

ORIGINAL ARTICLE

Comparing and combining studies of bronchial responsiveness

S Chinn

Thorax 2002;57:393–395

Correspondence to:
Miss S Chinn, Department
of Public Health Sciences,
King's College, 5th Floor,
Capital House, 42 Weston
Street, London SE1 3QD,
UK; sue.chinn@kcl.ac.uk

Revised version received
23 October 2001
Accepted for publication
13 November 2001

Background: There is no standardised protocol for the measurement of bronchial responsiveness. Results from different studies are difficult to compare and combine.

Methods: Analyses are divided between those of a continuous outcome, which can be directly standardised as effect size, and those based on a binary outcome. A published method is used to convert an odds ratio to equivalent effect size.

Results: The use of effect size allows comparison between studies using a continuous outcome but different protocols, provided the relevant standard deviation is reported. Effect size from a continuous outcome and that derived from an odds ratio from an equivalent analysis gave similar results.

Conclusions: Systematic reviews which include both continuous effect estimates and odds ratios can include both in one meta-analysis, provided relevant standard deviations are published for the former. Authors are encouraged to report these in all fields in which measurement protocols vary.

The measurement of bronchial responsiveness (BHR) is advocated in population studies of asthma.¹ While not synonymous with asthma, it is an objective measure that has particular advantages in multicultural studies.² BHR has also been used in the diagnosis of asthma and the measurement of severity in clinical settings.³

The results of a scientific study are of little use unless they are in a form that is useful to the readers and allows comparison with previous or future work. This normally requires that the measurement protocols are similar. While some protocol variation exists for many measurements, this is a particular problem for BHR.^{3,4}

Results of histamine or methacholine challenge are most commonly summarised in clinical studies by the dose (PD₂₀) or concentration (PC₂₀) that produces a 20% fall in forced expiratory volume in one second (FEV₁). A logarithmic scale is considered appropriate for the analysis,⁵ and results are frequently expressed in doubling doses or concentrations. For example, in a randomised controlled trial of two treatments for asthma, the result for BHR might be one doubling dose difference—that is, that after one treatment, compared to the other, twice as much of the provoking agent was required to produce a 20% fall in FEV₁ or, equivalently, that BHR was reduced by one doubling dose.

CONSEQUENCES OF PROTOCOL VARIATION FOR CLINICAL STUDIES

Results reported as PD₂₀, PD₁₀, PC₂₀, etc can be converted to doubling dose units, and authors of meta-analyses have used this method to combine results from different studies.^{6,7} This does not necessarily overcome the problem of protocol variation. The response to treatment may be more variable in one study than in another as a result of the protocol differences. Provided a standard deviation of the outcome on the doubling dose or other logarithmic scale is reported, this can be used to calculate a standardised effect for each study. The estimate and its associated standard error are divided by the standard deviation to give “effect size”. This does not affect the associated p value. When the estimate is a difference in means, effect size is also known as standardised difference. This approach was used by Abramson *et al* to combine results from 12 studies on immunotherapy in asthma with non-

specific BHR as outcome and 14 studies that measured allergen specific BHR.⁸

A FURTHER PROBLEM IN POPULATION STUDIES

In clinical studies on asthmatic patients, each participant is likely to have a measurable PD₂₀ on each occasion. In a population study the maximum dose of provocation agent will be limited by safety considerations, and at least 50% of the population will have a less than 20% fall in FEV₁ at the highest dose administered. For these participants, PD₂₀ is said to be “censored”. There are three ways of dealing with this in the analysis.

One is to use a method known as censored regression, implemented in at least one major statistical program.⁹ This produces estimates on the log(PD₂₀) scale, but is heavily dependent on the assumption that log(PD₂₀) is normally distributed. However, there is evidence for this assumption,¹⁰ and it has been used in the analysis of several population studies of the relation of BHR to urinary electrolytes.^{11,12}

A second method is to use an alternative summary statistic for BHR. Several measures of “slope” have been proposed^{3,4} which have the advantage that a value can be defined for each person. The simplest of these, due to O'Connor,¹³ is the fall in FEV₁ divided by the final dose. Provided a standard deviation is reported, estimates using censored regression of PD₂₀ or a slope outcome can be combined or compared with each other using effect size as described above.

A third method is the most common in population studies. Participants are simply divided into those “with” and those “without” BHR according to whether PD₂₀ is less than the maximum dose. Descriptive statistics are proportions of the sample “with BHR”, analysis is using logistic regression, and results are given as odds ratios. This analysis makes few assumptions and is relatively simple to understand. However, it is wasteful of information. In addition, BHR in the population has a unimodal distribution,^{14,15} so any division into subjects “with” or “without” BHR is purely arbitrary and usually determined by the maximum dose administered.

Cook and Strachan found 10 studies that reported an odds ratio for BHR for children exposed to environmental tobacco smoke compared with those who were not.¹⁶ Odds ratios can be combined on the log scale and antilogged to give a final

Table 1 Example of conversion of results in the relation of BHR to sodium excretion in two different forms to effect size

	Regression coefficient	Standard error	Residual SD
(1) Continuous outcome			
Regression coefficient of \log_{10} (PD ₂₀) on \log_{10} (sodium excretion)			
Published	-1.778	0.777	0.788
Standardised (/0.788)	-2.256	0.986	
(2) Binary outcome			
Logistic regression coefficient of proportion PD ₂₀ <8 mmol on \log_{10} (sodium excretion)			
	ln (odds ratio)	Standard error	
Published	3.168	1.809	
Standardised (/1.81)	1.750	0.999	

Results taken from Burney *et al.*¹¹

summary. Whether Cook and Strachan¹⁶ or van Grunsven *et al.*⁶ had to omit some studies from their meta-analyses is unclear. However, Abramson *et al.*⁸ found a mixture of results reported as a continuous or binary outcome. Their solution was to perform two separate meta-analyses.

COMPARING OR COMBINING ODDS RATIOS AND EFFECT SIZE

It is clearly desirable to be able to compare or combine results from studies using these different methods. A simple method was recently published¹⁷ which is illustrated in table 1. Results are taken from a study of the relation of BHR to sodium excretion.¹¹ The report published both a censored multiple regression analysis, giving a regression coefficient of \log_{10} (PD₂₀) on \log_{10} (sodium), and a multiple logistic regression with PD₂₀ dichotomised at the maximum dose of 8 μ mol histamine, giving a logistic regression coefficient for the proportion "with BHR" on \log_{10} (sodium). An estimated standard deviation of 0.788 was also reported from the first, allowing the calculation of effect size simply by dividing both the regression coefficient and standard error by this figure. The logistic regression coefficient is the natural logarithm of the odds ratio associated with a unit increase in \log_{10} (sodium). This is converted to effect size by dividing by the factor 1.81 which was derived from the properties of the logistic distribution.¹⁷ This allows for the fact that the logistic transformation "stretches" the scale more than the normal equivalent deviate or probit transformation. For example, a proportion of 0.025, or 2.5%, is a normal equivalent deviate of -1.96 but is transformed to -3.66 (natural logarithm of (0.05 divided by 0.95)) on the logistic scale. The ratio of these transformed values is slightly greater than 1.81, which is the average

over the whole scale, and good approximation over the range of proportions from 0.03 to 0.97.¹⁷ The conversion can thus be used provided the prevalence of BHR is not less than 3%.

The two estimates of effect size in table 1 differ in sign because a decrease in PD₂₀ is an increase in BHR. Allowing for this, the two estimates are close—within a difference that might be expected from the standard errors and assumptions of the two methods. Two further points should be noted. It is more common now to report an odds ratio with 95% confidence interval than a logistic regression coefficient and standard error. This requires that the natural logarithm of each value is taken before dividing by 1.81 to give effect size and its 95% confidence interval. The second is that, when the estimate, as here, is a regression coefficient rather than a difference in means, the independent variable also needs to be on the same scale. A unit increase in \log_{10} (sodium) represents a 10-fold increase; a more useful quantity might be the change in BHR associated with a 50% increase which can be derived by multiplying each estimate by \log_{10} (1.5).

Table 2 shows effect sizes derived from results in a paper that reported BHR summarised as a slope measure, devised for use in the European Community Health Survey,² and analysed by multiple regression (table 6 in reference), and equivalent results from logistic regression (table 7 in reference).¹⁸ The residual standard deviation from the multiple regression was not reported explicitly, but the total standard deviation was 2.10 and the variation explained 33.3%, so the residual standard deviation was calculated as $\sqrt{0.667(2.10)^2}$. As stated, the sign of each effect size derived from the multiple regression was reversed so that it was compatible with the comparable value from the logistic regression. The effect sizes derived from odds ratios generally have wider confidence intervals and

Table 2 Effect sizes for the relation of bronchial hyperresponsiveness (BHR) to serum total IgE, individual allergens, and lung function derived from a multiple regression analysis of a continuous measure compared with those expressed as odds ratios from the equivalent logistic regression model

Factor	Effect size derived from multiple regression*			Effect size derived from odds ratio from logistic regression		
	Effect size	95% confidence interval	p value	Effect size	95% confidence interval	p value
Total IgE (log kU/l)	0.08	-0.05 to 0.20	0.23	0.16	-0.06 to 0.38	0.16
Specific IgE to:						
Cat	0.34	0.05 to 0.63	0.02	0.55	0.11 to 1.00	0.02
House dust mite (HDM) titre (just positive)	0.42	0.22 to 0.62	<0.001	0.43	0.10 to 0.75	0.01
per 10 kU/l increase in HDM titre above 0.35 kU/l	0.13	0.03 to 0.23	0.01	0.09	-0.06 to 0.24	0.22
Timothy grass	0.29	0.08 to 0.49	0.005	0.64	0.33 to 0.95	<0.001
Cladosporium	0.84	0.37 to 1.30	<0.001	1.20	0.49 to 1.91	0.001
Birch	-0.28	-0.62 to 0.05	0.09	-0.74	-1.33 to -0.17	0.01
FEV ₁ as standardised difference from predicted value	-0.22	-0.31 to -0.12	<0.001	-0.35	-0.52 to -0.18	<0.001
FEV ₁ %FVC	-0.05	-0.07 to -0.04	<0.001	-0.06	-0.08 to -0.03	<0.001

*Derived from the slope measure of BHR used in analyses of data from the European Community Respiratory Health Survey.¹⁸ Decreasing slope is increasing BHR, so here signs are reversed so that positive effect size represents increasing BHR with presence or increasing level of factor.

greater p values due to the loss of information on dichotomising BHR. Where the two effect sizes do appear to differ, which was notable for the association with specific IgE to Timothy grass and, to a lesser extent, with specific IgE to birch, this is due not to a problem with the conversion of odds ratio to effect size but to the fact that the multiple regression and logistic regression did in this case give different results, with the p value for the latter less than that for the former.

DISCUSSION

The above method for conversion of a $\ln(\text{odds ratio})$ to effect size shows that the two are essentially equivalent. However, an analysis of a continuous outcome is generally more powerful than one based on an arbitrary division of the scale into two groups. This is not an argument for a slope measure of BHR over PD_{20} , as the existence of a value for each subject does not automatically imply more information.¹⁰ Although the result implies that an odds ratio may be little affected by the alteration of cut off point, other statistics will change. Peat *et al* reported that BHR has high specificity for asthma and low sensitivity,¹ but a greater maximum dose of provoking agent and cut off point would increase sensitivity and decrease specificity.

The method was illustrated using results from Burney *et al*¹¹ and Chinn *et al*¹⁸ because, in each study, the data were analysed in the two ways; in the former the residual standard deviation from the linear regression was reported and in the latter it could be calculated from published results, enabling the calculation of effect size in both cases. More often results are reported without the standard deviation¹² or only as an odds ratio.¹⁹ When a mixture of odds ratios and estimates based on continuous outcomes are to be compared or combined, it is essential to obtain an estimate of the residual standard deviation for the latter. Publication of residual standard deviations from every analysis of variance and multiple linear regression should therefore be encouraged. This also applies to any outcome in addition to BHR for which a completely standardised protocol does not exist.

A result expressed as effect size is not, of course, as useful clinically as one in doubling dose units. An alternative would be to assume an underlying standard deviation for studies reporting just an odds ratio and use this to convert the effect size to approximate doubling dose units. Again, only if authors report standard deviations will it be possible to know what standard deviation might be assumed.

It seems unlikely that a standardised protocol can be achieved for BHR, so there will be a continuing need to compare and combine results in the above manner. While any comparison may require this, clearly the need is greatest in conducting a systematic review and meta-analysis. Conducting two meta-analyses, one of odds ratios and the other of effect size, is unsatisfactory. Each will have reduced power to detect and explain heterogeneity between studies, they may give different answers, and when they do not the two confidence intervals will be wider than that for all studies combined. Curiously, Abramson *et al* realised in an earlier paper²⁰ that conversion of odds ratios to effect sizes or vice versa was possible by probit transformation, but dismissed the former because "to express essentially categorical outcomes such as ... BHR as effect sizes would make the results too difficult to interpret". The opposite is true as BHR is continuous, not categorical.

The above is not an argument for combining every outcome in a meta-analysis regardless of comparability. For example, specific and non-specific BHR should not be combined, and

provocation with direct and indirect stimuli may measure different aspects of asthma.²¹ Care must be taken to include each study once only when more than one analysis or BHR outcome has been included, and the choice should not be made on the grounds of "greatest significance" which can lead to publication bias. However, where two analyses have been performed, that based on a continuous outcome—whether linear regression of slope or censored regression of PD_{20} —should be preferred. The direction of each estimate must be determined, with signs reversed as necessary. Protocol variation should be a factor that is considered in examining heterogeneity, whether in effect sizes or of unstandardised estimates.

The simple method outlined here should enable better comparison between studies employing different methods of summarising BHR and promote meta-analyses which seem to be uncommon in respiratory medicine in comparison with other fields.

ACKNOWLEDGEMENTS

Susan Chinn is funded by the Higher Education Funding Council for England.

REFERENCES

- 1 Peat JK, Toelle BG, Marks GB, *et al*. Continuing the debate about measuring asthma in population studies. *Thorax* 2001;**56**:406–11.
- 2 Chinn S, Burney P, Jarvis D, *et al*. Variation in bronchial responsiveness in the European Community Respiratory Health Survey (ECRHS). *Eur Respir J* 1997;**10**:2495–501.
- 3 Chinn S, Sunyer J. Bronchial hyperresponsiveness. In: Annesi-Maesano I, Gulsvik A, Viegi G, eds. *Respiratory epidemiology in Europe*. European Respiratory Monograph 15. Sheffield: European Respiratory Society, 2000: 199–215.
- 4 Chinn S. Methodology of bronchial responsiveness. *Thorax* 1998;**53**:984–8.
- 5 Peat JK, Unger WR, Combe D. Measuring change in logarithmic data, with special reference to bronchial responsiveness. *J Clin Epidemiol* 1994;**47**:1099–108.
- 6 van Grunsven PM, van Schayck CP, Molema J, *et al*. Effect of inhaled corticosteroids on bronchial responsiveness in patients with "corticosteroid naive" mild asthma: a meta-analysis. *Thorax* 1999;**54**:316–22.
- 7 Verberne AAPH, Fuller R. An overview of nine clinical trials of salmeterol in an asthmatic population. *Respir Med* 1998;**92**:777–82.
- 8 Abramson M, Puy R, Weiner J. Immunotherapy in asthma: an updated systematic review. *Allergy* 1999;**54**:1022–41.
- 9 StataCorp. Stata Statistical Software: Release 6.0. Stata Corporation, College Station, Texas, USA. 1999.
- 10 Chinn S, Burney PGJ, Britton JR, *et al*. Comparison of PD_{20} with two alternative measures of response to histamine challenge in epidemiological studies. *Eur Respir J* 1993;**6**:670–9.
- 11 Burney PGJ, Britton JR, Chinn S, *et al*. The response to inhaled histamine and 24 hour sodium excretion. *BMJ* 1986;**292**:1483–6.
- 12 Devereux G, Beach JR, Bromly C, *et al*. Effect of dietary sodium on airways responsiveness and its importance in the epidemiology of asthma: an evaluation in three areas of northern England. *Thorax* 1995;**50**:941–7.
- 13 O'Connor G, Sparrow D, Taylor D, *et al*. Analysis of dose-response curves to methacholine. *Am Rev Respir Dis* 1987;**136**:1412–7.
- 14 Cockcroft DW, Bersheid BA, Murdock KY. Unimodal distribution of bronchial responsiveness to inhaled histamine in a random human population. *Chest* 1983;**83**:751–4.
- 15 Rijcken B, Schouten JP, Weiss ST, *et al*. The distribution of bronchial responsiveness to histamine in symptomatic and asymptomatic subjects. *Am Rev Respir Dis* 1989;**140**:615–23.
- 16 Cook DG, Strachan DP. Parental smoking, bronchial reactivity and peak flow variability in children. *Thorax* 1998;**53**:295–301.
- 17 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;**19**:3127–31.
- 18 Chinn S, Jarvis D, Luczynska C, *et al*. Individual allergens as risk factors for bronchial responsiveness in young adults. *Thorax* 1998;**53**:662–7.
- 19 Britton J, Pavord I, Richards K, *et al*. Dietary sodium intake and the risk of airway hyperreactivity in a random adult population. *Thorax* 1994;**49**:875–80.
- 20 Abramson MA, Puy RM, Weiner JM. Is allergen immunotherapy effective in asthma? *Am J Respir Crit Care Med* 1995;**151**:969–74.
- 21 van den Berge M, Meijer RJ, Kerstjens HAM, *et al*. PC_{20} adenosine 5'-monophosphate is more closely associated with airway inflammation in asthma than PC_{20} methacholine. *Am J Respir Crit Care Med* 2001;**163**:1546–50.