal of the previous relative risk, and also that there is no simple relationship between the relative risk of event E occurring and the relative risk of event non-E occurring.

The calculation of confidence intervals and the statistical tests derive from the fact that the logarithm of the relative risk is normally distributed with known standard deviations. The confidence intervals for the log(*RR*) is then calculated from the normal distribution, and then converted back in terms of *RR*. Details are available from Agresti (1996) or Bulmer (2005).

	Outcome$_1$	Outcome$_2$	Total	
Condition$_1$	n_{11}	n_{12}	$n_1 = n_{11} + n_{12}$	Risk of Outcome$_1$ given Condition$_1$ $p_{1,1} = n_{11}/n_1$
Condition$_2$	n_{21}	n_{22}	$n_2 = n_{21} + n_{22}$	Risk of Outcome$_1$ given Condition$_2$ $p_{1,2} = n_{21}/n_2$

Relative Risk of Outcome$_1$ (Condition$_1$, Condition$_2$) $RR_1 = p_{1,1}/ p_{1,2}$

Figure 1. Calculation of relative risk.

A final point cannot be observed from these articles, but may be hidden in the original sources. Many of the calculations for relative risk are actually not done as above, but use a related concept called the *odds ratio* (Agresti, 1996; Bulmer, 2005; Steinle & Stacey, in press). Relative risks are used for reporting because they are easy to interpret whereas odds ratios are rather difficult to express in common language. When the risks of an event under the conditions to be compared are low (e.g. less than 10%), the odds ratio is a good approximation to the relative risk and can be interpreted as such. In the epidemiological examples above, these conditions apply. Odds ratios are commonly used because they can be applied to a wider range of research designs than relative risks. For example, they can be applied in experimental designs where the number of occurrences of a given outcome is experimentally manipulated. In addition, the odds ratio has strong mathematical properties giving it a more robust role in other statistical testing. Details are given in standard reference works such as Agresti (1996). SPSS performs odds ratio calculations under the Crosstabs menu, as the *Mantel-Haenszel common odds ratio estimate.*

Use of Relative Risk and Odds Ratios in Reporting Educational Studies

In the remainder of this paper, we will show how these ideas of relative risk can be applied to educational data and briefly discuss the benefits and issues arising. We illustrate the methods and challenges by analysing some results of a cross-sectional and longitudinal study of students' understanding of decimals in this new way. This was a cohort study, which tracked the developing understanding of over 3000 students in Years 4 – 10 at 12 schools for up to 4 years, testing them with the same test at intervals of approximately 6 months. Details of the sampling, the test and its method of analysis and many results have been described elsewhere; for example, Steinle and Stacey (2003) and Steinle (2004). For the purpose of this paper, it is sufficient to know that students are classified by the test as *experts* (coded as A1) or as having various *misconceptions*:

- L1, where a student generally interprets a decimal number as a whole number of parts of unspecified size, so for example, thinking that 0.10 comes after 0.9;
- S1, where a student interprets the place value correctly but assumes that any number of hundredths (e.g. 0.34) will be smaller than any number of tenths (e.g. 0.2), etc;
- S3, where a student draws a false analogy of a decimal with a fraction (e.g. 0.4 is like ¼);
- A2, where a student may think that only initial values of a decimal number are meaningful (e.g. 0.12345 is really 0.12) by analogy with money, etc;

- U1, where a student cannot be classified by our test, generally because they do not respond according to a known pattern, possibly because they mix misconceptions or make sporadic errors when trying to follow a consistent idea, correct or incorrect.

Further details of the thinking that lies behind these codes are explained in Steinle (2004).

Our previous analysis of the data demonstrated that the learning paths of students with given misconceptions vary markedly. In particular, they vary in the likelihood that a student will become an expert by the time of the next test and in the likelihood that a student will stay trapped in the particular misconception until the next test (and possibly beyond). Table 1 shows this variation over the whole sample. The students most likely to be expert on the next test are those already experts (A1), followed by A2, U1 (interestingly the students who showed no consistent misconception), S1, S3 and finally L1. Table 1 also shows that the students most likely to stay in the same code are A1 (this is good!); then the remaining codes are ranked L1, S3, U1, A2 and S1 (so S1 students are the least likely to stay in the same code). Excluding those already experts on the test (the A1 students), A2 students have the greatest "risk" of becoming experts on the next test. Here we see that the language of risk in inappropriate, since becoming an expert is a benefit. Hence, it is preferable to speak of the "chance" instead of "risk". The chance of an A2 student becoming an expert is about 3.5 times (= 53/15) the chance of an L1 student becoming an expert.

Table 1

Percentage of Students Moving to Expertise (A1) on the next test by code, and Percentage of Students Staying in the same code at the next test

	L1 (n=853)	S1 (n= 245)	S3 (n= 385)	A2 (n= 280)	U1 (n= 757)	A1 (n= 3279)
Moving to expertise	15%	33%	22%	53%	37%	89%
Staying in same code	38%	13%	33%	19%	28%	89%

Note. Data from Steinle (2004) Table 5.18

To illustrate the terminology further: the chance of an A2 student staying in code A2 at the next test is only half (19/38) the chance of an L1 student staying in the same code, so the relative risk of an A2 student retaining their misconception is half that of an L1 student. Risk is appropriate here since retaining a misconception is not desirable. We can also say that the risk of an S1 student retaining the misconception is 66% less than an L1 student (13/38 = 34%, 100% - 34% = 66%): the risk reduction is 66%.

Relative risks require a comparison between two absolute risks. This will be illustrated by using the unclassified students (U1) as the reference group i.e. we will compare students with and without a definite misconception. The point estimates for relative risk/chance in Table 2 are calculated by dividing the absolute chances in Table 1 by 37% (in row 1) and by 28% (in row 2). To determine whether differences between these risks are statistically significant, a confidence interval can be constructed around each point estimate. The confidence intervals for the RR in row 1 of Table 2 (i.e. for the relative chances of students moving to expertise on their next test, compared with U1 students) are illustrated in Figure 2. We can see that only one confidence interval (S1) includes 1.00, so this is the only point estimate in row 1 of Table 2 that is not significantly different at 5% from U1. Non-overlapping confidence intervals show the paths of students with different misconceptions are statistically different from each other: a strong result. Confidence intervals were also used on the RR in row 2 (the relative chance of students staying in the same code compared with U1 students) to determine which results were significant to 5%, but are not provided graphically. These results indicate that, compared with U1 students, L1 students are 37% more likely to stay the same (1.37) and A2 students have a 31% reduced risk of staying the same (0.69).

Table 2
Relative Chance of becoming an Expert (A1) and Relative Chance of staying in Same Code, compared to unclassified (U1) students

	L1	S1	S3	A2	A1
RR to A1 (compared with U1)	0.40*	0.90	0.61*	1.42*	2.42*
RR stay same (compared with U1)	1.37*	0.47*	1.19	0.69*	3.24*

**Note*. Significant at 5% level

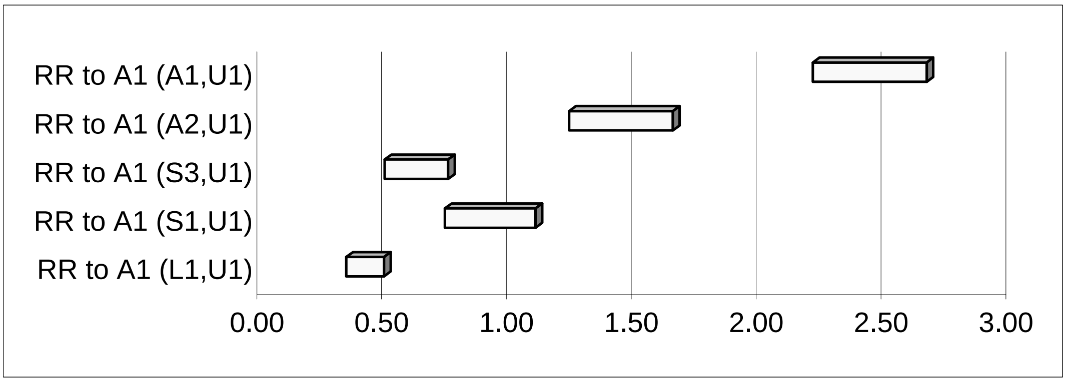

Figure 2. Confidence intervals (95%) for chance of a student in a given code becoming an expert, relative to an unclassified (U1) student.

We now show that there are variations in the future learning of younger and older students with given misconceptions. Data derived from Steinle (2004) Appendix 7 splits the data in Table 1 by grade: primary refers to grades 4 – 6 and secondary refers to grades 7 – 10. Row 1 of Table 3 shows the relative risk (RR) that a primary student with a given misconception will become an expert at the next test, relative to a secondary student. As noted above, since this is a benefit rather than a hazard, it is better to refer to it as the relative chance instead of relative risk. This information, along with the 95% confidence intervals is illustrated graphically in Figure 3. Table 3 also gives the relative risk of a primary student retaining their misconception compared to a secondary student, and the 95% confidence intervals are given in Figure 4.

Table 3
Chance that a primary student, relative to a secondary student in the same code, will be an expert (A1) at next test and relative chance of staying in same code at next test

	L1	S1	S3	A2	U1	A1
RR to A1 (pri, sec)	0.65*	1.73*	1.50*	1.36*	0.98	1.01
RR stay same (pri, sec)	1.18	1.21	0.71*	0.32*	0.97	1.01

**Note*. Significant at 5% level

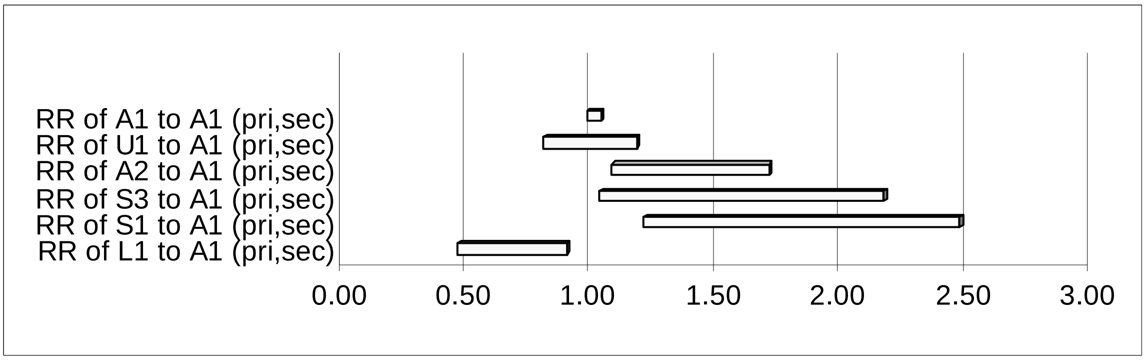

Figure 3. Confidence intervals (95%) for chance of a primary student in a given code becoming an expert, relative to secondary student.

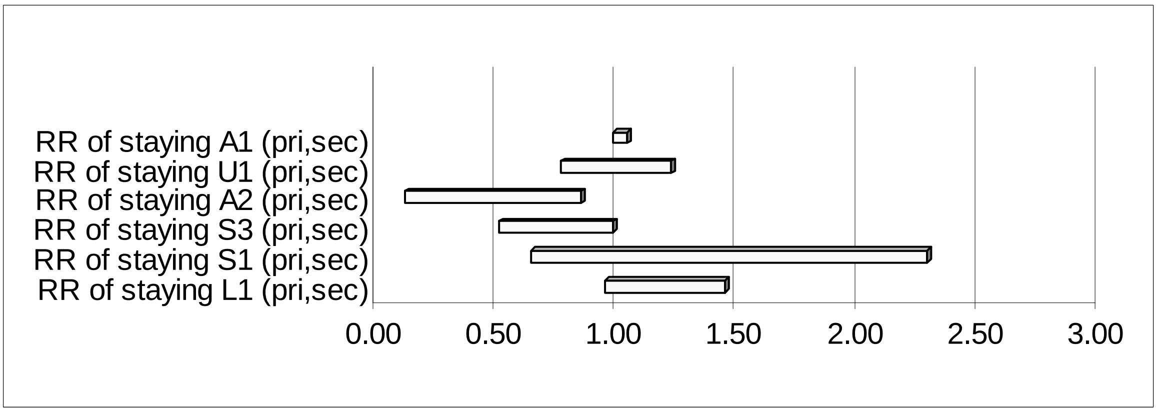

Figure 4. Confidence intervals (95%) for chance of a primary student in a given code staying in the same code, relative to secondary student.

From Table 3 we see that the primary school students in all of the four misconception groups L1, S1, S3 and A2 have a different chance of becoming experts on the next test than secondary students with the same misconception. However, for L1 the primary school student has less chance and for the other misconceptions, the secondary students have less chance. An L1 student in primary school has only 65% of the chance of a secondary L1 student of becoming an expert on the next test. In terms of risk reduction, the primary L1 student has 35% less chance than the secondary L1 student. Figure 3 shows that this result is statistically significant at the 5% level, since the confidence interval (from 0.46 to 0.90) does not include 1. The confidence interval can be interpreted this way: the primary L1 student has at least 10% less chance and up to 54% less chance of becoming an expert than the secondary L1 student. On the other hand, Figure 4 shows that the primary and secondary L1 students have a similar risk of staying in L1, as the confidence interval includes 1. Note that this risk is quite high (38% from Table 1).

In contrast, primary students in codes S1, S3 and A2 have a significantly greater chance than their counterparts in secondary school of becoming experts on the next test. The extreme case is that of S1, where a primary student has 73% more chance of becoming an expert than a secondary student. The confidence interval shows that the chance may be up to two and a half times as great. The extreme case of difference between primary and secondary students is in the relative chance of staying in code A2. Here a primary student has only a third of the risk of a secondary student of staying in code A2 at the next test. This demonstrates that secondary students hold on to the misconceptions associated with the code A2 (and similarly S3); there are likely to be new ideas and practices in the secondary curriculum which reinforce these ideas. The practice of habitually rounding calculations to two decimal places, as if further places have no meaning, may be a reason.

Conclusion

The aim of this paper has been to illustrate how educational data can be analysed and reported using relative risk; a statistic which is common in the popular press as well as in the scientific literature in some fields. There are several advantages, which relate to the ease of interpreting the change in risk and the way in which it provides an alternative presentation of results in possibly a more memorable form, and in a form which highlights the real meaning of differences which in absolute terms appear to be small. The relative risk analysis provides a tool which could be frequently used in reporting educational results, although the language of risk will often need to be changed when benefits or neutral outcomes are discussed rather than hazards.

As mathematics educators we need to be concerned with how educational researchers, teachers and the general public understand quantitative results. The common use of concepts related to relative risk indicates that it is likely that researchers feel these ideas are intuitively understood and hence are appropriate for communication to wide audiences. We recommend that research be carried out to establish whether this is indeed the case.

References

Agresti, A. (1996). *An introduction to categorical data analysis*. John Wiley: New York.

Bulmer, M. (2005). *A portable introduction to data analysis* (3rd ed.). Brisbane: The University of Queensland.

Capraro, R. M. (2004). Statistical significance, effect size reporting, and confidence intervals: Best reporting strategies. *Journal for Research in Mathematics Education*, *35*(1), 57 – 62.

Steinle, V. (2004). *Changes with age in students' misconceptions of decimal numbers*. Unpublished PhD, University of Melbourne, Melbourne.

Steinle, V., & Stacey, K. (2003). Grade-related trends in the prevalence and persistence of decimal misconceptions. In N.A. Pateman, B.J. Dougherty & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 259 – 266). Honolulu: PME.

Steinle, V., & Stacey, K. (in press). Analysing longitudinal data on students' decimal understanding using relative risk and odds ratios. *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*. Melbourne: PME.

HREF1 American Heart Association http://circ.ahajournals.org/cgi/content/full/111/2/e10 Accessed 2/4/2005.

HREF2 Yale-New Haven Hospital http://www.ynhh.org/healthlink/womens/womens_12_99.html Accessed 2/4/2005.