Industrial Health 2012, **50**, 463–465

Editorial

Measuring 'Risk' in Occupational Health Studies: Standard Methods and Some Alternatives for Epidemiological Research

Understanding the potential relationship between a workplace exposure and whether an employee subsequently contracts a disease represents one of the time-honoured cornerstones of occupational health research. Establishing whether a statistically significant relationship actually exists between these variables, and then quantifying it in terms of 'risk', has long been the domain of occupational epidemiology[1]. As human understanding of biological processes progressed throughout the 19th and 20th Centuries, clinicians and practitioners were increasingly able to draw intuitive links from anecdotal reports and individual case studies, thereby leading to disease causation hypotheses and research studies to investigate them[2].

For many years the accepted scientific paradigm has demanded that any proposed links between variables under study be mathematically validated via the use of appropriate statistical techniques. To reduce the chance element in scientific research it became necessary to calculate the probability of an outcome occurring due to the experiment, versus the likelihood of it occurring due to chance alone. As a result, most contemporary statistical techniques are based on early studies of games of chance[3], including two of the most well-known examples: Pearson's chi-squared test and Fisher's exact-test. By the mid 20th Century, the general understanding of statistics, medicine and clinical experimental design had all progressed to a point where simple chi-squared tests were no longer being seen as adequate for determining 'risk' in scientific research. A more reliable method was also being sought in the discipline of occupational health, or industrial hygiene as it was often called at the time, so that the 'risk' to an individual worker exposed to various substances could be calculated and appropriate protective measures determined[4].

Although it did not originate in the field of occupational health, the development of an Odds Ratio (OR) calculation as we know it today began to appear in clinical medicine during the mid 20th Century. Indeed, the original method is now so widely used that most published studies no longer reference the original source[5]. Credit for its discovery is generally given to Jerome Cornfield (1912–1979) − a pioneering American biostatistician, who in 1951 demonstrated that an OR calculated from data in case-control studies could be used to estimate the 'relative risk' of developing a disease[5]. His exposition of the now-famous *Cornfield's Odds Ratio* was part of a study that investigated links between smoking and lung cancer and reported the categories as 'relative risk'[6]. In 1956 Cornfield further demonstrated how to calculate confidence limits of the relative risk from Fisher's exact-test − although it is worth noting that Fisher never actually used the term 'odds ratio', and had instead referred to it as a 'cross-product ratio'.

Most contemporary studies which report ORs still use the basic measure proposed by Cornfield in the 1950s. While this calculation remains one of the most influential and simple techniques to quantify the association between sets of categories, few realise that it is only one of a number of different methods that may be appropriate for estimating 'risk' in epidemiological research. In occupational health studies, the concept of 'risk' with respect to a particular exposure has always been a fundamental one[7], and not surprisingly therefore, relative risks estimated by the OR have now become a de facto standard for 'hazard' in many research fields[8]. This is probably because the OR calculation exhibits convenient mathematical properties[9], especially in the interpretation of case-control studies[10]; and can provide (given certain conditions) a reasonable approximation of the relative risk[11]. The increasingly widespread use of logistic regression in modern research has also helped further popularise the method[12]. With a plethora of studies now using the OR, it is timely to reflect on some other statistical techniques that may be considered, and why they may be preferable to the 'standard' (Cornfield's) OR.

The first of these, *Haldane's Odds Ratio*, is named after John Burdon Sanderson Haldane (1892–1964), a world renowned geneticist and biologist[13], who devised it while working at University College, London during the 1950s. During a study of stomach cancer Haldane had calculated a positive OR, although what concerned him was the possibility of obtaining a zero cell value when selecting another random sample from the same population. This meant that, if at least one of the off-diagonal cell values were zero, the OR could not be defined (as opposed to its value being zero or infinite). To overcome the problem, Haldane suggested adding 0.5 to each cell value in the 2 × 2 table before calculating the OR using the Cornfield technique. With mathematical rigor, he further demonstrated that doing so also reduced some intrinsic biases of the original OR, especially when the cell numbers were very small in magnitude. His approach was endorsed early on[14], with more contemporary authors also finding that the addition of a reasonably small positive constant to each cell value still has its merits[15].

A second alternative adaptation of Cornfield's OR was proposed by Nicholas P. Jewell in 1986[16], which involved adding a value of 1 to the off-diagonal elements of the 2 × 2 table. By doing so, it was demonstrated that *Jewell's Odds Ratio* (similar to Haldane's OR) provided a more appropriate point estimate of the population OR than did Cornfield's OR. Amending Cornfield's OR in the manner suggested by Jewell leads to a reduced estimate of the population OR, which in turn, helps address the bias issue. Despite this fact, caution still needs to be taken when using the calculation as, unlike Cornfield's and Haldane's OR, Jewell's OR is not impervious to changes in the order of the row and column categories. Nonetheless, a detailed study published in 1991 compared all three OR choices, with the authors reporting that Jewell's OR has the smallest bias for analysing the association between rows and columns in a 2 × 2 table[17].

Analysis and improvement of the statistical calculations used in occupational epidemiology have continued since the pioneering works of Cornfield, Haldane and Jewell. Frequency of the outcome under study, for example, has long been a major consideration in epidemiological research. This is mainly because the more frequent the outcome, the less accurate an OR calculation is at estimating the risk ratio in a cohort study or clinical trial. To help adjust for this phenomenon, in 1998 Zhang and Yu[11] proposed a method of correcting the OR of cohort studies with common outcomes. Most recently, Tran and colleagues[18] have further explored a new statistical approach called the *Aggregate Association Index* (AAI) for undertaking risk calculations in occupational epidemiology − especially in studies where the 2 × 2 tables contain missing values.

Regardless of which method one chooses to employ, it can be seen that the development and subsequent use of ORs has become a fundamental aspect of modern epidemiological research. The impact of OR calculations to represent 'risk' are especially evident within occupational health research, as a convenient and easily understandable method for reporting and describing workplace hazards to researchers, legislators and workers themselves. These factors represent, perhaps, some of the most important issues to consider when choosing a method for the statistical analysis of research findings in modern epidemiology.

References

1) Smith DR, Beh EJ (2012) Asbestos kills − no matter how you cut the data. Arch Environ Occup Health **67**, 187–8.
2) Smith DR, Beh EJ (2011) Occupational epidemiology in the real world: Irving Selikoff, odds ratios, and asbestosis. Arch Environ Occup Health **66**, 63–4.
3) Beh EJ, Smith DR (2011) Real world occupational epidemiology, part 1: odds ratios, relative risk, and asbestosis. Arch Environ Occup Health **66**, 119–23.
4) Smith DR (2009) The historical development of academic journals in occupational medicine, 1901–2009. Arch Environ Occup Health **64**,(Suppl 1): 8–17.
5) Greenhouse SW (1982) Jerome Cornfield's contributions to epidemiology. Biometrics **38**, 33–45.
6) Cornfield J (1951) A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst **11**, 1269–75.
7) Beh EJ, Smith DR (2011) Real world occupational epidemiology, part 2: a visual interpretation of statistical significance. Arch Environ Occup Health **66**, 245–8.
8) Smith DR, Beh EJ (2011) Hirayama, passive smoking and lung cancer: 30 years on and the numbers still don't lie. Public Health **125**, 179–81.
9) Nurminen M (1995) To use or not to use the odds ratio in epidemiologic analyses? Eur J Epidemiol **11**, 365–71.
10) Bland JM, Altman DG (2000) Statistics notes. The odds ratio. BMJ **320**, 1468.
11) Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA **280**, 1690–1.
12) Robbins AS, Chao SY, Fonseca VP (2002) What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. Ann Epidemiol **12**, 452–4.
13) Pirie NW (1966) John Burdon Sanderson Haldane. 1892–1964. Biogr Mem Fellows R Soc **12**, 218–49.
14) Anscombe FJ (1956) On estimating binomial response relations. Biometrika **43**, 461–4.
15) Breslow N (1981) Odds ratio estimators when the data are sparse. Biometrika **68**, 73–84.
16) Jewell NP (1986) On the bias of commonly used measures

of association for 2x2 tables. Biometrics **42**, 351–8.

17) Walter SD, Cook RJ (1991) A comparison of several point estimators of the odds ratio in a single 2x2 contingency table. Biometrics **47**, 795–811.

18) Tran D, Beh EJ, Smith DR (2012) Real-world occupational epidemiology, part 3: an aggregate data analysis of Selikoff's "20 year rule". Arch Environ Occup Health **67**, 243–8.

Derek R. SMITH,
School of Health Sciences, Faculty of Health, University of Newcastle, Ourimbah, Australia

Eric J. BEH
School of Mathematical & Physical Sciences, Faculty of Science and Technology,
University of Newcastle, Callaghan, Australia